

Timo Kinnunen
Särkiniementie 16 A 41
70700 Kuopio
Finland

Ihminen ja tietokone - Taustamateriaalia

Käyttämäni pseudofilosofisia kirjallisuuslähteitä

Nämä lienevät aika lähellä filosofista kirjallisuutta, josta puuttuu useimmiten kaikki linkit esimerkiksi käytännön tieteisiin, tai referenssit tieteiden tutkimustuloksiin. Filosofointi on tautologiaa, ja vailla varsinaista, todellista sisältöä. Kun puhutaan tietokoneista, tulisi ainakin jossakin määrin tiedostaa niiden koneolemus, ja ymmärtää, etteivät ne ole samanlaisia kuin ihmisäivot ja elimistö, jotka tapaavat filosofoida. Koneet eivät filosofoi, koska niillä ei ole sellaiseen tarvetta. Kun kirjoitin väitöskirjaksi aikomaani tekelettä, olin työskentelevinäni itsenäisenä yrittäjänä, ja skannailin omia – ja muiden tekijä tekstinkäsittelyohjelmaa. Laite oli uusinta uutta, ja sangen kallis, mutta uskoin että sillä voisi tienata leipänsä. Väärä oli luuloni, sillä tuli lama ja äkkiä töitä ei enää ollutkaan. Seikkailin noihin aikoihin eräässä keksijäyhdistyksessä, ja eräs siinä vaikuttanut hemmo oli tekulla töissä, ja häneltä sain pinon englanninkielisiä MIT -tutkimuksia, jotka joutessani käänsin suomeksi. Käännökseni eivät ole virheettömiä, koska tekniikka on aina ollut minulle vierasta – vaikka teinkin sittemmin teknisiä käännöksiä Idioma -yhtiölle. Matkan varrella sain idean siitä, että filosofia on sinällään tyhjää, ja vaatii tuekseen empiiristen tieteiden uusimpia tutkimustuloksia, ja lisämausteita elävästä elämästä. Itse asiassa minulla on koko ajan ollut tendenssinä pyrkimys synteisiin, ja poikittaisten suhteiden etsimiseen. Filosofiasia ja tieteessä taasen esimerkiksi opinnäytteiden tuli olla kapea-alaisia ja vain yhteen tai kahteen teemaan keskittyviä. Niitä piti sahata edestakaisin, ja kai sitten uskottiin että totuus ja todellisuus paljastuisi siten parhaiten. En ole koskaan uskonut tuonkaltaisen paskaan. Näin jälkikäteen näitä tässä esiteltäviä tutkimuksia tarkastellessa tuntuu vahvasti siltä, että MIT -instituutissa suunniteltiin jo täyttä päätä sitä yhteiskuntaa, jossa parhaillaan elämme. Keneltäkään ei kysely mitään, vaan tulos annettiin ihmisille valmiina ja kaupparatsujen märehimänä ainoana ratkaisuna. Nyt sitten tekoäly tunkeutuu kaikkialle, ja tarjolla on karttasuunnistuslaitteita, vaikka kaikki eivät edes sellaisia tarvitse. Heidätkin manipuloidaan haluamaan näitä uusia hyödykkeitä, sillä kaupparatsujen miestä kaikki heidän märehimänsä ja suodattamansa aines on nimenomaan hyödyllistä (siitä kai termi *hyödyke*) Ololuone haluttiin täyttää tietotekniikalla, jossa kaikki laitteet kommunikoivat keskenään. Pian ihmiseen kytketään laitteita jotka mittaavat jatkuvasti sykettä ja suolentoimintoja, sun muita. Kaikkea halutaan valvoa, ja niinpä me elämme isoveljen valvomassa yhteiskunnassa. Olisin halunnut kertoa kaikesta tästä, mutta siinä en onnistunut, vaikka tarkoitukseni oli hyvä ja erittäin filantrooppinen.

TOWARDS AN INTEGRATED VIEW OF 3-D COMPUTER ANIMATION

David Zeltzer

Tr. into Finnish by **Timo Kinnunen**

*Computer Graphics and Animation Group, The Media Laboratory,
Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Jotta voitaisiin automatisoida animaatiota ja laajentaa sitä kolmiulotteiseksi (3-D) on luotava ja manipuloitava kolmiulotteisia hahmoja, kuten niitä maailmoja, joita ne tulevat "asuttamaan". Abstraktointi ja adaptoituva liike ovat avainmekanismeja käsiteltäessä vapausasteiden ongelmaa, mikä viittaa kontrollointi-informaation ohjausvoimaan koordinoitaessa keinoitekoisten hahmojen liikettä silloin, kun linkkien määrä on hyvin suuri. Animaation kontrolloinnin muotojen kolmeksi hierarkiseksi tasoksi on ehdotettu: OHJAUS (guiding), ANIMATOINTITASO (animator-level), ja TEHTÄVÄNRATKAISUJÄRJESTELMÄT (task-level systems). Ohjaus soveltuu parhaiten hienojakoiisiin yksityiskohtiin, mutta ei mutkikkaiden liikkeiden kontrollointiin. Animatointitason ohjelmointi on voimakas, mutta myös vaikea. Tehtävänratkaisujärjestelmät antavat soveltuvan kontrollointimahdollisuuden mutkikkaita liikkeitä varten, ja sen tehtävät sinällään mahdollistavat liikeosien kontrollin. Tässä tutkimuksessa keskustellaan näiden kontrollitasojen integroinnista. Avainsanoja: Tietokoneanimaatio - Simulaatio - Ihminen-tietokone vuorovaikutekenttä

ANIMATION AS SIMULATION

Tänä päivänä vallitsee erimielisyyttä kolmiulotteisen tietokoneanimaation luonteesta. Pitäisikö systeemin perustua simulaatioon, avainruutuihin (keyframing), tai animaation ohjelmoitukieleen. Pitäisikö vuorovaikutuksen tapahtua graafisten näyttöjen kautta, vai näppäimistöillä? Tässä tutkimuksessa tarkoitukseni on tuottaa käsitteellinen viitekehys kolmiulotteista animaatiota varten yleisesti, ja

hahmojen animaatiota varten erikseen. Automaattinen väliintulo (inbetweening) on ollut mielenkiinnon kohteena jo käytettäessä tavanomaista tietokoneavusteista kaksiulotteista animaatiota (Burtnyk and Wein 1976; Catmull 1978). Tästä näkökulmasta ajatellen tuntuisi luonnolliselta siirtyä automaattisen väliintulon (inbetweening) käyttöön myös kolmiulotteisessa animaatioissa, ja itse asiassa onkin kehitetty lukuisasti sentapaisia systeemejä (Chuang and Entis 1983; Gometz 1984; Magnenat-Thalmann and Thalmann 1983). Tässä tutkielmassani väitän, että kolmiulotteinen avainruudutus (keyframing) kuuluu ensimmäiselle kolmesta hierarkian tasoista animoitaessa keinotekoisia hahmoja. Tuotettaessa vakuuttavia henkilöahmo-animaatioita, referoivat kaksiulotteisen animaation käyttäjät jatkuvasti eläviä malleja, tai tutkivat elävistä malleista tehtyjä kuvasarjoja, tai siirtävät suoraan kuvasarjojen erillisistä, liikkumattomista kuvaruuduista niissä liikkuvia hahmoja (rotoscoping). Tämä tarkoittaa sitä, että henkilöahmoanimaatio ei ole prosessi, jossa käsiteltäisiin viivoja ja muotoja kaksiulotteisella pinnalla, eikä se ole myöskään yksinkertaisesti vain hahmojen litistämistä, venytystä, tai muulla tavoin hahmojen liioittelua ja karikatyyrien muodostamista niistä. Thomas ja Johnston (1981) tekivät täysin selväksi, että Disneyn palkkaamien animaattoreiden menestys perustui siihen, että he perehtyivät pitkiä aikoja ihmisten ja eläinten liikkumistapojen tarkkailuun ja opiskeluun valmistellessaan erityistä kuvajaksoa. Henkilöahmon onnistuminen riippui täysin siitä, kuinka hyvin animaattori ymmärsi kuvattaviin liittyvää kinematiikkaa, ja liikkeiden struktuuria ja ajoitusta, kuten myös sitä, kuinka liikkeen saisi näkymään myös henkilöahmojen ollessa vaetetettuina. Kun tällaiset elementit hallittiin, niin silloin animaattori saattoi kehittää henkilöahmoihin persoonallisuutta tietyillä liioitteluilla, tai tiettyjen attribuuttien jättämisellä vähemmälle huomiolle. Niin pitkään kun animointi vaatii lukuisten piirrosten valmistamista käsin, tulevat yksinkertaisuus ja valmistuksen taloudellisuus olemaan sen olennaisimpia elementtejä. Kolmiulotteinen tietokoneanimaatio on kuitenkin kokonaan erilainen medium. Animaattorin energia ei enää suuntaudu piirtämiseen, ja väliintulon monotonisuuteen (tedium of inbetweening). Sen sijaan kohteena on ympäristön luominen -mikromaailmojen luominen, ja niiden asuttaminen mielenkiintoisilla henkilöahmoilla. Koska kuvaruutuja ei enää luoda käsin, niin yksinkertaistettujen ja tyylieltyjen kuvitehahmojen sijasta saavutamme tietokoneanimaation suurimmalta osaltaan kiinnittämällä huomiomme komiulotteisuuden simulointiin, valaistukseen, taustaan, ja pintojen tekstuuriin. Animatoidut kuvat voivat saada lisää kompleksuutta ja tulla detaljoidummaksi; ne tulevat näyttämään realistisemmilta, mikäli niin halutaan, tai vakuuttavammin omaan maailmaansa kuuluvilta. Tämä koskee myös hahmojen käyttäytymistä, kuten myös niiden objektien maailmaa, ja niiden kaikkien fysikaalista ulkonäköä. Komiulotteinen tietokoneanimaatio on siten simulaatioprosessi sanan kaikkein yleisimmässä merkityksessä; objektien spesifointi ja niiden transformointi. Tietokoneen käyttö simulaation mediana ei ole uutta; Turing osoitti jo kaikkiaan, että tietokone voisi simuloida itseään, tai mitä tahansa muutakin -tätä teemaa esittivät usein mm. Alan Kay (kay and Goldberg 1977; Kay 1985). Varhaisimpien graafisten sovelteiden tutkijain päämääränä oli apprehensoida maailman visuaalista kompleksuutta opiskelemalla valon vaikutustapoja, varjoa, tekstuuria jne. Nykyinen synteettinen mielikuvitus on saavuttanut lähes valokuvallisen realismisuuden tietyissä geometrisesti ja stokastisesti kuvattavissa oleviin objekteihin ja ymähäisiin nähden. Nyt tehtävänä on yrittää apprehensoida maailman proseduraalista monimutkaisuutta, mikä vaatii, kuten tulemme näkemään, tietokonealle monista komplekseista, tai ainoastaan osittain ymmärretyistä artikuloitujen liikkeen dynaamisista ja inversiivin kinematologian prosesseista, kuten myös terveeseen järkeen perustuvaa suunnittelua ja ongelmanratkaisuja. Kuinka me voimme spesifioida ja koordinoita objekteja, joita tahdomme animatoida? On olemassa kolme perustavaa lähestymistapaa:

1. OHJAUSMALLISSA (guiding mode) voimme eksplisiittisesti kuvata ne käyttäytymiset, joista olemme kiinnostuneita,
2. ANIMATOINTITASOLLA (animator-level) voimme kuvata käyttäytymisiä algoritmisesti jollakin ohjelmoinnin notaatiolla,
3. TEHTÄVÄNRATKAISUSYSTEEMEISSÄ (task-level) voimme kuvata käyttäytymisiä implisiittisesti tapausten ja suhteisiinasetumisten termein.

Tarkoituksiamme varten soveltuu kiinteistä linkeistä artikuloitujen struktuurien kontrollointi ja koordinaatio hyvin ajatellen kolmiulotteista piirteistä animaatiota. Erityyppiset skaalaukset ja muut muotojen transformaatiot ovat tärkeitä, mutta sittenkin toissijaisia ajatellen motorista kontrolliota. Seuraavassa luvussa tutkimme sitä perustavanlaatuista ongelmaa, jonka kohtaamme yrittäessämme koordinoita kenotekoisien hahmojen liikkeitä, ja tarkastelemme mekanismeja, joilla käsitellä hahmojen

animaation mutkikasta kokonaisuutta. Myöhemmin havaitsimme, että mainitut kolme hierarkista tasoa ovat tärkeitä hyödynnettäessä mekanismeja.

THE DEGREES OF FREEDOM PROBLEM

Olellainen ongelma koordinoitaessa keinoitekoisten hahmojen liikkeitä liittyy soveliaiden liittymävariabelien (joint variables) arvojen generointiin, joilla kontrolloidaan jokaisen linkin asemaa ja orientaatiota. Liittymät voidaan muotoilla alemmiksi pareiksi (Denavit and Hartenberg 1955) kuten rotaatioksi, tai liukuviksi liittymiksi, tai ne voivat olla vielä kompleksisempia, kuten ihmisen polvea kuvaavassa detaljoidussa mallissa. N -liittymäiselle kuviolle voidaan ajatella n -dimensionaalinen asettumisen avaruus (pose space), johon me liitämme koordinaattiakselin jokaista n -vapausastetta varten, ja n -komponenttisen asettumisen vektorin, joka täydellisesti spesifioi erityisen konfiguraation. Jotta voitaisiin animatoida toisiinsa liittyvien kuvioiden liikettä, on asettumisen vektori spesifioitava jokaista kuvaruutua varten erikseen kokonaiskuvien sekvenssissä. Jotta voitaisiin animatoida yhden minuutin mittainen kokonaiskuvien sekvenssi, jossa liike on kompleksista, ja kuvio on detaljoituva niin, että siinä on 30 linkkiä, niin tarvitaan 10 000 arvoa, joilla spesifoidaan liittymäarvot, jotka mahdollistavat sen, että jokaisessa kuvaruudussa on toisistaan poikkeavat konfiguraatiot. Jopa silloin, kun kokonaiskuvien sekvenssi on avainruudutettu (keyframed) niin, että avainruudut (keyframes) seuraavat toisiaan kahden sekunnin välein, tarvitsee spesifoida 30 asettumisen vektoria (pose vectors) -lähes tuhat arvoa. Tämä olkoon esimerkkinä vapausasteiden ongelmasta (The degrees of freedom problem, DOF) (Turvey et al. 1982), mikä viittaa ohjausinformaation määrällisen volyymin tarpeeseen koordinoitaessa keinoitekoisten hahmojen liikettä, kun linkkien määrä on suuri, kuten esimerkiksi ihmishahmon kohdalla. Tässä on syy siihen, miksi animatoijat eivät pidä komiulotteisesta henkilöahmoanimaatiosta. Tietenkään emme ole kiinnostuneita satunnaisesta liikkeestä (random motion): kuvion liikkeiden tulee olla "oikeita" tietyssä mielessä, jotta niillä olisi käyttöä. Robottiohjelmoija saattaa toivoa energiankäytön optimointia, ja animatoija sitä, että kuviot liikkuisivat jollakin ekspressiivisellä tavalla. Ajatellen asiaa tältä kannalta, on animaation ongelma hakemisessa (search). Meidän ei ainoastaan pidä generoiman asettumisten vektoreita (pose vectors), vaan myös löytää erityinen variaabeleiden joukko välittömästi asettumisen avaruudesta (pose space). Jos jokaisella kolmestakymmenestä liittymästä on vain kaksi mahdollista asettumisen tapaa, niin sitä vastoin mahdollisia konfiguraatioita asettumiselle on yli miljoona! Komplisoidaksemme ongelmaa todettakoon, että monet kiinnostavista kuvioista ovat kinesteettisesti redundantteja; niillä on "ylimääräisiä" vapausasteita, jotka tarjoavat monia vaihtoehtoisia ratkaisumalleja. Niillä on ehkä lukematon määrä asettumisten vektoreita (pose vectors), joista kaikki antaisivat tyydyttävän ratkaisin liikkeen ongelmaan. Ihmisen käsivarsi on redundantti, koska sormilla on mahdollista poimia jokin esine pitämällä kyynärpäätä eri asennoissa. Tämä tarkoittaa sitä, että käsivarrella on monia konfiguroitumisen mahdollisuuksia, joissa sillä on silti tietty, kiinteä asema avaruudessa. Tämä redundanssi toisaalta mahdollistaa tietyn joustavuuden, tavoiteltaessa kädellä jotakin, tai tehtäessä sillä jotakin mutkikkaasti rakentuneessa ymärissä. Tästä syystä johtuen individuaalit voivat kehittää käsiensä liikkeisiin heille luonteenomaisia ilmaisullisia tyylejä.

ADAPTIVE MOTION

Adaptoituva liike liittyy kuvankontrolloijan kykyyn käyttää hyväkseen ympäristön informaatiota, kuten kuviota itseäänkin kontrollointiprosessissa. Tämä tarkoittaa sitä, että voidaan hyödyntää takaisinkytkentöjä, joilla liikeratojen etsintää ohjataan mahdollisten konfiguraatioiden valtavasta avaruudesta. Jotta tämä voitaisiin tehdä, ainakin koskien objektien ja niiden pintojen avaruudellista sijaintia ja liikettä siinä koskien, on takaisinkytkennän oltava saatavilla animatointiin käytettävässä ohjelmistossa, eikä juuri niin, että se mahdollistuisi aliohjemilla, kuten tavallista. Fyysiset interaktiot kuvioiden ja objektien välillä ovat mutkikkaita reaalisessa maailmassa: koskettamista, raaputtamista, työntämistä, puhumattakaan erilaisista mahdollisuuksista työstää erilaisia pintoja liittyen automaattiseen työstötekniikkaan, joka on kiinnostanut CAD/CAM:ia ja robotiikkaa. Niinpä tällaisten erilaisten fyysisten interaktioiden mahdollisuuksien tulisi tulla animatoinnilla luotujen ympäristöjen kiinteäksi osaksi. Adaptoituva liike mahdollistaa päämääräohjautuvan ja kurinalaisen käyttäytymisen, sillä se sallii animatoijan kuvata liikettä termein, jotka liittyvät kuvioiden ja objektien välisiin suhteisiin. Se tuottaa

yleistyvyyttä animatoituihin sekvensseihin, koska käytetty ohjelmisto voi mukautua liikesarjoihin erilaisissa kuvasekvensioissa. Tämä myös auttaa kätkemään sellaiset detaljit animatoijalta, jotka muodostuisivat hänelle taakaksi liiallisen animointityöhön sen vaiheina liittyvän kontrollointi-informaation muodossa, joka jäisi siten ohjelmoijan huoleksi. Reynolds (1982) on ehdottanut, että animatoijinen tulisi luoda "käyttäytymiselle säännöstöjä" koskien objekteja ja hahmoja kuvitteellisissa mikromaailmoissa. Kun alustavista olosuhde-säännöistä näissä simuloituissa kaikkeuksissa olisi päästy yksimielisyyteen, voisivat animatoijat istahtaa tuolilleen, ja antaa animaatiojärjestelmien generoida sekvenssejä. Tämä liittyy kaksiulotteisessa (2-D) animaatiojärjestelmään vallitsevaan "suoraan eteenpäin"-järjestelmään (straight ahead), jolle tuotettaisiin näin kolmiulotteinen (3-D) laajennus (Thomas and Johnston 1981). Adaptoituvaa liikettä tekee tämän tekniikan laajennuksen mahdolliseksi myös kolmiulotteisessa tietokoneanimaatiojärjestelmässä.

ABSTRACTION

Abstrahoinnin tärkeys käsiteltäessä tietokoneohjelmoinnin älyllistä kompleksisuutta on tunnettua (Shaw 1980), ja se onkin samalla tavoin perustyökalu käsiteltäessä kinesteettistä ja käyttäytymyksellistä artikuloitua liikkeen kompleksisuutta. Käyttökelpoisia abstrahoinnin tapoja hahmojen animaation kontrolloinnissa on viisi: strukturaalinen, proseduraalinen, funktionaalinen, ja hahmojen, sekä niiden maailman muotoiluun liittyvä.

Structural abstraction

Strukturaalinen abstrahointi kuvaa kinesteettisiä kuvion ominaisuuksia liittyen transformaatiohierarkian, saatavilla olevien yhdistettyjen liikkeiden luonteeseen, ja siihen, milloin linkit ovat rigidejä ja milloin ne eivät sitä ole (me käsittelemme tässä ainoastaan rigidiä liikettä). Transformaatiohierarkiaan liittyvä kaksiulotteisesta systeemeistä tuttujen hierarkioiden yleistäminen kolmiulotteisuuteen. Useimmat kolmiulotteisista animointijärjestelmistä sisältävät joitakin keinoja, joilla voidaan esittää transformaatiohierarkioita, kuten esimerkiksi Crown "scn-assmblr" (Crow 1982)), Blinnin "artic" (Blinn 1982) ja Reynoldsin "ASAS" (Reynolds 1982). Näissä systeemeissä liittymistransformaatiot esitetään yksinkertaisina rotaatioina ja käännösinä, jotka joskus sisältävät skaalausta, vaikkakin enimmäkseen artikuloitua liikkeen yleisiä representaatioita, kuten esimerkiksi Denavit-Hartenburgin (D-H) notaatio (Denavit-Hartenburg 1955; Lee 1982), jota on pitkään käytetty mekaanisessa suunnittelussa, ja nykyisin robotiikassa (Paul 1981). "Sdl", eli luurankokuvauskieli (skeleton description language) on työkalu, jolla spesifioidaan niitä strukturaalisia aspekteja, joita käytetään "sa":ssa, joka on artikuloitu liikejärjestelmä (Zeltzer 1984). Samantapainen työkalu on "mat", jollainen on käytössä New York Institute of Technology'ssa (Williams 1982). Media Laboratory'ssa olemme kehittäneet valikoiman graafisia työkaluja, jotka on istutettu Symbolics 3600 Lisp Machine'iin, joilla voi suunnitella ja muotoilla liittyvien (jointed) kuvien standardisoituja kinesteettisiä kuvauksia. Nämä kuvaukset on ulotettu koskemaan myöskin D-H representaatiota prismaattisten (prismatic) ja rotaaristen liittymien osalta. Representaatio on laajennettavissa niin, että siihen voidaan lisätä muitakin liittymistyyppisiä helposti. Tämöntapainen standardisoitu representaatio on keskeinen monikeskisessä tietokoneympäristössä, jossa työskentelee useita ohjelmoijia ja animoijia käyttäen monenlaisia tietokoneita, ja jotka haluavat jakaa tuotettua graafista dataa ja hyödyntää tuotettuja ohjelmistoja. Lisäksi voimme helposti integroida robottiohjautuvat kontrollialgoritmit animaatio-ohjelmistoon, koska D-H notaatio tarjoaa miellyttävän ja voimakkaan representaation tällaisille motorisen kontrollin tekniikoille.

Procedural abstraction

Proseduraalinen abstrahointi (Tennent 1981) on sellainen liikealgoritmin representaatio, joka on riippumaton sen kuvion struktuurista, jota se kontrolloi. Esimerkiksi DOF ongelma ei ole kovinkaan vakava robottimanipulaation tapauksessa, jossa on kuusi tai seitsemän liittymistä (joint). Jopa sellaisenaankin pelkistettynä eivät ihmiset kovinkaan hyvin onnistu laskemaan tarvittavia liittymiskulmia (joint angles) kontrollidessaan yksinkertaista manipulaattoria, ja liikeratojen laskennalliseen

määritykseen turvaudutaan usein (resolved motion) (Whitney 1972; Klein and Huang 1983). Tämä merkitseekin sitä, että syöttönä annetaan kohteen määränpääasema ja liikesuunta, ja että manipulaatiokontrolleri laskee automaattisesti tarvittavan asettumisvektorin (pose vector), jolla annettu määränpää saavutetaan. Liikkeen laskennallinen määrityskontrollointi on tärkeä esimerkki proseduraalisesta abstrahoinnista ratkaistaessa DOF ongelmaa: on määritelty se laskentatapa, joka transformoi syöttöparametreja, eli kohteen sijaintia ja orientaatiota tulosteobjektiksi, jossa on sellaiset liittymiskulmien arvot, jotka orientoivat lopputulosteen (the end effector) toivottuun sijaintiin työskentelytilassa, mikäli mahdollista. Laskennallinen määrityskontrolli on riippumaton siitä, millainen erityinen kinesteettinen struktuuri on vallitsevana, ja sitä voidaan soveltaa kuvioihin, joissa on 6 tai 8, tai useampiakin linkkejä, ja myös ihmishahmoja koskien sitä voidaan käyttää kontrolloimaan niin käsivarsien kuin raajojenkin liikkeitä (Girald and Masiejewski 1985; Ribble 1982). Muita esimerkkejä proseduraalisesta abstrahoinnista ovat putoavien esineiden liikeratojen laskeminen, törmäävien esineiden liikeradat, tai loivien käännösten laskeminen gereroitaessa pehmeitä liikkeitä. Tällaisia ominaisuuksia tarjotaan usein animatoijille valmiina pakettina, mutta niitä voi myös hän itse konstruoida korkean tason ohjelmointikielten animointisysteemeissä.

Functional abstraktion

Robottikädessä linkkien lukumäärä on pieni, ja käsi toimii yhtenä kinesteettisenä kokonaisuutena. Kuvioille, joissa linkkejä on enemmän me kuitenkin haluamme yhdistettäväksi niin ne strukturaaliset kuin proseduraaliset elementit, jotka ovat välttämättömiä tiettyyn luokkaan kuuluviin liikkeisiin. Vaihtoehtoisesti voimme pakottaa liikkeen ohjautumaan tiettyjen liittymisten mukaisesti. Kutsumme tällaista abstrahointia funktionaaliseksi, mikä on tärkeä siksi, että animatoija voi faktoroida asettumisavaruuden (pose space) motorisiksi taidoiksi (motor skills). Mikäli tunnemme liikkeen yleisen muodon, on meidän enää tarkasteltava asettumisen osa-alueita koko asettumisten avaruudessa. Sanokaamme, että haluamme kuvion käden tavoittelevan jotakin objektia, ja tiedämme, mitä liittymiä meidän on liikuteltava, tai ainakin suunnilleen tiedämme tämän, ja tiedämme myös, että kyseessä on hyödyllinen liike, joka toistetaan usein. Voimme ryhmittää tämän liittymisliikkeiden joukon tehtävään "tavoittaminen", ja sitten käyttää sovelluksessa yhtä tai useampia mahdollisuuksia (ehkä laskennallisesti määritelty liike). Kun tällainen motorinen taito on määritelty, voidaan itse taidon luomisen detaljit poistaa työskentelystä, mikä tarkoittaa sitä, että meidän tarvitsee käyttää vain tiettyjä, soveltuvia parametreja, kuten päämääräsijainti, nopeasti tai hiljaa, kovaa tai pehmeästi -luodaksemme motorisen ohjelman, jolla taito saavutetaan. Spesifioidessaan funktionaalisia abstraktioita "tavoittamisesta", tai muista tehtävistä, säästyy animatoija asettumisvektoreiden rakentamisen taakasta, ja voi sen sijaan muotoilla liikkeitä korkeammalla suoritettavien tehtävien ja tapausten tasolla. Funktionaalinen abstrahointi mahdollistaa meille implisiittisten päämäärinen saavuttamisen kuvioiden liikkeissä. Jäsentämällä kuvioiden liikkeit uudelleen potentiaalisiksi liikkeiksi taitojen repertuaariin voi animatoija assosoida tapahtumia ja suhteita taitoihin (funktionaalilla abstrahoinnilla), jotka kuvioidenkontrolleri "tuntee". Lisäksi voimme ulottaa funktionaalisten abstrahointien referoituvuuden muihin funktionaaliin abstrahointeihin, jolloin on mahdollista konstruoida käyttäytymisiä yksinkertaisempien liikkeiden valikoimasta.

Character abstraction and world modeling

Fysikaalisessa maailmassa kuviot ja objektit ovat vuorovaikutuksessa mutkikkaalla tavoin monilla tasoilla ja monin detaljein. Adaptoituva liike vaatii vähintään tehokkaan geometrisen representaation törmäystestejä ja polkusuunnittelua bvarnten: päämäärästäohjautuva animaatiokontrolli vaatii lisäksi sofistikoituneempia mekanismeja, jotta tieto voitaisiin esittää. Osa ongelmasta liittyy korkealaatuisen graafisen datan strukturointiin, jotta vältettäisiin pitkiä hakuja pintaelementtien luetteloista sanokaamme törmäystestien tapauksessa. Pikemminkin haluaisimme huomioida ainoastaan ne kokonaisuuden objektit, jotka ovat "lähellä" kuviota. Tämä tarkoittaa sitä, että data on organisoitava huolellisesti spatiaalisesti siten, että etsintä tapahtuu aina soveliaalla detaljitasolla. Lukuisista hierarkisista datanstrukturointimeteodeista, joilla hakua voidaan nopeuttaa, on esitetty raportteja (Clark 1976; Fuchs et al. 1980; Rubin and Whitted 1980). Franklin (1981) kuvaa joukkoa algoritmeja, jotka soveltuvat useamman alueen testaukseen (intersection testing) aivan yhtä hyvin.

Suurempi ongelma on esittää attribuutteja, funktionaalisuutta, ja objektien välisiä suhteita kokonaisuudessa (scene) niin, että voidaan yleisesti simuloida niiden välisiä interaktioita ja käyttäytymisiä. Me haluamme tämän representaation olevan yhdenmukaisen niin, ettei agenttien ja objektien välillä ole eroa. Tämä tarkoittaa sitä, että siinä missä ihmiset ovat jossakin suhteessa aktiivisia agenteja, he myöskin noudattavat lakeja, jotka ovat newtonilaisen mekaniikan mukaisia; joku henkilö putoaa alastönäistynä kalliolta juuri samalla tavoin kuin kivi. Toisaalta voi olla, että animatoija haluaa tuolien ja pöytien tanssivan ympäri huonetta isännän poistuttua. Olisi helppoa otaksua sellaisen käyttäytymisen tietoisiksi aiheuttajiksi näitä objekteja, jotka muutoin ovat inanimantteja. Meidän on siten spesifioitava kolme seikkaa objektista: mikä se on, millaisten objektien kanssa se esiintyy, ja kuinka se käyttäytyy. Fysikaalisten objektien representaatio-ongelma liittyy aktiivisesti keinoitekoisen älyn tutkimukseen (Wasserman 1985). Selvästikin objektit voidaan kuvata yleistyvyyden hierarkiassa, ja sellaisissa tapauksissa, joissa erityisiä seikkoja esiintyy, yleisempien objektiluokkien specialisaatioina. "Periytyvyys" (inheritance) on avaintermi, joka tarkoittaa sitä, että objekti-tapaukset voivat käyttää sellaisia attribuutteja ja proseduureja, joita käyttävät muutkin niiden luokkaan kuuluvat objektit. Periytyvyudet ovat asetusarvoja (default values), jotka voidaan korvata spesifioimalla erityisiä arvoja näitä lohkoja (slots) varten tapauksissa itsessään. Myös moninkertaista periytyvyyttä (multiple inheritance) tarjotaan, jossa objekti voi periä attribuutteja useammastakin kuin yhdestä sitä koskevasta vallitsevasta luokasta. Tästä syystä hierarkian yleistyvyyttä kuvaa pikemminkin monikerrostuva tasojakauma (lattice) kuin yksinkertainen puu. Useat ohjelmointisysteemit kannattavat tällaista näkemystä objektien representaatiosta, mikä on myös hyvin kuvattu kirjallisuudessa (Stefik et al. 1983; Tesler 1981).

Objektin paikantuminen sellaiseen yleistyvyyden tiettyyn tasopintaan kertoo siitä, mitä se on, ja miltä se näyttää (kuinka sitä voidaan muokata). Voimme kuvata objektin struktuurin assosioimalla jokaisen yleistyvyyden tasopintaleikkauksen transformaatiohierarkiaksi, kuten kuvattu. Leikkauspinta ihmishahmoa varten voidaan assosoida pinnalle puurakenteeksi, jotka kuvaavat liittymiä ja linkkejä, joista kuvio muodostuu. Jokainen linkeistä vuorostaan voisi olla tapaus, joka liittyy monoliitiseen objektiin, joka sijaitsee toisaalla yleistyvyyspinnalla. Ihmishahmojen luokan tapaukset perivät tällaisen strukturaalisen kuvauksen paikallisine variaabeleineen, jotka kuvastavat erityisen ihmishahmon sijaintia ja liikettä. Strukturaalisen kuvauksen tapaan tulisi jokaisella objektilla olla siihen assosioituva käyttäytymyksellinen kuvaus. Yksinkertaisia objekteja varten voisivat tällaiset kuvaukset olla myös yksinkertaisia. Prototyypimäiset fysikaaliset objektit esimerkiksi saattaisivat totella joitakin newtonilaisen mekaniikan lakeja. Artikuloituilla kuvioilla voisi olla jo koko joukko kykyjä, kuten käveleminen ja esineiden tavoittelu. Koska käyttäytymisiä voidaan periä, voivat ihmisfiguurit periä myös sellaisia ominaisuuksia, jotka liittyvät newtonin mekaniikan mukaiseen käyttäytymiseen, jota niiltä odotammekin. Lisäksi on välttämätöntä representoida objektien välisiä mekaanisia interaktioita rajoittunein, ja hyvin määritellyin suhteisiinasettumisten termein. Koska me haluamme esittää muuttuvaa maailmaa, on näiden suhteisiinasettumisten oltava dynaamisia, ja sisällettävä objektien välisiä linkkejä, jotka merkittävästi tukevat kontakteja, sisältöjenottoa, epsilon-proksimiteettia, ja myös tietoa siitä, milloin objekti on toisen osa, tai toisen objektin liikkuva osa. Tällaista suhdetta voidaan hyödyntää objekteihin, jotka itsessään ovat mutkikkaita liittymärakenteita, ja joilla jokaisella on oma strukturaalinen kuvauksensa, kuten on laita esimerkiksi auton moottorin ollessa osa autoa. Lopuksi, ratkaistaessa yksinkertaista motorista ongelmaa, haluamme liittää yleistä tietoa objektien kuvauksiin. Se tarkoittaa sitä, että kykenemme enkoodaamaan sentapaista tietoa, kuten "joku poistuu tavallisesti huoneesta etsien oven, jonka avaa". Näin ei ole laita aina animaatiossa, ja siksi representaatioiden tulisi olla helposti muotoiltavissa. Joka tapauksessa tällainen tieto edustaa kulttuurillisia individuaalien oppimia seikkoja, ja on yleistä tietämystä (common knowledge), jonka hallitsemme ovien avaamisista. On sopivaa assosoida tällaista tietoa objekteihin itseensä. Se tarkoittaa sitä, että muotoillessamme mikromaailman fysikaalisia ja geometrisia ominaisuuksia, täytyy maailmaan sisällyttää objektien saatavissa olevaa kulttuurista informaatiota. Tutkimme parhaillaan tekniikoita, joilla sellaista informaatiota voisi enkoodata käyttäytymyksellisen kuvausten yhdenmukaistamiseksi koskemaan kaikkia yleistyvyyden tasojen objekteja.

Text-mediated and device-mediated interaction

Animointisysteemin voimakkuus on riippuvainen täydellisesti siinä käytetystä abstrahointimekanismista

ja adaptoituvan liikkeen hallintavälineestä; mikä ei suoranaisesti liity animatojan käytettäväksi tarjoutuviin ohjausvälineisiin, kuten joystickiin. On tehty paljon, jotta tietokonegrafiikkaan voitaisiin liittää oheislaitteistovuorovaikutusta (device-mediated interaction), jollainen työ alkoi jo silloin, kun ohjelmointikielinä ainoina vaihtoehtoina olivat Fortran ja assemblykoodaus ihmisen ja tietokoneen välisessä kommunikaatiossa (Baecker 1969; Sutherland 1963). Kuitenkin ohjelmointikieli tulee todennäköisesti säilymään valinnan mediumina algoritmien ja mutkikkaiden spatiaalisten, temporaaristen ja käyttäytymyksellisten suhteiden kuvauksessa. Tekstimedia vuorovaikutus liittyy kirjoittamiseen. Kehitystä on tapahtunut, ja tulee edelleen tapahtumaankin hyvin pitkälle näppäimistön ergonomiassa, ja erityisesti puheentunnistuksessa (speech recognition), joita voidaan käyttää ihminen/ tietokone vuorovaikutekentässä (interface). Samaan aikaan on ilmaantunut useita muitakin funktioita, kuten poimiminen (picking), paikallistaminen (locating), ja luonnehtiminen (sketching), jotka edellyttävät graafista kuviota sinällään interaktion kanavaksi. Äärimmäisin esimerkki graafisesta interaktiosta lienevät tapaukset, joissa esimerkiksi lennetään lentosimulaattorilla, tai ohjataan kuusijalkaista science-fiction robottia joystickilla. Nämä ovat kuitenkin laajoja simulaatio-ohjelmia, jotka rakentuvat mutkikkaista proseduureista. Käyttäjä on interaktiossa systeemiin korkealla hierarkian abstrahointitasolla, ja nimenomaan organisaatio sallii systeeminkäyttäjän tekevän mm. kädellään pieniä liikkeitä, tai ohjausliikkeitä joystickilla, jotka sitten vahvistetaan merkityksellisiksi ja komplekseiksi kontrollisignaaleiksi, joilla sitten on siten myös voimakas vaikutus. Seuraavissa luvuissa tulemme näkemään, kuinka kolme kontrollioinnin tasoa liittyvät adaptoituvan liikkeen hallinnan työvälineistöön ja abstrahointimekanismeihin.

A THREE LEVEL HIERARCHY FOR CHARACTER ANIMATION

Voimme luokitella animointijärjestelmät ohjaukseen (guiding), animointitasoon (animator level) ja tehtävänratkaisusysteemeihin (task level systems). (Myös robottien ohjelmoinnissa käytetään samanlaista luokitusta, vrt. Lozano-Perez 1982).

Guiding

Ohjaussysteemit (guiding systems) ovat sellaisia, ettei niihin liity mekanismeja, joilla käyttäjä määritteli abstrahoituvuutta, tai ohjaisi adaptoituvaa liikettä. Ohjaussysteemejä on hyvin monenlaisia, kuten liikkeen nauhoittaminen (recording) (Calvert et al. 1980; Ginsberg and Maxwell 1983), muodon interpolointi (Gomez 1984), avain-trasformaatiojärjestelmät (Chuang and Entis 1983; Gomez 1984; Williams 1982) ja notaatioon perustuvat systeemit (Calvert et al. 1982; Weber et al. 1978). Liikkeen nauhoituksessa (motion recording) käytetään useammanlaisia välineistöjä (devices), joilla kinesteettistä dataa liikkuvista kuvista hankitaan. Kinesteettistä dataa käytetään sitten animoitujen kuvioiden kontrollointiin. Sellaiset systeemit rajoittuvat usein laboratorio-olosuhteissa suoritettuihin ihmisen liikkeitä koskeviin mittauksiin, mutta jotka silti tarjoavat potentiaalisesti tarkastellen runsaan ihmisen liikkeitä koskevan data-aineiston. Muodon interpolaatio (shape interpolation), tunnettu myös "metamorfoosina", on kaksiulotteisen avainruudutuksen (keyframing) kolmiulotteinen vastine. Jos vastaavuus on yhden-suhde-yhteen ajateltaessa erillisten objektien kuvapisteitä ja pintoja, voidaan väliintulokuvavuodet (inbetween frames) laskea interpoloimalla objektien välisten datapisteiden keskinäiset välimatkat. Avain-trasformaatiojärjestelmissä (key-transformation systems) manipuloidaan kokonaisia objekteja suhdetransformaatioilla. Väliintulokuvavuodet (inbetween frames) generoidaan interpoloimalla transformaatioparametrit, ja transformoimalla objekteja. Tällaiset systeemit mahdollistavat transformaatiohierarkioiden spesifioimisen ja siten myös artikuloitun liikkeen muodostamisen. Sellaisessa avain-asettumisen -systeemeissä (key-pose systems), kuten BBOP, tuotetaan myös "p-curve facility" (Baecker 1969) niin, että käyttäjä saattaa graafisesti spesifioida nopeuksia. Notaatiooperusteiset systeemit (notation based systems) ovat esimerkiksi tekstiohjauksesta, jossa käyttäjä kuvaa liikkeen koreografisena notaationa, tai sen alfanumeerisina vastineina (i.e. Calvert et al. 1982).

Limitations of guiding systems

Ohjaussysteemissä (guiding systems) animatoijan täytyy ennalta spesifioida liikkeen detaljit. Tämä on järkevää ainoastaan niukkapiirteisissä ympäristöissä. Otaksukaamme, että ihmishahmo kävelee luonnollisella tavalla muotoutuvassa maisemassa. Kävelyaskeleita ei ole vaikea gereroida käytettäessä avainruudutusta (keyframing) tai muodon interpolaatiota (shape interpolation), mutta tässä tapauksessa se vaatii suuren määrän intermediatiivisia konfigurointeja, jotta kävely näyttäisi luonnolliselta. Tämä johtuu siitä, että väliintulokuvavuodut (inbetween frames) lasketaan huomioimatta muita kuvakentän objekteja. Tällöin voi tapahtua, että jalka uppoaa lattiaan, tai kuvio kävelee seinän läpi. Pahempaa on, jos kuvio käveleekin toiseen suuntaan kuin on aiottu, koska silloin mitään aiemmista avainkonfiguraatioista ei voida käyttää. Ohjauksessa animatoijalla on lähes täydellinen kontrolli kuvion liikkeisiin. DOF -ongelmasta johtuen tämä on samalla sekä siunaus, että riesa. Animatoija on vapaa suunnittelemaan ekspressiivisiä liikesarjoja "in toto", mutta mutkikkaampien kuvioden, tai mekanismien tämä on vaativa-, tai kenties mahdotonkin tehtävä, vaikka käytettävissä olisikin hyvin suunniteltu oheislaitteohjattu vuorovaikutekenttä (device-mediated interface) (Lundin 1982). Useimmat ohjaussysteemit (guiding systems) sisältävät ennalta-asetettuja proseduraalisia abstrahointeja liikkeen pehmentämiseen, jotka perustuvat yhteen -tai useampaankin "spline" -tekniikkaan (Rogers and Adams 1976). Usein nämä työkalut sallivat animatoijan säätää "spline" -parametreja, kunnes jokin toivottu profiili on työstetty. Kysweinen tekniikka sallii animatoijan myös tarkemmin simuloida kiinteiden kehojen dynamiikkaa, kuten nopeutumista ja hidastumista inertia- ja painovoiman vaikutuksessa, sillä eihän sellainen liike näytä luonnolliselta, joka on hyppivä, vaikka onkin suuntautumiseltaan lineaarinen, ja sen katseleminen on myös epämiellyttävää. (Tavanomaisessa animaatiossa liikkeen nopeutuminen liittyy ilmaukseen "ease in", ja sen hidastuminen ilmaukseen "ease out", ja se on laskettava käsin, tai käyttäen taulukoita). Yleisesti tarkastellen "splining" tarjoaa soveliaan monenlaisten transformaatioiden nopeuksien kontrollointitavan, kuten kokojen-, muodon-, ja värien muutoksiin liittyviä keinoja, ja myös asettumisiin ja orientaatioon liittyviä kontrolleja. Parametrisoitujen (kuvaus) käyrien käyttämisen arvo animaation kontrolloinnissa todettiin jo aiemmin, ja se on yhä aktiivisen tutkimuksen kohteena (e.g. Kochanek and Bartels 1984). Samalla kun nämä "spline" -käyrät ovat voimakas simulaatiomekanismi, ei tämä tekniikka yksin riitä yleiseksi ratkaisuksi DOF -ongelmaan, koska monien transformaatioiden kontrollointi vaatii useiden kuvauskäyrien generoimista ja hienosäätöä. Tähän mennessä on tuotettu lukuisia mielenkiintoisia animaatioseksenssejä käyttäen ohjaussysteemejä mm. lukuisissa kaupallisissa tuotteissa ja yliopistolaboratorioissa. Koska niissä ei kuitenkaan tarjota voimakkaista abstrahointimekanismeja, ja koska adaptoituvaa liike ei ole lainkaan mahdollista, eivät ohjaussysteemit sovellu mutkikkaiden kuvioden kontrollointiin, ja niiden liikkeiden ohjaukseen komplisoiduissa ympäristöissä. (Voisi tietysti ajatella, että kuvauskäyrät, joita myöten kuvioden eri osat "liikkuvat" kolmiulotteisessa avaruudessa kullekin "annetuilla" nopeuksilla saataisiin "noudattamaan" tiettyjen käyrästönosien osalta mm. kuljettavan pinnan muodostusta, ja että ne ohjelmoitaisiin "väistämään" avaruudessa esiintyviä objekteja niin, että osa "väistöistä" olisi "kovia", ja osa "pehmeitä" -riippuen objektien "laadusta", suom. huom.).

Animator-level systems

On suunniteltu lukuisia animatointitason systeemejä (animator level systems), jotka sallivat animatoijan spesifioida liikettä algoritmisesti. Jotkut näistä systeemeistä tarjoavat jossain määrin sekä adaptoituvan liikkeen- ja abstrahoinnin hyödyntämistä, vaikka niitä ei ole erityisesti suunniteltu henkilöahmoanimaatioon.

GRAMPS, ASAS, and MIRA

GRAMPS-ohjelma (O'Donnel and Olson 1981) ei sisällä adaptoituvan liikkeen manipuloinnin mahdollisuutta, mutta sen avulla voidaan konstruoida liikemakroja (motion macros), jotka perustuvat funktionaaliseen abstrahointiin. Liittymät (joints) voidaan ryhmittää yhteen ja niiden syöttö derivoida ohjausvalikosta (dials), ja liittymien liike voidaan eksplisiittisesti asettaa tietyn arvojakauma-alueen sisälle. Tämä on hyvä esimerkki ohjausmekanismien interaktiosta (dials) ja funktionaalisesta abstrahoinnista (motion macros). Vaikka ohjelmaa ei olekaan suunniteltu henkilöahmoanimaatio-

systemiksi, on sitä käytetty generoitaessa mielenkiintoisia ihmishahmojen animaatioita. Craig Reynoldsin ASAS (Reynolds 1982) tarjoaa valikoiman alemman tason mekanismeja, jotka soveltuvat sekä abstrahointiin -että adaptoituvaa liikkeeseen. Näyttelijäparadigma (actor paradigm) tarjoaa eksplisiittisellä tavalla yleisen abstrahointimekanismin, ja sallii transformatiohierarkioiden määrittämisen (struktuurallisen abstrahoinnin), ja myös käyttäytymyksellisyyden (proseduraalisen ja funktionaalisen abstrahoinnin kautta). Viestinantomekanismi mahdollistaa adaptoituvan liikkeen työkaluohjauksen, koska animatoidut entiteetit raportoivat fysikaalisten attribuuttinsa osalta, kuten myös sisäisten tilojensa suhteen. Toinen äskettäin raportoitu animaatiojärjestelmä MIRA (Magenat-Thalman and Thalman 1983) perustuu lähesisesti näyttelijäparadigmaan liittyvään ohjelmoitparadigmaan, nimittäin data-abstrahointiin (Shaw 1980). MIRA tarjoaa valikoiman tärkeitä abstrahointiominaisuuksia, jotka muistuttavat ASAS:n vastaavia. MIRA ei ole viestienantossysteemiin perustuva, joten animatoija voi asettaa ja tutkia erilaisten datatyypin arvoja ja variaabeja niin, että kuvioita ja objekteja voidaan käyttää generoitaessa liikkeitä.

TEMPUS

Norman Badlerin johtama ryhmä Pennsylvanian yliopistossa on pitkään tutkinut inhimillisten liikkeiden representaatiota ja niiden esittämistä kuvana. He ovat kehittäneet TEMPUS -ohjelman (Koren et al. 1983; Badler 1982), jossa kyetään analysoimaan ja esittämään realististen ihmishahmojen liikkeitä työavaruudessa. Vaikka TEMPUS ei olekaan yleiseen tarkoitukseen valmistettu animaatiojärjestelmä, on siinä pitkälle kehitettyjä piirteitä inhimillisten hahmojen liikkeitä ja muodonmuokkauksia koskien, kuten myös liikeratojen laskennallisen määrityksen kontrolli (resolved motion control). Koska TEMPUS on rajoitettu, päinvastoin kuin MIRA ja ASAS, vain inhimillisten hahmojen asettumiseen ja orientoitumiseen, voi TEMPUS olla pitkälti oheislaitteohjattu (device-mediated). Käyttäjä poimii liikkeitä graafisen näytön luettelosta (menu), ja kontrolloi liikkeitä näyttöön simuloituilla potentiometreillä. Käytettävissä ovat koko hahmon rotaatio ja traslaatio, kuten myös valittujen liittymien (joints) rotointi, ja raajojen laskennallisesti määriteltävä liikkuminen (resolved motion of the limbs). TEMPUS ei sisällä adaptoituvaa liikettä, ja abstrahointimahdollisuudet, joita käyttäjä voi käyttää, ovat myös rajatut parametrattomiin makroihin (parametreless macros), joiden avulla käyttäjä voi ryhmittää liikekomentoja. Joustava laskennallisesti määriteltävä liikealgoritmi asettaessa ihmiskehon raajojen asetoja kolmiulotteisessa avaruudessa on askel kohti tehtävänratkaisutason animaatiota (task-level animation).

Discussion

Koska on mahdollista käyttää välineohjatusti adaptoituvaa liikkeensäätöä, ja määrittää struktuurallisia, funktionaalisia ja proseduraalisia abstrahointeja tarjoavat animointitasojen systeemit (animator level systems) merkittäviä parannuksia ohjaukseen DOF -ongelmaa ajatellen. Mutta kuten tavallista, on köyden pää äkkiä vetävän käsissä (trade off). Ohjaussysteemit (guiding systems) ovat suhteellisen helppoja oppia ja käyttää, mutta eivät riitä mutkikkaampien hahmojen animaatioon. Animointitasojen systeemit (animator level systems) tarjoavat taas toisaalta yleisten ohjelmointikielten käytön kautta tietojenkäsittelyllistä voimaa, mutta samanaikaisesti sitovat käyttäjät ohjelmankehittelyn ongelmiin. Thalman et al. (Magenat-Thalman and Thalman 1983). Huomattakoon, että "otti" peräti 14 kuukautta tuottaa [YKSI] 13 -minuutin pituinen elokuva, mikä selkeästi valottaa ongelmaa. Ongelma on siis siinä, että vaikka onkin mahdollista kehittää kompleksien liikkeiden hallintaa ASAS:lla tai MIRA:lla, se ei välttämättä ole helppoa, koska kumpikaan kieli ei eksplisiittisesti tarjoa tukeaa kehitettäessä funktionaalisia abstrahointimalleja, tai hallitessa adaptoituvaa liikettä. Thalman et al. pitivätkin tärkeänä integroida ohjaava järjestelmä, MUTAN (Fortin et al. 1983), joka liittyi heidän ohjelmaskeemaansa. Palaan integrointikontrollin muotoihin myöhemmin.

Task level animation

Tehtävänratkaisutasolla (task level)) täytyy animaatiojärjestelmän sisältää motoristen ohjelmien suorittamismahdollisuuksia, joilla hahmoja kontrolloidaan, ja näiden motoristen ohjelmien tulee kyetä

itsessään generoimaan tarvittavia asettumisen vektoreita (pose vectors). Jotta tätä voitaisiin niiltä odottaa, on käytettävissä oltava tietoa kuivioista, objekteista ja koko niiden ympäristöstä, joka sisältää tietoja niiden asemasta, fysikaalisista attribuuteista, ja niiden funktionaalisuudesta. Toisessa tutkimuksessani (Zeltzer 1983) luonnehdin erästä lähestymistapaa, jolla ratkaista tehtävänratkaisutason (task level) animaatiota, jossa motorinen käyttäytyminen generoidaan taitojen hierarkiasta (esitettyinä luetteloina (Minsky 1975) tai näyttelijöinä (Hewitt 1979) objekti-orientoituneessa systeemissä) valitsemalla ne sellaisten sääntöjen mukaan, jotka suhteuttavat meneillään olevan toiminnan seuraavaksi esitettäväksi toivottuun toimintaan. Albus The Bureau of Standards -yhtiöstä on suunnitellut robottikontrollisysteemin, joka perustuu hierarkiseen tauluvalintaohjautuviin laskentaelementteihin (table-driven computing system) (Albus 1981). Powers on suunnitellut käyttäytymiskontrolli-hierarkian, joka perustuu servomekanistiseen teoriaan (Powers 1973). Molemmat näistä jälkimmäisistä lähestymistavoista näyttävät toimivan hyvin alemmanasteisessa motorisessa kontrollissa, jollaista me voisimme kutsua vaisto-ohjautuvaksi käyttäytymiseksi (instinct-driven behavior), mutta näyttävät heikoilta, mikäli niiltä vaaditaan symbolista vuorovaikutusta ympäristönsä kanssa. Tehtävänratkaisutason (task level) motorinen kontrolli on vaikea ongelma, jota tutkivat parhaillaan tiedemiehet, robotiikan asiantuntijat, ja tietysti myös ne, jotka ovat kiinnostuneita korkean tason animaatiosysteemeistä. Lähitulevaisuudessa voimme odottaa kehitettävän yksinkertaisten käyttäytymisten generointiin kykeneviä prototyyppisiä systeemejä. Se, kuinka nopeasti sellaiset systeemit laajenevat riippuu siitä, kuinka hyvin opimme ymmärtämään motorista kontrollia itseään. Tehtävänratkaisutason (task level) kontrollissa animatoija voi ainoastaan spesifioida erityisten liikkeiden suhteen laajoja toiminnan linjoja, ja animaatiosysteemi on kuitenkin täynnä hienoja yksityiskohtia. Se, onko tällainen lähestymistapa sovelias riippuu erityisestä soveltamisesta. Asiantuntevan voi hyvinkin tyytyä asetusarvoliikkeiden (default movements) käyttöön, ja kuvioihin, joita systeemi tarjoaa huomioiden kohtuullisen ajankäytön ja kustannukset, ja siten animointi hakee jonkinlaisen kultaisen keskitien. Erittäin asiantunteva käyttäjä esimerkiksi esittävän taiteen piirissä saattaa haluta lähes täydellistä hienojenkin nyanssien kontrollia liittyen hahmojen liikkeeseen muodostaakseen kuvasekvensseistä ilmaisullisesti niin rikkaita kuin mahdollista. Kuitenkaan halu liikkeen ilmaisullisten laatuominaisuuksien kontrollointiin ei tarkoita pelkästään puhdasta systeeminohjausta generoimalla asettumisvektoreita (pose vectors). Animatoijan on kyettävä saavuttamaan eritasoisia kontrollihierarkioita generoidakseen uudenlaisia motorisia taitoja, tai muotoillakseen jo hallittuja taitoja.

Integration of control modes

Ohjaus (guiding) on vallitseva nykyisten interaktiivisten animaatiosysteemeiden muoto. Tarve yhdistää kaikkien kolmen tason muodot juontuu siitä, että erillisinä käytettyinä ne eivät, taloudelliselta kannalta ajatellen, tarjoa riittävää kontrollia. Ohjaus (guiding) soveltuu parhaiten hienojen yksityiskohtien käsittelyyn, mutta sen avulla ei voida kontrolloida komplekseja liikkeitä. Animointitason (animator level) ohjelmointi on voimakas, mutta vaikea hallita. Tehtävänratkaisutason (task level) systeemit mahdollistavat kompleksien liikkeiden kontrollin sulkeistamalla liikkeen detaljit ohjauksellisuuden ulkopuolelle. Osin mahdollinen ratkaisu piilee soveltamalla ohjaus (guiding) tekniikkoja liikkeen kontrolliin hierarkian erityisissä kohdissa. Avaimena on kyky jäsentää liikkeen repertoaari käsiteltävissä olevaksi, hierarkisesti organisoituneeksi motoristen taitojen kompositioksi. "Browsers" -notaatio, kuten sitä käytetään työkaluna "Smalltalk" issa (Tesler 1981) tai "Loops" issa (Stefik et al. 1983) tarjoaa voimakkaan metodin, jolla saavutetaan motoristen taitojen ohjauskontrolli. Otaksukaamme, että minulla on RGB monitorissani varjostettu ihmisen hahmo. Terminaalinäytöksäni minulla on representaatio, josta ilmenee hahmon struktuuri ja sen osaamat taidot. Otaksukaamme, että valitsen kuvauskäyrän grafiikkataulukosta. Jos nyt spesifioin tämän kuvauskäyrän niin, että se edustaa erityistä liittymän rotaatiota (joint rotation) l. osoitan sen kuvaavan pikkusormen liikettä terminaalinäyttövalikostani, niin minun tulisi välittömästi nähdä ihmishahmon sormen liikkuvan adekvaatilla tavalla RGB näytössä. Otaksukaamme edelleen, että osoitan terminaalinäyttövalikosta toimintaa "tavoittele jotakin vasemmalla kädelläsi", tulisi minun nähdä RGB näytön ihmishahmon avaavan ja sulkevan sormiaan nopeudella, jonka olen määrittänyt. Annan sitten valikosta toimintakäskyn "kävele", jolloin käskyn seurauksena ihmishahmo kävelee näyttöruudun poikki, ja tällä kertaa kuvauskäyrä, jonka olen jo edellä valinnut, voisikin kuvata kävelynopeutta. Tällainen modulinen, hierarkisoituva organisaatio sallii käyttäjän indentifioida ne liikevaliteetit, joita tulee säätää, ja samalla lokaisoida muutosten tuottamat vaikutukset. Tällainen on yhdenmukaistuva motoristen taitojen representaatio, johon kuuluu taitojen esittäminen, ja

niiden säätöjen spesifioinnin, jotka ovat mahdollisia, ja myös lisäksi yhdenmukaisen mekanismin, jossa p-käyrät interaktoituvat taitojen kanssa.

Conclusion

Olen esittänyt käsitteellisen analyysin kolmidimensionaalisesta tietokoneanimaatiosta simuloitavien objektien ja niiden käyttäytymisten prosessointina, jossa animatoija manipuloi niitä mikromaailmassa. Vapausasteiden ongelma (degrees of freedom) on keskeinen kysymys artikuloitujen kuvioiden koordinoinnissa. Tietokoneanimaatiosysteemien tulisi perustua soveltuviin peruskäsitteisiin, kuten adaptoituvaan liikkeeseen (adaptive motion), ja viiteen abstrahointimekanismiin, joilla animatoija voi manipuloida mielenkiintoisia piirteitä ja ympäristöjä ekspressiivisellä tavalla. Keskustelu kolmesta kontrolloinnin moodista viittaa hyvän ohjauksen (guiding) kriteeriin ja animointitason systeemiin (animation level systems). Ohjaussysteemit (guiding systems) ovat tänä päivänä saaneet eniten huomiota osakseen, ja interaktiivisia, oheislaitteohjattuja (device-mediated) vuorovaikutekenttiä (interface) on käytetty pelkästään standardilla tavalla kommunikoidessa tietokoneen kanssa, jota todistaa "hiiriohjattujen ikkunoiden" (mouse-and-window) käyttö tietokoneen ohjauksessa. Yleisesti tietokoneen ohjaus tulisi nähdä mekanismina, jolla kehitetään ja kontrolloidaan kompleksien systeemien käyttäytymistä, eikä pelkästään keinona poimia vaihtoehtoja, piirtää viivoja, tai generoida skaalariarvoja erilaisia transformaatioparametreja määriteltäessä. Kuten edellä jo on ehdotettu, me haluaisimme kyetä tavoittamaan fyysisen oheislaitteen kautta syötetyn tiedon monitoritulosteen muodossa, jossa käyttäytymykselliset hierarkiat (behavioral hierarchy) tulostuisivat erilaisina ilmaisutasona. Samalla kun hahmon merkitys riippuu yksittäisestä prosessista, joka sitä kuvaa, on vaikeaa löytää soveliaita parametreja kompleksien prosessien kontrollointiin, kuten kasvoin liittyvien ilmausten kontrolloiminen (cf. Parke 1982; Platt 1981). Kun on määritelty luonnollinen kontrollointivalikko, ei ole vaikeaa käyttää syöttöoheislaitteita (input devices) parametrien arvojen integratiiviseen generointiin. On olemassa kaksi toisiaan täydentävää suunnittelun teemaa: Kuinka me voimme "sukeltaa" ohjausmekanismeihin niin, että voimme tuottaa kompleksia käyttäytymistä? Kuinka syöttöoheislaitteet (input device modules) voivat toimia "hahmovahvistuksen" standardeina, joita helposti voidaan uudelleensuunnata kuvioidenkontrollihierarkian eri funktioihin? Animointitason (animator level) kielten tulisi sisältää muotoiluominaisuuksia ja piirteitä, joita odotamme voimakkailla ohjelmointikieliltä. Kätkemällä käyttäjältä ohjelmoinnin avulla tehdyt ratkaisut, tai päällelystämällä syntaksi sokerilla ei ole läheskään niin tärkeää kuin tarvittavan animaation ekspressiivisen voiman kehittäminen. Tämä ei ole oikea paikka keskustelulle automaattisen ohjelmoinnin tulevaisuudesta, tai viimeisimpien ohjelma-paradigmojen ansioista, koska tärkeintä on käyttäytymisen algoritminen kuvaus; "tee tämä, tee tuo" on olennaisin ja fundamentaalisin tapa kommunikoida suoritettavaan liikkeeseen liittyvästä toiveesta. Usein "naiivi käyttäjä" oppii nopeasti animaatiokielen syntaksin frustroitumalla siitä, ettei kieli olekaan riittävän voimakas. Animointitason (animation level) kielten ja systeemien tulisi siten kombinoida yhteen kaikkea sitä, mitä ohjelmateknologiasta tiedetään -liittyen mekanismeihin, joilla abstrahoidaan funktionaalisuutta ja kontrolloidaan adaptoituvaa liikettä. Lopuksi: Adaptoituva liike on eräänlainen muoto törmäystesteistä (collision testing), ja siten laskentaperusteisen liikekontrollin tulisi sisältyä ainakin osittain peruselementteinä kolmiulotteiseen animaatiosysteemiin. Kolmiulotteisen tietokoneanimaation taide ja tiede jatkaa kehittymistään kohti hypoteettisten maailmojen simulointia, täydennettynä fysiikan laeilla ja kuvioiden hallitsemilla käyttäytymisillä repertoareilla. On opittava konstruoida ja kontrolloida näitä simulaatioita, joilla tietokoneanimaatiolle annetaan sen ekspressiivinen voima.

References

- [1] J.S. Albus, Brains, Behavior and Robotics, Byte Books, Peterborough, NH (1981).
- [2] N.I. Badler, Design of a Human Movement Representation Incorporation Dynamics, Course Notes, Seminar on Three-Dimensional Computer Animation, ACM SIGGRAPH 82 (1982).
- [3] R.M. Baecker, Picture-driven Animation, Proceedings AFIPS Spring Joint Computer Conference, vol 34, pp. 273-288 (1969).
- [4] J.F. Blinn, Systems Aspects of Computer Image Synthesis, Course Notes, Seminar on Three Dimensional Computer Animation, ACM SIGGRAPH 82

(1982).

- [5] N. Burtnyk and M. Wein, Interactive Skeleton Techniques for Enchanting Motion Dynamics in Key Frame Animation, *Commun ACM* 19:564-569 (1976).
- [6] T.W. Calvert, J. Chapman, and A. Patla, The Integration of Subjective and Objective Data in the Animation of Human Movement, *Proc ACM SIGGRAPH 80, Comput Graphics* 14:198-203 (1980).
- [7] T.W. Calvert, J. Chapman, and A. Patla, Aspects of The Kinematic Simulation of Human Movement, *IEEE Computer Graphics and Applications*, 2:41-50 (1982).
- [8] E. Catmull, The Problems of Computer-Assisted Animation, *Proc ACM SIGGRAPH 78, Computer Graphics*, 12:348-353 (1978).
- [9] R. Chuang and G. Entis, 3-D Shaded Computer Animation -Step-by-Step, *IEEE Computer Graphics and Applications*, 3:18-25 (1983).
- [10] J.H. Clark, Hierarchical Geometric Models for Visible Surface Algorithms, *Comm ACM*, 19:547-554 (1976).
- [11] F.C. Crow, A More Flexible Image Generation Environment, *Proc ACM SIGGRAPH 82, Computer Graphics*, 16:9-18 (1982).
- [12] J. Denavit and R.B. Hartenberg, A Kinematic Notation for Lower-Pair Mechanics Based on Matrices, *J Appl Mech*, 23:215-221 (1955).
- [13] D. Fortin, F.J. Lamy, and D. Thalman, A Multiple Track Animator System for Motion Synchronization, *Proc ACM SIGGRAPH/SIGART Workshop on Motion*, pp. 180-186 (1983).
- [14] W.R. Franklin, 3-D Geometric Databases Using Hierarchies of Inscribing Boxes, *Proc Canadian Society for Man-Machine Interaction*, pp. 173-180 (1981).
- [15] H. Fuchs, Z. Kedem, and B. Naylor, On Visible Surface Generation by A Priori Tree Structures, *Proc ACM SIGGRAPH 80, Computer Graphics*, 14:124-133 (1980).
- [16] C. Ginsberg and D. Maxwell, Graphical Marionette, *Proc ACM SIGGRAPH/SIGART Workshop on Motion*, pp. 172-179 (1983).
- [17] M. Girald and A.A. Maciejewski, Computational Modeling for the Computer Animation of Legged Figures, *Proc ACM SIGGRAPH 85, Computer Graphics*, 19:263-270 (1985).
- [18] J.E. Gomez, Twixt: A 3-D Animation System, *Proc Eurographics '84, North-Holland* (1984).
- [19] C. Hewitt, Control Structure as Patterns of Message Passing, In: R.H. Brown (ed.) *Artificial Intelligence: an MIT Perspective*, MIT Press, Cambridge, MA, pp. 433-465 (1979).
- [20] A. Kay and A. Goldberg, Personal Dynamic Media, *Computer*, pp. 31-41 (1977).
- [21] C. Klein and C. Huang, Review of Pseudoinverse Control for Use with Kinematically Redundant Manipulators, *IEEE Transaction on Systems, Man, and Cybernetics*, Vol. SMC 13, pp. 245-250 (1983).
- [22] D.H.U. Kochanek and R.H. Bartels, Interpolating Splines with Local Tension, Continuity, and Bias Control, *Proc ACM SIGGRAPH 84, Computer Graphics*, 18:33-41 (1984).
- [23] J. Korein, J. Korein, G. Radack, and N. Badler, *TEMPUS User Manual* (Unpublished), Dept. of Computer and Information Science, University of Pennsylvania, Philadelphia, PA (1983).
- [24] C.S.G. Lee, Robot Arm Kinematics, Dynamics, and Control, *Computer*, 15:62-80 (1982).
- [25] T. Lozano-Perez, *Robot Programming*, AI Memo 698, MIT, Cambridge, MA (1982).
- [26] D. Lundin, 3-D Modeling, A Personal Orthodoxy, Course Notes, Seminar on Three Dimensional Computer Animation, *ACM SIGGRAPH 82* (1982).
- [27] N. Magenat-Thalman and D. Thalman, The Use of High-Level 3-D Graphical Types in the Mira Animation System, *IEEE Computer Graphics and Applications*, 3:9-16 (1983).
- [28] M. Minsky, A Framework for Representing Knowledge, In: P. Winston (ed.) *The Psychology of Computer Vision*, Graw-Hill, New York (1975).
- [29] T.J. O'Donnel and A.J. Olson, GRAMPS -A Graphics Language Interpreter for Real-Time, Interactive, Three-Dimensional Picture Editing and Animation, *Proc ACM SIGGRAPH 81, Computer Graphics*, 15:133-142 (1981).
- [30] F.I. Parke, Parametrized Models for Facial Animation, *IEEE Computer Graphics and Applications*, 2:61-68 (1982).
- [31] R. Paul, *Robot Manipulators: Mathematics, Programming, and Control*, MIT Press (1981).
- [32] S.M. Platt and N.I. Badler, Animating Facial Expressions, *Proc ACM SIGGRAPH 81, Computer Graphics*, 15:245-252 (1981).

- [33] W.T. Powers, Behavior: The Control of Perception, Aldine Publishing Co., Chigago (1973).
- [34] C.W. Reynolds, Computer Animation with Scripts and Actors, Proc ACM SIGGRAPH 81, Computer Graphics, 16:289-296 (1982).
- [35] E.A. Ribble, Synthesis of Human Skeletal Motion and the Design of a Special-Purpose Processor for Real-Time Animation of Human and Animal Figure Motion, M.S. Thesis, The Ohio State University (1982).
- [36] D.F. Rogers and J.A. Adams, Mathematical Elements for Computer Graphics, McGraw-Hill, New York (1976).
- [37] S. Rubin and T. Whitted, A 3-Dimensional Representation for Fast Rendering of Complex Scenes, Proc ACM SIGGRAPH 80, Computer Graphics, 14:110-116 (1980).
- [38] M. Shaw, The Impact of Abstraction Concerns on Modern Programming Languages, Proc of the IEEE, 68:1119-1130 (1980).
- [39] M. Stefik, D. Bobrow, S. Mittal, and L. Conway, Knowledge Programming in Loops: Report on an Experimental Course, AI Magazine, 4:3-13 (1983).
- [40] I.E. Sutherland, Sketchpad: A Man-Machine Graphical Communication System, Proc AFIPS Spring Joint Computer Conf., 23:329-346 (1963).
- [41] R.D. Tennent, Priciples of Programming Languages, Prentince-Hall, Englewood Cliffs, NJ (1981).
- [42] L. Tesler, The Smalltalk Enviroment, Byte, 8:90-147 (1981).
- [43] F. Thomas and O. Johnston, Disney Animation: The Illusion of Life, Abbeville Press, New York (1981).
- [44] M.T. Turvey, H.L. Fitch, and B. Tuller, The Problems of Degrees of Freedom and Context-Conditioned Variability, In: J.A.S. Kelso (ed.), Human Motor Behavior, Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 239-252 (1982).
- [45] K. Wasserman, Physical Object Representation and Generalization, AI Magazine, 5:28-42 (1985).
- [46] L. Weber, S.W. Smoliar, and N.I. Badler, An Architecture for the Simulation of Human Movement, Proc ACM Ann. Conf., pp. 737-745 (1978).
- [47] D.E. Whitney, The Mathematics of Coordinated Control of Prosthetic Arms and Manipulators, Transactions of the ASME, J Dynamic Systems, Measurement, and Control, 122:303-309 (1972).
- [48] L. Williams, BBOP; Course; Notes; Seminar on Three Dimensional Computer Animation ACM SIGGRAPH 82 (1982).
- [49] D. Zeltzer, Knowledge-based Animation, Proc ACM SIGGRAPH/SIGART Workshop on Motion, pp. 187-192 (1983).
- [50] D. Zeltzer, Representation and Control of Three Dimensional Computer Animated Figures, Ph. D. Thesis, The Ohio State University (1984).

News and Movies in the 50 Megabit Living Room

Andrew Lippman and Walter Bender
MIT Media Laboratory

suom. Timo Kinnunen 24.2.1989

Pelastettu 3.5 Levykkeeltä NU.EXE Ohjelmalla 12.4.1989 T. Kinnunen

Muotoiltu uudelleen 1.10.1997 T. Kinnunen

Abstrakti

Tässä artikkelissa kuvataan kahta demonstratiivista systeemiä joissa tietojenkäsittelyä sovelletaan informaationkäsittelyyn ja esittämiseen niitä kanavia käyttäen, joita on käsillä tavallisessa olohuoneessa. Ensimmäisessä niistä esitellään ohjelmoitavaa ja yksilöllistä sanomalehteä, jollaisen ryhmä tietojenkäsittelyn automaattisesti työskentelevää "toimittajaa" luovat päivittäin lukijan toiveiden mukaisesti. Toisessa esitetään kuinka luodaan digitaalisesti kompressoituja elokuvia käyttäen muistiyksiköjä niin, että muokkaus tapahtuu itse katselun aikana. Molemmissa projekteista tutkitaan kommunikaatiokanavien persoonallistamisen mahdollisuuksia niin, että tarvittava tietojenkäsittely tehtäisiin terminaalipääteissä. Vaikka ne toimivatkin tänä päivänä alemmilla taajuuksalueilla voidaan niiden välitön muokkauskyky nostaa korkeammalle siirtymällä ylemmille lähetystaajuuksille. Niinpä esitetyt mallit ovatkin demonstratiivisia esityksiä korkealla lähety- ja käsittelytaajuuksella kotona tapahtuvasta muokkauksesta.

1. Introduction

MIT Media Laboratoryn Electronic Publishing ryhmä on viime vuosien aikana tutkimustensa teemana

pitänyt traditionaalisten massamediakanavien persoonallistamista liittyen tiedonsaantiin. Näiden kanavien joukossa ovat olleet radio, televisio, ja kirjoitettu media, kuten sanoma- ja aikakauslehdet, sekä kirjat. Hypoteesina on, että tietojenkäsittelyä sovellettaisiin informaatiota luotaessa, sekä sitä vastaanottaessa. Useita lähetystapoja voidaan saattaa keskenään dialogiin. Ehkä tärkeintä edellisessä on se, että tietojenkäsittelyllä voidaan prosessoida ennemminkin sisällöllisiä ominaisuuksia kuin signaalia sinällään. Tässä keskustelussa mediat voidaan jakaa kolmeen luokkaan: Suoraan vastaanottoon perustuvat (directly accessible), epäsuoraan vastaanottoon perustuvat (indirectly accessible), ja tietojenkäsittelyllä tuotettuun vastaanottoon perustuvat (computationally accessible). Ensimmäisessä tapauksessa, jossa esimerkkinä painettu teksti, liittyy vastaanottoon jakelujärjestelmä, jonka kautta tuotteet ovat suoraan vastaanottajan saatavissa ilman käännoästä tai prosessointia. Mahdollisesti silmäläsejä lukuunottamatta esimerkiksi kirjan lukemiseen ei tarvita mainittavampia apuvälineitä. Epäsuoran vastaanoton saannissa, kuten television- ja radion tapaisten välineiden kautta tapahtuvassa tarvitaan väline, joka muuntaa lähetys-signaalin hyödynnettävään muotoon. Televisio ja radio ovat laitteina tällaisia välineitä. Kehitettäessä epäsuoraan vastaanottoon perustuvia välineitä on kokeiltu lukuisia koodituksen algoritmeja, jotta kanavia voitaisiin hyödyntää tehokkaammin, ja nostaa presentaation tasoa korkeammalle ainakin ajoittain. Useimmissa lähetyssovelluksissa kanavan käytön tavan voi sanoa viittaavaan laatutasoon. Lähes näihin päiviin saakka kanavan koodaus on ollut analyttistä, mikä tarkoittaa sitä, että lähetysdekoodaus on ollut signaalin prosessointia, eikä informaation prosessointia. Uusimmat adaptiiviset systeemit ovat hämärtäneet eroa signaalin enkoodauksen ja sisällönmuokkauksen välillä. NTSC on eräs varhaisimmista esimerkeistä psykofysikaalisesta optimoinnista; adaptiiviset systeemit, jossa kytketään televisiolähetys progressiivisesti skannautuvaan vastaanottomenetelmään on jo myöhempi tapaus. On ollut vaikeuksia suunnitella sellaista dekooderia, joka kykenisi työskentelemään kyllin laaja-alaisesti erilaisilla signaaleilla, josta syystä lähetysstandardit ovat nykyisenlaiset. Tietokonesysteemeissä terminaalisuunnittelijan päämääränä taas on ollut nimenomaan tuottaa keinoja hyödyntää mahdollisimman monia syöttömuotoja (input format), muodostavat televisio- ja radiolähetykset traditionaalisissa muodoissaan rajoittuneisuuden verrattuna välineisiin, jotka voivat valmistajasta riippumatta ottaa vastaan mitä tahansa. Kolmantena kategoriana ovatkin tietojenkäsittelyn avulla vastaanotettavissa olevat systeemit. Niille karakteristista on se, että sekä lähetyksessä, että vastaanotossa sisältöjä voidaan muokata, ja että sekä kooderi -että dekooderi ovat suunnitellut niin, että niissä huomioituu nimenomaan sisältö, sekä myös se, mitä vastaanottaja tahtoo sisällöstä hyödyntää. Työssämme otaksumana on ollut, että prosessointi, jolla persoonallistettua informaatiotulosteita tuotetaan, ei välttämättä ole liian kallista tai komplisoitua käyttöönotettavaksi. Hypotetisoimme joitakin rooleja, joita persoonallisella tietokoneella tulevaisuudessa voisi olla, ja myös piirteitä, joita sillä on. karakterisoimme niistä muutamia myöhemmin. Pyrimme työssämme välttämään extrapoloimasta potentiaalisesti käyttöönotettavissa olevia taajuuksia minkään kommunikaatioväylän kohdalla. Jotkut systeemeistä hyödyntävät puhelinliikennettä ja televisiolähetysten järjestelmiä, ja toiset optisten tallenteiden käyttöä korkeampien taajuusalueiden kanavien käytön simuloimisessa. Pyrimme myös välttämään taajuusalueen kompression ja saatavissa olevain kanavakapasiteettien välistä dikotomiaa. Ne eivät toimi toisiaan poissulkevassa mielessä, koska jonkun niistä läsnäolo ei sulje muita välttämättä käytettävyyden ulkopuolelle. Tässä tutkimuksessa kuvaamme kahta järjestelmää, jotka ovat potentiaalisesti mahdollisia käyttää tulevaisuuden 50 megabitin olohuoneessa. Ensimmäinen niistä on elektroninen sanomalehti (NewsPeek), jossa on mahdollista persoonalliseen tietojenkäsittelyyn niin, että käyttäjää voidaan pitää osallisena koostettaessa päivän uutismateriaalia. Toinen esimerkeistä niveltyy kuvanprosessointiin sen esityksen aikana; aiemmin tallennettujen kuvien varastoa käytetään luotaessa uusia muotoja. Tätä ohjelmaa kutsumme nimellä "Paperpack Movies", ja esimerkkinä sellaisesta on digitaalisesti enkoodaten tallennetut kokoillan elokuvat optisille levykkeille. Ohjelman nimi liittyy elokuvien teon kustannuksiin: originaali elokuva on huomattavasti halvempi kuin siitä tehty kopio. Digitoituja, toisiaan seuraavia kuvia voidaan tallentaa massamuisteihin, siirtää puhelinlinjoja myöten, tai kaapelitelesioverkossa. Niitä voidaan käyttää elokuvallisina, muokata niistä valokuvia, tai esittää elokuvan kuvia jossakin muussa järjestyksessä kuin mitä tallenteessa on laita, jos enkoodataan ei-reaaliaikaisesti. Kummatkin esitetyistä vaihtoehdoista toimivat matalassa taajuuskaistaympäristössä, jossa osa prosessoinnin energiasta tarvitaan taajuuskaista-alueen muokkaamiseen huomioiden vastaanottotaajuuden ja lähetystaajuuden. Suurin osa energiasta hyödynnetään individuaalisissa presentaatioissa. Esimerkit eivät niinkään ole esimerkkejä koodauksesta, vaan siitä, miten informaatiota on mahdollista jakaa, kun terminaalivälineet ovat aktiivisesti osallisina informaationsiirrossa. Mikäli kyseessä olisi 50 megabitin olohuone, olisi vaikutus huomattavasti välittömämpää. Elektroninen

sanomalehti, jota kuvaamme, voi operoida minkä tahansa aallopituuden kanavilla. Se vaatii korkeaa taajuusalueetta prosessorin ja lukijan välillä, mikä taas ei ole tarpeen prosessorin ja uutislähteen välillä. Työmme on siksi relevantti ajatellen tulevaisuuden puhelinsysteemejä ja saattaa auttaa ohjaamaan niiden sovellusalueiden kehittämistä. Molemmat esimerkkitapauksista toimivat korkean taajuskaistan alueella kun on kyse kommunikaatiosta käyttäjän ja paikallisen prosessorin välillä. Suuressa määrin on kyse inhimillisten tiedonvälitystapojen eksploraatioista informaatiojärjestelmien suhteen. Ne perustuvat korkealuokkaiseen kuvalliseen esitykseen, joihinkin suhteellisen tehokkaiisiin ja integroituneisiin vastaanottovälineisiin, sekä suureen paikallisesti käytettävissä olevaan massamuistiin. Ne esitetään tässä kokeina ja esimerkkeinä, ja potentiaalisina suuntaa-antavina mahdollisuuksina, joiden suuntaan tulevaisuuden kehitys voisi ohjautua.

2. Electronic News

Tässä luvussa kuvaamme kotimediaa koskeneita kokeita, joissa tutkitaan uudenlaista lähetystapaa: datan jakoa ja sen ohjelmointia pikemminkin kuin ennalta tietynmuotoiseksi pakatun materiaalin mekaanista vastaanottoa. Tällaista lähetystä ei suunnata ihmisille sinällään, vaan tietoja käsittelevälle agentille, joka työskentelee hänen puolestaan. Sekä lähetysten lähettäjän -että lukijan instruktioiden pohjalta tämä editoi dataa, ja esittää sen sitten tavalla, joka on tuttu traditionaalisista medioista. Jokaisessa kokeessa otaksumana oli, että on olemassa jokin datakanava ei-annetussa tietojenkäsittely-ympäristössä, jossa on tarvittava I/O. Sana "ei-annettu" viittaa tässä siihen, että kun vastaanotettu data on dekodattu, on olemassa riittävästi sylinteritilaa jotta datan "sisältöjä" on mahdollista käsitellä. Kokeiden välineenä on sellainen uutisia vastaanottava systeemi, jossa uutisten lähettäjien suorittama editointi korvautuu henkilökohtaisen tietokoneen avulla suoritetulla uutiseditoinnilla. Suuri joukko databaseja, niin paikallisia kuin lähetettyjäkin, passiivisia ja interaktiivisia saatetaan "reporttereille", ohjelmallisiin vuorovaikutusyhteemiin, jotka on nimetty ohjelmoitu uutisten kokoamista silmälläpitäen. Ideaalisessa muodossaan "ne" ovat laajakaistaisesti toimivia niin, että ne ovat valppaina jokaisella kanavalla ja lukevat jokaista sanomalehteä. Ne databaset, joita me käytämme sisältävät kaapelipalveluja; XPressin; auditiiviset uutispalelut Dow Phonelta; lehtiä; lukemistoja ja aikakauslehtiä Dow Jones Newsilta; Mead Data Central NEXISin, televisioiden uutislähetysiä; elektronista postiliikennettä, ja paikallista arkistoitua materiaalia, kuten 20 000 optiselle videolevyllä tallennettua värivalokuvaa (Associated Press). Tietokoneen editorin roolina on etsiä, suodattaa, ja formatoida sille syötettyjä databaseja lukijan toivomusten mukaisesti; valita "printattavaksi" soveltuvaa uutismateriaalia vastaanottajan ohjeiden mukaisesti, ja niputtaa ne niin, että niiden käsillesaatavuus paranee ja niitä voidaan analysoida. Tuloksena on persoonallinen uutispresentaatio, joka on asynkroninen suhteessa alkuperäiseen lähetykseen. Joissakin kokeissa elävät lähetykset on taitettu suoraan sivuiksi lähetetyistä tiedostomateriaalista, kun taas toisissa kokeissa lähetykset on tallennettu ensin levyllä tai nauhalle, joista on otettu peräkkäisiä otantoja. Molemmissa tapauksissa on tehty ero "elävän materiaalin" ja referoidun materiaalin välillä. Kuvatunkaltainen vastaanotettu data on multi-media, josta valintoja suoritetaan pikemminkin sisältöä huomioimalla kuin että vastaanotto olisi pelkkä tiedonsiirron tavanomainen media. Tämä vaatii vastaanotetun materiaalin sivuttamista niin, että materiaali niveltyy toisiinsa. Käytetyn välineen hallitsevat formatointimuodot ovat joustavia. Esittävä medium voidaan irroittaa lähetysvirrasta. Otetut kopiot voivat olla "kovia" tai "pehmeitä": kirjoitettuja, oraalisia ja kuvallisia, joista kuvat voivat olla yksittäiskuvia tai liikkuvia. Se, mitä näistä aineistoista korostetaan, riippuu opituista preferensseistä, fyysisistä rajoitteista vastaanottoon, ja kustantajan antamista ohjeista. Useimmissa tapauksissa presentaatioiden tavat ovat ekstrapolaatioita olemassaolevista mediatyypeistä: sanomalehdistä, aikakauslehdistä ja televisiouutisista.

2.1. NewsPeek

NewsPeek on interaktiivinen database-vuorovaikutuskenttä, jonka esitystapa muistuttaa sanomalehteä. Se on suunniteltu helpottamaan lukijaa tutkimaan multi-media databasea itse ilman, että lähettäjä vaikuttaa asiaan (non-directive manner). Siinä ei ole mieleöntä kysyä epämääräisiä tai huonosti muotoiltuja kysymyksiä täsmennettäessä oman mielenkiinnon kohdetta lukemisen aikana. Käytetään kahta mekanismia, jotka perustuvat lukijan tarpeisiin valikoida tulosteiden saantityyppiä vastaanotetusta

materiaalista: relaatiotabaset kertomuksista käännetään paikallisesti niin, että presentaatio ei ole sarjallinen, vaan rinnakkainen.

2.2. Programmable agents

Käytössä individualisoitu sanomalehti luodaan ajoittain uudelleen käyttämällä joukkoa tietojenkäsittely-agentteja, jotka jäljittelevät niitä toimintoja, joita reporterit ja taittajat suorittavat luodessaan painettua lehteä. Agentit skannaavat vastaanotettavaa ja varastoitua uutismateriaalia tiettyjen "artikkelinimikkeiden" alle, joiden on oletettu kiinnostavan lukijoita. Yleisesti, kuten useimmissa database-hauissa, löydetään useampia artikkeleita, joista on mahdollista esittää; siksi agenttien työn olennaisin osa koostuu erityisen painoarvon omaavien artikkelinimien esittämisestä etusivulla. Tämän vuoksi on kiinnitetty erityistä huomiota tämän sivun taitolliseen asuun, jotta artikkelien valinta olisi helppoa.

2.2.1. Keyphrase selection

Alunperin agentit valittiin käyttämällä sellaisia avainsanoja (keywords), joilla tiedettiin olevat lukijan kannalta mielenkiintoarvoa. Käytimme mallia, jonka Nexis on luonut, jossa kaikki niistä kertomuksista, joita systeemi vastaanottaa, tallennetaan invertoiduksi luetteloksi paikalliseen järjestelmään. Tämä avainsanaluettelo (dictionary) indeksoidaan kaikilla niillä sanoilla, joita ei ole määritetty "kohinaksi" (noise words); muokattaessa uutisaineistoa tämä luettelo on samalla referentiaalisissa suhteissa alkuperäisiin artikkeleihin, ja siinä on kuvailevia elementtejä, jotka auttavat artikkelien tärkeisyyden määrittelyssä koostettaessa niistä kokonaisuutta.

Table 1: Word Dictionary Entries

WORD: The ascii representation of the word or phrase.

WEIGHT: The relative importance of this word, calculated from the fields described below.

TIME 1: A timestamp of story most recently access to the data base which incorporated this word or phrase.

TIME 2: A timestamp of story most recently access from the data base which incorporated this word or phrase.

COUNT: The number of occurrences of this word.

FLAGS: An indication of special occurrences: headline, byline, dateline, first paragraph, search keyword, most frequent word.

LINK: A symbolic link to any synonyms.

LIST: A list pointers to each story, picture or sound where the word occurs.

Tämän avainsanaston avulla on mahdollista skannata kertomuksia, ja laskea niiden mielenkiintoarvo lukijalle, mikä riippuu artikkelin ajankohtaisuudesta ja niiden keskeisyysarvoista. Kun artikkelien keskeisyydet on laskettu, on niiden kategoriat määrätty. Mikäli kertomus tärkeistyy etusivulle, tämä kategoria määrää sen, mille sivun palstalle se asettuu. Nykyisessä ohjelmaversiossa kertomukset ovat topikaalisesti suhteutettu toisiinsa, ja siksi materiaalinvalinta perustuu siihen, mistä lähteestä se on peräisin. Palstoista mainittakoon teknisiä,- poliittisia,- lainsäädännöllisiä, ja populaareja teemakokonaisuuksia kokoavat kategoriat. Aineiston valikointi riippuu myös, millaisessa lehdessä tai lukemistossa ne on alunperin julkaistu.

2.2.2. Layout: The Front Page

Kun artikkeliervalinta jokaiselle palstalle on tehty, määrää presentaationmuokkain sen, miten ne

taitetaan. Etusivun taittotapa ohjautuu seuraavien sääntöjen mukaan:

* Palstat pysyvät samoissa kohdin päivästä päivään. Palstan järjestymisen määritellään suunniteltaessa systeemiä, ja se jää sellaiseksi jatkossakin. Siten aivan kuten tavanomaisessa sanomalehdessäkin samantyyppiset artikkelit asettuvat samalle alueelle,

* palstojen koot vaihtelevat riippuen siitä, kuinka paljon tärkeitä artikkeleita on saatavilla annetuissa kategorioissa. Mikäli artikkeleita on vähän tietyllä palstalla, kutistuu se suhteessa toisiin palstoihin.

2.2.3. Layout: Page 14

Lehden sisällöstä esimerkkinä olkoon arkkityyppisesti sivu 14. Kun kertomus jatkuu sivulta sivulle, se jatkuu täyden sivun eräänä palstana. Sivun vasemmassa reunassa esitetään siihen suhteessa olevat kuvat, kuten kartat, liitetiedostot, ja filmit. Lehden lukija voi käyttää näitä tarpeidensa mukaisesti hyväkseen. Sivun oikeassa reunassa esitetään luettavaa kertomusta tukevat ja sitä täydettävät kertomukset. Aivan kuten tavanomaisessa sanomalehdessäkin ovat toisiaan tukevat kertomukset samalla aukeamalla. Korrelaatiotaulun avulla on mahdollista saada näyttöön myös niitä kertomuksia, joilla on yhteyttä esitettyyn kertomukseen.

Table 5: Correlation Table

STORY: A pointer to story.

WEIGHT: A relative correlation between each of the stories. This is calculated from the Story Entries Table 2.

Table 2: Story Entries

WEIGHT: The relative importance of this story, calculated from the field described below.

SOURCE: The source of this story, e.g. NY Times.

SECTION: The section in source for this history, e.g. Sports.

HEADLINE: The headline of this story.

BYLINE: The byline of this story.

DATELINE: The dateline of this story.

WORD: A list of pointers into the word dictionary from which this story is reconstructed.

PICTURE: A list of illustrations associated with this story.

TIME 1: A timestamp from when this story was entered into the database.

TIME 2: A timestamp from when this story was accessed.

ARCHIVE: A flag indicating that story has been archived to optical storage.

FREQUENT: A list of pointers to most frequently occurring words in this story.

KEYWORD: A list of pointers to search keywords which caused this story to be added to the database.

2.3. Reading the newspaper

Lehti esitetään etusivun iskulauseilla painettuna suurikokoisin kirjaimin. Palstalla on pääotsikko, joka on topikaalinen referentiaalinen parametri, ja siten avainsana. Se voi olla tekninen, poliittinen, tai jokin muu. Näyttöön ei kerrallaan mahdu kaikkea, ja haluttujen teemojen saamisessa näyttöön voidaan käyttää kahta tapaa. Toisaalta voidaan esittää skemaattinen kokonaisnäyttö, jossa itse tekstejä ei voida lukea, mutta niiden yleiset kokosuhteet, ja asu ovat hahmotettavissa. Toisaalta kokosivun pienempää osiota voidaan kelata näyttöruudulla, ja hakea siten haluttua otsikkoa.

figure 1.

Otikonvalinta näyttöruudulta tapahtuu koskettamalla haluttua kohtaa ruudulla, joka sensitivioituna reagoi kosketukseen. Näytössä näkyvän sivun kokosiirto kuvaruudulla suoritetaan aloittamalla koskettava liike kuvaruudulla sen nurkasta, jolloin näyttöön ilmaantuu toisia osia koko sivusta. Jokaisella palstalla artikkelin alkuosa on palstaladottu; pystysuoran suunnan kosketusliike siirtää luettavaksi muuta osaa luettavaa artikkelia, joka tällöin yksin liikkuu omana kaistanaan muun osan sivua pysyessä liikkumattomana. Tällöin luettavaa artikkelia voidaan verrata viereisten palstojen artikkeleihin.

figure 2.

Luettavan artikkelipalstan poikki tehty liike vierittää muita palstoja eteenpäin. Haluttaessa voidaan helposti siirtyä lukemaan niitä, ja jättää luettavana ollut palsta silleen. Jokaisen artikkelin sisällä koskettaessa mitä tahansa kirjainta tai lausetta erikseen voidaan mitata sitä, missä määrin sitä esiintyy muissa artikkeleissa, joita etusivulle on valikoitu. Mikäli esimerkiksi tietty sana esiintyy muissa artikkeleissa, se korostuu muista sanoista erottuvaksi, ja luo artikkelienvälisen kytkennän. Mikäli sanaa ei löydy muista artikkeleista, vaan ainoastaan luettavasta artikkelista sen jossakin toisessa kohdassa, korostuvat luettavan artikkelin palstan reunat näytön taustasta. Tällainen malli on suunniteltu havainnollistamaan sitä, miten päivän lehteä luetaan ilman, että lähettäjä vaikuttaa asiaan (undirected manner); artikkeleja ei välttämättä lueta peräkkäin niin kuin ne on tavanomaisessa sanomalehdessä ladotut, vaan poimimalla luettavaa aineistoa sieltä, ja täältä (huomattakoon, että näyttöön voidaan saada aineistoa myös pitemmältä ajanjaksolta sen liittyessä luettaviin artikkeleihin, suom. huom.)

2.4. Programming system

Muokkaajat ja agetit NewsPeek -järjestelmässä eivät ole inerttejä, vaan reaktiivisia niin julkaisijain kuin lukijainkin suuntaan. Lukija asettaa mielenkiintonsa kohteet määrittämällä haluamiaan otsikkokokonaisuuksia.

Table 4: Transaction Entries

STORY: The story being accessed.

WORD: A significant word or phrase. This field is used only when access is by directed search.

TIME: A timestamp of access.

FLAGS: An indication as to whether access was from front page, related story or archive.

Henkilökohtaisten elektronisten postilaatikoiden ja paikallisten databasejen käyttö sanomalehden materiaalina liittyy persoonallisen kirjeenvaihdon arvostukseen, ja voi olla muokkaimen vastaaottamana aineistona. Inhimilliset lähetetyn datan muokkaajat ovat myös tärkeitä. Kustantaja voi vaikuttaa siihen, minkälainen aineisto voisi olla kiinnostavaa valikoimalla johtavia etusivun juttuja luokittelemalla tiettyjä artikkeleita "tärkeiksi tapauksiksi".

Sen jälkeen kun alustava työstöprofiili on luotu, jokainen lukutapahtuma muuntaa sitä mielenkiinnon suuntaumia toteuttamaan.

Table 6: Search Entries

DATABASE: The name of database to be search, e.g. NEXIS.

LIBRARY: The name of library to be searched, e.g. Wire Services.

WORDLIST: A list of search keywords, separated by booleans.

TIME 1: A timestamp of most recent search.

TIME 2: A timestamp of most recent addition to the database.

Sanat ja lauseet sanaluettelossa, joita lukija artikkeleista osoittaa, lisätään hakusanaketjuihin. Avainsanat, jotka eivät kytkeydy niihin artikkeleihin, joita luetaan, poistetaan. Niin hakusanojen profiili vähitellen adaptoituu lukijan intressien mukaiseksi; lukija kirjaimellisesti esiohjelmoi seuraavaksi lukemiensa lehtipainosten luonnetta, ja tapaa, millä muokkain artikkeleita valikoi. Niinpä lehden sisällöllinen olemus on tulevien painotusalueidensa suhteen tietyllä tavoin määrittelemätön ja avoin.

2.5. Network Plus

Eräs seuraus informaationprosessoinnista, jonka lähdeaineistona on toisistaan olennaisesti poikkeavia lähemateriaalityyppejä, on hälventämässä osaltaan niiden välillä aiemmin vallinnutta eroa. NewsPeek -ohjelmassa televisiota käytetään sanomalehden lukemiseen; Network Plus [3] varastoi paikallisesti lähetettyjä televisio-ohjelmia niin, että niitä voidaan liittää synteettisesti persoonalliseen sanomalehteen. Käytännössä tallennettaessa lähetettyjä televisio-ohjelmia niitä skannataan samoin agentein, jotka luovat NewsPeek -sanomalehteä sitä luettaessa. Ne artikkelit, josta ovat korrelatiivisessa suhteessa lähetettyjen televisio-ohjelmien teemoihin, voidaan löytää, ja esittää omalla palstallaan persoonallisessa sanomalehdessä. Halutut kuvat televisiolähetyksistä voidaan tallentaa omiksi tiedostoikseen. Tuloksena on database, josta voidaan tulostaa printtattu transkriptio välittömästi ohjelman päätyttyä, vaikka lehden lukemisen aikana siitä onkin ollut näkyvissä vain lyhyt abstrakti. Lisäksi voidaan juuri tavoitettua lähetystä katsoa sellaisenaan, ilman, että sitä vedostettaisiin.

figure 3.

Tällä hetkellä tutkimuksen kohteena on etupäässä puheen analyysi, ja tavoitteena on käyttää intonaatiota ja painotuksia valikoitaessa sitä, millaista informaatiota ensisijaisesti varastoidaan. Tällaista vapausastetta ei ole painetuissa teksteissä, eikä sitä voida toteuttaa traditionaalisissa databaseen nojautuvissa järjestelmissä.

3. Digital Movies

Tässä luvussa kuvaamme joitakin osoita tutkimusohjelmasta, jolla luodaan kuvien sekvensseille, joiden tuloksena on liikkuva kuva, digitaalinen representaatio. Ohjelman päämääränä on tehdä elokuvista tietokoneen kannalta "ymmärrettäviä". Tällä tarkoitamme, että

1. ne tallennetaan niin, että persoonalliset tietokoneet voivat niitä tulkita, ja että
2. representaation taso on sellainen, että elokuvaa voidaan manipuloida korkeamman tason komennoilla niin, että siitä muokataan tapahtumakokonaisuuksien elementtejä (on scene elements) pikemminkin kuin kuvan elementtejä tai yksittäisiä kuvaruutuja, ja että
3. kontrollirakenne toimii varioiden joitakin elokuvan aspekteja samalla kun elokuva etenee näytössä.

Merkittävä komponentti ohjelmasta on ohjattu varastoimaan filmiä täysin digitaaliselle audiollevylle, ja tulemme kehittämään edelleen tätä erityistä sovellusta. Compact disk on malli mille tahansa digitaaliselle kanavalle. On myös tulossa sovelteita, jotka hyödyntävät digitaalista ääninauhaa ja digitaalisia puhelinkanavia.

3.1. Digital Coding Requirements

On mahdollista digitoida NTSC videosaignaalia ilman kompressiota lähetyksen alkuperäisellä tasolla käyttämällä kahdeksanbittistä kvantisoijaa, joka toimii kolme kertaa värisignaalin kantaosion tasolla, arviolta noin 10,7 megaheziä, bittinopeus 85 megabittiä sekunnissa. Tämä on hieman vähemmän kuin kuusikymmentä kertaa kompaktin audiollevyn bittinopeus. Toisaalta on osoitettu demonstratiivisesti (teleconferencing video systems) voitavan käyttää bittinopeuksia 1.5 megabittiä sekunnissa. Kuitenkin kuva huononee jos siinä havaituksi tuleva liike on hyvin nopeaa, ja optimointi tehdään perustaen tehtyyn otaksumaan kuvan sisällöstä ja rajoista representoitavassa liikkeessä. Hitaasti

liikkuvat kasvot ovat tavanomaisin kohdekuva, ja joissakin systeemeissä huulten liikkeiden synkronointi on vaikeaa kohdentaa.

3.1.1. Asymmetrical Coding

Koska levy on julkaisumekanismi, joka masteroidaan kerran, vaikka esitetäänkin useita kertoja, voi enkoodaus olla mutkikasta. Dekoodausprosessin on oltava yksinkertainen. Prosessoinnin termein, ei enkoodauksen algoritmin tarvitse operoida reaaliajassa, ja se voi olla intensiivisten tietojenkäsittelyoperaatioiden kohteena. Tämä on vastakohtana videokoodaukselle, jota käytetään televisiolähetyksissä, tai telekommunikaation enkoodauksessa, joissa kummassakin tapauksessa enkooderin ja dekodeerin on toimittava reaaliajassa, ja joissa koodauksen on tapahduttava jo signaalia luotaessa, ja sitä vastaanotettaessa. Julkaisuympäristössä vain dekodeerin on oltava tällä tavoin optimoitu, jotta sitä voitaisiin laajasti käyttää. Kutsumme tällaista koodauksen muotoa asymmetriseksi, koska enkooderi ja dekodeeri eivät ole yhtä komplekseja.

3.1.2. Channel Characteristics

Toinen eroavaisuus tavanomaisen videokompressoinnin ja aineiston syöttö tallenteista eroaa kanavankäytön tavasta. Elävässä videokoodauksessa kuvaruudut on enkoodattava ja dekodeattava lennossa, joista jokainen on optimoitu käytetylle taajuudelle. Elokuvien esityksessä ei verkkokanavakapasiteettia ylitetä; välitön kanavan kapasiteetti voi ylittyä tarkasteltaessa dekodeerissa elokuvan vielä esittämättömiä jaksoja. Hinta tästä on dekodeerin tarvitsema lisämuistitila. Samoin on, kun elokuva enkoodataan, sillä koko kuvien sekvenssi on käytettävissä enkoodaukseen. Siten mahdollistuu prosessoida "globaalisesti" elokuvaa. Esimerkki tällaisesta enkoodauksesta on se, kun koodataan kaikki visuaalisesti toisiaan muistuttavat kokonaisuudet (scenes) yhteen, ja tuotetaan niistä liukuva kuvien jono (re-shuffling) näyttöön.

3.1.3. Image Characteristics

Kuvien, joita tallennetaan levyille, otaksutaan olevan osia kuvien muodostamasta sekvenssistä. Niinpä on tärkeää erottaa toisistaan yhden tunnin mittainen elävien liikkeiden sarja ja 108 000:n yksittäisen televisiokuvan muodostama joukko. Se, että on mahdollista tallentaa edellämainittu, ei vielä viittaa siihen, että se sisältäisi jälkimmäisen. Kuten kuvattu ohjelman yhteydessä, ei CD ole neljän tuuman videolevy, vaan se on elokuvien medium. Kuitenkin pyrittäessä pienempään kapasiteettiin on tosiasia, että kuva on digitaalinen, ja sitä voidaan käyttää useita kertoja niin, että siitä muovataan uusia versioita uusiin käyttöihin. Toisin sanoen on järkevää otaksua olevan tarvetta siihen, että voidaan valita levyiltä jokin "kuvaruutu". CD ei välttämättä varastoi kuvaruutuja sellaisenaan, vaan se sisältää tarpeellisen datan, jolla yksi tai useampia kuvaruutuja voidaan luoda.

3.1.4. Printing Images

Tärkeää työllemme on se tosiasia, että digitaalinen kuva on saatavilla niin, että se voidaan kääntää printtimedialle ilman, että kuvaa olisi sinällään tarpeen prosessoida, mikä on tavallista videoprinttauksessa. Synteettinen mustavalkokuvan tuottaminen paperille, jota useimmissa kuvantulostimissa käytetään voidaan ohittaa kokonaan digitaalisella ohjauksessa korkealaatuisella resoluutiolla alkuperäisen kaltaisena; se ei ole NTSC -kopio. Myöskin systeemi voi käyttää ruutujenvälistä korrelaatiota derivoidakseen soveltuvan masterin, joka printataan, eikä siten yksinkertaisesti vain printtaa yhtä ruutua.

3.2. Digital Coding: The approach

Useita eri lähetystapoja digitaaliseen kuvien sinällään enkoodaukseen on tutkittu. Joidenkin niistä

parametreja kuvataan seuraavassa.

3.2.1. Vector Quantization

Vektorikvantisointi on enkoodaussysteemi, joka soveltuu hyvin tähän erityiseen ongelmaan. Enkoodausprosessi on tietojenkäsittelyllisesti intensiivinen, kasvaa eksponentiaalisesti kuvan koon suuretessa, mikäli optimisointia ei ole suoritettu; se vaatii melko yksinkertaisen "table-lookup" enkoodauksen. Vektorikvantisoija toimii enkoodaamalla ryhmittäin signaalelementtejä niin kuin ne olisivat yksittäisiä, monidimensionaalisia vektoreita. Kompresio saavutetaan asettamalla vektorin paikalle koodisana, tai otsikko [7], joiden määrä ei ylitä suurinta mahdollista vektoreiden lukumäärää. Dekoodaus sisältää koodisanan katsomisen taulusta, ja sen asettumisen vektoriksi, joka aproksimoi alkuperäisen elementtiryhmän [8]. Kompresio saadaan aikaan käyttämällä pienempää vektorien määrää kuin mitä niitä maksimaalisesti voi olla olemassa. Tekniikka toimii hyvin yhdistyneenä input-signaalin todennäköisyysfunktion, joka on sekä "non-uniform", ja joka ei ole erillään mistään koodatuista dimensiosta. Kun dimensiot ovat lineaarisesti toisistaan riippumattomia, niin kooderi voi toimia kuten prediktiiviset kooderit. Aikaisemmin vektorikvantisoijia hyödynnettiin menestyksellä puheeseen ja liikkumattomiin värikuviin. [9,7,10] Sen ansiot juontuvat sen statistisesta tehokkuudesta, ja sen virhetoiminnot potentiaalisista virhesuhteutuksista (mismatchings), mikä on statistisessa mielessä tärkeää, ja mikä taas visuaaliselta kannalta merkityksellistä. Psykofysiologiset parannukset vektorikoodereissa ovat olleet työmme eräitä komponentteja. Esimerkiksi Architecture Machine Group kehitti 1970 -luvun keskivaiheilla adaptiivisen värienkoodausalgoritmin, jonka Heckbert [12] julkaisi. Tässä algoritmossa värit, kuten punainen, vihreä ja sininen koodattiin ryhmänä, eikä niin, että ne erillisinä muodostaisivat kuvan. Histogrammi tehtiin suosituimmista väreistä jossakin erityisessä kuvassa, ja erotetuista joukoista tuotettiin kolmiulotteinen koordinaatisto, joissa koordinaattien määrä määräytyi erilaisten algoritmien mukaan rajoitetuiksi joukoiksi värejä, jotka edustivat tätä tiettyä kuvaa. Kun kuva eritellään histogrammin luontia varten, muotouttavat valitut joukot (samples) eräänlaisen harjoitusjoukon, joka määrää sen koodikirjan, jota käytetään kuvan muissakin osissa. Kun jokaiselle kuvalle on oma koodikirjansa, on systeemi adaptiivinen. Olemme myös tutkineet muita samantapaisia systeemejä, joissa koodikirjaa kyetään hyödyntämään. Tällainen algoritmi antaa mielekkäitä tuloksia verrattuna menetelmään, jossa väriakseleita enkoodataan erillään toisistaan. Mikä tahansa kuva sisältää sekä tummia, että vaaleampia varjoja, mutta vain harvoin on niin, että samantapainen värien jakauma pätee koko kuvien joukkoon. Suurempaa kooditiheyttä käytetään jonkin erityisen kuvankohdan väritilassa. Mikäli kuvan värit ovat lähes olemattomat, niin tällöin tuollaisten kuvien köyhään väritilaan voidaan lisätä tonaalista vaihtelua korvaamalla joitakin sävyjä harvemmin käytetyillä sävyillä. Eräs lähetyksistapa tällaisessa tapauksessa on se, että kolmielementtinen neliöalue kolmessa peräkkäisessä kuvaruudussa koodataan ryhmänä tuottamalla 27 dimensionaalinen avaruus, joka voi sisältää 2 potenssiin 27 arvoa. Koodikirjaa, joka jakaa tällaisen avaruuden käsiteltävämpään muotoon 4-16K tutkitaan parhaillaan. Adaptiivinen koodikirja joidenkin kokonaisuusien entistämistä varten on tutkimuksen kohteena, jotta voitaisiin laajentaa sellaisten erityisten koodien hyödynnettävyyttä, jotka tuottaisivat takaisin derivoiduista kokonaisuuksista niiden alkuperäisen muodon. Eräässä tapauksessa elokuva järjestetään uudelleen niin, että samantapaiset tapahtumakokonaisuudet (scenes) jäsennetään samaan tauluun. Alkuperäinen järjestys tallennetaan myöhemmin, mutta siitä poikkeavalla koodikirjalla käsitellään jokaista erillistä tapahtumakokonaisuutta.

figure 4.

Tätä voidaan laajentaa koodittamaan erillisiä objekteja kuvajaksoissa käyttämällä vektorikvantisoijaa vaihtoehtoisena kuvarepresentaatioformaattina kuvaan itseensä. Erityinen tapa jolla tila voidaan polulatioida voi paljastaa kuvan ominaisuuksia, jotka voivat ohjata muita koodauskeemoja, ja potentiaalisesti tehdä erottelua kuvan eri entiteettien välillä. Esimerkiksi ne kuvan komponentit, jotka eivät liiku muodostavat erillisen alueensa vektorivaruudessa verrattuna liikkuvien komponenttien vektorivaruuteen. Kuvattaessa kameralla sen liikkeet voivat siirtää jakaumia, kuten myös valaistuksen muutokset. Tämä mahdollistaa tapahtumakokonaisuuden jäsentelyn individuaaleiksi objekteiksi ja skriptausparametreiksi.

3.2.2. Real Space Coding

Toinen koeasetelma liittyy kuvien sekvenssissä esiintyvän liikkeen koodaukseen syntetisoimalla ehdokkaaksi kolmidimensionaalinen database, josta kuva voidaan derivoida. Tämä database kontrolloi sitten ehdolla olevaa muuttunutta tapahtumakokonaisuutta (scene), minkä prosessori "valokuvaa", ja jota verrataan myöhempiin aktuaalisiin tapahtumakokonaisuuden kuvaruutuihin. Jos kolmidimensionaalisen datan estimaatti on oikea, ja siirto hyväksyttävissä, niin myös tapahtumakokonaisuudessa esiintyvät objektien liikkeet ovat myös ennustettavissa. Laajemmassa merkityksessä on tällainen lähestymistapa ekvivalentti siihen, kun ratkaistaan koneellista visioinnin ongelmaa. On määriteltävä niin objektit kuin niiden tekemät liikkeetkin. Yleisesti todettakoon, että tällaista ei ole mahdollista tehdä. Jonkinlaista menestystä on kuitenkin saavutettu käsiteltäessä tapahtumakokonaisuudessa esiintyviä liikkumattomia hahmoja [13]. Optimointi voidaan tehdä myös niin, että merkitään käsin tiettyjä objekteja tapahtumakokonaisuudessa, tai käyttämällä kameraa, jolla kuvaan lisätään syvyysinformaatiota kuvan tueksi. Työn alla on sentapaisen kameran kehittäminen.

4. Conclusions

Esitellyt kaksi näennäisen toisistaan erillistä ohjelmaa suuntautuvat kohti samaa päämäärää: kuinka me voimme käyttää tietojenkäsittelyä korkean taajuuskaistan kanavilla kotioiloissa, jonka kautta voidaan tarjota uudenlaista informaatiota ja kokeilun mahdollisuuksia. Tutkimuksemme toinen osa keskittyy uutissaantiin, ja vaatii lukijalta aktiivista osallistumista; toinen osa taas omistautuu kuvallisten esitysten luomiseen. Tosiasiassa ne ovat saman teeman kaksi eri näkökulmaa. Syvempi kysymys on se, kuinka media voi potentiaalisesti kehittyä, kun jotkut sen teknisistä perusteista perustuvat muutokseen. Voimme väittää lähetysten itsensä olevan derivoituja välineiden rajoitteista vuosisadan alkupuolelta. Vastaanottimien hinta oli riittävän korkea, jotta niiden kohdalla voitiin tietyt ominaisuudet jättää pois juuri hinnan vuoksi; hinta määräsi sen, millaiseksi niiden toiminnan eräät piirteet muodostuivat. Samalla tavoin voi sanoa kanavienvälityksessä hinnan määränneen sitä, ettei kuvanlaatu muodostunut korkeimmantasoisiksi. Kuitenkin, kun tietojenkäsittely ja korkeammat taajuusalueet tulivat halvemmiksi, tuli mahdolliseksi myös lisätä informaation määrää systeemeissä; niiden muodot ja ilmentymät voivat muuttaa ne aiemmista lähetysjärjestelmistä persoonallisiksi ja individuaalisiksi vuoro vaikutuskentiksi. Olemme vältäneet yrittämästä puolustaa prosessoinnin relatiivista arvoa niin vastaanotossa kuin kanavapapapiteetissa, koska emme näe korkeampien taajuusalueiden käytöllä olevan merkitystä informaation koodaukseen kannalta sinällään. Koodauksen tarkoituksena on ollut tuottaa representaatio sellaiseen muotoon, että käyttäjä voi sitä muokata. Se, että käyttämämme tapa säästää lähetystaajuudessa on pelkkä korollaari, ei päämäärä. Ero matalilla ja korkeilla taajuusalueilla on niiden käytöllisen välittömyyden asteessa. Sanomalehti voidaan tuottaa julkaisuksi 50 megabitin kanavilla, mutta sen saanti käsittelyyn nykyisiä puhelinlinjoja myöten kestää pidempään.

References

- [1] Gitta Salomon Design and Implementation of An Electronic Special Interest Magazine. MIT MSVS Thesis September 1986.
- [2] Aya Konishi Auto-Cassette: Automatic Clipping Service for TV News, MIT Media Laboratory BSCS Thesis June 1986.
- [3] Pascal Chesnais Network Plus. Media Lab Memo June 1986.
- [4] Judith Donath The Electronic Newstand: Design of an Intelligent Interface to a Variety of New Sources in Several Media. MIT MSVS Thesis September 1986.
- [5] Walid Hamdy Hardware Implementation of Dynamic Video in Personal Computers. MIT BCCS Thesis June 1986.
- [6] Uri Feldman Image Compression using Vector Quantization. Media Laboratory Research report, December 1985.
- [7] Goldberg and Sun Image Sequence Coding using Vector Quantization, IEEE Transaction on Communications, forthcoming.
- [8] Gray vector Quantization, IEEE ASSP Magazine, April 1984.

[9] Makhoul et al Vector Quantization in Speech Coding, Proceedings of IEEE, November 1985.

[10] Budge and Baker Compression of Color Digital Images Using Vector Quantization in Product Codes, IEEE 1985.

[11] Internal Architecture Machine Group Memo.

[12] Heckbert Adaptive Color Image Coding, ACM-SIGGRAPH.

[13] Netravali and Sals Algorithm for Estimation of Three Dimensional Motion, Netravali and Salz, AT&T Technical Journal, February 1985.

[14] V.M. Bove Imaging With A Range Camera, 1986, MIT Media Laboratory Report, Electronic Publishing Group.

Andrew B. Lippman is Associate Director of MIT's Media laboratory. He is currently Director of the Electronic Publishing Group within the lab and is principle investigator of a new research program addressing the "Future of the Movies". This program focuses on the intelligent interaction between computing and sequential visual images.

Since 1978, Walter R. Bender has been a member of the Architecture Machine Group at MIT, where he received his M.S. in 1980. Bender is currently a Principle Research Scientist at MIT Media Laboratory. He is Associate Director of the Electronic Publishing Group, and its Principle Investigator of the Personal Computing Program.

The Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

Adaptive color coding based on spatial/temporal features

Walter Bender

*Massachusetts Institute of Technology Media Laboratory
20 Ames Street, Cambridge, Massachusetts 02139*

(To be presented at SPSE Electronic Imaging Devices and Systems Symposium, Los Angeles, January 1988)

This work has been supported in part by International Business Machines.

Suom. **Timo Kinnunen** 10.3.1989

Pelastettu 3.5 Levykkeeltä NU.EXE Ohjelmalla 12.4.1989 T. Kinnunen

ABSTRACT

Tehokkaan tietokonenäyttöön tulostuvien värikuvien adaptiiviseen koodaukseen liittyviä ongelmia tutkitaan parhaillaan. Onkin luotu nykyisin käytössä olevien algoritmien modifikaatioita, joilla voidaan esittää paikantuvuuteen liittyviä kuvan karakteristikoita. Spatiaalisia, -temporaalisia, ja semanttisia piirteitä on käytetty luotaessa maskeja, jotka ilmaisevat virhepoikkeamaa (error masks). Näitä maskeja on vuorostaan käytetty korjaamaan vektorikvantisoijan käyttämiä tilastollisia histogrammeja.

1. INTRODUCTION

Tavallisin arkkitehtuuri tietokonenäytöissä rakentuu kuvaruutumuistista, jolla on yhteys grafiikkakontrollipiiriin, joka tuottaa värivalikon. Näytön kuvan laatu on suorassa suhteessa värivalikon laajuuteen, ja jokaiseen valittavana olevaan väriin liittyvään bittimäärään, kuten esimerkiksi 256 väriä 16 -miljoonan bitin paletista. Kahdeksan bitin intensiteettikontrollia primaaria kohden käytetään usein värinäytöissä, vaikka se vaatiikin 24 bittiä kuvaelementtiä kohden, jotta voitaisiin tuottaa sävyjatkuvuudeltaan yhtenäinen kuva. Monet värivalokuvat voidaan kompressoida 256:sta 1024:aan väriin ilman, että kuvan laatu juurikaan heikkenisi [1,2]. Alkuperäisten kuvaoriginaalien värirunsauden toistettavuutta ajatellen on selvää, että käytettävä pienemmän värien määrä käyttö vaatii tiheennystä (dithering) tai virhe-erottelua (error diffusion), jotta vältettäisiin virheellisyyksiä väripintojen rajakohdissa (contours), ja siten rajakohdat onkin korvattava korkeafrekvessisellä kohinalla [3,4]. Nykyinen tutkimuksen suunta liittyy adaptiivisen värikuvien koodauksen yleiseen problematiikkaan tietokonenäyttöä ajatellen. Useita nykyisten algoritmien modifikaatioita on ehdotettu, jotka sallivat

paikantuvien kuvakarakteristikoiden tuoton näyttöön. Kaikki kompressiotekniikat joko hyödyntävät tilastollisia signaalien redundansseja, tai observoijan psykofysiologisia ominaisuuksia. Kumpaa tahansa lähestymistapaa käytettäessä enkoodauksessa, jätetään signaalien käsittelyllisesti tarpeettomat komponentit huomiotta. Kuitenkaan nämä tekniikat eivät ole erityisen tehokkaita. Adaptiivisen värikooderin toiminta, oli se tilastollinen tai psykofyysinen, soveltaa kuvan alempiin elementteihinsä jaoittamista (subsampling) koodauksessa. Kolmiulotteinen avaruus, jonka akselit useimmiten edustavat punaista, vihreää ja sinistä, populoidaan otosarvojen (sample values) histogrammijakaumaan. Enkooderi muodostaa sitten väriavaruuden alajakaumat halutunlaajuisella värikoodimäärällä. Tätä monidimensionaalisen avaruuden alijaoittamista (subdivision) tutkitaan runsaasti vektorikvantisoinnin -nimellä tunnetulla koodaustekniikalla. Gray [5] mm. tutkii erilaisia alijaoittamisen sääntöjä. Jokaista originaalikuvan elementtiä verrataan tämän jälkeen sitä lähinnä olevaan vastaavaan koodiin [2]. On hyödyllistä ajatella enkoodauksen ongelmaa maskivaikutuksena (in terms of masking). Enkooderin tuotoksen ja kuvaoriginaalin eroa tarkastellaan tuottamalla eroa ilmaiseva virhekuva. Tehokasta koodausta voidaan siten tarkastella siten ongelmana, kuinka tehokkaimmin voidaan häivyttää (mask) virhekuva. Etuna maskivaikutuksen tutkimisen lähestymistavalla koodausongelmaan on siinä, että se johtaa lukuisiin ratkaisumalleihin, joita ei voida tavoittaa sellaisin menetelmin, joissa kuva jaotetaan huomioimatta sitä, mistä kohden kuvaa otokset siitä löytyivät. Tämä raportti keskittyy nimenomaan maskivaikutukseen. Kaikki tekniikat hyödyntävät jotakin tietoa kuvasta itsestään, jotta voidaan jäsentää ja modifioida kuvan tiheyksiä kuvaavia otoksia, painottaen otoshistogrammia käytetyn tietämyksen mukaisesti. Eräs tekniikoista keskittyy kuvan spatiaalisiin piirteisiin. Tällöin tekniikassa korostuu se, MISSÄ kohden kuvaa virheitä esiintyy. Toiset tekniikat nojaavat kuvan semanttisiin sisältöihin I. MITKÄ KUVAN OSISTA OVAT OBSERVOIJAN HAVAINNOINNIN KANNALTA KESKEISIÄ. Alustava tutkimustyö tällä alueella on hyödyntänyt kuvan temporaalista analyysia [18], kuten myös syvyysinformaation osoittamista (depth keying) [19]. Toiset analyttisistä, tai ei-analyttisistä kuvan tärkeiden piirteiden mittauksista voidaan lisätä malliin, jotta enkooderia voitaisiin edelleen optimoida.

2. PRECEDENTS

On useitakin hyvindokumentoituja havainnon ominaisuuksia, joita voidaan hyödyntää koodauksessa:

- * Herkkyys epäjatkuvuuksille spatiaalisessa frekvensspektrissä.
- * Herkkyuden väheneminen vastaanotettuun ärsykkeeseen läheisärsykkeen vuoksi, sekä spatiaalisen -että temporaalisen.
- * Lisääntynyt herkkyys luminanssikohinalle kuvan tummilla alueilla.

Kuvanhäiriöiden mittaukset perustuvat johonkin, tai kaikkiin manituista havaittavuus-ominaisuuksista, joita ovat kuvanneet mm. Sakrison [6], Watson [7], Glenn [8], Limb [9], Lukas [10], Sharma [11], Legge [12], ja Stromeyer [13]. Häiriömittauksia käytetään usein visuaalisten eroavaisuuksien kvantisoinnissa; onhan se hyödylliseksi havaittu myöskin kuvien perseptuaalisessa koodauksessa, jossa soveltumaton informaatio siivilöityy pois, ja jossa sitten evaluoidaan resultanttikuvia. On tehty systemaattisia tutkimuksia häiriömittauksen hyödynnettävyydestä kuvankoodauksessa taajuusaluekompressiossa. Spatiaalisten luminanssimuutosten peittovaikutusta (masking) ovat tutkineet mm. Netravali ja Presada [14], sekä Ramaurthi ja Gersho [15]. Yamaguchi [16] on tutkinut nonlinearisten transformatioiden koodausta puna-vihreä-sininen -väriavaruudessa. Swaszek ja Thomas [17] käyttävät nonlinearisia funktioita kvantisoidakseen ja laajentaakseen luminanssisignaalia. Vektorikvantisoinnin käyttö kuvankoodauksen mekanismina näyttöön on lukuisia eroavaisuuksia koodattaessa kuvia tiedonsiirtoon, koska graafisen näytön resoluutio on kiinteä (yleisesti yhden suhde yhteen näyttömuistin ja kuvaelementtien välillä), kun taas tonaliteetin suhteen jää tietty vapaus. Ne mittaustavat, joissa hyödynnetään tonaliteetin riippumattomuutta sijainnista ovat hyödynnettävissä olemassaoleviin algoritmeihin käyttämällä häiriömittausta ennen kvantisointia, sillä kooderit kvantisoivat luminanssia hienojakoisemmin kuin krominanssia. Heckbert [2] soveltaa "suuntautuneeseen toimintaan liittyvän raapustelun" (doodle) mekanismia, jossa "käsin piirrettyä" maskia käytettäisiin lisäotosten syöttöön otoshistogrammiin. "Suuntautuneeseen toimintaan liittyvällä raapustelulla" tuotetaan enemmän keskeisyyttä kuvan mm. pienille kasvojen yksityiskohdille, kuten esimerkiksi silmille ja suulle suoralla ihmisen väliintulolla. Visuaalisten mallien käyttöön ja kuvan semantiikkaan liittyvänä on tämä tekniikka

saanut yleistävyyttä.

3. PROCEDURE

Tässä luvussa kuvataan kuvan kvantisoinnin proseduuria. Perusmuodossaan [2,3] on vektorikvantisoinnin algoritmilla kolme vaihetta:

- * Analyysi, jossa värienjakauma kvantisoitavasta kuvasta esitetään histogrammina kuvan jokaista kanavaa koskien. Harjoitujosetti generoidaan koodattavaksi aiotun kuvan satunnaisjakaumasta.
- * Koodikirjaa varten valitaan histogrammista edustava vektorien joukko (koodisanat), ja kd-puuta käytetään tietojenkäsittelyllisesti tehokkaana keinona koodikirjan generoinnissa.
- * Kvantisoinnissa kuvan jokainen otos liitetään sitä lähinnä olevaan koodisanaan. Virhe-erottelun (error diffusion) tekniikoita käytetään prosessin tässä vaiheessa.

Tässä raportissa esiteltävä koodausalgoritmin modifikaatio on maskivaikutusfunktion inkluusio kuva-analyysissa. Maskivaikutusfunktiota käytetään kontrolloimaan laatutason vaihteluja alueelta toiselle poikkeuttamalla (biasing) kuvan otoksia. Generoidaan sellainen maskivaikutuskuva, joka suoraan vastaa kuvaoriginaalia. Se generoidaan liittyen kuvahistogrammin generoimiseen, jotta luotaisiin joustava testausalusta (test bed), eikä pyrkimyksenä ole niinkään luoda tietojenkäsittelyllisesti tehokasta ratkaisua. Maskikuvan arvo $m(i,j)$ pisteessä (i,j) määrittää kaksiulotteisen otosalueen koon. Kaikki pisteet, jotka sisältyvät ympyrään, tai diametriin $m(i,j)$, ja jotka ovat keskittyneet pisteen (i,j) ympärille, lisätään histogrammiin. Siitä johtuen poiminnan lopputuloksena on suuri maskiarvo saadaan suhteellisesti suuremmalta alueelta kuin pieni maskiarvo. Poiminta ei ole enää satunnaista, vaan on suuntautunut erityisiin kuvan komponentteihin. Saatua histogrammi ei siten pelkästään heijasta sitä, mitä värejä löydetään, vaan myös jossain määrin sitä, MISSÄ KOHDEN ne kuvaan poimittiin. Koodikirjan valinta liittyy otoksen edustavuuteen (representative of sampling), sillä kvantisointivirheet korreloituvat maskivaikutusfunktiota (katso fig. 1 ja 2.). Maskivaikutusfunktion ja kuva-analyysin välinen suhde määritellään kokeellisesti, mutta määrittelytapa ei ole mielivaltaisen. Maskivaikutusfunktion vaihtelualue rajattiin 3 -bittiin. Resultanttiotosalueiden säde vaihteli 1-8 kuvaelementtiin. Kaikille otosalueille annettiin samanlainen painotusarvo. Lukuisat maskiarvofunktiot todettiin hyödyllisiksi. maskiarvofunktioita varioitiin niin, että niitä voitiin matemaattisesti seurata, sekä käyttää ad hoc -tekniikoita.

3.1. Spatial masking functions

On hyödyllistä erottaa toisistaan matala- ja korkeafrekvenssiset alueet, koska kohina on heikommin erotettavissa korkean- kuin alhaisen frekvenssitason alueilla, ja siten kvantisointikohina on (visuaalisesti) havaittavampaa alhaisen kuin korkean spatiaalisen frekvenssin alueilla. Maskivaikutuskuva voidaan generoida vertaamalla toisiinsa alkuperäisen luminanssin kuvaa ja samaa kuvaa ajettuna alitaajuussuotimen kautta (low-pass filter). (Fig. 1b). Resultantti maskivaikutuskuva sisältää matalia arvoja niissä kohdissa, missä alkuperäiskuvassa vierekkäisten kuva-alueiden ero on jyrkkä, ja korkeita arvoja siellä, missä siirtymät ovat loivia. Signaali-kohinasuhteessa (S/N) on vähän eroja verrattaessa kvantisoituja kuvia spatiaalisen maskin kanssa ja ilman sitä. Subjektiiivinen kuvien analyysi viittaa siihen, että virhettä liittyy kuitenkin vähemmän kuviin, jotka on kvantisoitu spatiaalisen maskin kanssa.

3.2. Semantic masking functions

Yritettäessä tuottaa maskikuvia siten, että huomioitiin kuvien semanttisesti "merkityksettömät" alueet sovellettiin lukuisia heuristiikkoja. Rekursiivista blokkienvertailun algoritmia sovellettiin kuvien sekvensseihin silloin kun haluttiin erottaa toisistaan ne kuvanosat, joissa oli liikettä niistä kuvan osista, jotka olivat kuvaavan kameran suhteen liikkumattomia. Tällainen algoritmi työskentelee etsimällä maksimaalisia spatiaalisen sijainnin korrelaatioita tietyn, valitun kuvaruudun ja sitä edeltävien ja seuraavien kuvien yksityiskohtien välille [18]. Maskivaikutuskuva luodaan liittämällä korkeita arvoja

niihin kuvanosiin, joissa on havaittu liikettä (katso fig. 3). Kuten odotettuakin, on kvantisointikohinaa eniten kuvan taustassa. Syvyyskameran (range cameras) kuvia voidaan myös käyttää maskien luomiseen, joissa erotetaan kuvan etuosa sen taustasta. Kameraa lähellä oleviin objekteihin liitetään suurempia arvoja kuin kauemaksi asettuviin objekteihin. Tulokset ovat samansuuntaisia kuin mitä on saavutettu blokkiensuhteutusalgoritmillä (block matching algorithm). Syvyysetäisyysmitaukset, joita käytetään syvyyskameran kuvissa [19] ovat soveltuvia fokuosintimaskien generointiin (focus mask). Fokusoituid objektit painotetaan eri tavalla kuin sen ulkopuolella sijaitsevat objektit.

4. EVALUATION

One man's enchainment may be another man's noise [21]

Signaali-kohinasuhde jää kaikkein yleisimmäksi koodauskeemojen punnitsemiskriteeriksi. Tässä raportissa luonnehdituid tekniikat vaihtelivat suurestikin S/N -suhteen huomioonottaen. Spatiaalisen maskin käytöllä oli S/N:aan vähäinen vaikutus. Testikuvassa (fig. 1.) S/N oli 16.2 db ilman maskia, ja 16.0 db sen kanssa. Tämä ei ole yllättävää, koska virhe on tasaisemmin jakautunut. Tausta huonontui etualalla olevien objektien hyväksi. Kvantitatiivinen ja kvalitatiivinen muutos virhejakaumassa viittaa siihen, että maskikuvan käytöllä on tiettyä potentiaalista etua.

5. CONCLUSION

Yksikään tarjolla olevista visuaalisista malleista ei ole adekvaatti yleiseksi työkaluksi kuvien evaluaatioon, mutta niitä malleja, joita meillä on käytettävissämme, voidaan soveltaa menestyksellä maskivaikutusongelmaa ratkaistaessa. Maskivaikutusvirheen (masking error) arviointi kuvansisällön perusteella on käyttökelpoinen tapa, mutta siihen liittyvät tekniikat eivät ole matemaattisesti jäljitettävissä, tai riittävästi yleistettävissä. Maskivaikutusfunktion tuottaminen vektorikvantisoinnin yhteyteen on ollut menestys siinä mielessä, että se tarjoaa vapausasteita koodausprosessiin. Kun uusia visuaalisia malleja kehitellään, esimerkkinä Adelsonin pyramidimalli [20], voidaan niitä soveltaa maskivaikutusongelman ratkaisuun edellä kuvatulla tavalla. Kun kuvien kontekstuaalinen ymmärtäminen kehittyy edelleen, voidaan sitä myös soveltaa värikuvien koodaukseen tietokonenäyttöjä varten.

6. ACKNOWLEDGEMENTS

Andrew Lippman, Pascal Chesnais ja Judith Donath saavat kiitoksen tämän raportin kirjoittajalta osallistumisestaan tutkimustyöhön. International Business Machines Inc. tuki osittain tätä työtä.

7. REFERENCES

- [1] T. Boyle, Coding Color by Popularity, Architecture Machine Group Memo (1978).
- [2] P. Heckbert, An Adaptive Color Encoding System for Computer Graphics, Computer Graphics, 16(3) (1982).
- [3] G. Houle, Quantization des Images Couleurs Pour Appareils Graphiques, INRS-Télécommunications Technical Report no. 85-37 (1985).
- [4] G. Braudaway, A Procedure for Optimum Choice of a Small Number of Colors From a Large Color Palette for Color Imaging, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, (UNDATED).
- [5] R. Gray, Vector Quantization, IEEE ASSP Mag. (1984).
- [6] D.J. Sarkinson, On the role of observer and a distortion measure in image transmission, IEEE Trans. Commun., 25(11):1251-1267 (1977).
- [7] A. Watson, A. Ahumada, Jr, and J.E. Farrell, The window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays, NASA Technical Paper 2211 (1983).
- [8] W.E. Glenn, K.G. Glenn, and C.J. Bastian, Imaging system design based on psychophysical data, SID Int. Sym. Digest of Tech. Papers, 15 (1984).

- [9] J.O. Limb, Distortion criteria of the human viewer, *IEEE Trans. Syst. Man & Cybern.*, 9(12):778-793 (1979).
- [10] F.X.J. Lukas and Z.L. Budrikis, Picture quality predictions based on a visual model, *IEEE Trans. Commun.*, 30(7):1679-1692 (1982).
- [11] D.K. Sharma and A.N. Netravali, Design of quantizers for dpcm coding of picture signals, *IEEE Trans. Commun.*, 25(11):1267-1274 (1977).
- [12] G. Legge and J. Foley, Contrast Masking in human vision, *J. Opt. Soc. Am.*, 70(12) (1980).
- [13] C. Stromeyer III and B. Julesz, Spatial-frequency masking in vision: critical bands and spread of masking, *J. Opt. Soc. Am.*, 62(10) (1972).
- [14] A.N. Netravali and B. Presada, Adaptive quantization of picture signals using spatial masking, *Proc. IEEE*, 65(4) (1977).
- [15] B. Ramaurthi and A. Gersho, Image coding using segmented codebooks, *Proc. Int. Picture Coding Sym.* (1983).
- [16] H. Yamaguchi, Vector quantization of differential luminance and chrominance signals, *IEEE Trans. Commun.*, 33(5) (1985).
- [17] P. Swazek and J.B. Thomas, Design of quantizers from histograms, *IEEE Trans. Commun.*, 32(3) (1984).
- [18] E. Devlin, High Resolution Video Stills, (SB Thesis), MIT (1987).
- [19] V.M. Bove, Pictorial applications for range-sensing cameras, *SPIE Proc.*, 901 (1988).
- [20] E.H. Adelson, C.H. Anderson, J.R. Bergen, P.J. Burt, and J.M. Ogden, Pyramid methods in image processing, *RCA Eng.*, 26(9) (1984).
- [21] H. Andrews, comment in *IEEE Tutorial on Digital Image Processing*, IEEE (1978).

DIRECTION ASSISTANCE

**James Raymond Davis and,
Thomas Frank Trobaugh, December 1987**

*Speech Research Group Technical Memo 1
The Media Laboratory
Massachusetts Institute of Technology*

Abstrakti

Direction Assistance on integratiivinen ohjelma, joka antaa englanninkielellä ajo-ohjeita autoilijoille Bostonin kaupungin alueella. Ohjelmassa on vastaanotto (input) äänitajuuspuhelimelle ja se antaa pyydytetyt ajo-ohjeet puhesyntetisaattorilla (output). Ehdotetut reitit ovat sekä lyhyitä, että myös helppoja seurata. Ajo-ohjeet annetaan sujuvalla englanninkielellä, ja ohjelmaa on menestyksellä kokeiltu vastamuuttaneiden opastukseen.

1. Johdanto

Direction Assistance koostuu viidestä modulista. The Location Finder pyytää käyttäjää antamaan reitin alkukohdan ja sen loppukohdan. Ne voivat tulla määritetyiksi annettuina katuosoitteina, tai puhelinnumeroina. The Route Finder etsii yksinkertaisen ja lyhyen reitin alkukohdan ja loppukohdan välille. The Descriptor luo kuvauksen reitistä erinomaisella englanninkielellä, ja The Narrator esittää sen ajo-ohjeiden pyytäjälle. Lisäksi ohjeiden pyytäjällä on mahdollisuus saada ohjeet graafisena esityksenä, koska ohjelmassa on tarvittava tulostusvalmius (interface). Edelläesitetyt modulit hyödyntävät databasea, joista tärkeimmän osion muodostaa katukartasto (yhdenoista neliömailin suuruinen alue, jonka keskus on Charles Riverillä). Toinen database on invertoitu puhelinluettelo, joka kääntää katuosoitteet puhelinnumeroiksi.

2. Databases

2.1. Kadut

Aloimme katukarttamme työstön DIMETiedostosta (Dual Independent Map Encoding), jonka on laatinut the United States Geological Survey. DIMETiedosto koostuu joukosta suoralinjaisia segmenttejä, joista jokaisella on tietty nimi, tyyppi, lopetuskohdat vaaka- ja pystysuorassa dimensiossa, ja jossa on myös jotakin muutakin informaatiota tämän lisäksi. Segmenttityypit sisältävät luonnollisia maisemapiirteitä (pääasiassa veden ja maa-alueen rajoja), rautatielinjoja, kaupunginosa- ja katulinjoja, sekä eräitä muita ominaisuuksia kuvaavia linjoja. Kadut identifioidaan osoitenumeroina, joilla määritellään kadun alku- ja loppupiste. On tosin mahdollista estimoida minkä tahansa katusoitteen koordinaatit interpoloimalla 1. otaksumalla korttelit samanpituisiksi. Aloitimme 11 neliömailin alueen osiosta, jonka keskus oli Charles River. Se sisältää osioita Bostonista (Charlestown, Allston, Back Bay, South End, North End), Brooklinesta ja Cambridgesta (Cambridgeport and Harvard, Inman, Central and Kendall Squares). Kartassa on yhteensä 279 mailia katuja, sisältäen 6163 segmenttiä, joista 5506 vastaa katuja. Tiedoston koko on 447 Kb.

Figure 1: Steet Database:

Tämä DIMETiedosto ei ollut läheskään sopiva meidän tarkoitukseemme, koska se sisälsi useitakin virheitä. Katuja puuttui; ne oli nimetty virheellisesti, tai liiden liittymät eivät olleet oikeita, ja kadunnimien tavutukset olivat virheellisiä. Joissakin tapauksissa useatkin segmentit sijoituivat samalle kohdalle karttaa, ja jotkut segmenteistä olivat yhdistyneet toisiinsa. Niinpä kirjoitimme useita ohjelmia, joiden avulla paikallistimme tällaisia virheitä ja poistimme niitä; teimme sen automaattisesti niin suuressa määrin kuin mahdollista. Sen lisäksi, että paikallistimme virheitä meidän täytyi myös lisätä uutta informaatiota databaseen, koska tärkein informaatio liittyi siihen, olivayko kadut yksisuuntaisia. Niinpä luokittelimme kadut myös laadullisesti, ja kirjoitimme tekstimuotoisia kuvauksia joihinkin erityisiin kohtiin karttaa. DIME tiedoston segmentit liittyvät toisiinsa mikäli niillä on yhteinen lopetuspiste. Kutsumme tällaista liittymistä fysikaaliseksi liittymiseksi (physical connectivity). Jokaisella segmentillä on kaksi lopetuspistettä, ja jokaiseen niistä liittyy luettelo segmenteistä, jotka ovat niihin fysikaalisesti liittyneet. Fysikaalinen liittyvyys ei kuitenkaan ole riittävä reitinlöytämässä, koska saattaa olla laitonta ajaa mm. joltakin kadulta toiselle sen yksisuuntaisuuden vuoksi, tai koska kääntyminen tiettyyn suuntaan on kielletty, tai koska keskikoroke estää kääntymisen. Tässä tapauksessa fyysinen este on rajoitteena, eikä niinkään laki tai ajosääntö. Koska sellaiset kuitenkin ovat eräissä tapauksissa rajoitteina lisäsimme liittymisten valikoimaan legaalisen liittymisen (legal connectivity). Kaksi segmenttiä ovat liittyneet legaalisti toisiinsa mm. silloin kun ei voida ajaa suoraan kadulta toiselle rikkomatta lakia. Legaalinen liittyvyys täydentää, mutta ei kuitenkaan korvaa fysikaalista liittyvyyttä. Autolla ei ole lupa ajaa rautatietä, vaikka se liittyisikin ajolinjaleikkauksena fysikaalisesti tiehen, mutta tällainen liittyvyys voi olla mm. ajo-orientoitumisen kannalta merkittävä yksityiskohta. Kaikki kadut eivät ole valinnallisesti samanarvoisia. Halusimme ajoreitin ohjautuvan niin, että valituiksi tulisivat leveimmät, nopeimmat, ja helpoimmin paikallistettavat ajoväylät. Annoimme siksi jokaiselle kadulle tietyn hyvyysarvon (goodness). Määriteltäessä useimmat kadut ovat valinta-arvoina keskinkertaisia (average). Pääliikenneväylät (super) voivat olla ohitusväyliä, kaupunkienvälisiä pääliikenneväyliä, tai muutoin sellaisia, että niille pääsy kaupunkialueelta on liittymiltään rajoitettu. Ne onkin luokiteltu pelkästään helpoiksi löytää ja seurata, koska ne ovat ruuhkaisia ja hitaita. Toisena äärimmäisyytenä ovat kapeat ja huonokuntoiset kadut (bad). Tällaisia katuja databasessamme on kolmisen mailia, ja niitä pyrimme välttämään. Käsite "parempi kuin keskimäärin" (better than average) oli vaikeammin määriteltävissä, koska halusimme identifioida ne kadut, jotka olisivat helpot löytää ja seurata. Päätimme, että ne kaduista, jotka olivat pitkiä, olisivat valinnallisesti tärkeimpiä. Niinpä merkitsimme yli puolen mailin pituiset kadut hyväksi (good), ja sitten lisäsimme niihin muin arviointiperustoin manuaalisesti joitakin (lyhyempiä), mikäli ne vaikuttivat tärkeiltä liikennöinnin joustavuuden kannalta. Tuloksena oli katuverkko, jonka pituus oli 105 mailia yhteensä, ja se muodostaa yksinkertaistetun koko katukartaston luurangon. Kolmas laajennus määritelmiin liittyy kadustoskeeman laajennukseen. Lisäsimme uusia segmenttityyppejä, kuten siltoja, alikulkukäytäviä, liikenteenjakkajia ja pääväylien liityntäramppeja. Tämä lisäinformaatio on tärkeää ajo-orientoitumisen kannalta, kuten myös The Describerin toiminnalle, kuten tulemme huomaamaan.

Lisäsimme myös edelläkuvattuja tiedonsisältöjä deskriptiivisesti luonnehtivaa informaatiota nimikkeinä, jotka kertoivat, minkälaisia tiedonsisältöjä tiettyihin segmenteille kuuluu. The Describer käyttää juuri tätä informaatiota hyväkseen.

Figure 2: Network of good streets:

Teimme kaikki korjaukset ja täsmennykset tekemiemme kenttähavaintojemme perusteella. Emme löytäneet yhtään paperille tehtyä karttaa, jossa olisivat näkyneet kaikki Bostonin yksisuuntaiset kadut, ja siksi meidän oli ajeltava ympäri kaupunkia niitä kartoittaen. Tällaiseen pohjatyöhön kulunut aika muodostaa systeemin kustannuksista valtaosan, mutta työ onkin tehtävä vain kerran. Graafinen database-muokkain oli erityisen kätevä, koska sen avulla databasea saattoi editoida nopeasti tarpeen vaatimalla tavalla. Me suosittelemme tämäntapaisten karttasto- ja ohjelmistojen muokkaajille The Lisp Machine window-systeemiä, jotta tehtävä tulisi helpoksi.

2.2. Neihborhoods

Relaatiotietokannassa luetteloitaan Bostonin ympäristöalueet niihin liittyvine zip -koodeineen. Tarvitsimme tällaista databasea siksi, että annettu katu voi sijaita useissakin kaupungeissa. Esimerkiksi kartassamme esiintyy kolmasti kadunnimi "Washington"; Bostonissa, Cambridgessa ja Somerwillessa. Ja vielä pahempaa: Cambridgessa kahdella eri kadulla on nimenä "Elm". The Location Finder käyttää tätä databasea tarkempaan paikantamiseen. Mikäli käyttäjä tarjoaa kadunnimeä, joka voi kuulua useammallekin eri kadulle on tarpeen esittää lisäkysymys, kuten: "Tarkoitatko mahdollisesti Cambridgen Bacon Streetia vai Bostonin Bacon Steetia?" Jotta tämä olisi niin helppoa kuin mahdollista on parasta käyttää kaikkein yleisimpiä paikanteita, jotta sen kautta mahdolliset vaihtoehdot suljettaisiin pois. Mikäli jokin tietty nimi esiintyy kahdessa kaupunginosassa yhtäaikaan on mielekästä käyttää kaupunkialueen nimeä, ja jos sama nimi esiintyy kahdesti tai useammin eri kaupunkien alueella on mielekästä käyttää kaupungin nimeä. Naapuruus määriytyy kadunnimen zip -koodista, vaikkakaan tällainen määrittely ei ole täydellistä. Kuitenkin se soveltuu hyvin tarkoituksiimme. Useimmiten naapuruussuhdenimikkeitä käytetään "paikallisissa" postitoimistoissa.

Figure 3: Central Square, 02139:

2.3. Inverted Phonebook

Invertoidun luettelon avulla voimme liittää puhelinnumeroita kadunnumeroihin. Rakensimme tämän datan itse inversoimalla "valkoisten sivujen" datan. Tämä vaati katuosoitteiden valkoisten sivujen alkuperäisen tekstimateriaalin uudelleenasettelua, mikä oli vaikeaa useistakin syistä. Valkoisilla sivuilla oli runsaasti tavumuotoja ja poikkeavia ilmaisuja. Esimerkiksi me löysimme 23 erilaista "Massachusetts" -variaatiota. Formaatti ei myöskään ollut vakioinen. Toisinaan listaus sisälsi ammattinimikkeitä (atty, arch), tai toisen puhelinnumeron tapauksia varten, jolloin puheluun ei vastattaisi, tai muunlaista informaatiota, kuten "toll free" ja "childrens phone". Meillä ei ollut typografista informaatiota, jolla olisimme kyenneet erottamaan nimet muusta informaatiosta. Lopuksi, osoitteet olivat epätäydellisiä niin, että niissä oli toisinaan vain kaupungin nimi, kadunnimi, tai jokin sellainen nimi, joka ei liittynyt millään tavalla kadunnimeen, kuten mm. "ostoskeskus". Uudelleenasettelun jälkeenkin oli työlästä määrittellä informaatiopaikanteita listauksesta, koska parhaimmillaankin saatavilla olivat vain kadunnimi, puhelinnumero ja kaupungin nimi. Kuten jo aiemmin todettiin, voi sama kadunnimi esiintyä useassakin kaupungin osassa. Sijainti voidaan kuitenkin määrittellä sen ylimääräisen informaation avulla, jonka saamme huomioimalla osa-alueenimikkeen. Niissä tapauksissa, joissa tämä ei ole mahdollista, on ohjelman käyttäjää pyydyttävä suorittamaan valinta.

3. Route Finder

The Route Finder löytää reitin kolmea pääperiaatetta noudattaen. Reitin on oltava helpposeurantainen,

järjellisen lyhyt, ja sen on oltava löydettävissä ennenkuin ohjelmankäyttäjä menettää kärsivällisyytensä. Pääperiaatteet konfliktioivat keskenään. Hyvin harvoin nimittäin on löydettävissä täysin suoraa reittiä, ja sellainen voisi löytyessäänkin vastata käytännössä umpikujaa. Joudumme ristiriitaan geometrisen yksinkertaisuuden periaatteen kanssa, koska emme halua asiakkaamme joutuvan eksyksiin. The Route Finderin tuloste on polku (path); järjestykseen asetettu luettelo katusegmenteistä, jossa ensimmäinen segmentti on polun alkukohta, ja viimeinen segmentti määräänpää, ja jossa jokainen segmenteistä on myös legaalisesti kytkeytynyt toinen toiseensa. Ohjelman ajankäyttö määriytyy ensimmäisestä välimatkanmäärittelystä. Ohjelma tarjoaa vaihtoehtona parasta, ensimmäisenä löytyvää vaihtoehtoa, joka on kohtuullisen hyvä ajatellen varsinaiseen autolla ajoon kuluvaa aikaa.

Figure 4: A Sample Route:

Paras ensin löytynyt vaihtoehto on parannettu versio karkeasti määritellystä lyhyimmästä välimatkasta. Itse haussa koeistetaan simulaatiivisesti joukkoa mahdollisimman lyhyitä polkuja, joista kehitetään versio, jonka kustannukset (costs) ovat alhaisimmat. Simulaatiossa jokaiselle polkukandidaatille on kaksi arvoa, joista toinen on sen tunnetun osan kustannusten summa, ja estimaatti-arvo, jota ei vielä tunneta, ja jolle haetaan alinta mahdollista arvoa legaalisia normeja noudattaen. Estimoinnin funktiona on kartesiolaisen välimatkan funktio, koska mikään reiteistä ei voi olla lyhyempi kuin suora viiva pisteestä toiseen.

Figure 5: All segments touched by search:

Kuten Elliot ja Lesk osoittavat, ei ole toivottavaa löytää reittejä, joissa välimatka on lyhin mahdollinen, koska tällöin on käännyttävä uusille katuosuuksille liian usein. Eliot ja Lesk ehdottavat 1/8 mailin kustannusta oikealle kääntymiseen, ja 1/4 vasemmalle kääntymiseen. Laajensimme heidän kustannusten systeemiään monin tavoin. Ensinnäkin me huomioimme kadun hyvyuden (goodness). Matkustaminen "super" -katua ei ole yhtä kallista kuin "average" -katua matkustaminen -puhumattakaan "bad" -kaduista. Toisaalta me pidimme jyrkkiä käännoiksi oikealle yhtä huonoina vaihtoehtoina kuin kääntymistä vasemmalle, koska niitä on vaikea perua, kun ne kerran on valittu. Kolmanneksi me vähensimme kääntymisestä aiheutuvia kustannuksia joissakin tapauksissa, kuten kääntymisessä yksisuuntaiselle kadulle tai sieltä pois, ja myös tapauksissa, joissa kääntyminen on legaalisti pakollista. Kääntyminen sillalle on myös vapaa, koska sillat ovat pääasiallisimpia suunnistautumisen ohjaimia, ja tukevat siten reitin seuraamista. Emme ole tutkineet näiden "reittien" reiteillä vaikutusta, emmekä sitä, ovatko ohjelman ehdottamista vaihtoehtoista poikenneet reitit sittenkin parempia kuin ehdotetut.

4. Describer

The Describer luo tektipohjaisen ohjeiston reitin seuraamista varten. Generoimme tekstiä kartan sijasta kahdesta syystä. Ensinnäkin systeemiä käytetään puhelimen välityksellä, mikä rajoittaa tulosteen puheeseen. Ja jos käyttäjillä olisikin kannettava graafinen terminaalimodeemilla, asettaisimme tekstin etusijalle, koska jotkut ihmiset eivät osaa lukea karttoja. Streeter ja Vitello suosittelivat tutkimuksessaan karttojenlukukyvystä erityisesti tekstiä. The Describer luo uuden representaation reitistä, käyttämättä polkua (path) itseään tähän tarkoitukseen. On kaksikin syytä siihen, että luodaan uusi sanallinen representaatio. Polun elementit (segmentit) ovat liian hienojakoisia tekstimuotoiseen representaatioon; ovathan segmenttien väliset suhteet yksiselitteisiä niin, etteivät ne vastaa käsitystämme "kadusta", joka mielessämme on kuitenkin osa kokonaista katujen verkostoa, tai jolla on kokemuksellisesti muitakin liittymiä kuin olla pätkä lyhyemmän ajoreitin muodostamaa jonoa. Lisäksi segmentit ovat suorista linjoista koostuvia, joten katu, jolla ei ole yhtään liittymää muihin katuihin alku- ja loppukohtaansa lukuunottamatta voi tulla esitettyksi segmenttien jonona, mikäli siinä on laajoja mutkia alku- ja loppukohtansa välillä.

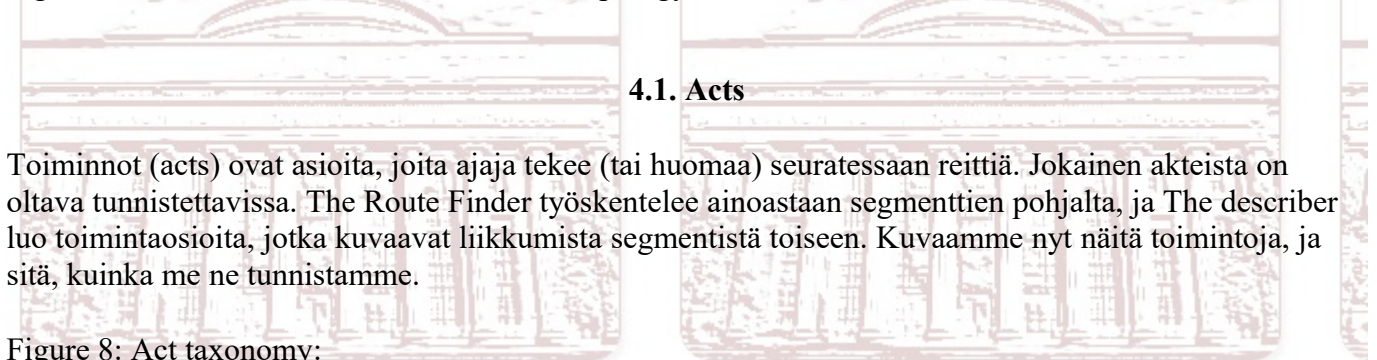
Figure 6: sample of directions:

"If Your car is on the same side of the street as 20 Ames street, turn around, and start driving. Drive all the way to the end, about eight of a mile. make a left onto Memorial Drive. Drive about one eighth of a

mile. After you pass Wadsworth street on the left, take the next left. It's an easy left. merge with main street. Stay on Main street for about ninety yards, and cross the Longfellow Bridge. You'll come to a rotary. Go about half way around it, and turn onto Cambridge street. Drive all the way to the end, about three quarters of a mile. Make right onto Tremont street. Drive about one half of a mile. After you pass Avery street on the left, take the next onto Boylston street. Stay on Boylston street for about one eighth of a mile. After you cross Washington street, it becomes Essex street. Keep going. Drive about one eighth of a mile. After you pass Ping On Street on the right, take the next right onto Edinboro Street. Number 33 is about one eighth of a mile down on your right side."

Me haluamme kuvata kuitenkin kadun kokonaisuutena. Toinen syy on se, että polku on vain topologinen struktuuri, vaikka luonnollisten ohjeiden tulisikin olla ilmaistavissa geometrisin termein ja kadun tyyppejä kuvailevina ilmauksina. Esimerkiksi T-liittymällä, tienhaaralla, ja vinoliittymällä on sama topologinen rakenne, mutta ne on silti kuvattava eri tavalla. Describerin stuktururi on itse matka, jonka varrella suoritetaan tiettyjä toimintoja.

Figure 7: T, fork, and exit all have the same topology:



4.1. Acts

Toiminnot (acts) ovat asioita, joita ajaja tekee (tai huomaa) seurattessaan reittiä. Jokainen akteista on oltava tunnistettavissa. The Route Finder työskentelee ainoastaan segmenttien pohjalta, ja The describer luo toimintaosioita, jotka kuvaavat liikkumista segmentistä toiseen. Kuvaamme nyt näitä toimintoja, ja sitä, kuinka me ne tunnistamme.

Figure 8: Act taxonomy:

* Boundaries

- Start
- Stop

* Straight

- Name Change

* Turn

- Enter
- Exit
- Merge
- Fork
- U Turn
- Rotary
- Ordinary

Ensimmäisin akteista on aloittaminen (start) ja viimeisin lopettaminen (stop). Ne on helppo erottaa toisistaan. Nimen muuttuminen (name change) toiminto vaatii ajajaa huomioimaan nimen muuttumisen, mutta ei enempää. Ero nimen muuttumisen ja käännöksen (turn) välillä on siinä, että edellinen sisältää kahden kadun leikkaamisen kymmenen asteen kulmassa toisiaan, ja että tällöin intersektiossa niihin ei ole toista sen nimistä katua kuin mitä leikkaava katu on. Kriteeri ei pidä sikäli paikkaansa, että on olemassa leikkauksia, jotka näyttävät meistä viittaavan nimenmuutokseen, vaikka kyse olisikin jyrkästä mutkasta (ainakin, mikäli asiaa tarkastellaan kartasta). Karttaamme onkin muutettu sikäli, että mutkien jyrkkyyttä on oioitu, millä ei olle merkitystä itse ajo-orientoitumisen kannalta. Käännöksistä (Turn Acts) "ENTER" ja "EXIT" viittaavat liittymistä ajoväyliin, joille on rajoitettu pääsy kaupunkialueelta. Usein kyse on nimettömistä ajorampeista, jotka me olemme kuitenkin halunneet identifioida mm. siksi, että ne ovat usein ainuita tapoja liittyä pääajoväyliin. Ajoramppi eroaa siten kaduista yleensä. Erilaiset, mm. useamman kuin kahden kadun väliset liittymät olemme myös halunneet identifioida erikseen siksi, että ajokäyttäytyminen niissä riippuu suuresti lähestymissuunnasta. Niihin liittyvät seuraavat karakterisoivat

piirteet.

1. Ajettavilla ja hakeuduttavilla kaduilla on eri nimet
2. Vain yksi katu on legaalisti mahdollinen.
3. Kääntökulma on pieni.
4. Ainakin yhdeltä vaihtoehdoista kaduista on mahdollista päästä päämääränä olevalle kadulle.
5. Kaikki kadut tekevät vain vähäisiä käännoiksi kohti päämääränä olevaa katua.

Y- tyyppisissä risteyksissä (FORK) on vähintään kaksi tarjolla olevaa vaihtoehtoa, ja yleensä käännot ovat niissä loivia. Huomattakoon, että mikäli liittymän vaihtoehtoihin kuuluu rampeja, ovat ne identifioituna "EXIT" -karakteristikalla. U-käännoiksi (U Turns) on Bostonissa kahtakin tyyppiä, joista toinen voidaan tehdä keskellä yksisuuntaista katua. Ohjelman tarjoama ajo-ohjevaihtoehto ei sisällä tällaisia illegaalisia ajotapahtumia. Toinen U-käännotyyppi on sellainen, jossa siirrytään keskiviivan yli vastakkaisuuntaiselle ajokaistalle. Tavallisimmin ajaja menettelee kuitenkin niin, että hän kääntyy vasemmalle jollekin lyhyelle, nimettömälle tienpätkälle, ja kääntyy siitä vasemmalle uudestaan; tätä tapaa me nimitämme U-kääntymisaktiksi (U Turn Act). On hyvin tärkeää, että käännon suoritus tapa tunnetaan, koska muutoin voi olla vaikea mieltää, mitä merkitsee kaksinkertainen vasemmalle kääntymisen samalla kadulla. Ehkä ongelmallisimmat Bostonin tieliittymiin kuuluvista ominaisuuksista ovat liikenneympyrät (ROTARY), joka on ympyränmuotoinen, yksisuuntainen risteyksen osio. Liikenne ohjautuu liikenneympyrään tavallisimmin tangentiaalisin liittymä, ja ympyrästä ohjautetaan ulos toiselle, tangentiaalisesti ympyrään liittyvälle kadulle. Liikenneympyrät ovat ajajalle sikäli ongelmallisia, että niissä liittymä ja erkaneminen tapahtuvat hyvin lyhyellä matkalla, ja niissä vaaditaan nopeita ohjausliikkeitä. Olemme merkinneet karttaamme kaikki liikenneympyrät.

4.2. Cues

Samalla kun The Describer koostaa ajotoimenpiteiden jonoa se myös liittyy niihin ajoa helpottavia vihjeitä (cues), joiden avulla ajajan on helpompi suunnistautua. Erotamme neljä erilaista vihjetyyppiä. Toimintavihjeet (Action cues) kertovat ajajalle sen, milloin ajotoimenpide on suoritettava. Konfirmatoriset vihjeet (Confirmatory cues) kertovat seikoista, joita ajaja voi nähdä reitin varrella. Varoittavat vihjeet (Warning cues) kertovat ajajalle mahdollisista, odotettavissa olevista virheellisistä toiminnoista. Varoittavat vihjeet ovat vaarinotettuina konfirmatorisia vihjeitä. Epäonnistumisvihjeet (Fairlure cues) kertovat, mitä toimenpiteiden epäonnistumisista seuraa, kuten: "Jos näet tätä tai tuota, olet ajanut ohi". Tavallisin toimintavihje on annettu kadunnimi, kuten: "Käänny oikealle Tremont Steetille". Vihje voi olla vaikea noudattaa siksi, että katuviitta puuttuu. Hyvin voimakas vihje liittyy kadun päättymisiin, ja se ilmaistaankin yleensä: "Sinun on pakko kääntyä...". Edettyä matkaa kuvaavaa vihjetä on vaikea käytännössä hyödyntää, koska ihmisillä on yleensä heikohko käsityskyky kuljettujen matkojen pituuksista. Silti kuljettua välimatkaa tarjotaan alemman asteisena vihjeenä, koska se on helposti laskettavissa, ja se voi auttaa joitakin ihmisiä. Kuljettu matka ilmaistaan jaardeina, mikäli se on vähemmän kuin 1/16 mailia, ja muissa tapauksissa mailin osina, koska ihmiset ovat tottuneet ilmaisemaan välimatkoja siihen tapaan. Emme käytä mailin kymmenyksiä, koska useimmat ihmiset eivät sellaista osittelua käytä, ja koska sellainen edellyttää aritmeettisia päässälaskusuorituksia, jotka saattaisivat olla vahingollisia ajon kannalta. Emme jaa segmenttejä osiinsa, koska kyseessä eivät olisi määritellyt käsitteet. Kyseessä olisivat tällöin tapaukset, joissa huomioitaisiin kaikki mahdolliset intersektiiviset liittymät, ja joille ei tulisi kuitenkaan kääntymään, ja joista matka jatkuisi kenties edelleen kohti risteystä, josta käännettäisiin lopulta toiselle kadulle.

Figure 9: Is the distance between "A" and "B" one block, or two?":

Ajajan olisi työläästä laskea mielessään kuljettuja blokkeja sekaantumatta laskuissaan. Emme luota etäisyyteen tai kuljettujen blokkien määrään hyvinä vihjeinä, vaan ennen kääntymistä tarjottuun kadunnimeen, joka saa ajajan valppaammaksi. Vihjeenä tämäkään ei ole riskitön, koska ohittaessaan oikean risteyksen vahingossa ajaja saattaa kuitenkin hakea edelleen sitä kaikkialta katseellaan, ja harhautua siten reitiltä. Jotta riski olisi mahdollisimman vähäinen, tarjotaan ajajalle se kääntymistä edeltävä kadunnimi, joka on hänen käyttämänsä ajokaistan puolella, ja siten ajaja voi tarkkailla vain

ajokaistansa puoleisia katuviittoja. Mikäli ajaja ei voi vihjeenä tarjottua, kääntymistä edeltävää kadunnimeä havaita, niin hän voi silti nähdä sen risteuksen, josta tulee kääntymään. Konfirmatoriset vihjeet (Confirmatory cues) liittyvät suurten katujen risteyksistä tiedottamiseen, kuten myös rautateiden, alikäytävien jne. rekisteröintiin matkan varrella. Tapauksessa, jossa ajajan on käännettävä valtaväylältä rampille, ei vihjeitä lisätä, koska rampi on käytetyn kaistaston puolella, ja siten tapauksiin ei liity tulkinnallisuutta.

4.3. Generating text

Jokaiselle aktille rutiini kokoaa pari- kolme lausetta valikoiden ajon- ja reitin kannalta soveliaimmat. Kuvaamme joitakin tekstin generoimiseen liittyviä аспектеja:

Figure 10: Generator for rotary:

(defun disc-seg-rotary (act)

(list

(make-sentence

"you'll" "come" "to"

(make-np-constituent ('("rotary") :article :indefinite))

(make-conjunction-sentence

(make-sentence

"Go" (rotary-angle-amount (get-info act rotary-angle))

"way" "around" (make-anaphora nil "it"))

(make sentence

"turn" "onto" (make-street-constituent (move-to-segment act act))))

(defun rotary-angle (angle)

(selector angle <=

(45 ('("just" "a" "little"))

(135 ('("about" "a" "quarter"))

(225 ('("about" "three" "quarters" "of" "the"))

(360 ('("almost" "all" "the"))))

Matkan aloituskohtaa (START) varten tekstin generoiminen on vaikeaa, koska siitä on vaikeaa spesifioida suuntaa. Emme käytä absoluuttisia suuntia, koska useimmat ihmisistä eivät niitä tunne. Jos meillä olisi maamerkkeihin perustuva database, voisimme joissakin tapauksissa käyttää relaatiotietokantaa, ja antaa esimerkiksi ohjeen: "Aja joen suuntaan" (Riippuu kuitenkin suuresti ajajan sijainnista, näkeekö hän tiettyjä maamerkkejä vai ei, ja siksi niiden näkyvyys olisi kenttäolosuhteissa ennakoitettava, tai annettava ohje: "Jollet jo näe, niin tulet pian näkemään oikealla puolellasi joen", suom. huom.). Käytämme lähtöosoitetta, koska se myös määrää käytetyn kadunpuoliskon, ja siten myös ajosuunnan. Olisimme voineet käyttää esimerkiksi: "Jos autosi on samalla puolella kuin... aja menosuuntaasi". Tällainen vaikuttaa kömpelöltä, joten valitsimme negatiivisen ohjeen: "Jos autosi on samalla/vastakkaisella puolella kuin... vaihda ajosuuntaa". Yksisuuntaisia ajoväyliä varten voisi sanoa: "Jatka ajoa", joskin ehdottomasti lisäyksellä "tiedätkö, että katu on yksisuuntainen", sillä muussa tapauksessa ajaja ei kenties luottaisi ohjaavaan systeemiin, koska kehoitus "jatka ajoa" voisi monien mielestä merkitä sitä, ettei opastava ohjelma kykene tunnistamaan yksisuuntaisia liikenneväyliä, ja aloittaa ohjauksen

kaikille samalla tavoin (Ohjetta ei kenties voitaisi antaa, mikäli ajaja ajaa liikenneympyrässä, tai lisätä ilman koomista sävyä huomautusta: "Tiedätkö, että ajat liikenneympyrässä"; ja ehdottomasti mahdotonta olisi, että ohjelma kysyisi ajajalta: "Miksi autosi on tällä hetkellä hinauksessa?", suom. huom.). Eräs yksinkertaisimmista generaattoreista on rotaatiopistegeneraattori; se on vaikeasti kuvattavissa ja seurattavissa, koska siihen ei liity lisäinformaatiota. Emme voi odottaa ajajien suorittavan kulmamittauksia rotaatiopisteen ympäristöä koskien. Rotaation segmentit voivat olla nimettömiä, tai niihin voi liittyä useitakin nimiä, ja rotaatiopisteeseen itseinkin voi liittyä useita eri nimikkeitä. Se voi olla alueen nimi, kuten Leverett Circle, mutta tällaista nimeä ei ole varsinaisessa databasessa, eikä se ole yhdenkään kadun nimenä. Generaattorin tuloste esitetään kuviossa 6 (Figure 6.) Generaattori tuottaa kaksi lausetta, joista toinen on kahden lauseen yhdiste. Välimatkat rotaatiopisteestä sen ympäristöön on konvertoitu absoluuttisesta karttakulmasta englanninkieliseksi approksimaatioksi. Luoduilla ohjeilla on syntaktinen rakenne vain siksi, että lauseiden koostuma on generoitu yleisrakenteeseen. Niinpä funktio "make-np constituent" luo artikkelin ja substantiivin välille yhdenmukaisuuden, ja funktio "make-sentence" varmistaa isojen alkukirjainten ja välimerkkien oikeellisuuden. Teksti lähetetään suoraan puhesyntetisaattoriin, joka vaatii välimerkkejä oikean sanojen intonaation vuoksi. Funktio "make-anaphora" ei palvele mitään tehtävää tällä hetkellä, mutta on ajateltu, että sitä käytettäisiin tulevaisuudessa luomaan luonnollisessa keskustelussa esiintyvää intonaation vaihtelua.

4.4. Comparison

Vertaamme kuvaustamme Streeterin ja hänen kolleegojensa esittämiin kuvauksiin. Streeterin kuvauksen on tarkoitettu ymmärrettäväksi ja reagoitaviksi reaaliajassa, aivan kuin ne olisi lausunut takapenkillä istuva opastaja. Tosiasiassa ohjeet kuitenkin nauhoitetaan, ja ajaja kuuntelee tarvittavan osan nauhaa ajaessaan. Tällainen vuorovaikutusmuoto vaatii kuitenkin uudenlaisia ohjeita, koska ne tulevat kuulluksi ajotapahtuman aikana, ja siksi ohjeiden olennaisia osia on voitava tarvittaessa toistaa ajajalle, jotta hän voisi muistaa kokonaisuuden. Meidän kehittelemässämme vuorovaikutuksen muodossa oletamme, että ajaja kirjoittaa ohjeet ylös ennenkuin aloittaa ajamisen, josta syystä toisto ei ole tarpeen. Käyttäjä voi pyytää toistoa, mikäli hän ei ole ohjeita ymmärtänyt. Streeter ja hänen kolleegansa luokittelevat ajokäännökset T-käännöksiin (T Turns), mutkikkaisiin katuliittymiin (Complex intersections), sarjoihin liittyviä lyhyellä matkalla (turns in short succession). Heidän T-käännöksensä on vihjeistetty (meidän tapaan) "pakotettu kääntymään". Meillä ero tavallisen käännöksen ja T-käännöksen välillä on siinä, että emme liitä jälkimmäiseen vihjetä koskien mahdollista epäonnistumista, kuten eivät hekään liitä. Me emme erottele mutkikkaita katuliittymiä yksinkertaisemmista, vaikka pitäisi. The Route Finderin tulisi osata karttaa sellaisia, ja The Describerin osata niistä varoittaa. Heidän ohjeensa ovat toisinaan strukturoituneempia kuin meidän. He asettavat ryhmäksi peräkkäiset käännökset, ja heidän "jatkakaa" -vihjeensä on sama kuin meidän "nimen muutos", joskin se liittyy heillä usein ajossa seuraavaan käännökseen. Pidämme tärkeänä, että luotaisiin korkeamman tason struktuureita, ja huomautamme, että Streeterin työryhmä työskenteli käsin, eikä käyttänyt ohjelmaa apunaan, ja siten heillä oli paremmat mahdollisuudet luoda korkeamman asteen struktuureita kuin meillä oli. Väitämme, että meidän ohjeemme ovat luonnollisempia kuin Elliotin ja Leskin ryhmällä, vaikka meillä ei olekaan siitä todisteita. Jätämme tämän seikan lukijain arvioitavaksi. Ovatko ohjeemme selkeitä? Tiedämme ihmisten olevan kykeneviä niitä seuraamaan, vaikka emme ole tehneetkään systemaattisia kokeita. Christopher Riesbeck kirjoitti ohjelman (MCMAP), jolla ohjeiden selkeyttä voi testata. Ohjeemme eivät ole hyväksyttäviä sen kriteereiden pohjalta. Osin tämä johtuu siitä, että me käytämme ilmauksia, joita kyseinen ohjelma ei tunne, kuten liikenneympyrät. Osin tämä johtuu myös siitä, että me käytämme välimatkojen ilmauksessa maileja, joita Riesbeckin ohjelma ei hyväksy epämääräisinä. Käytämme maileja kuitenkin vain approksimaationa, vihjeenä siitä, milloin tiettyjä maamerkkejä tulisi etsiä katseella. MCMAP:n heikko syntaktinen kyky ei voi huomata tätä merkitysdimensiota. Käytämme myös ilmaisua "Aja tietä perille saakka", jota Riesbeck pitää proseduraalisena operaattorina. Koska ihmiset kuitenkin hyväksyivät ohjeemme, se osoittaa Riesbeckin asettaneen tulkintasääntönsä liian ankariksi, tai ehkä hän ei ole kyennyt muokkaamaan välineestään riittävän voimakasta.

5. Discussion

Direction Assistance kaltaisia tuotteita on alkanut ilmaantua markkinoille. On raportoitu, että Sunnyvalessa, Kaliforniassa on tuotettu ETAK, jolla kyetään arvioimaan liikkuvan auton asema kytkemällä matkalaskin auton pyörän tekemien kierrosten lukumäärään, ja rekisteröimällä sen ohjauspyörän kääntökulmat, jota informaatiota verrataan tallennettuun karttaan. Auton ohjaspaneelista on nähtävissä auton kulloinenkin asema. Tämä Navigator-ohjelma ei anna ajo-ohjeita, mutta se epäilemättä voitaisiin muokata sellaisia antavaksi. Meidän tuotettamme muistuttaa enemmän DriverGuide, jonka ovat luoneet Karlin ja Collins (myös Sunnyvalessa), joka tulostaa ajo-ohjeet paperille California Bayn alueella.

5.1. Better databases are required

Kirjoittajat viittaavat ohjelman käyttämän maantieteellisen aseman pienuuteen, vuorokaudenajasta riippuviin liikennesäännöskäytäntöihin, osoitteistojen kirjavuuteen, kaksinkertaisiin kadunnimiin tapauksissa, joissa kadunnimi riippuu lähestymissuunnasta jne.

5.2. Applications

Suunnittelimme alunperin Direction Assistance -ohjelman matkailijoiden opastukseen, koska Bostonin kaduilla vierailija joutuu hämmennyksen valtaan. Matkailijoiden ohjausta menetelmällämme voisi ajatella sovellettavaksi kaupungeissa. Kuitenkaan läheskään aina matkailija ei tiedä tarkasti päämääräosoitetta tai puhelinnumeroa. Saatta olla, että päämääräosoite on naapurikaupungissa, josta syystä matkailijat voisivat hyötyä enemmän haettujen henkilöiden nimien pohjalta tapahtuvasta hakumenettelystä. Tällainen ei ole mahdollista muokkaamatta ohjelmansisäistä vuorovaikutusta mutkikkaammaksi. Elintarvikekuljetuksissa käytettyjen ajoneuvojen ohjauksessa on ongelmana Bostonissa se, että tietyt kadut ovat suljetut kaupalliselta liikenteeltä. Databasea voi laajentaa niin, että tämä seikka tulee huomioiduksi. Julkiseen ohjaukseen sovellettuna ongelmana olisi se, että tavoiteltujen ihmisten kotipuhelinnumerot ovat yksityisiä. Ihmiset saattavat hyvinkin toivoa, etteivät heidän kotipuhelinnumeronsa tule julkisuuteen, sillä niiden kautta ei-toivotut henkilöt saisivat heidän katuosoitteensa tietoonsa. Ei-toivottu puhelimenkäyttäjä on huomattavasti helpompi sulkea kuulumattomiin kuin ovelle jo ehtinyt kotirauhan häiritsijä.

Acknowledgements

Ohjelman perusversion kirjoitti Dinarte R. Morais talvella 1985. Palkkasimme hänet dekoddaamaan DIMETiedostot, tekemään window -systemiin soveliaan vuorovaikutusverkon, ja testaamaan käytetyt käsitteet. Loimme database -kokonaisuuteen laajennettuja käyttömuotoja ja muokkasimme siihen soveliaaksi Craig Stanfillin kirjoittamaa string -vertailuohjelmaa. Charles Lieserson teki pääasiallisimmat muutokset algoritmien etsimiseen Route Finderiin. Fletch McCellan Phonebook Corporationista lainasi meille käyttöömme raan puhelinluettelo-databasen. Työmme ei olisi voinut onnistua ilman Brewster Kahlen ohjausta ja myötätuntoa. Tätä raporttia selkeyttivät Janet Cahnin, Mike Hawleyn, Margaret Minksyn ja Chris Schmantdin tekemät huomautukset. Kiitämme heitä kaikkia. Työn ohjelmatukena olivat MIT DARPA Space ja Naval Warfare Systems Command kontaktinumeroineen N00039-89-C-0406 ja N00039-89-PRDX002 ja Nippon Telegraph ja Telephone Public Company. Hardware -tukea tarjosivat Symbolics ja Digital Equipment Corporation. Vaikka tälle työlle onkin ilmoitettu vain kaksi tekijää, niin työ oli lähinnä ryhmätyötä, ja kirjoittajista eniten vaikutti Tom, joka menehtyi ennen kuin näki työnsä kuvattuna. Vaikka kuvaus onkin muistolauseeksi vaatimaton on se parasta, mitä voin tällä hetkellä antaa.

References

[1] Geographic Base File GBDF/DIME:: 1980 Technical Documentation. U.S. Department of Commerce, Data Users Service Division, 1980.

[2] James R. Davis. Giving directions: a voice interface to an urban navigation program. In Proceedings of 1986 Conference, pages 77-84, American Voice I/O

Society, sept 1986.

- [3] James R. Davis and Julia Hirschberg. Automatic generation of prosodic support for discourse structure. In Proceedings of the Association for Computational Linguistics, page (submitted), 1988.
- [4] R.J. Elliot and M.E. Lesk. Let Your Fingers Do the Driving: Maps, Yellow Pages, and Shortest Path Algorithms. Technical Report unpublished, Bell Laboratories, 1982.
- [5] R.J. Elliot and M.E. Lesk. Route finding in street maps by computers and people. In Proceedings of the National Conference on Artificial Intelligence, pages 258-261, 1982.
- [6] W. Daniel Hillis. The Connection machine. MIT Press, 1985.
- [7] Kevin Lynch. The Image of the City. MIT Press, 1960.
- [8] Ronald Rosenberg. Mapping out a new idea. The Boston Globe, 39, 1987. February 17.
- [9] Lynn A. Streeter and Diane Vitello. A profile of driver's map reading abilities. Human factors, 28:223-239, 1986.
- [10] Lynn A. Streeter, Diane Vitello, and Susan A. Wonsiewicz. How to tell people where to go: comparing navigational aids. International Journal of Man/Machine Systems, 22(5):549-562, May 1985.

Strategies for selecting a fixed palette of colors

This work was supported in part by International Business Machines, Inc., **Nathaniel Jacobson** and **Walter Bender**

*Consultant on chromatology, 131 Kilsyth Road, Brookline, Massachusetts 02146
Massachusetts Institute of Technology, Media laboratory
20 Ames Street, Cambridge, Massachusetts 02139*

suom. Timo Kinnunen 3.3.1989

Pelastettu 3.5 Levykkeeltä NU.EXE Ohjelmalla 12.4.1989 T. Kinnunen

Abstract

Tämä tutkimus tarjoaa strategioita valitessa visuaalisesti artikuloituille väreille kiinteitä paletteja. Jotta voitaisiin saavuttaa värienluomiskapasiteetti tietokoneessa millä tahansa tasolla, on luotava jonkinlainen looginen näytöllinen esitystapa tietokoneelle mahdollisesta värien valikoimasta. Useissa näyttösovelteissa väripaletti on rajoitettu. Tavallisin ratkaisu on laajentaa näytönohjaimen suoritteen laajuutta yhdenmukaisesti (uniformly) kvantisoidulla lisärakenteella; sellainen on yhdenmukainen vain väriprosessorin suhteen, eikä ole inhimillisen visuaalisen systeemin mukaisesti järjestynyt esiintyvien värien keskinäiset suhteet huomioiden. Niinpä tuloksena on paletteja, jotka eivät ole tässä suhteessa yhdenmukaisia (non-uniform). Aloitimme lähestymisemme visuaalisesti tehokkaasta väriavaruudesta, keinoamme "the Logical Visual Display", joka on kehitelmä "Munsell" -systeemistä, joka perustuu visuaalisen kvantisoinnin yhdenmukaisuuden approksimointiin. Tällaisen väriavaruuden ortogonaalinen rakenne mahdollistaa kaikkien esiintyvien väridimensioiden suhteellisen resoluution säätämisen. Meidän käyttämässämme näyttötilassa siten on mahdollista säätää valoisuusarvoja niin, että esiintyvät eri valoisuustasot ovat keskenään yhdenmukaisia, ja mikä siten eroaa tavasta, jossa värejä sekoitetaan keskenään, tai lisätään paikallisesti. Niin järjestyneinä värit ovat perustavanlaatuisia rakenne-elementtejä suunniteltaessa kommunikatiivisessa yhteydessä toimivaa väriaistimusta. Olemme havainneet, että tietojenkäsittelyllisesti tällä tavoin jäsenyvien väripalettien käyttö ei vaadi laskennallista lisätilaa, koska tehokkaat hakustrategiat, kuten totuuspuurakenteet (binary trees), voivat toimia riippumatta yhdenmukaisesta, tai symmetrisestä näytönohjaimen lisärakenteesta toteutetusta kvantisointitavasta.

1. INTRODUCTION

Tavallisin tietokoneen näytön arkkitehtuuri on ruutumuistitila, joka on yhteydessä graafisen kontrollikortin piiriin, ja joka tuottaa näyttöön tietyn värien valikon. Näytön laatu on suorassa suhteessa niin valikkossa esiintyvien vaihtoehtojen lukumäärään kuin myös jokaisen valikkoväylän (entry) sisältämään bittimäärään, esimerkiksi 256:n värien paletti/ 16 -miljoonaa bittiä. Kahdeksan bitin intensiteetti primaaria kohti on usein käytetty väriäytöissä, ja siten se vaatii 24 bittiä peruselementtiä kohti tuottaakseen jatkuvana koettavan värikuvan. Vaikka onkin usein toivottavaa työskennellä paletilla, jonka koko ja laajuus ovat rajoitettuja, niin palettien adaptoitumista kuvien mukaisesti on vaadittu jo vuosikausia. Nykyään sekä hardware,- että software -arkkitehtuurit kehitetään pääsääntöisesti kiinteille, eli ei-adaptiiviselle periaatteelle rakentuviksi. Tutkimuksemme koskeekin siten yleistä ongelmaa liittyen kiinteiden palettien tehokkaaseen valintaan yleiseen näyttökäyttöön.

1.1. Application of color

Aina 70 -luvun puolestavälistä lähtien väristä on tullut näytöllisesti helposti esitettävä ja siinä käsiteltävä. Teknologian kehitys lupaa, että pian värejä voidaan yhtä helposti myös tuottaa. Värien käyttäminen on varmasti ollut leimaa-antavin piirre tietokonegrafiikkateollisuudessa. Sen sovelle on organisationaalinen elementti, mittaamisen yksikkö, kirjallinen kuvaus, ja toisinaan metafora. Kolmekin kiinnostunutta taho suuntautuu värien mahdolliseen tietokonetyöstöön: taide, fysiikka ja psykologia. "Taitelijat ovat olleet tekemisissä värikombinaatioiden ekspressiivisten piirteiden kanssa, kuten myös puhtaiden värien. Fysiologit etsivät selityksiä ja taksonomioita, kun taas psykologit ovat suuntautuneet tutkimaan värien perseptuaalisia ominaisuuksia ja niiden herättämiä affektiivisia kokemuksia [1]. Edellämainittujen perspektiivien toiveita ei ole toteutettu yleisesti tämän päivän tietokonegrafiikan sovelluksissa. Taiteilijoiden on nähty ikään kuin työskentelevän tiettyjä avainteemoja seuraten, kuten että he ehdottomasti kaipaavat ohjelmalta ominaisuutta, jolla luoda värejä ja varjoja annetusta kiinteästä valikosta. Fyysikkojen toivomaa mahdollisuutta analysoida värien spektriä luotettavasti ei voida mitenkään suhteuttaa tapaan, jolla normaalikansalainen värejä havaitsee. The Logical Visual Display, jota me tässä kohden suositamme, voisi auttaa integroimaan väriprosessoinnin kenttää niin, että kaikki voisivat hyötyä tarjoutuvasta proseduurista. Paletinvalinnan kriteereitä valitessa tärkeitä tekijöitä ovat yhdenmukaistuvuus, ortogonaalisuus, ja seurantakyky. Yhdenmukaistuva väriavaruus luonteensa mukaisesti mahdollistaa saatavissa olevien palettien tehokkaamman käytön. Väriavaruuden ortogonaalisuus paletinvalintaprosessiin liittyen lisää palettien hyödynnettävyyttä. Mikäli käytössä ei ole tehokasta soveltuvan väriavaruuden etsintämekanismia on palettien käytettävyys, kuten vaihtoehtoisten mallien saatavuus, vähäistä.

1.2. Optimization of visual detection

Hyvin tutkittuja värihavainnon aspekteja ovat detektio ja identifikaatio, jos tällainen ero tehdään. Mikäli tarjolla oleva värivalikoima on hyvin rajallinen, se merkitsee valtavaa akuuttien erottelutapahtumien lisääntymistä. Palettivalikko voisi olla esimerkiksi Boyntonin "yksitoista väriä, jotka eivät ole koskaan riitasoinnussa keskenään" [2]. On varmasti tärkeää tarjota riittävästi visuaalista kontrastiutta, vaikka eräät tutkimukset viittaavatkin siihen, että visuaalisen kontrastin maksimointi ei ole optimaalinen tapa ratkaistaessa esimerkiksi sellaisia tehtäviä, kuten lukeminen [3,4] ja hajanaisen visuaalisen massan rypäisiin koostaminen (decluttering) [5].

1.3. Rendering of natural images

Luonnolliset kuvat voidaan muokata sisäisesti oikeissa sävysuhteissa tietokoneelle (render) käyttämällä pientä määrää värejä hyödynnettäessä adaptiivista vektorikvantisointia [6,7,8,9]. Useimmat värivalokuvat voidaan redusoida 256:sta 1024:ään väriin tällaisessa muokkauksessa, ilman, että kuvan laatuoso huomattavassa määrin heikkenisi. Houle huomioi yhdenmukaisten väriavaruuksien tehokkuuden kvantisointiprosessissa [7]. Harvempien värien käyttö vaatii tihennyksen hyödyntämistä (dithering), ja (kvantisointi) virheen erottamista, jotta ei luotaisi virheellisiä kuvansisäisiä rajakohtia, ja ne korvataan siksi korkean frekvenssitason kohinamalleilla. Kohinamallit voidaan minimoida välttämällä niiden (keskinäisten) arvojen muutoksia, mikä tarkoittaa kuvassa akromaattisia muutoksia.

1.4. Expressive color communication

On tehty suuri määrä tutkimuksia koskien sitä, mikä olisi ergonomisesti soveliainta valitessa värejä erottelua silmälläpitäen. On myös suuri määrä kirjallisuutta, joka kuvaa fyysikaalisen energian ja siitä saadun aistimuksen välistä eroa [10]. Näytöissä käytettyjen värien psykofysiologisia vaikutuksia koskevaa tutkimusta, erityisesti kirjaimistoon liittyvää, ei ole juurikaan ennen tehty. Robertson [5] huomioi, että väri voi olla "sopiva tai stressaava valintatilanteessa". Onko siten olemassa mekanismeja, joita tulisi huomioida valitessa värejä, joita sovelluksissa käytetään? Kuinka sisältöihin tai kommunikaatioon soveltuvia värejä valittaisiin, jotta voitaisiin ilmaista mielialoja tai elämänsenteitä? Postuloimme, että on mahdollista valita juuri sellaisia värejä, jotka soveltuvat lähetettävän viestin

sisältöön; asettuvat sen laajuuteen, keston, ja kehonkieleen. Esimerkkinä olkoon vahvistussuhde, jollainen esiintyy mm. tien vieressä sijaitsevien suurten mainostaulujen ja samanaikaisesti niiden ohiajavissa autoissa kasvokkain käytyjen keskustelujen välillä. Tahtoisimme lisätä näitä näkökohtia relevantteina faktoreina värien ergonomisuuden määrittelyyn, ja valittaessa rajoitettuja väripaletteja.

1.5. Our strategy: visual color

Tutkimuksessamme korostamme eroa, jollainen on värillisten ingredienttien ja vastaanotetun visuaalisen väriaistimuksen mittauksellisuuden välillä, ja kuinka tästä voidaan päätyä paletteihin, jotka ovat sekä tehokkaita -että hyödyllisiä. Mitä tietokoneella tehtyihin peruskartoituksiin tulee, eroaa meidän tutkimuksemme aiemmin tehdyistä töistä tällä alueella. Nojaamme voimakkaasti Munshellin [11] esittämään visuaalisten värien systeemin kuvaukseen, joka on myös mainiosti esitetty sellaisissa lähteissä, kuin Meyer ja Greenberg [12]. Vain vähän yli vuosikymmen sitten Munshellin systeemi muotoiltiin tietokoneelle, joskin sen käyttö on jonkin verran rajoitettua. Se, mikä on estänyt Munshellin systeemin yleisempää käyttöä on sen epätarkkuus ja että sen käyttö on tietokoneessa kallista lukuunottamatta joitakin sen esoteerisimpia sovelteita. Kuitenkin on sen periaate (honing in) värien tuottamisessa hyvin adekvaatti. Munsell saattoi olla vaikea kontrolloida vielä kymmenen vuotta sitten, mutta tämän päivän standardeilla voidaan vähentää systeemin ylikuormitusta huomattavasti tuottaessa mitä tahansa väreihin liittyviä ratkaisuja. Se, että kyetään ottamaan käyttöön tehtävienratkaisuissa tarvittava tietojenkäsittely valitessa ja käytettäessä värejä, tekee menetelmästäme "kypsan hedelmän" otettavaksi käyttöön. Olemme laajentaneet Munshellin esittämää systeemiä kolmella tärkeällä tavalla: (1) laajentaneet visuaalisen värin kartografiaa, (2) muuntaneet laskennallisia malleja, ja (3) laajentaneet vastaanotettavissa olevien vastakkaisten värien tuottamien seuraamusvaikutusten ohjailtavuutta. Työstäessämme ideaa loogisesta visuaalisesta näytöstä, mikä meidän termeissämme liittyy visuaalisten värien kartografiaan, ovat ehdottamamme ja testaamamme muodot vaatimattomia, ja riittämättömiä tarkan kvantitatiivisen mittauksellisuuden tarpeisiin. Kuitenkin saamamme värinkäsittelyn tulokset ovat hyvin kaukana siitä, mitä voidaan saavuttaa visuaalisesti disorientoituneilla perusnäytöillä (default displays) tämän päivän tavallisessa käytännössä. Olemme varmoja siitä, että tutkimustoiminta tulee juuri tällä alueella kasvamaan. Toivomme, että meidän- ja visuaalisen väriaistimuksen pioneerien tekemä työ (joiden käsitteet stimuloivat meitä) tulisi vuorostaan stimuloimaan tietokonegrafiikan insinöörejä ja ohjelmansuunnittelijoita yhteistyössä estetiikan,- psykologian, ja kromatologian opiskelijoiden kanssa, jotta voitaisiin kehittää muunnettuja malleja ja käytännön työvälineitä.

2. LOGICAL VISUAL DISPLAYS

"Everything should be made as simple as possible, but not simpler."
- A. Einstein

The Logical Visual Display soveltuu (1) tuottamaan käyttökelpoista informaatiota valituista väripositioista suhteessa etäisyyteen ja suuntaorientaatioon; ja (2) tuottamaan informaatiota sellaisista praktisista formaateista, jotka eivät ole liian mutkikkaita luettaviksi silloinkin, kun se merkitsisi kuvakompressiota ja kuvainformaation laajennusta koskien kuvatilaan tehtyjä lisäyksiä. Analogia tällaiseen löytyy sellaisista maantieteellisistä kaksiulotteisista kartoista, joihin on lisätty referenssejä globaaleista suunnistautumista helpottavista etäisyyksienmäärittelyistä ja ilmansuuntiin kytkeytyvistä orientaatioista. Onnistuessaan karttakuvassa ei ole havaittavissa häiritsevästi muotoihin liittyvää rikkoutuneisuutta; The Logical Visual Display pyrkii myös regulariteettibalanssiin ja konsistenssiin, sekä minimaalisiin värinasemoinnin virheisiin.

2.1. Irregularity of process spaces

Tavanomaisimmin paletti on derivoitu näytönohjaimen lisämuistin suoran kvantisoinnin menetelmällä. Kolmidimensionaalinen avaruus, jossa akselit vastaavat useimmin punaista, vihreää ja sinistä on jaettu halutunlaisiksi koodinumeroiksi. Tällainen tapa on tyydytetty niin sanotuksi 3-3-2 paletiksi; kolme bittiä

punaiselle, kolme bittiä vihreälle, ja kaksi bittiä siniselle. Myös toisenlaisia kombinaatioita, kuten 2-4-2 on yleisesti käytössä. Tällaisten palettien etuna on yksinkertaisuus; paletin haku voi tapahtua konstantissa ajassa. Jotta lähin palettihakuväylä löytyisi vaatii tarvittava RGB arvonmääritys viidestä seitsemään boole- ja bittiarvon erillistä laskutoimitusta (ANDs, ORs, Shift). Vaikka tällaiset paletit tekevätkin värien käsittelyn helpoksi tietokoneelle ovat sen tuottamat haitat käyttäjän kannalta lukuisat: visuaalisesti paletti ei ole yhdenmukaistuva, ja siksi värien hakeminen on vaikeaa. Tällaiset paletit eivät sovellu hyvin harmaasävyjen määrittelyyn. Ne ovat epäadekvaatteja kuvien muokkaamiseen tietokoneelle sisäisesti oikeissa sävy-suhteissa (render) niin, että gradientit vastaisivat tyypillisesti luonnollisissa kuvissa esiintyviä gradientteja, ja epäjatkuvia (anti-aliased) linjoja tai tekstiä. Tavallisimmat tilaprosessoinnin variantit [13], kuten yksinkertaiseen/ kaksinkertaiseen hex- muotoon perustuvat (täyden- tai tyhjän tasopinnan tilan ilmaisevat) kuvausmallit eivät tuo ongelmaan juurikaan parannusta (katso kuva 1).

2.2. "Visual" spaces

Kuvaamme nyt hieman edellistä kunnianhimoisemmin observoijaan liittyviä psykofysiologisia piirteitä. Olemme valinneet lähtökohdaksi suhteellisen yhdenmukaisen tilan, joka koostuu ortogonaalisista akseleista. Jälkimmäinen attribuuteista mahdollistaa Munsellin valinnan mielenkiintoisempana vaihtoehtona kuin esimerkiksi mallin CIE Lab. Vaikka molemmat malleista ovatkin "yhdenmukaistuvuuteen perustuvia" (uniform), ja molemmissa on käytössä luminanssinsäätelyyn kytketty akseli, on ainoastaan Munsell kykenevä erottamaan toisistaan kuvan ominaisuudet: spektriasemo (hue) ja väripigmentoituminen (chroma). Tämän vuoksi Munsell kykenee hyödyntämään kromaattisten piirteiden hakua työskentelyssään. Seuraavassa joitakin perustavanlaatuisia visuaalisuuteen liittyviä suhteita, joista rakennamme uusia struktuureja: (1) Värien eroavaisuudet olennaistuvat kolmeen toisistaan erotettavaan laadukeeseen tai variaatioon: värispektriin liittyvä spektriasemo (hue), valoisuuden arvo, ja väripigmentoitumisen kylläisyysaste (chroma saturation). (2) Jokainen spektriasemo (hue) monokromaattisen vaihtelualueensa sisällä voidaan löytää myös kaikissa valoisuusarvoissa (light values) valoisien ja mustan muodostamien äärimmäisyyspisteiden välillä. (3) On olemassa konsistentti luonnollinen järjestys toisiinsa suhteutuvien spektriasemojen (hues) sijainnissa värispektrissä. (Keltainen on aina jossakin vihreän ja punaisen välimailla, ja sitä ei koskaan esiinny sinisen ja purppuran välisessä tilassa). (4) On olemassa erityinen komplementaarinen visuaalinen suhde erityisten spektriasemoparikkien muodostamana (by specific pairs of hues). (5) Aistimukselliselta kannalta analoginen ja komplementaarinen spektriasemosuhde ovat riippumattomia siitä, mitä ne ovat mitattuina nanometreillä. (6) Kaikkien värien spektraaliset (maximum chroma) spektriasemopisteet eivät ole samanlaisissa suhteissa keskenään mustan ja valkoisen välisellä vaihtelualueella kuin ne ovat värispektrissä. (7) Kaikkien värien väriaseointumat eivät myöskään ole yhdenmukaiset niihin välimatkasuhteisiin, jotka vallitsevat keskeisellä vertikaalisella akromaattisella akselilla. Värienspesifiointityökalujen tulisi ylläpitää niitä värisuhteita, jotka ovat visuaalisesti ortogonaalisia. Gradientit ja värienvaihtelun mahdolliset vaihtelutasot tulisi voida tuottaa näyttöön ilman, että päämääränä olisi jokin vakioinen, ennalta-asetettu laadullisuus. Tämä voisi tarkoittaa sitä, että näyttöjen ei tulisi kompressoitua tai venyttää väripinnoissa ilmeneviä muutosalueita tarkoituksena pyrkiä populoimaan "taloudellisesti" kaikki käytettävissä olevat vakiomääräiset positiot symmetrisessä, kiinteässä formaatissa (katso kuvio 2).

2.3. Spanning the "natural" gamut

Samalla kun näytöt ovat laajalla kromaattisuuden ja luminanssin vaihtelualueella soveltuvia inhimilliselle katselijalle, on meidän vaateemme, että värinäyttöjen normin tulisi ohjautua tietyllä tavoin muotoutuvalle vaihtelualueelle. Käyttökelpoinen värikorttimuisti on rajoitettu. Inhimillinen visuaalinen systeemi ei voi pitkään sopeutua värien muodostamille äärimmäisyyksille. Voimakkaiden luminanssin ja krominanssin kontrastin ollessa havaittavuudessa menettää silmä nopeasti kykynsä nähdä vaikeuksitta tavanomaisempien ja lievempien väri vaihteluiden keskuudessa tapahtuvaa vaihtelua.

2.4. Quantitization of perceived hue-chroma factors

Spektriasemojen (hue) erottelua mitataan tavallisimmin pienimmän mahdollisen kahden väriasemion erolla, josta voidaan vielä tulkita kyseessä olevan kaksi erillistä väriä. Kuitenkin riippuu väriasemojen (hues) toisistaan erottaminen myös niiden arvon tasosta (value level) [14]. Esimerkiksi arvojakauman yläpäässä voidaan erottaa toisistaan enemmän keltaisia sävyjä kuin sinisiä. Niinpä olemme laajentaneet Munshellin mallissa variaation inkluusiolla huomioiden väriasemion ja toisaalta myös arvojakauman (value grid). Tuloksena on "tulppaaninkukan muotoinen jakauma" pikemminkin kuin rektilineaarinen jakauma. "Tulppaaninkukkamainen" esitys tilanjaosta on tulosta väriasemojen avaruuden skaalauksesta yhdistyneenä suurimman mahdollisen kromaattisuuden määrään jokaisessa väriasemo/ arvojakauma -positiossa. Mittauksemme väriasemion konstanssiudesta liittyy suoraan kromaattisuuden varioituneisuuteen annetulla arvojakauman arvolla. Tämä mahdollistaa sen, että värisensitiivisyys voidaan suhteuttaa luminanssiin (katso kuvio 3).

3. PROCEDURE

Perustavanlaatuisessa menettelyssämme paletinvalinnassa on kolme vaihetta:

- * Käytetty Munsellia RGB konversioon, joka on tyypillisesti kaksivaiheinen prosessi. Konversio Munsellista CIE:en suoritetaan taulun seurantaan perustuen ja lineaarista interpolaatiota käyttäen. Konversio CIE:sta RGB:hen suoritetaan matrix -operaatiolla.
- * Kvantisointi suoritettu valitsemalla värien valikko jaoittamalla avaruus kolmelle akselille: väriasemo,- arvojakauma-arvo, ja kromaattisuusakseleille. Alalajittumien (subdivisions) lukumäärä, ja asemat riippuvat paletin koosta, ja käyttöyhteydestä, jota varten kuvaa prosessoidaan. Väriasemojen alalajittamisten tulisi olla yhdenmukaistuvia, mutta niiden ei tarvitse olla juontuvia yhteenkään "käytetyn kuvan alkuperäisistä väreistä". Pikemminkin niiden tulisi juontua mihin tahansa relevanttiin ulkoiseen referenssiin, kuten esimerkiksi yhtiologossa esiintyviin väreihin. Kromaattisuuden spesifioinnin tulisi vastata kromaattisuusprofieileita. Olemme havainneet hyödylliseksi skaalata tilan kvantisoinnin suhteuttaen sen suurimpaan mahdolliseen esiintyvän kromaattisuuden määrään jokaisessa väriasemo/ / arvojakauma-arvon pisteessä. Niissä esiintyvien värien määrän tulisi juontua kromaattisuuden määrästä kyseisessä väriasemossa, joka on suhteessa totaaliseen kromaattisuuden määrään kyseisellä arvojakauman tasolla. Samoin tulisi näillä tasoilla esiintyvien värien määrän juontua kromaattisuuden määrästä tasolla suhteutettuna kromaattisuuden määrään kaikilla tasoilla. Tällainen algoritmi approksimoi hyvin "tulppaaninmuotoista" mittauksen kuvainta väriasemosta/ arvojakauma-arvosta. Äärimmäisen vahvistuksen alueilla, kuten värikuution nurkissa tulisi väri-inkluusion olla kohtuullinen.
- * Etsintäavaruuden konstruoinnissa muotoillaan binaarinen totuuspuu, joka on juonnos staattisesta paletista. Tällainen puu voidaan lajitella sen mukaan, millaisia dimensioiden mahdollisuuksia on käytettävissä, kuten RGB tai HVC. Sovelluksesta riippuen voi olla tarpeen käyttää useitakin puita. Lisäksi etsintäavaruus voidaan konstruoida niin, että se organisoii paletin perusväriasemoitumisiin, mikä mahdollistaa sovelluksen, jossa suoraan spesifioidaan värien-suhde-väriin -suhderakenteita, kuten on laita mustavalkoisissa kuvissa.

4. EQUAL VALUE

Usein huomiotta jäänyt aspekti Munshellin systeemissä on yhdenmukaistuvan valoisuuden käsite, joka korrespondoi kaikkien väriasemojen (hues) gradientteihin. Munsellin "yhdenmukaistettavien leveyssuunnan arvojen" ortogonaaliset linjat ovat käyttökelpoisia navigatiivisessa mielessä tutkittaessa värin kolmiulotteista rakennetta. Olemme havainneet sillä olevan myös muita tärkeitä seuraamuksia: (1) Se laajentaa värien optisesti havaittavissa olevaa keskinäistä sekoitettavuutta. (2) Se vapauttaa vastakohtaistuvien kromaattisten värien interaktion energiaa.

4.1. Enchanged hue to hue relationships

Useimmat insinöörit eivät ole tietoisia siitä, että kromaattisia värejä voidaan tuottaa näyttöön loogisilla visuaalisilla skaalauksilla. Sellaiset asiantuntijat, kuten Edward Tufte -väitettyään, "että useimmat värigrafiikat eivät sovellu katseltaviksi" päättete, että "huolimatta kokemuksestamme spektrin väreistä, ei silmämme toiminta mahdollista värien jäsentämistä" [15]. Syy täällaiseen yleiseen intensiviteettiin yritettäessä jäsentää värien visuaalista järjestyneyttä juontuu prosessista, jolla tietokone luo kromaattisten värien sekvenssejä. Tietokone voi käsitellä kuvien harmaasävyjä, ja jäsentää niiden pohjalta oikein värien visuaalisia arvoja -niin kuin ne inhimillinen observoija vastaanottaa. Sitä vastoin tietokoneen suorittama luminanssinsäätely, jota kromaattiset sekvenssit johtavat, tuottavat vinoutuneita värin-suhde-väriin -orientoitumisia (katso kuvio 4). Kehitettäessämme The Logical Visual Display -systeemiä olemme huomioineet sen tosiasian, että kaikilla väriasemoilla voi olla avaruudellisesti samalla tapaa asettuvat arvoskaalat. Nämä kromaattisten arvojen skaalat voidaan yhdentää akromaattisten harmaaskaalojen kanssa. Voimme muodostaa samanlaisuutta ja kontrastiutta kuvaavia lineaarisia polkuja tasolle, jonka dimensioina ovat väriasemo ja arvojakauma. Voimme vaihtaa keskenään useita väriasemoja menettämättä kontrollia liittyen niiden suhteellisiin harmaaskaala-arvojen tason vakioisuuteen. Voimme myös avata yhä uusia optisten värisekoitusten mahdollisuuksia samanaikaisesti kontrastoiden niitä interaktioita, jotka ovat kontingenteja värien visuaalisen arvon vastaavuuden suhteen.

4.2. Communitability of hues

Useimmissa tapauksissa keinotekoiset kromaattiset värit suhteutetaan harmaaskaalaan laajennettaessa näytön informaatio- erottelua. Täällaisessa prosessissa kromaattisten arvojen rinnakkaisten skaalojen käyttö saattaa tuottaa kommunikaatiota väriasemojen välille. Värillisen informaation käyttäjä voi hyvin laajentaa väriasemojen välistä vaihdeltavuutta ilman, että puuttuisi annettuihin arvoihin perustuvaan värien suhteellisen järjestyneisyyteen. Koska "värisokeus" viittaa oikeastaan vain spektrin rajoitettuun osaan, voi värien mahdollisuus kommunikoidua hyvinkin vaikuttaa myös niihin väreihin, jotka ovat aiheutettuun vaikutukseen nähden fysikaalisessa "verenkierto-vaikutussuhteessa".

4.3. Enchanged optical mixtures

Mitkä tahansa kaksi väriasemoa (hues) voidaan sekoittaa tuottamalla niiden välille asettuvia väriasemoja. Mitä lähempänä väriasemot (hues) ovat spektraalisessa järjestyneisyydessään toisiaan, sitä suurempi on resultanttivärin kromaattisuus. Se spatiaalinen skaala, jossa värit tulevat sekoittumaan, on täydellisesti riippuvaista siitä, kuinka lähellä sekoitettavat värit toisiaan ovat. Mitä lähempänä alkuperäiset väriasemot ovat toisiaan, sitä yhdistetympi on niiden sekoitus. Värien erot voidaan selvästi nähdä akromaattisten muutosten ollessa havaittavissa, mikä implikoi siihen, että kun suhteutetaan toisiinsa spatiaalisia (tai temporaalisia) detalleja, on hyvä huomioida myös asianmukainen muutos luminanssissa [16]. Siitä seuraa, että kromaattisia värejä on helpompaa luoda silloin, kun akromaattinen eroavaisuus on pieni, mikä tarkoittaa sitä, että samanarvoiset väriasemot myös helpoimmin sekoittuvat visuaalisesti. Akromaattisten eroavaisuuksien minimointi heikentää katsojan herkkyyttä kuvan yksityiskohdille, ja kuvassa esiintyvälle kohinalle niin temporaalisella kuin spatiaalisella spektrialueella. Yhdenmukaisesta arvottumisesta kiinnipitäminen aiheutettaessa kromaattisia muutoksia voi laajentaa tihennyksen (dither) käyttöaluetta ja virheenpaikannustekniikoita.

4.4. Energy of color in juxtaposition

Taiteilijan kokemus optisesta värien sekoittamisesta impressionismissa ja pointillismissa, kuten myös värinvalmistajien visuaalinen integrointi tuotettaessa teollisia värejä, on jo pitkään ollut demonstraatiota kromaattisten sekoitusten samanarvoisuuden tehokkuudesta. Väriensekoittajat ovat myös osoittaneet vastakkaisvärien sekoituksen tehon tapauksissa, joissa väriasemon-suhde-väriasemoon ei sisällä kontrastiarvon eroa. Tällä tavoin he tuottivat värisävyjä ilman, että olisivat jäsentäneet samalla kromaattisia äärimmäisyyksiä. Seuraamalla tätä esimerkkiä valittaessa tietokonegrafiikan väripalettia, voimme saavuttaa aistimuksia korkeaenergisistä väreistä ilman, että värinäyttöjen teknologiset rajoitteet

olisivat esteenä. Energiavaraus, joka näin saavutetaan, voidaan käyttää käännettäessä informaatiota sellaiseksi, jota käytetään hard-copy vedoksia tulostettaessa.

5. COLOR SENSATIONS

"Color is colors, plural".
- J. Albers

Tekemällä eron prosessin ja visuaalisen värin välille, on tärkeää huomioida äärimmäinen tavoite palettien valinnassa: väriaistimusten tuottaminen. Lisäyksemme edelliseen on se, että voisimme tuottaa sellaisia paletteja, jotka tuottasivat värisovelteita, jotka olisivat luonteeltaan tulosta hyvin ennustavia ekspressiivisiä signaaleja. Tietokoneen käyttö on avannut mahdollisuuden saavuttaa taiteen vaatima taso värienvälinnassa. Koskaan aikaisemmin ei ole ollut käsillä tällaisia tarkkoja työkaluja peruselementtien käsittelyyn, jotka ovat visuaalisen skaalan "alapuolella", kuten pigmentit ja fosforiheijastepisteiden kolmiot, joista tuotetaan värien optiset sekoitukset; taiteilijan sivellin käsittelee ja ryhmittää siten kuvan elementtejä. menetelmä on hyödyllinen kuvien taltioinnissa ja siirtämisessä, jossa perseptuaalisesti konsistentti väriavaruus seuraa loogisesta kuvan kompressoinnista. Kuitenkin, jos värien optiset sekoitukset vastakkainasetetaan toisten värialueiden kanssa, saadaan ekspressiivisiä laatuominaisuuksia värienvälisen aktiviteetin seuraamuksena, jota fyysikot ja psykologit eivät ole tutkimuksissaan kvantifioineet. Tällaista värien ekspressiivisyyttä on pidetty liian subjektiivisena, että sitä tutkisivat fyysikot, tai että insinöörit soveltaisivat sitä, ja siten alue on jätetty psykologien tutkittavaksi. Niinpä tietokonegraafikan käyttäjällä ei ole formaalista systeemiä tai opastajaa, joka voisi tarjota kokemuksiaan käytettäväksi värien kombinaatioita muodostettaessa. On tehty epäonnistuneita yrityksiä koodata dyadisia värien välisiä suhteita käyttämällä värien väliseen harmoniaan perustuvia lakeja hyväksi. Ne on kuitenkin arvioitu subjektiivisiksi, ja olevan tulosta yksilökohtaisista emotionaalisista responseista. Väriharmonian pseudostruktuurit on erheellisesti yhdistetty värien peruselementtien löytämiseen, joiden ainut funktio on kuitenkin toimia värienmuodostuksen komponentteina. Ei ole käytännössä onnistuttu erottamaan väristimuluksen kapasiteettia siirtää häivää emotionaalisesta responsista ja virheellistä semanttista kavennusta, joka kutistaa harmonian polariteetterihin "miellyttävä" ja "epämiellyttävä". Ehdotamme metodia, jossa kvantisoidaan värikombinaatioiden suhteelliset ekspressiiviset laatuominaisuudet. Sekä fysikaaliset- että perseptuaaliset värien eroavaisuudet voidaan organisoida tunnistettavissa oleviksi struktuureiksi, joista käsin voidaan mitata värienvälisiä aktiviteetteja. Ei-arvioperusteisia objektiivisia laatuominaisuuksia voidaan siten juontaa. Värienvälisen kommunikaation ei siten tarvitse olla uhkapeliä, ja meillä on predeterministinen keino stimuloida väri-ilmaisuja.

5.1. Color Dyads: expressive controls

Jotta voisimme ilmaista perusvärien välisiä suhteita, olemme kehittäneet sanakirjan, joka perustuu väri-suhde-väriin -asteittaiseen muutokseen, kuten monokromisuuteen, tai analogiaan. Aloitusvärin, tai locus-väriasemon valinta viittaa siihen, että värin kombinoitumista toisiin väreihin spesifoidaan (katso kuvio 5). Vaihtelut ekspressiivisissä karakteristikoissa ovat primaaristi väriasemoiden asettumisia toistensa suhteen (spacing of hues). Reaktio väriasemo-kombinaatioihin säätyy annetusta arvosta ja kromaattisuuden vahvistumisesta. Kvalitatiiviset karakteristikat voidaan erottaa kombinoituista väriasemo-asettumis moodeista, annetusta arvosta ja kromaattisuuden vahvistumisesta. Ryhmittämällä mainitut seikat kykenemme tuottamaan värienkäsittelytyökalun, joka sallii suoran perseption seuraamusvaikutusten manipulaation.

6. COMPUTATIONAL EFFICIENCIES

Perustavanlaatuinen seikka koskein tietojenkäsittelyllistä tehokkuutta paletinvalinnassa on sen avaruuden etsiminen l. lähimmän palettiväylän löytäminen keinotekoiselle värille. Monet sovellukset, kuten kuvan renderointi, vaativat sen, että tehtävä suoritetaan pixeli pixeliltä. Tyypillisesti tällainen tapa vaatii sen, että todella löydetään useita värejä, jotta etsintäproseduuri olisi tehokas. Käsitettäessä ongelmaa ennalta-

asetetulla kiineellä paletilla avaruus on huomattavassa määrin yksinkertaistettu. Me voimme valita sitävastoin sellaisen järjestäytymisen, jollainen parhaiten soveltuu tietokoneen arkkitehtuuriin ja käytettyihin sovelluksiin.

6.1. Search strategies

Organisointimalli, jota voidaan käyttää tehokkaalla tavalla on binaarinen totuuspuu -rakenne. Staattisia tauluja varten, joissa värit eivät muutu, voidaan puuhakurakennetta käyttää, joka huomioi tunnetun jakauman ja taulunelementtien (värien) tavoitettavissa olevat frekvenssit. Etuna on se, että etsintä on logaritmista, jopa dynaamisilla paletteilla [17]. Vaikka tämä tapa on tietojenkäsittelyllisesti intensiivisempi kuin boolelainen tapa, ja luvussa 2.1. kuvattu bittikäsittely, se silti soveltuu tämän päivän persoonallisten tietokoneiden mahdollisuuksien rajoihin. Binaarinen totuuspuu on tarpeen luoda vain kerran (mahdollisesti offline), vaikka vaatiikin työtä luoda puuhakurakenne, jollainen lyhentää keskimääräistä haku-aikaa. Tosiasiassa tasapainoista puuhakurakennetta, joka on lähes optimaalinen, voidaan laajentaa ajan myötä paletin värivalikoiman mukaiseksi [17]. R. Gray esittää tutkimuksessaan lukuisia metodeja, jolla voi alajaoittaa monidimensionaalisia avaruuksia [18]. Käytännössä tekniikat tarjoavat tehokkaan hakumenetelmän myös meidän hakemillemme "epätavallisille" väriavaruuksille. Kun nimittäin etsitään avaruuksia isoloiduille väreille (kuten typografioihin valittaville väreille) on tehtävä tietojenkäsittelyllisesti insignifikantti. Kun taas on löydettävä useita värejä, kuten on laita sävytettäessä kuvia jatkuvin väripinnoin, on etsintä mutkikkaampi, mutta silti hallittavissa.

7. CONCLUSIONS

"Useful structures are not made by squaring telephone numbers".
- Bar Yosef

Tavanomaiset kiinteiden palettien valintametodit, jotka eivät huomioi inhimillistä värien perseptiota eivät voi toimia värigrafiikassa esteettisessä mielessä. Systemaattisella "visuaalisen" värien sovelluksella voidaan paletit suunnitella niin, että niitä käyttämällä saavutetaan maksimaalinen hyödynnettävyys. Olemme tutkimuksessamme ehdottaneet värienvälyntäsystemiä, joka antaa tasonmäärityskontrollin, joka huomattavasti ylittää individuaalien värien postioiden suhteen mahdollisen säätelyn. Koska muutos on kvantisoitavissa, tarjoaa systeemi työkalun värikoostojen muodostamiselle. Kun on valittu värialoituspiste (color starting point) on vain spesifioitava värikoostojen kvaliteetit. Tällä tavoin mahdollistetaan kuvaa käsittelevälle se, että hän voi tutkia aiuttamiaan kontekstuaalisia seuraamuksia liittyen mihin tahansa annettuun väripositioon. Mitä enemmän on saatavilla relevanttia kuvaimprovisaatioiden näyttöön kytkeytyvää esittävyttä, sitä suurempaa vapautta se merkitsee tehtävien valintojen suhteen. Laskennallinen mutkikkaus, jota systeemimme vaatii ei muodosta ongelmaa nykyiselle tietokoneteknologialle. Väriavaruusmallien välttämättömyysperiaatteet on kehitelty käyttämällä komplekseja, mutta silti konsistentteja formulaatioita. Sellaisten kompleksuuksien käsittely olisi liikaa yksittäiselle suunnittelijalle. Kuitenkin tietokone voi hyväksyä tällaisen mutkikkouden tavoiteltaessa erilaisia vaihtoehtoisuuksia sävyissä, kunhan se rakentuu konsistentille perustalle. Olemme siksi antaneet värienmanipulaation taakan koneelle toivoen samalla, että käyttäjät sen kautta voisivat artikuloitua työskennellä värisovelluksien parissa.

8. REFERENCES

- [1] N. Negroponte, "Color dynamics and decision making", Proposal to Office of Naval Research, (1980).
- [2] B. Boynton, "Eleven colors that are never confused", Proceedings SPSE (1989).
- [3] T. Lippert, "Color-difference perception of legibility performance for CRT raster imagery", SID Digest, (1986).
- [4] R. Martel, "Luminance and chrominance contrast in emissive displays", Proceedings of SPSE 901, (1988).
- [5] P.J. Robertson, "Guide to using color on alphanumeric displays", IBM Technical Report, (1980).

[6] P. Heckbert, "Color image quantization for frame buffer display", *Computer Graphics* 16(3), (1982).

[7] G. Houle, "Quantization des images couleurs pour appareils graphiques" INSR-Télécommunications Technical Report no. 85-37, (1985).

[8] G. Braudaway, "A Procedure for optimum choice of a small number of colors from a large palette for color imaging", IBM Thomas J. Watson Researcher Center, Yorktown Heights, NY. (UNDATED).

[9] W. Bender, "Adaptive color coding based on spatial/ temporal features", *Proceedings SPSE 901*, (1988).

[10] T. Cornsweet, *Visual Perception*, Academic Press, (1970).

[11] A. Munshell, *A Color Notation*, Munsell Color Co., Inc., (1946).

[12] G. Meyer and D. Greenberg, "Perceptual color spaces for computer graphics", *Computer Graphics* 14(3), (1980).

[13] H. Levkowitz and G. Herman, "Towards a uniform lightness, hue, and saturation color model", *Proceedings SPSE 901*, (1988).

[14] U. Feldman, "Hue discrimination in Munsell color space", Unpublished report, MIT Media Lab, (1988).

[15] E. Tufte, *The Visual Display of Quantitative Information*, Graphics Press, (1983).

[16] W. Cowan, "Colour psychophysics and display technology: avoiding the wrong answers and finding the right questions", *Proceedings SPSE 901*, (1988).

[17] E. Reingold and W. Hansen, *Data Structures*, Little Brown Computer Systems Series, (1983).

[18] R. Gray, "Vector Quantization", *IEEE ASSP Mag.* (1984).

9. FIGURES

Figure 1.

(a) The double cone farbenkubel. Despite a regular geometry, it is visually irregular.

(b) The Munsell tree. Organically irregular, but uniform.

Figure 2.

Two Chroma/Value grids. The horizontal axis represents the progression from achromatic to chromatic, while the vertical axis represents the value scale.

(a) The forced stretching of chroma for the sake of geometric symmetry results in a neat box, but the path of equal chroma is a weird route.

(b) Allowing some of the areas to remain unfilled reflects the irregularity of visual color perimeters. A weird box, but a neat route.

Figure 3.

The format of "tulip". Its organic irregularity yields straight line distances between colors.

Figure 4.

The placement of spectral hues between black and white. In figure (a), value is equated for all hues, without consideration of the fact that each colorant responds differently to the same amounts of luminant amplification. As a result, there is no visual correspondence between the value levels of different hues. In figure (b) spectral yellow is light while spectral blue is dark.

Figure 5.

The alignment of hues varies along the circumference of the outer circle. The central core is neutral. M is the monochrome region. A is the analogous region. C is the complementary region.

THE INTEGRATED MULTI-MODAL INTERFACE

Rickhard A. BOLT, Nonmember

The author is with The Media Laboratory, Massachusetts Institute of Technology, Cambridge, USA.

The reference of this article is:
Bolt, Richard A. "The integrated multi-modal interface".

*The Transactions of the Institute of Electronics,
Information and Communications Engineers (Japan),
November 1987, Vol. J70-D, No. 11, pp. 2017-2025.*

Suom. Timo Kinnunen 6.3.1989

SUMMMARY

Multi-Modaaliset vuorovaikutekentän muodot (interfaces), erityisesti sellaiset, joissa kombinoidaan puhetta, eleitä, ja katsetta, voivat tehdä ihminen/ tietokone -vuorovaikutuksesta enemmän keskustelunomaista luonteeltaan. Siinä kun "keskustelu" ihminen/ tietokone -vuorovaikutuksen metaforana ei ole sovelias sellaisten tehtävien ratkaisulle, kuten paperitulosteiden ja sanojenkäsittelyn käyttöön, se saattaa hyvinkin soveltua ideoiden kiireettömään paranteluun ja niiden tutkimiseen. "Itsestääntiedottavaa" (self-disclosing) systeemiä, jota tässä kuvataan, tutkitaan parhaillaan. Tämä järjestelmä tarkkailee tietokoneenkäyttäjän erilaisia aktiviteetteja, kuten silmien tekemää kohdistusliikettä graafisella näyttöpäätteellä, ja järjestelmä vastaa kysymyksiin syntesoidulla puheella, ja graafisen näytön erilaisilla aktiviteeteilla, joita käyttäjä haluaa toteutettaviksi.

Introduction

Olen johtajana Massachusetts Institute of Technology's Media Laboratory'ssa, ja hyvin kiinnostunut multi-modaalisista vuorovaikutekentän muodoista (interfaces): vuorovaikutekentän muodot hyväksyvät ja tulkitsevat niille annettua syöttöä (input) kahdessa tai useammassa muodossa yhtäaikaaisesti. Tämäntapainen vuorovaikutus (interaction) eroaa tavanomaisesta interaktiosta tietokoneen ja sen yksittäisen käyttäjän välillä (esimerkiksi näppäimistön kautta tapahtuvasta). Erityisen kiinnostunut olen sovittamaan ihminen/ tietokone vuorovaikutekenttään muodoista (interface) kolmea primaarimuotoa jäljitellen tapaa, jossa ihminen kohtaa ihmisen kasvokontaktissa: puheen, eleet ja katseen. Mahdollistamalla inhimilliselle käyttäjälle jonkin näistä muodoista käytön, joko yksittäisenä, tai kombinoituneena muiden mainittujen muotojen kanssa. Siten toivon, että kanssakäyminen tietokoneen kanssa voisi muodostua enemmän ihmistenvälistä keskustelua muistuttavaksi. On tosin niitäkin, jotka pitävät "keskusteluyhteys-systeemin" metaforaa soveltumattomana ihmisen ja tietokoneen väliseen kommunikaatioon (esimerkiksi, ref. [16]), ja että työkalujen- tai instrumenttienkäytön metafora olisi suositeltavampi. Tämä on totta koskien monia, tai ehkä useimpia tilanteita tänä päivänä, joissa ihmiset istuvat tietokoneiden ääressä. Silti minusta tuntuu, että on olemassa valtava tietokoneenkäytön tutkimaton alue, jossa vuorovaikus olisi enemmänkin "keskustelua muistuttavaa". Ajatelkaamme esimerkiksi sellaista vuorovaikutusta, jossa päämääränä on ideoiden muokkaus ja tutkiminen, ja jossa tarvitaan kiireetöntä tutkimusta ja laajaa informaation moniyhteyksellistä vertaamista, ja jossa "ääneen ajatteleva" tuosta tai tästä teemasta on keskeistä. Konkreettisesti tällainen ideoija voisi olla suunnittelija, joka harkitsee sopivia erityisten kasvien viljelijöitä; fysiikanopiskelija, joka haluaa tarkastella optiikkaan liittyviä teemoja; kotonaan oleva henkilö, joka suunnittelee ohjattua matkaa Eurooppaan tutustuakseen sen katedraaleihin, ja joka haluaa käyttää opastajanaan tietokonetta. Tällaisten pyrkimysten henki, ja sävy eroaa huomattavasti tietokoneliuskoihin, tai sanaprosessoreihin perustuvasta työskentelystä, ja liittyy enemmänkin tavanomaiseen keskustelullisuuteen, jossa joku tietää jotakin keskusteltavista teemoista, ja haluaa kuluttaa aikaansa keskustelemalla niistä jonkun kanssa, joka myös niistä tahtoisi jotakin tietää.

2. Technologies and Intelligence

On olemassa kaksikin seikkaa ratkaistavana luotaessa sellaista keskustelullista kvaliteettia käyttäjän ja tietokoneen vuorovaikutuskenttään (interface). Yhtäältä on huomioitava teknologia, jolla kyetään poimimaan tarvittavia aineksia käyttäjän puheesta, katseesta ja eleistä. Nämä teknologiat liittyvät automaattiseen puheensisältöjen tunnistukseen (speech recognition), silmien liikkeiden rekisteröintiin (eyetracking), ja manuaalisiin informaationsyöttölaitteisiin. Voimme odottaa näiden teknologioiden parantuvan ajan myötä, ja tulevan käytöllisyydeltään mukavammiksi, sekä myös halpenevan paremman insinöörisuunnittelun myötä, ja parempien materiaalien tulevan saataville niin, että laitteiden koko pienenee (miniaturization), sekä lisäksi sitä, että kehitetään parempia tiedonsiirtolaitteita (transducers).

Se, mitä kehityksen myötä kaikkiaan jää jäljelle on se, että puhuminen, katsominen ja osoittaminen tulee jäämään pääasiallisimmaksi tavaksi, jolla ihmiset ilmaisevat itseään. Toisaalta kehityksen kuvaan kuuluu koneäly (machine intelligence), joka tulkitsee informaatioesityöt, joita vuorovaikutekentän muodoilla laitteistoon syötetään; koneäly kartoittaa mm. graafiselle näytölle annetut vastaukset, puheella annetut ohjeet, tai hyödyntää niistä molempia. Voidaanko mainitunlaisilla puhumisen, katsomisen ja osoittamisen muotoja sitten päästä eroon näppäimistöllä tai "hiirellä" tehtyihin valintoihin, joita nykyään käytetään? Ei, koska näppäimistö on hyödyllinen työkalu syöttöön, jossa symbolista informaatiota annetaan merkkien ketjuina. Eikä myöskään "hiirestä" haluttane päästä eroon, sillä se on hyvä työkalu. Vaikka ihmiset puhuvat, katsovat ja osoittavat me käytämme silti yleensäkin työkaluja: vasaraa naulojen lyömiseen, saksia paperinleikkaukseen. Pääpainotusalueena ovat tässä ne ihmisryhmät, joiden ei ole mahdollista sellaisia käyttää, tai joiden ei ole tarpeen käyttää työvälineitä puhuessaan, osoittaessaan, tai katsoessaan.

3. Multi-Modalities

Mitä etuja multimodaalisesta lähestymistavasta sitten on? Moniakkin, mutta ehkä lähinnä kolme erityisen tärkeää etua seuraa siitä, että käytetään kahta- tai useampaa muotoa yhdistyneenä ja rinnakkain: taakasta vapautuminen (unburdening), summautuminen, ja redundanssi.

3.1. Unburdening

Kun meillä on käytettävissämme ainoastaan yksi muoto, me olemme taipuvaisia "konttorirottavaikutukselle" (one-armed paperhanger effect) niin, että pyrimme työskentelemään yksinomaan ja korostetusti yhdellä saatavissa olevalla muodolla. Entä mykät elokuvat? Äänen puuttuminen merkitsi sitä, että näyttelijöiden oli ilmaistava kaikki visuaalisessa muodossa. Tämä johti ylikorostuneiden kasvojenilmeiden ja kehonliikkeiden käyttöön mykän kauden elokuvissa. Ilmaistakseen suuttumusta näyttelijän oli heiluteltava villisti käsivarsiaan, vääristää kasvojenilmeensä suuttumusta ilmaisevaksi, ja auottava suutaan tehostetusti ikään kuin sanoakseen uhkaavia sanoja. Suun aukominen oli samalla tavoin tehostettua kuin puhuttaessa toiselle ääntä eristävän lasiseinän takaa. Kun filmiin liitettiin ääniraita hävisivät mainitunlaiset korostuneet ilmaisut. Auditorisen muodon mukaantulo kevensi ilmaisuun liittyvää taakkaa, koska sekä visuaalinen- että auitiivinen muoto yhdistyivät, ja tasapainottuivat. Etua ei koitunut pelkästään katsojalle, joka nyt saattoi nyt käyttää sekä korviaan että silmiään, vaan myös näyttelijälle, joka saattoi käyttää kumpaa tahansa näistä ominaisuuksista ilmaisussaan aiempaa luonnollisemmalla ja spontaanimmalla tavalla.

3.2. Information summation

Toinen merkittävä etu multi-modaalisten muotojen käytöstä on se, että vajavaisemman tai köyhemmän informaation (minkä tahansa muodon kautta vastaanotettuna) yhdistyminen summautuu rikkaampisisältöiseen informaatioon, mikäli muut muodot sitä tukevat. Otaksukaamme, että olet kahden kolmasosan tehokkuudella mahdollisesta jokaisella muodolla vastaanottamassa informaatiota, jota sinulle annan puhumalla, osoittamalla, tai katsomalla. Tämä merkitsee sitä, että vastaanotat 66 %:n tehokkuudella niistä jokaisella informaatiota. Katson ylös kattoon, ja osoitan kädellä sitä sanoen: "Tuolla noin". Jos työskentelet jokaisella muodolla toisistaan erillään (ei rinnakkaisesti), niin todennäköisyys, että ymmärrät vihjeeni on parhaimmillaankin vain yksi kolmannesta mahdollisesta, mutta jos muodot yhdistyvät kommunikaatiossa, niin todennäköisyys, että ymmärrät vihjeeni kasvaa suunnilleen 96.3%:iin. Helppo tapa havainnollistaa edellistä liittyy tunnettuun Rubikin kuutioon, joka osittuu jokaisella sivullaan kolmeksi nelikulmioksi, jotka ovat pitkänomaisten kappalten päätysivuja, jolloin kokonaissummaksi muodostuu 27 nelikulmiota. Mahdollisuus, että ET onnistu tulkitsemaan oikein kolmella kommunikaatiomuodolla antamaani tiettyä sijaintia koskevaa vihjetä on 1/27, eli onnistumistodennäköisyytesi on 3.7%. Tässä tapauksessa eri kommunikaatiomuodoilla samanaikaisesti annetun vihjeen teho on huomattava verrattuna niillä kullakin yksittäisesti annettuun instruktioon.

3.3. Redundancy

"Informaation summautumiseen" kykeäytyy läheisesti redundanssi, joka liittyy niin siihen kontekstiin, josta puhutaan, jota katsotaan, ja josta jotakin osoitetaan, kuin näiden muotojen käyttöön rikkakkain. Tällaiseen liittyy puheen tulkitseminen. Käytetyltä laitteistolta vaaditaan vähintään 99.7 %:n tarkkuutta. Mainitut korkean tason suoriteominaisuudet liittyvät mm. koulutettuihin puhujiin, korkeainformatiivisiin sanaluetteloihin, ja häiriöttömiin ympäristöihin. Mikäli niitä testataan realistisemmissä ympäristöissä putoaa puheen tunnistamisen tarkkuus 60-65 %:iin. Kuitenkin tunnistettujen puhesisältöjen interpretaatio on yhäkin tarkkaa, vaikka puheen tunnistaminen ei sitä olekaan -silloinkin, kun me tuskin tunnemme käytettyä kieltä, joka annetaan oikeassa ympärisössä. Ajatelkaamme seuraavaa: "Päivälliskutsuilla vieraassa maassa saatat ymmärtää käytyä keskustelua jollakin tavoin, mikäli puheet koskevat leipää, voita, ruokailuvälineitä, viiniä jne. Mikäli puheenaiheet siirtyvät historiaan tai politiikkaan, voit osallistua siihen vain siinä tapauksessa, että hallitset heidän kieltään täysin sujuvasti. On väärin otaksua, että ero olisi jonkin keskustelunaiheen välisessä erossa toiseen, tai jonkin sanaluettelon erosta toiseen nähden. Tosiasiassa kyse on siitä, että leipä ja voi ovat saman tapahtuma-avaruuden osia, johon itsekin kuulut, ja joita voit osoittaa, tai joiden suuntaan voit nyökätä vahvistaen siten samanaikaisten kanavien redundanssia yhteisvaikutteiseksi kommunikaatiokimpuksi". Käyttäjän ja tietokoneen vuorovaikutekentässä tietokoneen graafinen näyttö on analoginen sen välittömän ympäristön kanssa, jollaisen kaksi tai useampia ihmisiä jakaa käymässänsä tavallisessa keskustelussa. Se, mikä tekee puhumisesta, osoittamisesta ja katsomisesta niin tehokkaan kasvokkain käydyssä keskustelussa silloinkin, kun kieltä osataan huonosti, on jaettu konteksti. Puhumme, katsomme ja osoitamme objekteja ja seikkoja ympäristössämme. Samalla tavoin graafinen näyttö esittää ja sisältää sen, mitä tietokone meille "tarjoaa"; sen, mistä se (tietokone) on halukas "keskustelemaan". Puhumisemme, katsomisemme ja osoittamisemme seuraa sitä, mitä tietokone meille näyttöruudussa esittää. Formuloimme ja suoritamme toimenpiteitä sen mukaisesti, mitä näyttöruudun esityssisällöt ovat.

4. Speaking

Automaattinen puheentunnistus (speech recognition) on ollut kehitettynä jo vuosikausien ajan, ja sen hinta on pudonnut dramaattisesti. Ensimmäiset saatavilla olleet kaupalliset laitteet puheen tietokoneellesyöttöön maksoivat vuoden 1979 markkinoilla suunnilleen 75,000 dollaria. Nykyään niiden hinta liikkuu 1000:n dollarin tienoilla. Käytetty teknologia on silti suunnilleen sama (mallit perustuvat siihen, että tietokoneille on opetettu tietty sanaluettelo, johon ne vertaavat uusia, syötettyjä sanoja). Huulilitalukeminen tietokoneen avulla saattaa tuottaa keinon selvittää puheentunnistuksessa meluisassa ympäristössä. Taustamelu häiritsee automaattista puheentunnistusta, mikä ei koske ihmisten välistä kommunikaatiota. Miksi? Siksi, että ihmiset kykenevät hyödyntämään huulilitalukemista mainitunlaisissa olosuhteissa. Huulilitalukemisessa on olemassa tiettyjä rajoja, koska esimerkiksi vain 40 % puhutusta englannista on havaittavissa huultenliikkeissä. Silti tutkimukset ovat osoittaneet ihmisten ymmärtävän toisiaan paremmin silloin, kun he näkevät toisensa [22], mikä viittaa siihen, että kykenemme ymmärtämään kanssaihmissämme meluisissa ympäristöissä (juhlien väentungoksessa) tarkkailemalla heidän kasvojaan ja huultensa liikkeitä. Syy ymmärtämisen helpottumiseen ei ole selvä. Ehkäpä puheen visuaalisesti nähtävissä oleva osa jollakin tavoin virittää kuulemiseen liittyviä huomionkeskittämismekanismeja niin, että voimme suodattaa tehokkaammin signaalin taustasta. Ehkäpä me tiedostamattomasti olemme oppineet yhdistämään tietyt suun muodot tiettyihin ääniin niin, että varmistuessamme kuulemastamme käytämme tällaista näkyvää informaatiota hyväksemme. Mikä tahansa tämän perustana sitten onkaan, niin puhujan näkeminen auttaa. Kuinka tämä liittyy tietokoneisiin? Petajanin äskettäiseen väitöskirjan liittyvässä proseduurissa yritettiin yhdistää huulilitalukemista automaattiseen puheentunnistukseen syöttämällä tietokoneelle yksittäisiä sanoja. Monokromaattisen suun alueen lähikuvainformaatiota puhumisen aikana suhteutettiin aiemmin luotuun kokoelmaan tutkittavan henkilön suunliikkeitä (hän lausuttui kokeen valmistavassa vaiheessa tiettyjä sanoja erityisen huolitellusti, jotka koottiin kokeen referenssimateriaaliksi). Siten puheentunnistaminen perustui sekä auditiiviseen -että visuaaliseen informaatioon, ja tuloksena olikin, että puheentunnistamisen oikeellisuus audiovisuaalisessa puheentunnistuksessa oli parempi kuin käytettäessä niitä kumpaakin erillisinä [18]. Samantapaisia töitä on tehty myös MIT Media Lab -yksikössä [17]. Tällä hetkellä meneillään on eräs projekti, jossa pyritään parempaan informaatioisisältöön suun alueelta liikkuvalla värilähikuvalla siitä. Toinen meneillään oleva

Media Lab -projekti Mr. Chris Schmandtin ohjauksessa etsii "paraverbaaleja" malleja liittyen intonaatioon ja sanapainotukseen, kuten myös sentapaisiin nonverbaalisiin seikkoihin, kuten huudahdukset, huokaukset, hymähdykset ja voihkaisut. Sellainen lause, kuten "Minä pidin juustosta" voi saada erilaisia merkityksiä riippuen siitä, mitä sanaa lauseessa painotetaan: "MINÄ pidin juustosta" -ilmaisussa painotetaan sitä, kuka juustosta piti, kun taas lauseessa "Minä pidin JUUSTOSTA" korostetaan sitä, mistä pidettiin, ja lopulta painotuksessa "Minä PIDIN juustosta" on edellisistä poikkeavia implikaatioita. Tämä kaikki kuuluu osana puheella tehtyyn syöttöön, joka saattaa tulla ymmärretyksi väärin, mikäli ei kiinnitetä huomiota siihen, kuinka on puhuttu -sen lisäksi, mitä sanoja on käytetty. Nonverbaaliset äänet ovat intonaatioita, joilla on merkitystä ihmisten välisessä kommunikaatiossa. Kuvitelkaa hymähdystä "Joopa joo". Tavanomaisessa puheentulkinta-kommunikaatiossa tietokone saattaa ilmoittaa "error message", ja lopettaa vuorovaikutuksen tyyten siihen paikkaan.

5. Gesture

Nykyiset manuaaliset syöttötekniikat sisältävät mm. kuvaruudun kosketukseen perustuvan käskynannon ja hiiriohjauksen. Tavallisimmin käytössä lienee hiiri. Kuvaruutukosketusohjaus on ainut, jota käytettäessä ei tarvitse hakea jotakin apuvälinettä toimenpiteen suoritusta varten. Aiemmassa Architecture Machine Group laboratoriossamme käytimme ranneohjattavaa magneettista tilatunnistinta (space sensor), jolla osoitettiin seinäkokoiselta näyttöruudulta haluttuja kohtia (Ref. [4]). Teimme myös kehonliikkeitä rekisteröivän puvun, jossa sijaitsevilla LED-antureilla ohjattiin stereokameroita, joilla tuotettiin kolmiulotteisia tikku-ukkohahmoja, joiden liikkeet ohjautuivat tehtyjen kehonliikkeiden mukaan. LED-liitäntäni kyettiin tuottamaan detaljoitu seurantamahdollisuus sormituntumatasolla. Kaikki nämä menetelmät, paitsi kuvaruutukosketus-ohjaus, vaativat jonkinlaisen käyttäjään liittyvän laitteen käyttöä. Tulevaisuudessa tällainen taustakirjoitavuus voidaan poistaa. Eräs mahdollinen tekniikka, jolla voidaan tavoittaa käden liikkeiden aistivaikutelma vapaassa tilassa, on että käyttäjä ohjaa stereovärikameraa kuva-analyysimenetelmällä, jolla hän paikallistaa käsiensä asennon, ja niiden muutokset; kamera on opetettu nimenomaan hänen ohjaukseensa. Väreistä muodostuu lisäyksenomainen ja voimakas vihjeistö, jolla paikallistetaan ja ekstraktoidaan käsienasento-kuvaa; stereovaikutelmaa hyödyntäen voidaan orientoitua vapaassa tilassa. Jotta tämä saataisiin tapahtumaan reaaliajassa tarvitaan nykyistä voimakkaampia prosessoreja niin, että tuloksena on jonkinlainen rinnakkaistoimintainen laite. Eksoottisempi tapa suoritettujen liikkeiden syöttöön liittyy kasvojenilmeiden seurantaan, koska niiden kautta voidaan syöttää tunnesisältöjä, kuten tunnettua; syöttöön liittyy myöskin semanttista informaatiota. Tietokone voi oppia rekisteröimään ja tunnistamaan kasvojenilmeiden viestimiä emotionaalisia sisältöjä aivan samalla tavoin kuin termografinen video on herkkä eri kasvojenalueiden lämpötilanmuutoksille (ks. Ref. [5]). Se voi myös perustaa analyysinsä kasvojenalueiden ja kasvojenpiirteiden analyysiin.

6. Looking

Viime kuukausina olen ollut erityisen kiinnostunut muokkaamaan tietokoneen vuorovaikutekenttää (interface) niin, että se olisi tietoisempi siitä, milloin käyttäjä katsoo graafista näyttöä. Ihmisellä silmä on dominoiva aistin [20], ja tietokoneelle pääasiallinen tulostustapa on näyttö. Silmänliikkeet heijastavat henkilöiden mielenkiinnon suuntautumisia, kuten ajattelunkin [9]. Niinpä silmänliikkeet ja fiksoitumiset, yksittäin tai kombinoituvina puheen ja osoittamisen kanssa, tarjoavat ainutlaatuisia vihjeitä käyttäjien mielenkiinnon suuntautumisista ihmisen ja tietokoneen välisessä kommunikaatiossa. Kun tietokoneille perustuvat informaatiojärjestelmät yhä enenevässä määrin ovat kyenneet tunnistamaan puhetta tai osoittamisia, ne yleisesti jättävät silmänliikkeet huomiotta ja eivät siten "pidä" niitä olennaisina vastavuoroisen informaation lähteinä. Tähän on ollut vaikuttamassa niin ergonomiset -kuin taloudelliset syyt. Kaikki menetelmät, joilla mitataan silmien kohdistumispisteitä vaativat laitteita, jotka on sijoitettu hyvin lähelle tietokoneen käyttäjää (Cf. Ref. [24]); en tiedä kehitetyn ainuttakaan laitetta sitten Youngin ja Sheenanin artikkelin, joka muutti silloista tilannetta. Ainut poikkeus lienee silmärefleksien seurainlaite, joka on toteutettu käyttämällä pientä videokameraa, joka on zoomattu niin, että vain näkyvän sarveiskalvon alue mahtuu kuvaan. Silmänliikettä mitaus on toteutettu niin, että kamerasta suunnataan silmään infrapunasäde, joka heijastuu takaisin, ja jota verrataan mykiön kulloiseenkin kääntymiskulmaan;

siten pienet pään liikkeetkin ovat sallittuja. Mittauksessa on virhettä vain noin asteen verran, ja siten pään ollessa kunnolla tuettuna mittaus on riittävän tarkka. Käytettävä teknologia on kallista, sillä silmänliikkeiden etäismittarin perusmalli maksaa n. 35 000 dollaria. Malli, johon on lisätty mekanismeja, jotka sallivat päänliikkeet mittauksen tarkkuuden kärsimättä maksavat jo 65 000 dollaria, tai jopa 100 000 dollaria. Hinnoissa on tapahtumassa putoamista, sillä uusi Cambridge-etäisyysmittain maksaa n. 14 000 dollaria. On mahdollista, että laitteen hinta putoaa integroitujen ohjauspiirien käytön myötä, ja yksinkertaisempien objektiivien käyttöönotolla n. 1000-2000 dollariin. Soveltuvan videokameran hinta on n. 800 dollaria nykyään, tai vähemmänkin. Boston Company on kehittänyt pelilijärjestelmiä, jotka adaptoituvat päänliikkeisiin systeemin hinnaksi muodostuessa n. 500 dollaria, ja automaattitarkennuksella 400 dollaria lisää. Yhteenvetona todettakoon, että kompaktin järjestelmän hinnaksi tulisi ehkä muodostumaan 2700 dollaria, tai vähemmän. Kun otetaan huomioon se, että tietokonemuistin ja keskusprosessorien hinnat laskevat jatkossakin, niin tulevaisuuden mikrotietokoneissa tällainen järjestelmä tulisi maksamaan kaikkiaan n. 3800 dollaria. Järjestelmän karkeatekoisuuden ja hinnan vähetessä tulevat sellaiset systeemit, joissa katsominen huomioituu, muuttamaan ihmisen ja tietokoneen välistä kommunikaatiota. Ajateltakoon vaikkapa niitä monia rooleja, joita katsekontaktilla on inhimillisissä ilmaisuissa ja vuorovaikutuksessa.

7. The Information in Eyes

Kuvittele tapausta, jossa sinun suosikkisetäsi tulee vierailulle uuteen asuntoosi: Hän tulee ja katselee samalla kun sinä kommentoit sisustusta. Hän tutkii pöydällä olevia tietokoneprinttejä. "Sain ne Lontoosta", sanot hänelle, ja kerrot niistä. Hän kiinnittää huomionsa yhteen niistä kysyen "Mikä tämä on?" Vastaat hänelle "Covent Garden vuonna 1770". Istuudutte molemmat, ja hän yrittää sytyttää savukkeen, muttei löydä tulitikkuja. Ojennat hänelle sytyttimen näyttäen, kuinka se toimii (se ei ole helppoa). Koska huomaat hänen olleen selittämisesi aikana hajamielinen demonstroit sytyttimen toimintaa uudelleen. Hän kysyy "Miten autosi on pelannut viime aikoina?", mutta katselee kysyessään asuinkumppaniasi, eikä sinua. Asuinkumppanisi ylistää Volvojen paremmuutta. Setäsi katsoo sinuun uudelleen, ja toistaa kysymyksensä "Miten autosi on pelannut viime aikoina?" Vastaat kysymykseen "Eipä mitenkään, koska ajoin sillä kolarin viime viikolla". Selität edelleen "Olin juuri ajamassa Main Streetillä...", ja oikeaa kättäsi katsoen ja sitä liikuttaen näytät ajoreittiäsi, "...ja juuri silloin tuli se koliseva trukki", sanot tehden vasenta kättäsi katsoen sillä ajoneuvon liikettä kuvaavia eleitä. Manaat ihmisiä, jotka ajavat päin punaisia, heiluttaen asiaa korostavalla tavalla käsiäsi. Koko ajan silmienliikkeet kuvastavat mielenkiinnon suuntautumisia, huomionkeskittämisiä ja referenssejä.

7.1. Interest

Kun setäsi astuu huoneeseen, ulkoinen huomionkiinnittyminen sinuun ei vaikuta ainoastaan siihen, mitä sanot, vaan myös käyttämäsi kehonkieleen, kuten myös se, mihin HÄN huomionsa kiinnittää. Kun hän katselee ympärilleen, kommentoit huoneesta jotakin yleistä, ja kun huomaat hänen tutkivan tietokoneprinttejä sinä lausut niitä koskien jotakin. Kun hän poimii niistä jonkun erityisen liuskan, kommentoit jotakin juuri siitä. Niinpä et ainoastaan poimi vihjeitä siitä, mihin hän huomionsa kiinnittää, vaan myös säädät kommenttiesi sisältöä sen mukaan, mistä arvelet hänen olevan kiinnostunut. Responsisi "työskentelevät" ja ovat soveltuvia siksi, että ihmiset yleensä tapaavat katsella heitä kiinnostavia asioita, tai jotka ovat heistä erikoisia, tavallisia, tai odottamattomia [2,3,10]. Silmänliikkeet katseltaessa kokonaisuutta pyrkivät poimimaan siitä erotettavissa olevia osasia riippuen mielenkiinnon yleisestä suuntautumisesta. Neuvostoliittolainen Alfred Yarbus (klassisessa silmänliikkeitä koskevassa tutkimuksessaan) pyysi tutkittavia perehtymään kuuluisan venäläisen maalarin teoksen kopioon. "He eivät odottaneet hänenlaistaan" sanoi muuan nuorimies, joka oli juuri palannut poliittisesta maanpakolaisuudesta. Ennen maalauksen katsomista (kolmen minuutin ajan) kysyttiin tutkittavilta joukko kysymyksiä: Minkä ikäisiä tutkittavan omaiset ovat?, Millaisissa materiaalisissa olosuhteissa tutkittavan perhe elää?, Mitä perhe oli tekemässä ennen kuin tutkittava palasi? Katsojien tarkastelussaan käyttämät tulkintamallit erosivat riippuen siitä, miksi niiden päämäärä oli asetettu etukäteisillä kysymyksillä. Jos kysyttiin jotakin perheenjäsenen ikää koskevaa suuntautui tutkittavien mielenkiinto maalauksen henkilöiden kasvoihin, joista voi saada ikää koskevia vihjeitä, ja jos kysyttiin perheen sosiaalista taustaa,

suuntautuivat katseet lähinnä maalauksen henkilöiden vaatetukseen, joka kasvoja paremmin heijastaa sosiaalista asemaa [23].

7.2. Attention

Kun olet näyttämässä sedällesi savukkeensyöttimen toimintatapaa huomaat, ettei hän seuraa tarkasti, joten näytät asian uudelleen. Huomiosi perustuu siihen, että tarkkaat hänen katseensa suuntautuneen muualle. Siten voit havaita, "seuraako" hän esitystäsi, ja kiinnittää hänen- ja omaa lisähuomiotaasi demonstraatioon. On mahdollista kiinnittää visuaalista huomiota johonkin, vaikka ei kiinnitäkään siihen suoraa katsetaan [19]. On myös mahdollista katsoa jotakin suoraan ilman, että havaitsee todella silmin tarkkailtavaa kohdetta, kuten päiväuneksinnassa, tai tuijotettaessa muutoin vain eteensä. Kuitenkin on mahdollista tällöinkin päätellä suuntautuuko katse tarkkaavaisessa mielessä kohteeseen huomioimalla silmien katselukulman pysyvyys (Ref. [9], pp. 50-65). Silmien huomioiminen avaa uuden kanavan ajateltaessa sitä, mihin mielenkiinto on kohdistunut. Vaikutusta voi verrata siihen kun lapsi tekee havainnon, että vanhempien katseen suuntautuminen paljastaa luotettavasti heidän olevan todella kiinnostuneita lasten puuhista, ja vanhempien katseen harhailu taas antaa aiheen olettaa, että he eivät huomioi lasten tekemisiä. Psykologi Jerome Bruner kuvaa tätä seuraavasti: "Se, mitä on opittu hallitsemaan on sen seurantamekanismin hallinta, jolla toisten huomion kiinnittymistä tarkkaillaan: Se, mihin tulee katsoa, kun halutaan seurata toisten ihmisten huomion keskittymistä...". Äiti sanoo "kissii" katsoessaan kissaa, ja "mene ulos" katsoessaan ovea, ja niin edelleen (Ref. [6], p. 269).

7.3. Reference

Setäsi vierailuun liittyen sisältyy silmillä hahmotettavaan vaikutelmaan useita referentiaalisia seikkoja. Kun setäsi sanoo "Mikä tämä on?" katsoen samalla jotakin erityistä esinettä, on kyseessä lingvistiksessä mielessä deistinen ilmaus (deictic); sana "Deixis" juontuu kreikankielestä, ja tarkoittaa osoittavia tai indikaatiivisia sanoja. Sellaiset sanat taas, kuten "tuoli" ja "pöytä" ovat nondeistisiä, koska niiden tavanomaisesti ymmärretty referentti on osa niiden merkitystä; sanalla "tämä" taas ei ole mitään erityistä kiinteää referenttiä, vaan ne saavat erityismerkityksensä henkilöiden käyttäessä niitä puheessaan (Ref. [13], p. 128). Kun setäsi kysyy "Mikä tämä on?" täydentyy ilmaisun merkitys tarkatessasi sitä, mihin hänen katseensa on suuntautunut (tiettyyn kuvaan). Toisenlaatuinen referentiaalisuus liittyy siihen, KENELLE me olemme puhumassa. Setäsi tiedustelee kahdesti auton kuntoa käyttäen täsmälleen samanlaista ilmausta. Eroavaisuus on siinä, katsooko hän sinua vai huonekumppaniasi. Kysymyksillä on erilainen osoitus (addressee). Molemmipuolinen katseensuuntautuminen ja silmät jaksoittavat keskustelua yleisesti, indikoiden siihen, kenen vuoro kulloinkin on kyseessä, ja siten synkronoiden eri puhujien osallistumista, ja niin edelleen [1,8]. Silmät voivat toimia myös intermodaalisen vuorovaikutuksen funktioissa, ja meidän esimerkissämme silloin, kun kuvaat autokolaria TAPAHTUMANA käsiesi liikkeillä katsoen samalla sekä huonekumppaniasi -että setääsi. Kun setäsi huomaa sinun katsovan häntä "käsiesi liikkeillä" hän samalla kiinnittää NIIDEN liikkumistapaan huomiotaan. Kun sitten huitaiset kädelläsi voimakkaasti ilmaistaksesi vastenmielisyyttäsi punaista päin ajavia kohtaan, katsot vuorostasi setääsi vahvistaaksesi, ettei tämä liike ole varsinainen informaatiolisä, vaan ainoastaan eräänlainen puheen "alleviivaustapa".

8. Eye-Responsive Graphics and Sound

Silmänliikkeitä käsittelevä kirjallisuus kokeellisessa ja sovelletussa psykologiassa ei ehkä ole erityisen käyttökelpoista suunniteltaessa tietokoneiden suorittamaa silmänliikkeiden seuranta. Tehtyjen tutkimusten pääpainopistealue on siinä, miten ihmiset katsovat, eikä siinä, kuinka katsotut seikat voisivat tuottaa katsomisiin responseja. Eräs koeasetelmista, joissa tarkasteltiin katsottujen seikkojen mahdollista reagoivuutta katseisiin liittyi erääseen aiemman MIT Architecture Machine Groups'n projektiin, jonka nimi oli "Gaze-Orchestrated Dynamic Windows", jossa dynaamista näyttöä käyttäen sovellettiin tarkkailijoiden silmäkontrollointia useisiinkin videoituihin episodeihin (Ref. [4], Chapter 4). Tarkoituksena oli luoda visuaalinen analogia erälle informatiiviselle maailmalle, jossa ikään kuin

jäljiteltiin tilannetta, jossa ihminen kommunikoi aluksi tavallaan ei-inhimillisen ja ei-eläimellisen keskustelukuppanin kanssa (one of brevity), joka sisältää niin fragmentaariutta kuin vaihteluakin [14] niin, että observoija kykeni "suodattamaan" SEN ottaman kontaktin TÄHÄN maailmaan SEN omasta maailmasta; observoija kykeni tässä hyödyntämään sisäänrakennettua visuaalisen valinnan mekanismia. Kokeessa observoija istui seinäkokoisien värinäytön edessä, jossa samanaikaisesti esitettiin nelisenkymmentä TV-kuvaa niin, että jokaisen äänet sulautettiin yhdeksi "kokkarikutsunomaiseksi" taustaaääneksi. Observoija käytti erityisiä silmälaseja, jotka rekisteröivät silmänliikkeet. Mikäli observoija kiinnitti huomionsa johonkin näkemäänsä episodiin tietyn ajan (tätä aikaa varioitiin), kaventui kuultava äänialue yhdestä lähteestä kuuluvaksi (eräänlainen auditiivinen zoom -vaikutelma), ja jos katsekontakti jatkui systeemi "jäädettiin" tarkasteltavan videokuvan, samalla kun muut jatkoivat esitystään keskeytyttä, ja lopulta näytti jäädetytyn ruudun kokoseinän esityksenä. Jos observoija halusi palata usearuutuiseen näyttöön hän käänsi tuolinsa käsinojassa sijaitsevaa kytkintä. Olennaiselta osin tällainen vastaa tilannetta missä toiminta tapahtuu väentungoksessa (esimerkiksi toimistossa), jossa jokainen kilpailee huomiostasi usein niin, että joku toinen tulee poimituksi keskeiseksi hahmoksi, jota muut seuraavat. Tavallisinta on, että huomiota saadaan niiltä, jotka eivät meneillään olevaa "esitystä" seuraa, sekä myös siltä, joka on huomion keskipisteenä (Presidentti Reaganin ja Valkoisen Talon lehdistötilaisuudessa vallinnut vuorovaikutus TV- uutisoinnin aikana on tällaisesta tyypillinen esimerkki). Kun olet saanut asioitua asiiasi poimimasi henkilön kanssa katselet tavallisesti yleisesti ympärillesi, implisiittisesti "jättämällä kentän vapaaksi" muille ihmisille, ja aloittamalla HEHKUTUKSESI uudelleen. Systeemissämme käsinojakytkimen käyttö oli analogista sille, kun fiksaatio tiettyyn henkilöön loppuu, ja ryhdytään tarkastelemaan ympäröivää ihmisryhmää yleensä. Laajemmin: mahdollisuus käyttää sellaista graafista näyttöä, joka vastaa katseesi suuntautumisiin, nostaa esiin näköaloja uudeltaisesta tietokonegrafiikasta -"katseeseenreagoivasta grafiikasta". Mutta eivätkö graafisen alan taiteilijat ole aina otaksuneet, että heidän luomuksiaan katsellaan? Eivätkö he aina olekin olleet erityisen tietoisia siitä, että he käyttävät "visuaalisia kuvia"? Eivätkö he puhu viivoista ja kulmista, joilla silmänliikkeitä ohjataan, ja joiden avulla "silmiin piirretty kuva"? Kyllä, kaikkea tätä on sanottu, mutta silti on traditionaalisesti oltu tekemisissä "esineiden katsomisen" kanssa, eikä sen kanssa, mitä katsotut esineet voisivat tehdä KATSOMISELLE. Observoijan kannalta ovat silmienliikkeisiin reagoivat grafiikat ennustamattomia, vaikka katsomisen kohteina olisikin ihmisiä tai eläimiä. Myös äänilähteet kolmiulotteisessa tilassa voidaan samanaikaisesti hyvinkin paikantaa silmin. On tunnettua, että katsomme siihen suuntaan, josta meihin suuntautuva ääni herättää kiinnostuksemme. On havaittu, että olemme herkistyneet kuulemaan tiettyjä asioita katseemme suuntautumisen myötä [21]. Niinpä silmienliikkeille herkkien grafiikkojen tulisi sisältää niin auditiivisia kuin visuaalisiakin elementtejä.

9. A Multi-Modal "Self-Disclosing" System

Tutkimuksemme kohteena on multimodaalinen interaktio (multi-modal interaction); se, miten puhuminen, liikkeet, ja katsominen työskentelevät yhdessä. Ajatelkaamme sellaista erityistä sovellusta, jossa meillä on katse-kontingetti tietokonenäyttö, joka on "itsestääntiedottava" (Self-Disclosing) (Ref. [4]), Chapter 6). Tällainen systeemi on varustettu aistimaan läsnäolosi ja normaalin käyttäytymisesi. Kyseessä on täydellinen värinäyttö, johon liittyy silmiänliikkeiden seuranta, jolla systeemi paikallistaa katseesi kohdistumisen näyttöruudun alueella. Voit puhua tietokoneen kanssa automaattisen puheentunnistimen välityksellä, ja koskettaa sekä osoittaa näytöllä tiettyjä asioita. Laite kykenisi vastaamaan teksti-, ja grafiikkamuodossa, sekä tuottamalla syntesoitua, tai nauhoitettua puhetta. Laite voisi rajoittaa tuottamaansa informaatiota mielenkiintosi suuntautumisten mukaan, joita osoitat sille kanssakäymisesi muotojen avulla (vis-a-vis) paikantamalla seikkoja näytön alueella tavalla, joka soveltuu sinun työtahtiisi. Laite ei eroaisi juurikaan aiemmasta esimerkissämme esitetystä suosikkisedästä.

9.1. The Computer as Obliging Host

Otaksukaamme, että tietokonenäytöllä olisi esitettynä aidon huoneen seinä maalauksineen, takkoineen, laivan pienoismalleineen, kynttilänjalkoineen, sekä takan edustalla olevine kipinäsuojineen. Tietokone emuloi tässä mallissa isäntää, joka kommentoi näkymää ja näyttää siitä tarkemmin selittäen seikkoja, jotka sinua kiinnostavat. Tietokone seuraa silmiesi liikkeitä huomioiden katseensuuntautumisiin kuluvan

ajan, ja antaa sen mukaisesti selityksiä. Tietokoneessa on tallennettuna tekstimuotoinen database, jonka organisoituminen noudattaa näytössä esitettyjä kuvallisia aspekteja, kuten sen strukturaalisia aspekteja, kuten "maalaukset", "laivojen pienoismallit" jne., joista muodostuu se "käsikirjoitus", josta tietokone tuottaa syntesoitua puheinformaatiota. Mikäli katse vaeltelee huoneen seinällä erityisesti mihinkään kohdistumatta kertoo tietokone yleisiä asioita: "Tämä on suosikkiseinämme talossa, johon olemme koonneet esineitä matkoiltamme kaikkialta maailmasta...", jne. Tietokone ei tällöin syvenny mihinkään erityiseen teemaan. Mikäli tietokoneen käyttäjä katsoo kiinteästi jotakin erityistä seikkaa, alkaa tietokone kertoa siitä. Mikäli käyttäjä katselee tovin kynttilänjalkoja takan reunustalla, kertoo kone: "No joo, me ostimme nämä kynttilät viime vuonna Philadelphiasta. Ne on tehty suunnilleen vuonna 1760; niin meille ainakin kerrottiin...". Mikäli käyttäjä jatkaa katsomistaan, esittää tietokone kynttilänjaloista lähikuvan, ja kertoo yksittäisempiä tietoja esineistä. Jos käyttäjä vaihtaa katseensa suuntaa toistuvasti esimerkiksi kynttilänjaloista kipinäsuojukseen, ja takaisin, kertoo tietokone näiden esineiden väliseen inferenssivaikutukseen liittyviä seikkoja, jotka eivät enää liity esineisiin sinällään, vaan jollekin yleisemmälle niitä kumpaakin selittävälle tasolle, kuten siihen, että ne ovat antiikkiesineitä, tai että molemmat ovat kiiltäviä. Se, miten tietokone esineiden välisen inferenssin kategorioi, riippuu siitä, mihin katsoja katseensa keskittää, ja millaisia sanallisia tietokoneelle suunnattuja kysymyksiä hän samalla esittää. Siten "lopputulokset" riippuvat siitä, millaisia mielenkiinnon suuntautumisten alueita käyttäjällä on, ja millaisia katseensuuntautumisia niihin liittyy.

9.2. The Locus of Initiative

Millainen suhde tällaisessa kanssakäymisessä tietokoneen ja käyttäjän välillä vallitsee? Missä on alullepaneva ja käynnistävä keskus (the center of initiative)? Ajatelkaamme tapausta, jossa tennispelaaja tarkentaa lyöntiään seinäharjoittelulla. Kun pelaaja lyö pallon seinään riippuu sen suuntautuminen takaisin pelaajaa kohti käytetystä lyöntikulmasta ja lyönnin voimakkuudesta, kuten myös palloon mahdollisesti lyönnillä aiheutetusta kiertestä. Otaksukaamme, että meillä olisi sellainen "seinä", joka voisi antaa siihen osuvalle pallolle sysäystä niin, että paluukulma ja pallon nopeus muuttuisivat. Otaksukaamme edelleen, että "seinä" oppisi pelin kuluessa poimimaan pallon hallitusti, ja lyömään sen takaisin mielekkäällä tavalla. Esittelemässämme systeemissä tietokone toimii paljolti samalla tapaa kuin tällainen hypoteettinen "reagoiva seinä". Tietokoneen näyttö on käyttäjälle implisiittinen kutsu jollakin tavoin vastata; samalla tavoin myös tennisharjoitteluun käytetty "reagoiva senä" kutsuisi tennispelaajaa harjoittelemaan. Käyttäjä katsoo tietokonenäyttöä ja ohjaa katseensa suuntautumisilla tietokoneen responseja. Tietokone toimii alullepanevana tekijänä silloin, kun se, sekä käyttäjä joutuvat kanssakäymisen kannalta "kuolleeseen pisteeseen", ja se pyrkii tällöin tarjoamaan tuoreita vaihtoehtoja "joutuvat ulalle" (lull). Kyseessä olisivat tällöin vähemmät käytetyt katseensuuntautumisasettelut, tai sellaiset katseensuuntautumisten mallit, jotka vähemmän korreroituvat käytettyjen mallien kanssa, ja siten ne viittaavat myös vielä käyttämättömiin mielenkiinnon suuntautumisten alueille. Inhimillinen vuoropuhelu on samalla tavoin episodista. Ihmiset vaihtelevat keskusteluissaan puheenaiheita, ja jatkavat sitä niin pitkälle kuin voivat "joutumatta ulalle" (jutut tulevat aivan UFOIKSI, tarinat ovat HUUHAA, puheet alkavat olla RONALDEJA jne., suom huom.), tai keskustelun tyrehtyttävään kuolleeseen pisteeseen. Keskustelu viriää uudelleen, mikäli tuoreita näkökulmia löytyy. Yleisesti, käyttäjän ja koneen välinen vuorovaikutus on molemminpuolista provosointia ja virikkeidenantoa, jossa käyttäjä on alullepanevana (initiative) siinä mielessä, että hänellä on tiettyä uteliaisuutta näytönsisältöön, ja tietokoneella tapauksissa, joissa käyttäjä tavalla tai toisella jähmettyy paikalleen (stalled). Mikä tahansa erityinen muutos minkä tahansa seikan suhteen loppuu, kun käyttäjä ei enää katsele seikkaa, tai tietokone on sanonut kaiken seikkaa koskevan sanottavansa, kumpi tahansa näistä pikemmin tapahtuukaan. Niinpä kun tietokone ON itsestääntiedottava systeemi (self-disclosig system), ovat käyttäjän toimenpiteet systeemin olennaisia komponentteja. Molemmat dialogin osapuolet vapauttavat toisistaan informaatiota; toinen tekee sen seuraamalla vastapuolen silmänliikkeitä ja fiksaatioita niihin liittyvine sanallisine kommentteineen, ja toinen taas osallistuu prosessiin graafiseen näyttöön vaikuttaessaan ja tietokoneelle syntesoidulla puheella.

9.3. Changing the Subject

Systeemi saattaisi asettaa alustavan dialoginalkamiskohdan yksinkertaisesti esittämällä näytössä jotakin.

Käyttäjä voisi myös aloittaa dialogin esimerkiksi sanomalla: "Kerro minulle 16 -vuosisadan japanilaisesta arkkitehtuurista". Kun tällainen pyyntö esitetään systeemille, jolla on tarvittava database kyseisestä aiheesta, tietokone generoi alustavan esityksen aiheesta näyttöön, ja vuorovaikutus voi alkaa. Ongelmana olisi tällöin se, että voiko alisysteemi antaa käskyn toisen alisysteemin tuottamisesta, tai kokonaisjärjestelmää koskevan muutoksen tekemisestä. Eräs tapa muuttaa aihetta voisi perustua johonkin merkittävään muutokseen tavassa, jolla käyttäjä näyttöä tarkastelee. Tarkkaavaisina ihmiset ovat taipuvaisia katelemaan näytöstä sellaisia seikkoja, joille on annettavissa nimi [7]. Otaksukaamme, että käyttäjä tarkastelee kynttilänjalkaa, ja että tietokone kertoo hänelle kynttilänjalkojen kuuluneen aikanaan jollekin kuuluisalle henkilölle, kuten Thomas Jeffersonille (kuten laivan pienoismallinkin). Otaksukaamme edelleen, että systeemi toistuvasti huomaa käyttäjän tarkastelevan esimerkiksi laivan pienoismallia kiinteästi, mutta niin, ettei hänen katseensa kuitenkaan keskity laivan osiin sinällään. Tästä tietokone päättelisi, että käyttäjä on kiinnostunut jostakin osia yhdistävästä tekijästä, kuten kytkeytymisestä Thomas Jeffersoniin, ja tarjoaisi näyttöön valokuiva, jossa Thomas Jefferson seisoo kotonaan Monticellossa, istuu senaatissa, jne. Riippuu systeemin informaatiobasiksesta, onko siinä anekdoottista tai kuvallista informaatiota Thomas Jeffersonista, ja tietokone on kyvytön tietämään jotakin, jota se ei ole oppinut.

9.4. The Conversational "Contract"

Se, kuinka käyttäjän ja tietokoneen välisen kommunikaation joustavuus toteutetaan, riippuu tavasta, jolla osapuolet tekevät sitä koskevan "sopimuksen" (Cf. Ref. [12]). Kysyikö käyttäjä nimenomaan näyttämään esimerkissä esitetyn huoneen näyttöön? Onko etukäteisesti olemassa sopimus, että juuri kyseinen huone tuotetaan näyttöön? Jos sopimuksia on olemassa, tietokoneen mahdollisuudet tuottaa joustavasti näyttöön (vapaasti assosioituvasti) materiaalia Thomas Jeffersonista aina osavaltionlipusta vapaudenpatsaaseen New Yorkin satamalla (ja niin edelleen loputtomiin) ovat rajatut. Etukäteisestä sopimuksesta kiinnipitäminen on tärkeää tapauksissa, joissa pyydetään tietoja määräytyistä aiheista; systeemin on huomioitava sopimus ainakin niin pitkään kunnes käyttäjä ei enää ole aiheesta kiinnostunut. Se, missä määrin systeemi sietää poikkeamia riippuu pitkälti siitä, kuinka suurella määrällä se on personoituva niin, että se huomioi käyttäjän reaktiot ja keskenään ristiriitaiset reaktioiden kohdistuvuuden vaihtelut.

10. Ongoing Research

Kirjoittaessani tätä raporttia osallistun juuri tutkimusohjelmaan, jonka tarkoituksena on tutkia ja arvioida silmänliikkeitä tietokoneen ja käyttäjän välisessä dialogissa erillisen alueenaan, sekä yhdistettynä puheella ja manuaalisella osoittamisella tapahtuvaan vuorovaikutukseen. Kuten todettu, on korostus katsomiskäyttäytymisellä tapahtuvassa mielenkiinnon osoittamisessa ja huomionkeskittämisessä, ja referenssissä eräänä keinoista. Vuorovaikutekentän (interface) laitteisto koostuu silmänliikkeiden kaukoseuranta-laitteistosta (remote eyetracker); kosketusherkästä värinäytöstä (touch screen color display); automaattisesta puheentunnistussysteemistä (automatic speech recognition system); sekä puhesyntesoija-systeemistä (speech synthesis system). Ohjelmat on kehitetty tulkitsemaan ja integroimaan käyttäjän katsomista, puhumista ja osoittamista, sekä kartoittamaan näyttöön/ puhesyntesoijan kautta annettaviksi soveliaita responseja. Tutkimme kahta silmänliikkeiden avulla osoittamisen mahdollisuutta: Intentionaalista, jossa silmiä käytetään näytössä esiintyvien seikkojen soittamiseen, ja insidentaalista (incidental), jossa silmien kohdistusliikkeitä tarkastellaan pikemminkin silmillä suoritettuun spontaaniin katseluun liittyvänä. Silmäkontaktien notaatiota tietokoneen ja ihmisen kommunikaatiossa tutkitaan. Inhimilliset observoijat tarjoavat tutkimuksellista palautetta, jonka mukaan kehittää ohjelmia, ja edistää käyttäjien hyväksymien tapojen saatavuutta tietokonekommunikaation modaaliteettina. Työn päämääränä on laajentaa niitä tapoja, joilla ihmiset ovat interaktiossa tietokoneiden kanssa, ja joiden avulla tietokoneita olisi miellyttävämpi käyttää.

11. An "Expert Conversationalist"

Kun Asiantuntijajärjestelmiä tarjoutuu käytettäväksi lääketieteessä, öljynetsinnässä, ja

tietokonejärjestelmien konfiguroinnissa, saattaa tietokoneesta tulla "asiantunteva keskustelija" (expert conversationalist). Ajatelkaamme inhimillistä keskustelua politiikasta, viime kesälomasta, vastaostetusta autosta, mistä tahansa. Yksilöllisten keskusteluaiheiden piirissä, ja riippumattomina erityisistä teemanasetteluista ne näyttävät käytännön taitonsa käsitellä mitä tahansa puheenaihetta. Inhimillistä keskustelua sinällään ohjataan seuraavasti: katsekontakti katkeaa, kun haluat puhua; ei tapahdu mitään, mikäli joku toinen henkilö katselee oikeaa kohtaa samalla kun sinä katsot jotakin muuta; kuvaat tapahtumia ja asioita käsilläsi (kuten reittiä kotiin ja pyydystetyn kalan kokoa). Tällaiset kommunikatiiviset taidot ihmisellä ovat suurelta osin tiedostamattomia, ja kehittyvät pitkän harjoituksen myötä. Voidaanko tällaisia taitoja siirtää tietokoneelle? Meneillään oleva työmme ei välttämättä nivelly "asiantuntijajärjestelmiin", ja keskustelykykyisen tietokoneen älykkyys ei myöskään välttämättä liity nykyisiin psykologian, lingvistiikan, ja keinotekoisien älyn (AI) tieteenaloihin. Tarvittavat näkemykset eivät välttämättä siten tule tämänhetkisestä psykologiasta, lingvistiikasta, tai keinotekoisien älyn teoriasta. Sitä vastoin näillä aloilla suoritettava kehitystyö tulee vaikuttamaan siihen, miten saavutetaan tietokoneen ja ihmisen välisessä kanssakäymisessä hienoja inhimillisiä kvaliteetteja.

Acknowledgement

The research described herein was supported in part by Grants IST-8414574 and IRI 8615741 from the National Science Foundation Division of Information, Robotics, and Intelligent Systems.

References

- [1] M. Argyle and M. Cook: "Gaze and mutual gaze", Cambridge University Press, Cambridge, England (1975).
- [2] D.E. Berlyne: "Curiosity and explanation", *Science*, 153, pp. 25-33 (1966).
- [3] D.E. Berlyne: "The Influence of Complexity and novelty in visual figures on orienting responses", *Journal of Experimental Psychology*, 55, pp. 289-296 (1958).
- [4] R.A. Bolt: "The human interface", Van Nostrand Reinhold, New York (1984). Translated into Japanese and distributed in Japan through the Tuttle-Mori Agency, Inc., Tokyo.
- [5] B. Bower: "The face of emotion", *Science News*, 128, pp. 12-13 (July 6, 1985).
- [6] J.S. Bruner: "From communication to language - a psychological perspective", *Cognition*, 3, 3, pp. 255-287 (1974/1975).
- [7] R.M. Cooper: "The control of eye fixations by the meaning of spoken language", *Cognitive Psychology*, 6, pp. 84-107 (1974).
- [8] G.D. Cumming: "Eyemovements and visual perceptions", *Handbook of perception: Vol. IX, Perceptual processing*, E.C. Carterette and M.P. Friedman, Eds., Academic Press, New York (1978).
- [9] D. Kahneman: "Attention and effort", Prentice-Hall, Englewood Cliffs, New Jersey (1973).
- [10] G.R. Loftus and H. Mackworth: "Cognitive determinants of fixation location during picture viewing", *Journal of Experimental Psychology*, 4, 4, pp. 565-572 (1978).
- [11] N.H. Mackworth and A.J. Morandi: "The gaze selects informative details within pictures", *Perception and Psychophysics*, 2, 11, pp. 547-552 (1967).
- [12] A.P. Martinich: "Communication and reference", Walter de Gruyter, New York (1984).
- [13] G.A. Miller: "Language and speech", W.H. Freeman and Company, San Francisco (1981).
- [14] H. Mintzberg: "The nature of managerial work", *Theory of Management Policy Series*, Lawrence Erlbaum Associates, Publishers, Englewood Cliffs, New Jersey (1980).
- [15] N. Negroponte: "The sensory apparatus of computers", *Prochure for the Media Lab Dedication Ceremonies*, MIT Media Lab (Oct. 1985).
- [16] R.S. Nickerson: "Using Computers: The Human factors of Information Systems", MIT Press, Cambridge, Massachusetts (1986).
- [17] S. Nishida: "Speech Recognition Enhancement by Lip-Information", *CHI '86 Conference Proceedings*, pp. 198-204 (April 1986).

[18] E.D. Petajan: "Automatic Lip-reading to Enchange Speech Recognition", Ph. D. Thesis, University of Illinois at Urbana-Champaign (1984). (Dr. Petajan is now at Bell Labs.).

[19] M.I. Postner: "Orienting of attention", Quarterly Journal of Experimental Psychology, 32, pp. 3-25 (1980).

[20] M.I. Postner, M.J. Nielszen and R.M. Klein: "Visual dominance: an information-processing account of its origins and significance", Psychological Review, 83, 2, pp. 157-171 (1976).

[21] D. Reisberg, R. Schreiver and L. Potenzen: "Eye position and the control of auditory perception", Journal of Experimental Psychology: Human Perception and Performance, 7, 2, pp. 318-323 (1981).

[22] W.H. Sumby and I. Pollak: "Visual contribution to speech intelligibility in noise", journal of the Acoustical Society of America, 26, 2, pp. 212-215 (March 1954).

[23] A.L. Yarbus: "Eyemovements and vision", Translated by B. Haigh, Plenum Press, New York (1967).

[24] L.R. Young and D. Sheena: "Eye-movement measurement techniques", American Psychologist, 30, 3, pp. 315-330 (1975).

(Manuscript received July 15, 1987)

Richard A. Bolt is the Principal Research Scientist at the Massachusetts Institute of Technology Media Laboratory and is Director of the Laboratory's Human Interface Group. He holds a Doctorate in Experimental Psychology from Brandeis University. He has lectured nationally and internationally on human/computer interaction, and is author of THE HUMAN INTERFACE published by Van Nostrand Reinhold and distributed in Japan by Tuttle-Mori Agency, Inc.

Voice Interaction in a Integrated Office and Telecommunications Environment

Christopher Schmandt, Barry Arons, and Charles Simmons

*Media Laboratory, Massachusetts Institute of Technology
Proceedings, American Voice Input/Output Society Conference
1985, AVIOS, Palo Alto, CA, pp. 51-61 (1985).*

Suom. **Timo Kinnunen** 13.3.1989

Introduction

CONVERSATIONAL DESKTOP -projekti tutkii puheensyöttö- tulostusteknologioita (speech input/output technologies) laitteistovälitteiselle (machine mediated), äänen avulla tapahtuvalle kommunikaatiolle toimistoissa ja teleliikenneympäristöissä. Työn keskeinen alue liittyy vuorovaikutekentän (interface) suunnitteluun, jossa huomioidaan useita inhimillisen keskustelukäyttäytymisen Aspekteja. Tämä sisältää kyvyn jatkaa keskustelua silloinkin, kun syöte (input) on epämääräinen, ja kyvyn käyttää hyväksi syntaktista ja akustista kontekstia keskustelun edetessä, kuten myös laitteen herkkyyden tunnistaa erilaista äänellistä informaatiota. Tämä on hyödyllistä ympäristöissä, joissa puhetta käytetään erilaisiin tarkoituksiin, kuten audiomemoihin, puhelinkeskusteluun, ja hälytysmerkkifunktioihin -sen lisäksi, että ääni liittyy laiteohjauksen komentokanavaan. Aikaisempi työ on osoittanut syntaktiseen analyysiin perustuvan dialogin olevan käyttökelpoisen ajateltaessa sitä kopioimistapahtumana, johon liittyy tunnistamisvirheiden paikantaminen [Schmandt 82], vaikka käytetty kieliopillinen kuvaus olikin karkeasti koodattu tiettyä sovellusta varten, ja laajennettavissa ainoastaan suunnittelullisesti. PHONE SLAVE [Schmandt 84, Schmandt 85] kykeni hyödyntämään menestyksellä ihmisten halukkuutta keskusteluun tietokoneen kanssa mutta se oli siinä mielessä passiivinen systeemi, ettei se kyennyt käyttämään hyödykseen tietoa keskusteluun liittyvistä muista aktiviteeteista, joita laitteen käyttäjällä oli. Tässä projektissa tarkoituksena on yhdistää nämä molemmat lähestymistavat.

The Environment

Tämä projekti perustuu integroituun toimistotyöasemaan, jossa yhdistyvät tehokkaan henkilökohtaisen tietokoneen, ja älykkään teleliikennöintisysteemin ominaisuudet. Sen lisäksi, että käytettävissä ovat

tavanomaiset henkilökohtaiseen tietokoneeseen liittyvät sovellukset, voi tämä työasema olla aktiivisena tekijänä digitaalisessa verkossa. Työasema käsittelee omistajansa asiakirjoja, matkasuunnitelmia, puhelinluetteloita, ja myös vastaanottaa viestejä ja aktivoi auditiivisia muistioita, jotka liittyvät käsiteltäviin tapauksiin. Kuten tulee myöhemmin esille, niin mitä enemmän työasema tietoinen omistajansa aktiviteeteista, sitä suurempi on sen kyky muodostaa oikeita inferenssejä oman käyttäytymisensä ohjaimina vastatessaan ulkoisen maailman ärsykeisiin. Mitä teleliikenteeseen silmukohtiin (node) tulee, perustamme tämän työmme näkemykselle "point-to-point" -tyyppisestä kommunikaatiosta, johon samanaikaisesti liittyy ääni- ja datalinkkejä, joista jälkimmäisen ei tarvitse perustua suurnopeuksiselle tiedonsiirrolle. Lisäksi silmukohdat (nodes) kykenevät hyödyntämään niihin liittyviä aktiviteetteja, jotka vaativat paikallisten databasejen käyttöä, kuten tapaamisten järjestelyä eri työasemilla työskentelevien välille, tai käsittelemään älykkäästi puhelin yhteyden kontrollisignaaleja. Kun jotakin työasemaa pyydetään "yhdistämään puhelu X:lle", se ensin ottaa yhteyden digitaalisesti X:n työasemaan määritelläkseen, voiko X vastata puheluun, ja jos haluaa, niin mihin osoitteeseen (puhelinnumeroon) äänilinkki yhdistetään. Samoin voidaan muodostaa digitaalinen kytkentä jonkun toisen silmukohdan (node) prosessiin peruutettaessa jotakin tapaamista huomioiden kunkin työasemankäyttäjän aikataulut. Meidän CONVERSATIONAL DESKTOP sovelluksessamme käytetään Sun Microsystems -työasemia, joissa on Internet -protokolla ja jossa datalinkkinä on Ethernet (hardware); tavanomaisia analogisia puhelin yhteyksiä käytetään audiolinkeinä. Digitaalisen puhelinliikenteen ja siihen liittyvien apuprotokollien (kuten ISDN) kehittyessä on otaksuttavaa, että ääni- ja datakanavat tulevat saataville yhtenä integroituneena puhelinliikennesysteeminä jo lähitulevaisuudessa. Jokainen työasemista on varustettu lukuisilla puhe-periferaaleilla, kuten puheentunnistus, sen syntesointi, puheen digitaalinen tallennus ja toisto asianmukaisilla laitteilla. Laitteet, joita käytetään, voidaan konfiguroida ajoaikaan (run-time); systeemiä voidaan ajaa siten haluttaessa myös vähäisemmällä kapasiteetilla ja käyttäen pelkästään sen eräitä osa-alueita. Työmme pääpainoalue on näiden puhe-tekniologioiden välisessä synergiassa, erityisesti sellaisissa konteksteissa, joissa ääntä hyödynnetään tehtäväratkaisuisissa, jotka sisältävät sisäisten moniyhteyksien huomioimista (interrelated) ja sisäisten moniliittymisten muodostamista (interconnected). Työasema on suunniteltu niin, että käyttäjä voi ohjata sitä puheellaan kun hän liittyy väliintulevana monikeskeiseen keskusteluun (interleaved) muiden silmukohtien (nodes) kanssa. Saataville tulleiden operaatioiden määrä tätä kirjoitettaessa sisältää: aikataulujärjestelyt yksilöiden ja ryhmien välille, otettujen puhelujen valmistelun (placing), tulevien ääniviestien vastaanoton ja tallentamisen äänimemoiksi, jotka suhteutetaan tiettyihin työtapauksiin. Tulevien viestien käsittelyn pohjana on ANSWERING MACHINE, jota on kuvattu jo PHONE SLAVE:n yhteydessä.

CONVERSATIONAL ASPECTS

DESKTOP on luonteeltaan keskustelullinen: dialogi on yhtä hyvin tasainen takaisinkytkentöjen virta kuin keino paikantaa epäselvyyksiä ja virheitä puheentunnistimella (speech recognizer). Puheentunnistimen tuloste on taipuvainen sisältämään kohinaa, jolle luonteenomaista on väärät sanojen asettamiset (insertion), sanojen väärät korvaamiset (substitution), ja erilaiset määrittelemättömät virheen sanojen käsittelyssä. On välttämätöntä rakentaa joustava kielioppitulkki (robust parser), jolla peilataan (scan) puheentunnistimen tuottamaa aineistoa, kuten myös rakentaa datastrukturi, johon sisältyvät tiedot siitä, mitä osaa ko. aineistosta käytetään gereroitaessa dialogia. Tavanomaiset kieliopintulkintatekniikat (parsing techniques) liittyen luonnollisen kielen prosessointiin [Winograd 83] ovat yleisesti riittämättömiä, koska ne perustuvat siihen, että sanalliset syötöt (tavallisesti kirjoitettuja) ovat oikein asetettuja. Sovellettuna ratkaisuna on kontekstivapaa kielioppi (context free grammar) ja kielioppitulkki (parser), joka perustuu Unix YACC (YET ANOTHER COMPILER-COMPILER) kieliopintulkintagenerointiin, jossa jokainen merkki (token) on eräs tietyn syntaktisen luokan tapaus, kuten esimerkiksi "käsky, jonka antamiseen tulee liittyä annettava päivämäärä". Kielioppitulkki (parser) käyttää puheentunnistimen aineistoa ja ajaa kaikki jäsenytyneet alaketjut (substrings) YACC:n kautta, joka laskee niille arvot vaihteeltaisesti karsien kieliopin jokaisessa silmukohdassa (node), ja hyppää aina sellaisten kohtien yli, jossa arvot ovat pienempiä kuin todettavissa olevat suuremmat jäljelläolevat arvot. Esimerkiksi sellainen alaketju, kuten kirjainryhmä ABC voisi tulla kieliopillisesti tulkituksi tapauksista, jossa mahdolliset puheentunnistimen antamat vaihtoehdot olisivat ABC, AB-, -BC, A-C, A--, -B-, --C. Arvojen antaminen perustuu siihen, että tunnetaan jo ennalta minkätyyppisiä virheitä puheentunnistin tekee, kuten konnektoidussa puheessa virheet ilmenevät purskeina (bursts), mikä on tulosta tunnistimen

suorittamista virheellisistä segmentointiratkaisuista. Kielioppitulkki antaa pisteitä lukuisille tunnistetuille sanoille, joista työstetään täydellisiä lauseita karsintaprosessissa, jossa annetaan tavallaan lisäbonusta aina niille sanoille, jotka jäävät jäljelle, ja jossa prosessissa säätyy sanojen oikeellisuus. Dialogi, mikä usein tunnistuksen yhdeydessä käydään, on koneen yritys täyttää aukkokohdat muodostetussa kielioppipuussa, joka prosessina perustuu korkeimpien pistemäärien omaavien vaihtoehtojen joukon muodostukseen. Esitettävien kysymysten muodostaminen on kriittistä monistakin syistä. Dialogi soveltaa kaikutekniikoita (echoing techniques) [Hayes 83] varmistaakseen implisiittisesti alkuperäistä kommunikaatiota. Esimerkiksi: "Sovi aikatauluun tapaaminen Walterin kanssa aa.. <muminaa>" voisi tuottaa täydentävän kysymyksen "milloin haluat tavata Walterin?". Esitetyt kysymykset suuntautuvat täydentämään annettuja sanallisia responseja niin kuin ne parhaimmalla todennäköisyydellä ovat tulleet tunnistetuiksi, aina kun se on mahdollista. Toinen keskustelullisen kyvykkyyden aspekti liittyy systeemin metodiin ottaa vastaan puhelimesta jätettäviä viestejä. Soittajia tervehditään nauhoitetulla äänellä, joka kysyy joukon kysymyksiä, ja nauhoittaa saadut vastaukset, samalla kun adaptoitua taukokohdan määrittävä algoritmi (adaptive pause detection algorithm) laukaisee seuraavan kysymyksen. Vastaukset sellaisiin kysymyksiin, kuten "kuka soittaa?", "mihin tämä liittyy?", ja "mistä numerosta teidät voi tavoittaa?" tallennetaan individuaaleiksi äänitiedostoiksi. Peräkkäisten nauhoitusten sekvenssi tarjoaa kontekstin, tai tavan käsitellä audiodata-sisältöä. Jopa silloinkin, kun kone ei tunnista yhtään sanaa vastauksesta esimerkiksi kysymykseen soittajan henkilöllisyydestä, se tietää, että juuri tämä vastaussegmentti on toistettava, kun laitteen omistaja esimerkiksi kysyy: "Kuka jätti tämän viestin?".

Addressability

Huomioimalla lukuisat vihjeet, kuten erityisesti silmäkontaktit, voi henkilö pientä ryhmää koskien määrittellä, onko jokin viesti juuri heille suunnattu. Toivomme voivamme hyödyntää samanlaista tekniikkaa niin, että tietokone voi määrittää, milloin sille puhutaan, vastakohtana puheluille ja toimistossa työskenteleville ihmisille yleensä. Jotta tämä helpottuisi, olemme hyödyntäneet sijoitusjärjestelyissä spatiaalista sijaintia niin, että tietokone sijaitsee käyttäjän (taka) oikealla puolella, ja puhelin (etu) vasemmalla. Systeemin näyttö, jossa näkyvät DESKTOPin tulostamat aikataulu ja tiedot jätetyistä puhelinviesteistä, ovat sijoitetut (suoraan) käyttäjän oikealle puolelle, ja DESKTOPin kovaääninen käyttäjään nähden (etu) oikealle. Puhelimeen liittyvä mikrofoni on kiinnitetty käyttäjän päähän (hands-free), ja puhelimeen yhdistetty kovaääninen sijaitsee käyttäjän (etu) vasemmalla puolella. Käyttäjän TAAKSE on sijoitettu kaksi mikrofonia suunnilleen tasakylkisen kolmion kantoihin, ja niiden avulla määritellään suunta, johon käyttäjä kulloinkin puhuu; näiden mikrofoniin vastaanotto on minimissään silloin, kun käyttäjä puhuu puhelimeen. Takana sijaitsevien mikrofoniin avulla saavutetaan suuri herkkyys suunnanmäärityksessä, koska se hyödyntää ihmisen pään asentojen kulmanmuutoksia [Flanagan 60]. Sekä laitteisto -että ohjelmisto IBM PC:ssä hyödyntävät tätä informaatiota kommunikaatiossaan SUN -työasemaan (Sun Workstation). Samainen laitteisto myöskin kontrolloi puheen taukokohdista ohjautuvia kytkimiä (noise-free ramped switches) suorasta audiosta aina oheislaitteisiin. Silloin kun tietokone itse puhuu analysoimalla nauhoitusta, tai syntesoimalla tekstiä puheeksi, on syöttö puheentunnistimeen käyttäjän kanavalla poikki, jotta käyttäjän puhuminen ei sekoittuisi tunnistettavana olevaan puheeseen tai tekstiin. Systeemi tunnistaa käyttäjän puhetta hänen osoittaessaan puhetta puheentunnistimen suuntaan. Puhelimeen puhuessaan käyttäjälle saattaa tulla tarvetta yksityiseen vuorovaikutukseen DESKTOPinsa kanssa, jollaisen systeemi tunnistaa käyttäjän oikealle suuntautuneesta pään asennosta, ja katkaisee puhelun väliaikaisesti (ei kokonaan).

Context

Suuntaherkät mikrofonit on suunniteltu myös määrittämään taustalta suuntautuvia ääniä (määriteltyinä signaaleiksi, jotka eivät ole peräisin käyttäjän päähän kiinnitetystä mikrofonista), jolloin huomioituu se, että toimistossa työskentelee mahdollisesti muitakin. Tätä läsnäolevaa taustapuhetta käytetään muodostettaessa sitä operaatioiden luokkaa, joiden avulla määritellään tietoa DESKTOP systeemin toimintaympäristön akustisesta kontekstista. Esimerkiksi kun on aika toistaa jokin audiomuistio, niin systeemi ensin tarkistaa TÄMÄN TAUSTASIGNAALIEN LAADUN, ja voi esimerkiksi määrittää sen kuuluvaksi ajanjaksoon, jolloin käyttäjä oli yksin toimistossa. Yleisesti systeemi seuraa sellaista sääntöä,

että käyttäjää ei keskeytetä silloin, kun hänellä on jokin määriteltävissä oleva työtapahatuma meneillään; tulevaisuudessa tutkimustyö tulee pyrkimään priorisoimaan läsnäolevuuteen ja työstä tiedoittaviin elementteihin kuuluvia ääniä. Esimerkiksi systeemi ei keskeytä käyttäjää ohjaamalla hänelle puhelua silloin, kun hän tekee jotakin erityistä työtehtävää, vaan tallentaa viestin. Mitä enemmän systeemi tietää käyttäjän aktiviteettien laadusta, sitä enemmän se kykenee hyödyntämään kontekstia puhesyöttöjen ymmärtämisessä ja siihen liittyvien aktiviteettien ohjailussa. Useimmiten nämä liittyvät puhelinkeskusteluihin. Mikäli huomioidaan kaikki mahdolliset aktiviteetit, viittaa komento "järjestä meille molemmille tapaaminen" sekä käyttäjään, että toiseen osapuoleen; systeemi, joka tuntee henkilöstösuhteet, voi asettaa puhelun ensisijalle. Kun käyttäjä kertoo DESKTOPille menevänsä lounaalle, se tietää asettaa ulossuuntaamiinsa vastauksiin tästä tiedotteen. Samalla tavoin systeemin aktiviteetit voivat ohjautua eksternaalisista tapauksista. Audiomuistuttaja (audio reminder) voi esimerkiksi tuottaa sanasekvenssin: "Kun puhun Barrylle, muistuta minua...". Ja vaikka systeemi ei suoritakaan varsinaista sisällöllistä muistutteen tunnistusta, se tietää silti siitä tarpeeksi tuottaakseen sellaisen automaattisesti soittamalla tietyn puhetiedoston, mikäli sille esitetään pyyntö "Soita Barrylle". Saman puhetiedoston systeemi huomioi myös silloin, kun Barry soittaa käyttäjälle. Muistuttaja voi myös kytkeytyä toimintaan, jos käyttäjä ottaa Barryyn suoran puhelinyhteyden, on menossa kokoukseen, tai kytkeytyä sellaisesta käskystä, kuten "Lähden kotiin".

Future Work

Tällä hetkellä olemme laajentamassa DESKTOPin kyvykkyyttä useillakin tavoilla. ENSIMMÄISEKSI olemme lisänneet fuktionaalisuutta, kuten lentolippujen varausta, ennalta tapahtuvaa säätilan tarkistusta ennen töiden loppumista, etc. Käytettävissä on lukuisia elektronisia databaseja, jotka ovat DESKTOPin tavoitettavissa digitaalisilla puhelinyhteyksillä, ja joilla automaattisesti päivitetään tietoja maailmassa vallitsevista olosuhteista. TOISEKSI olemme kehittämässä personoituvuutta, kuten se, että jokaisen silmukohdan (nodes) aikataulua ohjataan huomioimalla henkilökohtaisia preferenssejä: joku saattaa tuntea vastenmielisyyttä aamuisia kokouksia kohtaan, kun taas joku toinen saattaa pitää vastenmielisenä viiden jälkeen pidettäviä kokouksia. Vaikeampi on haasteena preferenssien yhteensovittaminen huomioimalla jollakin tavoin sen, että kokous saattaa olla jollekin osapuolelle tärkeä. Esimerkiksi minä en tapaisi opiskelijoita mielelläni ennen kymmentä aamupäivällä, kun taas joku tärkeä vieras saattaisi hyvinkin tulla tätä aikaisemmin. Samalla tavoin voisivat monet saapuvat puhelut voisivat keskeyttää joitakin meneillään olevia töitäni, mutta en hyväksyisi mielelläni työtovereideni tai paikalla asioivien henkilöiden suorittamia keskeytyksiä.

Acknowledgement

Työtä on tukenut NTT, (The Nippon Telegraph and Telephone Corporation)

References

- [Flanagan 60] J.L. Flanagan, Analog Measurements of Sound Radiation from the Mouth, J. Acoust. Soc. Am., 32(12) (1960).
- [Hayes 83] P. Hayes and R. Reddy, Steps Toward Grateful Interaction in Spoken and Written Man-Machine Communications, Int 'l J. Man-Machine Studies, 19:231-284 (1983).
- [Schmandt 82] C. Schmandt and E. Hulthen, The Intelligent Voice Interactive Interface, In Human Factors in Computer Systems, NBS/ACM (1982).
- [Schmandt 84] C. Schmandt and B. Arons, A Conversational Telephone Messaging System, IEEE Trans. on Consumer Electr, CE-30(3):xxi-xxiv (1984)
- [Schmandt 85] C. Schmandt and B. Arons, Phone Slave: A Graphical Telecommunications Interface, Proc. of the Soc. for Information Display, 26(1) (1985).
- [Winogard 83] T. Winogard, Language as a Cognitive Process-Syntax, Addison-Wesley (1983).

Mr. Schmandt received his B.S. in Computer Science and his M.S. in computer graphics from MIT. He has Continued his work as a Principal Research Scientist at the Architecture Machine Group, a component of the Media laboratory. His research interests there are focused on interactive systems and human-interface issues, with emphasis on voice interaction and telecommunications.

Barry Arons:
Research Associate;

Mr. Arons received his B.S.C.E. and M.S. in computer graphics and interactive systems from MIT. His research interests include speech input/output, raster graphics, and interactive video.

Charles Simmons:
Undergraduate Researcher;

Mr. Simmons is completing his senior year at MIT. The topic of his Bachelor's thesis is the design and implementation of speech direction sensing hardware and software.

The authors can be contacted at:

Media Laboratory
Massachusetts Institute of Technology
20 Ames Street, Room E15-327
Cambridge, MA 02139