

**Puhepohjaisten keskustelujärjestelmien inhimillistäminen ja
arviointi**

Mikko Hartikainen

Tampereen yliopisto
Tietojenkäsittelytieteiden laitos
Pro gradu -tutkielma
Kesäkuu 2003

Tampereen yliopisto

Tietojenkäsittelytieteiden laitos

Mikko Hartikainen: Puhepohjaisten keskustelujärjestelmien inhimillistäminen ja arviointi

Pro gradu -tutkielma, 67 sivua, 6 liitesivua

Kesäkuu 2003

Tiivistelmä

Tässä työssä tutkitaan puhepohjaisia keskustelujärjestelmiä inhimillisinä järjestelminä. Luonnollista kieltä ja puhetta käyttävä järjestelmä on inhimillisempi kuin esimerkiksi tekstiä käyttävä järjestelmä, sillä puhetta käytetään yleensä ihmisten väliseen vuorovaikutukseen. Tutkimuksessa perehdytään puhekäyttöliittymien suunnitteluun ja arviointiin huomioiden järjestelmien inhimillistäminen. Tässä työssä tarkastellaan SERVQUAL-asteikkoa ja sovelletaan sitä puhepohjaisten keskustelujärjestelmien arviointiin. Asteikko on alkuaan palveluiden markkinoinnin tutkimukseen kehitetty subjektiivisia odotuksia ja havain-toja kartoittava menetelmä. Se perustuu asiakkailta pyydettyihin arvioihin. Työssä kokeillaan sovellettua arviointimenetelmää tapaukseen, jossa verrataan puhelinpohjaisen bussiaikataulujärjestelmän puhe- ja kirjakielistä versiota. Tulokset osoittavat, että versioiden ero ei ole tärkeä, kun järjestelmä ei muuten toimi riittävän hyvin. Voidaan päätellä, ettei puhekielisyys ainakaan paranna arvioita järjestelmästä. Metodien soveltamisessa oli ongelmia, joiden poistamiseksi esitetään ryhmien tasapainottamista ja parempaa ohjeistusta.

Avainsanat ja -sanonnat: Puhekäyttöliittymät, puhepohjaiset keskustelujärjestelmät, antropomorfismi, inhimillisyys, keskustelullisuus, puhekieli, SERVQUAL, palvelun laatu

University of Tampere

Department of Computer and Information Sciences

Mikko Hartikainen: Spoken Dialogue System Humanizing and Evaluation

M. Sc. Thesis, 67 pages, 6 appendix pages

June 2003

Abstract in English

This work studies spoken dialogue systems as anthropomorphist systems. Natural language and speech is more human-like than for example text-based system, since speech is generally used in human-human interaction. Speech user interface design and evaluation is reviewed considering humanizing. This work introduces and applies SERVQUAL scale to speech dialogue system evaluation. The scale is developed in services marketing. It is used to collect subjective expectations and perceptions from clients of the service. This work experiments the applied method on a case that compares spoken and literary language versions of a phone based bus time table system. Results show that the versions difference is not essential in a case where system functionality is not efficient otherwise. It's concluded that spoken language does not improve evaluations of the system. There were some problems applying the method. Group balancing and better advising are suggested for improvement.

Keywords: Speech User Interface, SUI, Spoken Dialogue System, anthropomorphism, conversationality, spoken language, SERVQUAL, service quality

Kiitos

Kiitos ennen kaikkea vaimolleni Marialle ja pojallemme Antonille ajasta ja tuesta. Tämän työn edistymistä auttoivat Timo Partala, Veikko Surakka ja Scott MacKenzie opetustyön puitteissa. Koko SPI-ryhmä auttoi monin eri tavoin, työpisteestä kommentteihin. Ohjaajani Markku Turusen kannustus motivoi työhön. Kiitos kokeen rakentamisessa ja kokeessa olleille vapaaehtoisille. Laitoksen taloudellinen tuki tuli tarpeeseen. Kiitos kaikille.

Sisällys

1.	Johdanto.....	1
2.	Puhekäyttöliittymät	4
2.1.	Luonnollinen puhuttu kieli ihmisen ja tietokoneen vuorovaikutustekniikkana ...	4
2.2.	Puhekäyttöliittymien toteuttaminen.....	6
2.2.1.	Dialogikontrolli	7
2.2.2.	Keskustelumalli.....	8
2.2.3.	Puhetulosteiden suunnittelu	9
2.3.	Esimerkkejä puhepohjaisista keskustelujärjestelmistä.....	10
2.3.1.	Puhepohjaisten keskustelujärjestelmien kaupalliset sovellukset.....	10
2.3.2.	Puhepohjaisten keskustelujärjestelmien kehitys tutkimusprojekteissa	12
2.4.	Tulevaisuuden näkymiä	12
	Luku 2 lyhyesti	15
3.	Tietokoneen inhimillistäminen	16
3.1.	Inhimillistämisen hyödyt ja haitat.....	16
3.2.	Inhimillistämiseen vaikuttaminen.....	18
3.2.1.	Modaliteetit.....	19
3.2.2.	Minä-muoto	19
3.2.3.	Puhekieli.....	20
	Luku 3 lyhyesti	22
4.	Puhekäyttöliittymien arviointimenetelmiä	23
4.1.	PARADISE.....	24
4.2.	Paperiprotot, Wizard of Oz ja System in a Loop.....	24
4.3.	Käytettävyyystestaus.....	25
4.4.	Heuristinen arviointi.....	26
4.5.	Käyttäjäkeskeinen koe.....	28
	Luku 4 lyhyesti	30
5.	Palvelu	31
5.1.	Palvelun laatu.....	31
5.2.	Palvelun laadun mittaaminen.....	32
	Luku 5 lyhyesti	33
6.	SERVQUAL-asteikko.....	34
6.1.	Laadun ulottuvuudet	34
6.2.	Suosio, kritiikki ja siihen vastaaminen	36
6.3.	Aineiston analysointi.....	39
6.3.1.	MSA ja MSS.....	39
6.3.2.	Laadun tarkastelu tärkeys-laatu -taulukossa.....	39

6.4. Puhepohjaisten keskustelujärjestelmien arviointi SERVQUAL-asteikolla	42
Luku 6 lyhyesti	43
7. SERVQUAL-tutkimuksen puhepohjaisiin keskustelujärjestelmiin soveltaminen	44
7.1. Kuvausta arvioiva SERVQUAL-metodi.....	45
7.2. Koekäytettävää järjestelmää arvioiva SERVQUAL-metodi	47
7.3. Vapaassa käytössä ollutta järjestelmää arvioiva SERVQUAL-metodi	48
Luku 7 lyhyesti	50
8. Interact-bussiaikataulujärjestelmän versioita vertaileva koe	51
8.1. Järjestelmän kuvaus.....	51
8.2. Versioiden erot.....	52
8.3. Koehenkilöt.....	54
8.4. Koejärjestely.....	54
Luku 8 lyhyesti	56
9. Tulokset ja keskustelu.....	57
9.1. Versioiden arviointi ja vertailu	57
9.1.1. MSA- ja MSS-mittarit.....	57
9.1.2. Vertailu tärkeys-laatu -asteikoilla.....	59
9.1.3. Vapaamuotoiset kommentit.....	61
9.1.4. Yleinen tyytyväisyys	62
9.2. Metodien kritisointi.....	63
Luku 9 lyhyesti	65
10. Yhteenveto.....	66
Viiteluettelo	68
Liitteet 5 kpl 6 sivua	
Liite 1: Odotuslomake.....	75
Liite 2: Havaintolomake	77
Liite 3: Esimerkkidialogit	78
Liite 4: Taustatietolomake	79
Liite 5: Jälkikyselylomake.....	80

1. Johdanto

Tämä työ käsittelee puhepohjaisia keskustelujärjestelmiä eli tietokonesovelluksia joille puhutaan ja jotka vastaavat takaisin puheella. Historiallisesti puhepohjaiset keskustelujärjestelmät ovat tutumpia sci-fi -viihteestä kuin käytännöstä. Tutkimuksen lähtökohta onkin usein ollut uutuudenviehätys – kun puhuva tietokone on innostava jo pelkän puheen takia, sitä suurempi innostus herää mitä ihmismäisempään vuorovaikutukseen se kykenee.

Puhepohjaisia keskustelujärjestelmiä ja niitä mahdollistavia tekniikoita kehitetään koko ajan. Nykyään onkin monia valmiita puhepohjaisia keskustelujärjestelmiä, jotka ovat joko kaupallisessa tai tutkimuskäytössä. Esimerkiksi Fonectan puheohjatussa numeronhakupalvelussa puhelinnumeroita voidaan tiedustella palveluautomaatilta sanomalla haettavan henkilön nimi ja paikkakunta automaatin niitä kysyessä [Fonecta 2003b].

Tietokoneen inhimillistäminen on yleisempää puhepohjaisissa keskustelujärjestelmissä kuin tekstipohjaisissa tai graafisissa käyttöliittymissä, sillä olemme tottuneet pitämään keskustelua yksinomaan ihmisten välisenä vuorovaikutuskeinona. Näin ollen tietokone, joka kykenee keskusteluun, koetaan inhimillisemmäksi kuin mitä se onkaan. Inhimillistämisestä voi olla hyötyä, mikäli järjestelmä koetaan ystävälliseksi ja luontevaksi, mutta myös haittaa mikäli käyttäjälle muodostuu liian korkeat odotukset järjestelmän kyvyistä.

Puhepohjaisten keskustelujärjestelmien yleistyminen on nostanut esiin uusia kysymyksiä: Miten tietokoneen kanssa puhutaan? Ja miten tietokoneen pitäisi puhua takaisin? Näitä kysymyksiä on tutkittu varsin vähän. Erilaisia tyylejä, kuten tokaisujen käyttämistä [Kotelly 2003], on kokeiltu, mutta tyylien arviointi on perustunut vaillinaisille tutkimusmenetelmille.

Tyylierot tarkoittavat puhepohjaisten keskustelujärjestelmien yhteydessä järjestelmän keskustelutyilien eroja. Keskustelutyöli tarkoittaa käytännössä lähinnä sananvalintaa ja puheen sävyä. Nykyään yleisesti käytetyt puhepohjaisten keskustelujärjestelmien arviointimenetelmät eivät ole sopivia tyylierojen vertailuun. Tyylierot eivät lähtökohtaisesti vaikuta järjestelmän suoriutumiseen sillä keskustelutyöli ei vaikuta prosessoinnin kestoon tai varmuuteen. Siksi yleensä tutkittu objektiivinen suoriutuminen ei ole oikea tapa vertailla järjestelmien tyylieroja. Toki tyylierot vaikuttavat käyttäjän ja siten myös järjestelmän suoritukseen, mutta suorituksen mittaaminen olisi epäsuoraa tutkimusta. Tyylierot vaikuttavat suoraan vain siihen, miten käyttäjä kokee järjestelmän. Siksi tyylierojen vertailussa mitataan käyttäjien subjektiivisia kokemuksia. Nykyään yleiset menetelmät, jotka mittaavat käyttäjien subjektiivisia kokemuksia ovat liian raskaita tähän tarkoitukseen tai ne eivät mittaa kokemuksia tarvittavassa laajuudessa.

Tässä työssä on kolme keskeistä tutkimusongelmaa:

- Sopiiko puhe- vai kirjakielisyyys paremmin puhepohjaisille keskustelujärjestelmille?
- Miten palvelun laatua mittaava subjektiivinen SERVQUAL-arviointimenetelmä voidaan soveltaa puhekielisten keskustelujärjestelmien arviointiin?
- Miten SERVQUAL-asteikkoa tulisi edelleen kehittää sekä sitä kohtaan yleensä esitetyn kritiikin pohjalta että puhepohjaisiin keskustelujärjestelmiin soveltamisen kannalta?

Näiden ongelmien lisäksi tässä työssä käsitellään monia muita kysymyksiä, jotka ovat oleellisia aihepiirin määrittämiseksi: Millaista puhepohjaisten keskustelujärjestelmien toteuttaminen on? Miten yleisiä puhepohjaiset keskustelujärjestelmät ovat? Mitä inhimillistä tarkoittaa puhepohjaisten keskustelujärjestelmien suunnittelun kannalta? Miten inhimillistämiseen voidaan vaikuttaa? Miten puhepohjaisia keskustelujärjestelmiä yleensä arvioidaan? Mikä on SERVQUAL-asteikko? Mitkä ovat sen vahvuudet ja heikkoudet? Miksi se sopisi puhepohjaisten keskustelujärjestelmien arviointiin? Kaikkiin näihin kysymyksiin vastataan siinä laajuudessa, kuin on keskeisten ongelmien kannalta tarpeellista.

Keskeisiin ongelmiin vastataan tiettyjen oletusten pohjalta. Tämän työn keskeinen ajatus on, että viestinnässä käytettävä kieli noudattaa viestintätilanteesta riippuvia sääntöjä. Tarkoitus ei ole kirjata sääntöjä – se olisi liian työlästä puhepohjaisten keskustelujärjestelmien kehittämisen kannalta, jollei jopa mahdotonta.

Itse kukin meistä kielenkäyttäjänä osaa soveltaa kielen sääntöjä vaistonvaraisesti – poikkeuksen muodostaa kirjakieli, jolle on olemassa säännöt joiden soveltamisessa useimmilla on vaikeuksia. Kuitenkaan vaistonvaraisessa sääntöjen soveltamisessa ei pidä tukeutua yhden ihmisen mielipiteeseen tilanteeseen sopivasta viestintätyylistä. Yksi tapa määritellä sopiva viestintätyyli puhepohjaisille keskustelujärjestelmälle on vertailla järjestelmiä, jotka käyttävät eri viestintätyylejä.

Puhepohjaisissa keskustelujärjestelmissä vuorovaikutus tapahtuu keskustelemalla, eikä keskustelu noudata kirjakielen sääntöjä [Hakulinen 1989]. Järjestelmiä ei siis kannata suunnitella sen varaan, että vuorovaikutus noudattaa kirjakielen sääntöjä. Toisaalta kaikki yleisimmät tietokonesovellukset välttävät puhekielen sanoja ja sanontoja. Tässä työssä selvitetään, mitä eroa käyttäjän vaikutelmien kannalta on sillä, puhuuko järjestelmä kirjakieltä vai puhekieltä.

Tutkimus keskittyy suomenkielisten puhepohjaisten keskustelujärjestelmien viestintätyyleihin. Kielen merkitys on keskeinen viestinnässä, joten lähinnä englanninkielisille järjestelmille tehtyjä tutkimustuloksia ei voi suoraan yleistää

suomenkielisille järjestelmille – joitain ratkaisuja olisi vaikea edes kääntää. Koska suomenkielisistä puhepohjaisista keskustelujärjestelmistä ei ole tehty viestintätyylitutkimusta, työhön sisältyy empiirinen osuus. Empiirinen osuus on vertailu puhe- ja kirjakielen käytöstä samalla järjestelmällä.

Työ jakaantuu kymmeneen lukuun. Tämä johdanto on luku yksi.

Puhepohjaisia keskustelujärjestelmiä käsitellään käyttöliittymätasolla luvussa kaksi. Luku sisältää yleisen kuvauksen siitä, mitä puhepohjaisten keskustelujärjestelmien suunnittelu käyttöliittymätasolla tarkoittaa. Samassa luvussa kerrotaan, minkälaisia puhepohjaisia keskustelujärjestelmiä on toteutettu ja pohditaan järjestelmien kehityssuuntia.

Kolmannessa luvussa tarkastellaan miten järjestelmien inhimillistäminen vaikuttaa puhepohjaisten keskustelujärjestelmien suunnitteluun. Järjestelmälle sopivan viestintätyylin valinta liittyy keskeisesti siihen, koetaanko tietokone sosiaalisesti toimijaksi. On olemassa keinoja, joilla käyttäjien kokemuksiin järjestelmän inhimillisyydestä voidaan vaikuttaa. Luvussa esitellään kolme tapaa, joilla inhimillistämistä voidaan tukea tai heikentää. Puhe modaaliteettina vahvistaa inhimillistämistä, joten puhepohjaiset keskustelujärjestelmät ovat lähtökohdiltaan inhimillistämistä tukevia.

Luvussa neljä esitellään puhepohjaisten keskustelujärjestelmien arviointimenetelmiä. Samalla motivoidaan luvussa viisi esitettävän uuden arviointimenetelmän soveltaminen puhepohjaisille keskustelujärjestelmille.

Luvussa viisi määritellään palvelut ja palvelun laatu ja perustellaan puhepohjaisten keskustelujärjestelmien tarkastelu palveluina. Luvussa käsitellään myös palveluiden laadun tutkimisen periaatteita.

Luvussa kuusi esitellään uusi tapa arvioida puhepohjaisia keskustelujärjestelmiä, SERVQUAL-asteikko. Järjestelmiä arvioidaan palveluna käyttäjien subjektiivisen näkemyksen perusteella. Arviointimenetelmä on sovellettu palveluiden markkinoinnin tutkimusalalta. Luvun loppuun pohditaan, mitä lisäarvoa SERVQUAL-asteikko tuo puhepohjaisten keskustelujärjestelmien arviointiin.

Luvussa seitsemän esitellään laatimani SERVQUAL-asteikkoa soveltavat arviointimetodit. Metodit soveltuvat puhepohjaisten keskustelujärjestelmien arviointiin palveluina. Metodeita on kolme, joista sopiva valitaan järjestelmän valmiuden perusteella.

Empiirinen osuus on luvuissa kahdeksan ja yhdeksän. Luvussa kahdeksan esitellään automaattinen bussiaikataulupalvelu, jonka puhe- ja kirjakielisiä versioita vertailtiin soveltamalla luvussa seitsemän esiteltyä metodologiaa. Luvussa yhdeksän kerrotaan tulokset – toisaalta arvioinnin itsensä tulokset, mutta samalla kerrotaan havainnoista, joita soveltamistyössä syntyi.

Luku kymmenen on yhteenveto koko työstä. Siinä pohditaan myös, miten tätä työtä voidaan käyttää hyödyksi tulevissa tutkimuksissa.

2. Puhekäyttöliittymät

Puhepohjainen keskustelujärjestelmä (englanniksi spoken dialogue system) tarkoittaa tietokonesovellusta, joka sallii vuorovaikutuksen luonnollisella puhutulla kielellä [McTear 2002]. Luonnollinen puhuttu kieli tuo järjestelmien kehittämiseen haasteita, joihin ei voida vastata pelkän järjestelmän mahdollistavan teknologian kehittämällä. Käyttöliittymäsuunnittelu on tärkeä osa puhepohjaisten keskustelujärjestelmien kehittämistä.

Puhepohjaisissa keskustelujärjestelmissä on puhekäyttöliittymä. Toki keskustelujärjestelmä voi olla multimodaalinen, eli vuorovaikutus voi tapahtua monta eri vuorovaikutustekniikkaa käyttäen. Tässä työssä käsitellään kuitenkin keskustelujärjestelmiä, joilla on joko pelkästään puhekäyttöliittymä tai puhe on ainakin avainasemassa järjestelmässä.

Tässä luvussa käsitellään puhepohjaisia keskustelujärjestelmiä. Ensin keskitytään luonnolliseen kieleen käyttöliittymänä. Seuraavaksi selvitetään millaista puhepohjaisten keskustelujärjestelmien käyttöliittymäsuunnittelu on. Sitten esitellään olemassa olevia järjestelmiä muutamien esimerkkien avulla. Luvun lopuksi tarkastellaan tulevaisuuden näkymiä.

2.1. Luonnollinen puhuttu kieli ihmisen ja tietokoneen vuorovaikutustekniikkana

Luonnollinen puhuttu kieli on käsite, joka menee helposti sekaisin puhekielen kanssa. Esimerkiksi Hintikka [2003] kertoo luonnollisen kielen järjestelmästä käyttäen siitä nimitystä ”puhekielinen järjestelmä”. Puhekieli tarkoittaa kieltä, jota ihmiset käyttävät vapaamuotoisessa puhevuorovaikutuksessa toisten ihmisten kanssa. ”Puhekieli”-ilmaisua käytetään usein vastakohtana ”kirjakielelle”, sillä ihmiset soveltavat usein erilaisia normeja kieleen viestivälimestä (puhe vs. kirjoitus vs. joku muu) riippuen. Puhe- ja kirjakielen ero on hyvin selvä ainakin nykysuomessa. Puheen ja kirjoituksen ero ei ole kuitenkaan ainoa kielenkäyttöä jakava kriteeri. Toinen kriteeri on suunnitellun ja suunnittelemattoman kielenkäytön ero. Kolmas, olennainen ero on dialogisen ja monologisen kielenkäytön välillä [Hakulinen 1989]. Lisäksi kieleen sovelletaan erilaisia normeja riippuen siitä, kenen kanssa viestitään – ainakin koneen kanssa viestiminen on erilaista kuin ihmisen kanssa viestiminen [Doreen ym. 2001].

Luonnollinen kieli on muutenkin hankala käsite: yleensä luonnollinen kieli tarkoittaa ihmisten välisessä kommunikoinnissa käytettävää kieltä, kuten siis esimerkiksi suomen kieli. Kirjallisuudessa luonnollisuus kuitenkin toisinaan sekoittuu sen arkikielen merkitykseen eli luontevuuteen, ja luonnollinen kieli tarkoittaakin luontevaa kieltä (eli luontevaa luonnollista kieltä). Tässä työssä tarkoitan käsitteellä luonteva kieli kieltä, joka soveltaa viestivälisestä ja viestikumppanista riippuvia normeja, eli on luonnollisen kielen tilanteeseen sovellettu versio. Luonteva puhuttu kieli soveltaa siis viestikumppanista ja

suunnittelun tasosta riippuvia normeja tilanteessa, jossa viestivälineenä on puhe. Luontevan kielen vastakohtana voidaan pitää sellaisen luonnollisen kielen käyttöä, joka syystä tai toisesta ei sovi tilanteeseen eli ei vastaa viestintäkumppaneiden odotuksia viestintätäydyistä.

Historiallisesti tarkasteltuna puhepohjaisten keskustelujärjestelmien kehitys on alkanut 1950-luvulla, jolloin tekoälytutkimus kiinnostui keskustelevista käyttöliittymistä. Kuitenkin vasta 1990-luvulla teknologinen kehitys on mahdollistanut laajamuotoisen järjestelmien tekemisen ja markkinoinnin. [McTear 2002]

Vaikka tekoälytutkimus olikin luonnollisesti kiinnostunut luontevasti puhuvista tietokoneista, luontevuus ei kuitenkaan ole ainoa syy puhekäyttöliittymän valitsemiseksi. Rosenfeld ym. [2001] esittävät kritiikkiä sellaisia tutkimuksia kohtaan, jotka perustelevat puhepohjaisten keskustelujärjestelmien käytön pelkästään luontevuudella – heidän mielestään tärkeämpiä perusteluja ovat seuraavat kolme hyötynäkökohtaa:

- Puhe on ympäröivä viestiväline, ei keskittyvä: puhetta voi käyttää vuorovaikutukseen ja tehdä samalla jotain muuta.
- Puhe on kuvailevaa, ei osoittavaa: puhuessamme kuvailemme puheen kohteen ominaisuuksia. Puhuminen ja osoittaminen (esimerkiksi motorisesti sormella) ovat toisiaan täydentäviä modalityetteja.
- Puheelle tarpeelliset fyysiset resurssit ovat vaatimattomia: puhevuorovaikutus voidaan toteuttaa pienemmällä ja halvemmilla fyysisillä komponenteilla kuin visuaaliset ja käsiohjatut käyttöliittymäteknikat.

Tämä kritiikki on aiheellista, mutta mielestäni Rosenfeldin ym. [2001] esittämät perustelut puheen käyttämiseksi eivät ole vastakkaisia luontevuudelle. He esittävät perustelut puheen käyttämiseksi hyötynäkökulmasta, mutta mielestäni samat perustelut paljastavat, miksi puheen käyttäminen on luontevaa joissain tilanteissa. Siinä mielessä heidän vastakkainasettelunsa on turha. Esimerkiksi kun haluaa käsillään tehdä jotain muuta ja viestiä samanaikaisesti, puhe ei ole pelkästään käytännöllinen vaan myös luonteva valinta, koska puhe on ympäröivä viestiväline.

Toisaalta puhekäyttöliittymien yksi kohderyhmä ovat näkö- ja motorisesti vammaiset, joiden kohdalla puheen käyttäminen tietokoneen kanssa käytävässä vuorovaikutuksessa voi olla ainoa todellinen vaihtoehto. Esimerkiksi näkövammaisille on olemassa pistenäyttöjä, joissa näyttöruudulle tuleva teksti esitetään pistekirjoituksena, mutta kaikki näkövammaiset eivät käytä pistekirjoitusta [Näkövammaisten keskusliitto 2003].

Laajemmin ajateltuna luontevin vaihtoehto on se, mihin on totuttu. Tottuminen tapahtuu käyttämällä, joten luontevin vaihtoehto on eniten käytetty. Usein eniten käytetty vaihtoehto on alun perin valittu sen takia, että se on hyödyllinen. Siis luontevuus ei ole

missään nimessä hyödyllisyyden vastakohta, vaan päinvastoin usein käytännössä hyödyllisyyden seuraus.

Mielestäni puhekäyttöliittymien hyödyllisyysajattelua tärkeämpi havainto Rosenfeldin ym. [2001] näkökulmissa on puhepohjaisten keskustelujärjestelmien luokittelu älykkäisiin koneisiin ja yksinkertaisiin koneisiin. Kun lähtökohtana on puhekäyttöliittymä luontevana valintana, saattaa jäädä huomaamatta, että yksinkertainen kone voisikin hoitaa tehtävän paremmin kuin älykäs kone. Esimerkiksi mikroaaltouunin säätäminen on niin suoraviivaista, ettei puhesyötekään voi olla monimutkaista. Yksinkertaisen puhesyötteen käsittelemiseksi puhepohjaisen keskustelujärjestelmän pitää olla myös yksinkertainen, ettei tilanne johtaisi väärinkäsityksiin.

2.2. Puhekäyttöliittymien toteuttaminen

Puhepohjaisten keskustelujärjestelmien toteuttamiseen tarvitaan puhekäyttöliittymä. Teknisesti puhekäyttöliittymän toteuttamiseen tarvitaan puheen koodaus, puheen tunnistus ja tekstisynteesi [Kamm ym. 2003]. Puhetta voidaan tuottaa myös ääninauhotteita käyttäen. Puheen tunnistuksen apuna käytetään hienostuneissa järjestelmissä luonnollisen kielen ymmärtämis Komponenttia. Luonnollisen kielen ymmärtäminen voi toimia puheen tunnistuksesta itsenäisenä (eli puheen tunnistus tuottaa tekstimuotoisen esityksen jota luonnollisen kielen ymmärrys analysoi) tai yhdessä puheen tunnistuksen kanssa (esimerkiksi *n-best* -listojen avulla eli puheen tunnistus tuottaa tietyn määrän todennäköisimpiä tunnistustuloksia joista luonnollisen kielen ymmärrys valitsee sopivimman) [McTear 2002].

Kuten Salonen [2002] toteaa, teknologia asettaa kaikessa sovelluskehityksessä mahdollisten toteutusten rajat. Teknologian kehityksestä huolimatta tämä on edelleen erityisen huomion kohteena puhejärjestelmien suunnittelussa. Puheentunnistus on tärkein mahdollisuuksia rajoittava teknologia. Puheentunnistus on virhealtista. Puhetta tunnistetaan ennalta määritellyn kieliopin perusteella, eli puheentunnistus valitsee mahdollisista vaihtoehdoista lähinnä puhunnosta olevan. Tunnistusvarmuutta voidaan parantaa pienentämällä kielioppia. Esimerkiksi järjestelmän varmistuksissa voi olla kahden sanan kielioppi, kuten ”kyllä” ja ”ei”. Jos tähän lisää kolmanneksi vaihtoehdoksi sanan ”ehkä”, on tunnistus jo epävarmempaa, varsinkin kun ”ei” ja ”ehkä” alkavat samalla äänneellä.

Puhesynteesi ei täysin vastaa ihmisääntä, mutta on varteenotettava vaihtoehto ihmisäänen nauhoittamiselle. Tässä työssä ei käsitellä puhepohjaisia keskustelujärjestelmiä kokonaisvaltaisesti rakentamisen kannalta, vaan keskitytään käyttöliittymäsuunnitteluun ja siihen vaikuttaviin tekijöihin. Esimerkiksi Salosen [2002] työ on ajantasainen suomenkielinen johdatus toimivien puhekäyttöliittymien rakentamiseen.

Vaikka teknologia asettaakin joitain rajoituksia, hyvä käyttöliittymäsuunnittelu pelastaa paljon. Yksi tärkeimpiä suunnittelun painopisteitä on dialogikontrolli. Toinen

keskeinen suunnittelualue on keskustelumalli. Kolmas suunnittelun kohde on puhetulosteet eli mitä järjestelmä käytännössä sanoo.

2.2.1. Dialogikontrolli

Dialogikontrolli tarkoittaa lähinnä sitä, miten järjestelmä edistää keskustelun kulkua. Dialogikontrollimalli määrää pitkälle myös sen, miten järjestelmä tukee virheiden korjausta.

McTear [2002] luokittelee puhepohjaiset keskustelujärjestelmät kolmeen erilaiseen dialogikontrollimalliin.

- Äärelliset tila- (graafi-) perusteiset järjestelmät (eng. finite state- (graph-) based).
- Kehysperusteiset järjestelmät (eng. frame-based).
- Agenttiperusteiset järjestelmät.

Äärelliset järjestelmät perustuvat ennalta määrättyihin keskusteluvaiheisiin [McTear 2002]. Esimerkiksi aikatauluja kysyttäessä vaiheet voisivat olla lähtöpaikan selvitys, tulopaikan selvitys, halutun tiedon selvitys (esim. lähtö- vai tuloaika) ja halutun tiedon kertominen. Järjestelmä edistää keskustelua siirtymällä aina seuraavaan vaiheeseen kun edellinen on ohitettu onnistuneesti (selvityksessä saatu hyväksyttävä tunnistustulos, varmistukseen saatu myöntävä vastaus). Vastaavasti mikäli vaihetta ei voida ohittaa, siinä joko pysytään (esimerkiksi toistamalla kysymys) tai mennään taaksepäin johonkin aiempaan vaiheeseen.

Kehysperusteiset järjestelmät perustuvat tietokehykseen, jota järjestelmä täyttää keskusteluvaiheesta riippumatta [McTear 2002]. Jos edellinen esimerkki olisi kehysperusteinen, tietokehyksen alkioita olisivat lähtöpaikka, tulopaikka, halutun tiedon nimi ja haluttu tieto. Tällaista järjestelmää voi käyttää äärellisen järjestelmän tavoin, mutta halutessa käyttäjä voi kertoa vaikka heti aluksi lähtöpaikan, tulopaikan ja halutun tiedon nimen. Tällöin järjestelmä sijoittaa annetut tiedot tietokehykseen. Järjestelmä valitsee toiminnan tietokehyksessä olevien tietojen perusteella. Tässä tapauksessa tietokehyksessä olisi lähtöpaikka, tulopaikka ja halutun tiedon nimi täytettyinä tietoalkioina, jolloin järjestelmä hakee halutun tiedon kyseisten tietoalkioiden sisällön perusteella tietokannasta.

Agenttiperusteiset järjestelmät eli agenttijärjestelmät suunnitellaan monimuotoisen viestinnän sallimiseksi. Järjestelmä pyrkii tehtävän suorittamisen lisäksi auttamaan ongelman selvittämisessä tai tehtävän suorittamisessa. [McTear 2002] Voidaankin ymmärtää, että agenttijärjestelmä ei pyri ainoastaan vastaamaan kysymyksiin tai suorittamaan tehtäviä käyttäjän antaman syötteen perusteella, vaan toimimaan yhteistyössä käyttäjän kanssa. Edellisissä esimerkeissä kuvattu järjestelmä voisi olla agenttijärjestelmä, jos se osaisi vaikkapa ehdottaa jatkoyhteyttä mikäli suoraa linjaa ei ole.

Vielä agenttimaisempaa olisi, jos järjestelmä ehdottaisi jatkoyhteyttä suoran linjan lisäksi mikäli arvelisi tämän soveltuvan jostain muusta syystä käyttäjälle – vaikka toki agenttimaisuutta on moni muukin ominaisuus kuin jatkoyhteyden ehdottaminen.

Agentin toiminnallisuus riippuu järjestelmän tarjoamasta viitekehystä, mutta yleensä kaikki agenttijärjestelmät sisältävät ongelmien tunnistus ja –korjaustoimintoja. Lisäksi ne osaavat yleensä käyttää odotuksia (esimerkiksi aiempiin vuorovaikutuskertoihin perustuen) käyttäjän tulevien ilmaisujen ennustamisessa ja tulkinnassa. [McTear 2002]

Salonen [2002] mainitsee myös neljä muuta dialogikontrollimallia: tehtäväpohjainen dialogikontrolli, agendapohjainen dialogikontrolli, HMIHY:n dialogikontrolli ja komentopohjainen dialogikontrolli. Tehtävä- ja agendapohjaiset dialogikontrollimallit ovat dynaamisia eli dialogin tai kyselyn rakennetta ei ole rajoitettu tiettyihin vaihtoehtoihin – mielestäni nämä sijoittuvat McTearin jaossa ainakin osittain agenttiperusteisiin järjestelmiin, sillä dynaamisten rakenteiden toteuttaminen ei ole triviaalia ja tekoälymenetelmien käyttäminen on perusteltua. HMIHY (”How May I Help You”) on AT&T:n automaattinen puhelinkeskus, joka osaa vaihtaa keskustelun aihetta ja palata edellisiin aiheisiin. Tällainen järjestelmä voi olla periaatteessa monen tavoitteen äärellinen järjestelmä, jossa on vain laajennettu edellisten keskustelukierrosten hyväksikäyttö. Komentopohjainen dialogikontrolli on samoin monen tavoitteen äärellinen järjestelmä, jossa aloite on selkeästi käyttäjällä. Vaikka nämäkin dialogikontrollimallit on mahdollista luokitella McTearin [2002] kolmijaon mukaisesti, on varsinkin komentopohjaisen dialogikontrollin erikseen jaottelu mielestäni perusteltua sillä sille pohjautuva käyttöliittymä poikkeaa muiden äärellisten järjestelmien käyttöliittymistä huomattavasti.

2.2.2. Keskustelumalli

Dialogikontrollimallin lisäksi keskustelumallin suunnittelu on avainasemassa puhekäyttöliittymien suunnittelussa. Keskustelumalli määrittelee, kenellä on aloite keskustelussa.

Järjestelmän aloite on usein suora kysymys. Esimerkiksi ”Sano lähtöpaikka” on suora kysymys. Järjestelmän aloitteen etu on, että tunnistettavien sanojen lukumäärä eli kieliopin laajuus voidaan pitää pienenä. Pieni kielioppi parantaa tunnistusvarmuutta. Toisaalta pienet kieliopit hankaloittavat väärin tunnistettujen sanojen korjaamista, sillä ongelmien korjauskin tapahtuu järjestelmän aloitteen sallimissa rajoissa [McTear 2002]. Järjestelmän aloite saattaa usein johtaa pitkiin vuorovaikutuksiin, joissa järjestelmä kerää tietoa usean kierroksen ajan, vaikka käyttäjä voisi ilmoittaa tarvittavan tiedon yhdellä vuorovaikutuskierroksella.

Käyttäjän aloite voi olla komentoihin perustuvaa tai luonnollista kieltä. Luonnollisen kielen järjestelmän ei kuitenkaan tarvitse tunnistaa kaikkia kielen sanoja, vaikka aloite

olisi käyttäjällä. Järjestelmä tietää kuitenkin kontekstin ja siten voi rajoittaa puheentunnistuksen kielioppia selvästi tilanteeseen kuulumattomien asioiden osalta. Lisäksi järjestelmä voi valita, mitä tietoja antaa käyttäjälle ja näin johdatella implisiittisesti käyttäjää johonkin suuntaan. Esimerkiksi tervehdytystä ”Tervetuloa. Mitä haluat tehdä?” ei johdattele käyttäjää kuten ”Tervetuloa. Sinulla on kaksi uutta viestiä. Mitä haluat tehdä?”. Tämä liittyy suunnittelussa myös puhetulosteiden suunnitteluun, josta lisää seuraavassa osiossa.

Käyttäjän aloitteen etuna on vuorovaikutuksen lyheneminen; ainakin siihen pyritään, vaikkei se aina toteudukaan. Vuorovaikutus lyhenee, kun käyttäjä voi heti alkutervehdyksen jälkeen ilmoittaa asiansa. Käyttäjän aloitteen toteuttava järjestelmä on hankalampi toteuttaa kuin järjestelmän aloitteen toteuttava järjestelmä, sillä kieliopit ovat väistämättä suurempia ja puheentunnistus on epävarmempaa.

Aloitetta voi myös vaihdella järjestelmän ja käyttäjän välillä. Esimerkiksi oletusarvoisesti käyttäjän aloitteen järjestelmä voi ottaa aloitteen virhetilanteessa. Aloitteen suunnitteleminen on tapauskohtaista eikä voida sanoa, että jokin tapa olisi aina paras vaihtoehto.

2.2.3. Puhetulosteiden suunnittelu

Vaikka puhetulosteiden eli järjestelmän puheenvuorojen suunnittelu saattaa vaikuttaa pelkältä hienosäädöltä, puhetulosteet ovat kuitenkin käyttäjän kannalta keskeinen asia. Varsinkin puhelinpohjaisessa tai vastaavassa järjestelmässä, jossa puhe on ainoa modaliteetti, ovat puhetulosteet ainoa tulostustapa. Käyttäjä päätelee puhetulosteiden perusteella mitä järjestelmälle voi sanoa. Käyttäjä saa virheilmoitukset, vastaukset ja varmistukset puhetulosteiden kautta. Käyttäjä muodostaa mielikuvan järjestelmästä puhetulosteiden sisällön (ja taukojen keston) perusteella.

Siis jo puhetyyli, jota järjestelmä käyttää, on selkeä vihje käyttäjille siitä, miten kattava järjestelmän tunnistus ja ymmärrys on. Esimerkiksi sama asia voidaan sanoa näin ”Kaksi yhteyttä löytyi. Haluatko mennä yhdeksältä vai kymmeneltä?” tai näin ”Minä löysin tietokannasta kaksi yhteyttä. Busseja lähtee yhdeksältä ja kymmeneltä. Kummalla haluat mennä?”. Jälkimmäisestä on vaikeampi tietää, mitä järjestelmä ymmärtää: Riittääkö, että käyttäjä sanoo ”jälkimmäinen” tai ”se toinen”? Ensimmäisestä käy epäsuorasti ilmi, että järjestelmä hyväksyy vastauksen ainakin muodossa ”kymmeneltä”, vaikka puheentunnistus käyttäisikin laajempaa kielioppia.

E erityisen tärkeää puhetulosteiden suunnittelussa on se, mitä sanotaan. Esimerkiksi kun kerrotaan, että bussi keskustasta lähiöön lähtee kello 12.00, on usein riittämätöntä sanoa vain ”kaksitoista nolla nolla”. Jotta käyttäjä voisi olla varma, että järjestelmä puhuu samasta bussista kuin hän itse, täytyy järjestelmän kertoa, mistä bussista se puhuu. Tämä voidaan tehdä eksplisiittisesti esimerkiksi varmistuskysymyksellä ”kysytkö koska lähtee bussi keskustasta lähiöön? Sano kyllä tai ei” tai implisiittisesti, esimerkiksi näin ”bussi

keskustasta lähiöön lähtee kaksitoista nolla nolla”. Toisaalta uudet käyttäjät saattavat tarvita sellaista tietoa, jota kokeneet käyttäjät eivät tarvitse. Käyttäjän tunnistava järjestelmä voikin valita mitä sanoo sen perusteella, miten kokenut käyttäjä on [Yankelovich 1996].

Vaikkei järjestelmää suunniteltaisikaan kaupalliseen käyttöön, on merkitystä myös sillä, miten asiat sanotaan. Yleisenä suunnitteluperiaatteena voidaan pitää ytimekkyyttä – käyttäjä voi turhautua jos järjestelmä selittää asioita liian tarkasti tai puhuu muuten vaan liikaa. Toisaalta järjestelmän inhimillistäminen on joskus perusteltua, ja tällöin voidaan käyttää laveampaa kieltä. Inhimillistämisestä kerron lisää luvussa kolme.

2.3. Esimerkkejä puhepohjaisista keskustelujärjestelmistä

Kaupallisia puhepohjaisia keskustelujärjestelmiä on toteutettu monissa maissa. Puhepohjaisten keskustelujärjestelmien kielisidonnaisuudesta (jos järjestelmä on norjankielinen, palvelu on saatavilla norjaksi) johtuen järjestelmistä on harvoin esittelysivuja internetissä (tai esittelyjä muutenkaan kansainvälisissä julkaisuissa) muilla kuin järjestelmän kielellä. Niinpä minun oli vaikea löytää muita kuin suomen-, ruotsin- ja englanninkielisiä järjestelmiä. Larsson [2003] ja Bohus [2003] pitävät yllä englanninkielisiä linkkiloistoja puhepohjaisista keskustelujärjestelmistä. Nämä listat tarjoavat hyvän lähtökohdan löytää esimerkkejä puhepohjaisista keskustelujärjestelmistä. Listoja on kuitenkin ylläpidetty harrastusmielessä, ja niistä löytyy lähinnä samoja järjestelmiä kuin hakukoneella hakemalla, eli tutkimusjärjestelmiä ja englanninkielisiä (lue: yhdysvaltalaisia) järjestelmiä. Mielestäni olisi perusteltua, että jokin puhepohjaisista keskustelujärjestelmistä kiinnostunut kansainvälinen yhteisö (kuten ISCA [2003], AVIOS [2003]) tai muu taho ylläpitäisi avoimesti saatavilla olevaa listaa muistakin kuin tutkimuslaitosten kehittämistä järjestelmistä, jotta käytössä olevien puhepohjaisten keskustelujärjestelmien yleisyydestä eri puolilla maailmaa saisi todellisen kuvan. Tässä kappaleessa luettelen muutamia puhepohjaisia keskustelujärjestelmiä, jotka ovat kuvaavia käsitykseni mukaisesta puhepohjaisten keskustelujärjestelmien nykytilasta. Lisää esimerkkejä löytyy ainakin Larssonin [2003] ja Bohusin [2003] linkkiloistoilta sekä internet-hakukoneilta.

2.3.1. Puhepohjaisten keskustelujärjestelmien kaupalliset sovellukset

Suomessa tunnetuin kaupallinen puhepohjainen keskustelujärjestelmä on varmaankin Fonectan (entinen Sonera Info Communications) puheohjattu numerohaku. Järjestelmä on kuten mikä tahansa kilpaileva puhelinnumerotiedustelu, eli se sisältää kaikki Suomessa luetteloidut puhelinnumerot. Numeroita voi hakea nimen perusteella tai nimeä numeron perusteella. Koneen kanssa käytävä keskustelu on hyvin tilaperusteinen ja järjestelmä on vahvasti aloitteellinen. [Fonecta 2003a; Phonetic 2003a] Palvelu perustuu Phonetic Systemsin Voice Search Engine -teknologialle ja se on kansainvälisesti palkittu

uraaurtavana puhepohjaisena keskustelujärjestelmänä [Phonetic 2003b; Rantanen 2003]. Järjestelmän rinnalla toimii ihmisoperaattorien hoitama numerotiedustelu, johon kyselyt ohjataan heti, mikäli järjestelmä ei tunnista tai löydä tietoa [Fonecta 2003b].

Hollannissa on toiminnassa kattava julkisen liikenteen puheohjattu neuvonta. Se perustuu kolmessa tutkimusprojektissa tehdyille työlle: MAIS – Multilingual Automatic Inquiry Systems, ARISE - Automatic Railway Information Systems for Europe ja The Dutch National R&D programme “Language and Speech Technology”. [OVIS 2003] Palvelua ylläpitää OV Reisinformatie B.V. (OVR), joka tarjoaa liikenneneuvontaa myös henkilökohtaisena palveluna, internet-hakuna ja kännyköihin [OVR 2003].

Myös suomen junaliikenteen aikatauluista on puheohjattu puhelinpalvelu [VR Henkilöliikenne 2003]. Kokeilin VR Aikataulut –automaattia. VR Aikataulut –automaatissa järjestelmällä on aloite ja kieliopit vaikuttavat hyvin rajoitetuilta. Järjestelmä pitää kuitenkin huolen, että käyttäjällä on oikea käsitys dialogin tilasta varmistuksilla ja luettelemalla sallittuja sanoja. Vaikuttikin siltä, että järjestelmän rakentamisessa on kiinnitetty huomiota toiminnan varmuuteen hieman liikaakin nopeuden kustannuksella. Esimerkiksi haettava aikataulu (lähtöpaikka, tulopaikka, päivä, kuukausi, kellonaika mistä eteenpäin) varmistetaan vaikka kyseessä on vain aikataulun listaaminen eikä kriittisempi toiminto kuten lipun varaaminen. Tämä ei välttämättä ole suunnittelun virhe, vaan kuvaa VR:n prioriteetteja palveluissa.

Yhdysvalloissa on useita toiminnassa olevia puhepohjaisia keskustelujärjestelmiä. Kotelly [2003] luettelee kuusi erilaista liikealaa, joilla on käytössään puhepohjaisia keskustelujärjestelmiä. Suluissa esimerkkejä puhepohjaisten keskustelujärjestelmien käyttämisestä näillä aloilla.

- Liikennöinti (Kotellyn esimerkit tosin vain lentoliikenteestä: henkilökunnan matkavaraukset, asiakkaille lentotiedotus, kadonneen matkatavaran seuranta).
- Logistiikkayritykset (FedEx ja UPS, lähetysten seuranta)
- Rahoituslaitokset (esimerkiksi pankit, tiedonanto, tilisiirrot)
- Yritykset, joiden ydintoimintaa on tiedon välitys (sähköpostin kuunteleminen puhelimella, säätiedotukset tietokoneen lukemana)
- Ääniportaali eli puhelinpalvelu, joka tarjoaa esimerkiksi sää-, osake-, ajo-ohje- ja elokuvatietoja sekä
- Puhelinyhtiöt (puhelureititys ja tiedonanto ääniportaalien tapaan)

Lisäksi Kotellyn työnantajan Speechworksin sivuilla [Speechworks 2003] mainitaan seuraavat liikealat, joille yhtiö markkinoi puhepohjaisten keskustelujärjestelmien osaamistaan:

- Autot (auton kunnan tiedottaminen, ajo-ohjeet, asioiden tekeminen matkalla, esim. sähköpostin lukeminen)
- Julkishallinto (tuki-info, sakkojen maksaminen)
- Terveystieteidenhuolto (reseptin uusinta, perusterveyden tietojen tarkkailu)
- Vähittäiskauppa (tilaukset)
- Turismi (liikennetiedot, hotellin varaus)

Näistä kaikista ei tosin ole mainittu esimerkkejä, mutta kaikki mainitut esimerkkipuuhjelmät ovat käyttötasolle toteutettavissa. Speechworks ei toki ole ainoa yritys, joka markkinoi puhepohjaisia keskustelujärjestelmäteknologioita. Esimerkiksi myös Nuancen verkkosivuilla [Nuance 2003] kerrotaan heidän osaamisestaan samoilla toimialoilla, joille puhepohjaiset keskustelujärjestelmät sopisivat.

2.3.2. Puhepohjaisten keskustelujärjestelmien kehitys tutkimusprojekteissa

Vaikka varsinaisia kaupallisia puhepohjaisia keskustelujärjestelmiä on melko vähän, toimivia keskustelujärjestelmiä on kuitenkin kehitetty myös tutkimusprojekteissa. Huomionarvoinen tutkimusprojekti Yhdysvalloissa on DARPA (Defense Advanced Research Projects Agency) Communicator [DARPA Communicator 2003]. Projektissa kehitystyöryhmät puhepohjaisia keskustelujärjestelmiä toteuttavista ja tutkivista asiantuntijayksiköistä ovat saaneet koulutuksen puhepohjaisen keskustelujärjestelmän käytännön toteutuksesta. Kaikkien ryhmien toteutukset tehtävänannon mukaisesta puhepohjaisesta keskustelujärjestelmästä perustuvat Galaxy-arkkitehtuurille [Galaxy 2003; Seneff ym. 1998]. Projektin tarkoituksena onkin, että itse järjestelmän toteuttaminen olisi mahdollisimman helppoa ja kehitystyöryhmät saisivat itsenäisesti ratkoa muita keskustelullisuuden liittyviä ongelmia, kuten aloitteen siirtäminen, joustava tiedon antaminen, selvennyksen pyytäminen, ymmärtämättömyyden ilmaiseminen ja niin edelleen. [Aberdeen ym. 1999]

Huomattavaa on myös, ettei tutkimustyö ole keskittynyt pelkkiin yksittäisiin sovelluksiin, vaan esimerkiksi Tampereen yliopistossa kehitetty Jaspis [Turunen & Hakulinen 2000a] on avoin arkkitehtuuri mukautuville puhekäyttöliittymille. Jaspiksen arkkitehtuurille perustuu mm. puheohjattu sähköpostijärjestelmä Postimies [Turunen & Hakulinen 2000b], puheohjattu läsnäoleva ovenavaus- ja opastusjärjestelmä Ovimies [Mäkelä ym. 2001] ja myöhemmin tässä työssä tarkemmin esiteltävä puhelinpohjainen puheohjattu bussiaikatauluautomaatti Interact [Jokinen ym. 2002].

2.4. Tulevaisuuden näkymiä

Kun puhepohjaiset keskustelujärjestelmät jaetaan älykkäisiin ja yksinkertaisiin koneisiin [Rosenfeld ym. 2001], myös kehityksen lähtökohdat ovat kahden suuntaiset. Älykkään

puhepohjaisen keskustelujärjestelmän toteuttaminen vaatii mittavaa tekoälytutkimusta ja järjestelmä on teknisesti haasteellinen etenkin puheentunnistuksen osalta. Toisaalta yksinkertaisten koneiden ongelmat ovat lähinnä tottumuskysymyksissä: ensivaikutelma on tärkeä tekijä puhepohjaisten keskustelujärjestelmien hyväksymisessä, ja ensivaikutelma on huono mikäli jokaista järjestelmää varten joutuu omaksumaan uuden toimintamallin. Toisaalta yksinkertaiset puhepohjaiset keskustelujärjestelmät saattavat olla älykkäitä järjestelmiä hyödyllisempiä ja jopa luontevampia, kun konetta on perinteisesti totuttu pitämään työvälteenä.

Toimintamallien yhdistämiseksi puhepohjaisille keskustelujärjestelmille tarvitaan yleispäteviä käyttöliittymäperiaatteita graafisten käyttöliittymien tapaan [Rosenfeld ym. 2001]. Shriver & Rosenfeld [2002] ovatkin aloittaneet kehitystyön käyttöliittymäperiaatteiden luomiseksi. He ovat käyttäjien mielipiteitä kysymällä selvittäneet, mitä yleispäteviä avainsanoja pidetään erilaisiin vuorovaikutustilanteisiin sopivina. Shriverin & Rosenfeldin tutkimuksessa [2002] selvisi esimerkiksi, että käyttäjien mielestä tilanteeseen, jossa vuorovaikutus ei etene ja käyttäjä haluaisi aloittaa alusta, sopii parhaiten yleispäteväksi avainsanaksi (tai -sanonnaksi) ”start over” eli suomeksi ”aloita alusta”. Vaikka kehitystyö on vasta alussa eivätkä englanninkieliset avainsanat sovi suoraan suomeen, on hyvä saavutus jo se, jos puhepohjaisten keskustelujärjestelmien kehittäjät pääsevät yhteisymmärrykseen siitä, minkälaisiin tilanteisiin avainsanojen käyttäminen soveltuu. Tällöin käyttäjä osaa aiempien kokemusten perusteella (vaikka kokemukset olisivat eri järjestelmistä) käyttää avainsanoja sopivissa tilanteissa ja mahdollisesti käyttää järjestelmän tukemaa avainsanaa, vaikka se poikkeaisikin muiden järjestelmien avainsanoista.

Cox ym. [2000] näkevät tulevaisuudessa viisi kehityssuuntaa puhepohjaisille keskustelujärjestelmille.

- Agentit, jotka hallitsevat viestintää.
- Automaattiset asiakaspalvelijat IVR-järjestelmien tilalle.
- Puhelukeskusten automatisointi yksinkertaisten varaus- tilaus- ja lomakkeen täyttämistoimintojen hoitamiseksi.
- Tietokoneen ja puhelimen yhteensovittaminen eli viestinnän täydentäminen tietokoneen kortistosta, käyttöön soveltaminen (eli oppiva puhelin) ja internetin puheohjattu selailu.
- Puhesanelun yhdistäminen sähköpostiin vastaamiseksi tekstinä.

Pidemmälle menevissä tulevaisuuden pohdinnoissaan Cox ym. [2000] visualisoivat puhelinpalveluita multimedialliseksi ja multimodaaliseksi uudistavan älykkään agentin, jota he kutsuvat nimellä VoiceTone. Sen toimintoja olisivat esimerkiksi seuraavat:

- Puheohjattu numerovalinta, esimerkiksi sanomalla ”Yhdistä Jane Doelle Memphisiin” eri luetteloita (keltaiset sivut, kotinumerot, omat kaverit jne.) käyttäen.
- Mediakonversio eli esimerkiksi tekstiviestien lukeminen puheena ja puheen esittäminen tekstinä.
- Suoran yhteyden tiedonhaku eri tietokannoista eli haun yhdistäminen sopivaan palveluun; esimerkiksi pyyntö ”Millainen sää on Bostonissa” yhdistäisi käyttäjän säätietokantapalveluun.
- Muistiinpanojen tekeminen keskustelun aikana.
- Viestien tulkkaus (kielestä kieleen) ylimpänä päämääränä keskustelu suorassa yhteydessä eri kielisten välillä.
- Kuulo-, puhe- ja näkövammaisten avustaminen viestintätarpeissa.
- Viihteen vuorovaikutuksellinen ohjaus esimerkiksi sanomalla ”Ehdotapa jotain hyvää James Dean –leffaa” tai ”Näytä se maali eri kulmasta”.

Jotkut näistä palveluista ovat hyvinkin toteutuskelpoisia teknologian valmiuden puolesta, kuten mediakonversio. Cox ym. [2000] edustavat kuitenkin sitä koulukuntaa, joka näkee tulevaisuuden puhepohjaiset keskustelujärjestelmät lähinnä älykkäinä ihmisen kaltaisina (tai taitavampina) järjestelminä sen sijaan, että muodostettaisiin käyttöliittymäperiaatteita yksinkertaisille järjestelmille.

Osiossa 2.3.1 mainittu Hollannin julkisen liikenteen puheohjattu puhelinneuvonta on projektina mielestäni yksi todennäköinen kehityssuunta puhepohjaisille keskustelujärjestelmille. Koska keskustelujärjestelmien kehitystyö on ollut mittavaa tieteellisissä tutkimusprojekteissa, näiden projektien tuloksia halutaan hyödyntää. Julkisen liikenteen neuvontapalvelut ovat yksi tällainen sovellusalue, jossa on tehty laajamittaista kehitystyötä tutkimusprojekteissa.

Hyvä esimerkki tutkimusprojektissa tehdyn työn soveltamisesta käytössä olevaksi järjestelmäksi on Tampereella kehitteillä oleva Jaspis-arkkitehtuurille [Turunen & Hakulinen 2000a] perustuva puheohjattu bussiaikataulujärjestelmä. Järjestelmä kysyy vuorovaikutuksen aluksi pysäkin nelinnumeroista tunnusta, joka on näkyvillä pysäkillä. Tämän perusteella luetellaan pysäkiltä seuraavaksi lähtevät linjat lähtöaikoineen (luettelon pituus vaihtelee sen perusteella, montako linjaa pysäkin kautta kulkee). Tämän jälkeen käyttäjä voi hakea pysäkkiaikatauluja joko käyttöhetkestä eteenpäin tai jostain annetusta ajankohdasta eteenpäin. Myös ajankohtaa edeltäneet vuorot voidaan selvittää. Järjestelmä voi myös listata kaikki linjat, jotka ajavat kyseisen pysäkin kautta. Aikataulutietoja voidaan listata myös pelkästään tietyn linjan perusteella; tämä on käytännöllistä etenkin jos pysäkin kautta ajaa monia linjoja. Järjestelmä on tätä

kirjoitettaessa (kesäkuu 2003) testikäytössä ja se aiotaan julkaista yleiseen käyttöön loppuvuodesta 2003. [Helin 2003]

Luku 2 lyhyesti

- Puhepohjaisissa keskustelujärjestelmissä ihminen ja tietokone puhuvat keskenään luonnollista kieltä, kuten suomea, käyttäen.
- Puheen käyttäminen ihmisen ja tietokoneen vuorovaikutuksessa on perusteltua silloin, kun siitä on jotain hyötyä. Esimerkiksi jos kädet ja silmät on varattu johonkin muuhun tehtävään, voi puhua keskeyttämättä muita tehtäviä. Myös näkö- ja motorisesti vammaiset voivat käyttää tietokonetta puheen avulla. Hyöty ei kuitenkaan ole ristiriidassa luontevuuden kanssa, vaan usein onkin luontevaa käyttää puhetta juuri silloin kun siitä on eniten hyötyä.
- Puhepohjaisen keskustelujärjestelmän toteuttamiseksi tarvitaan puheentunnistus, puheen tuottaminen ja kielen käsittely. Puheentunnistuksen heikkoudet tuottavat ongelmia, joita paikataan hyvällä puhekäyttöliittymäsuunnittelulla. Käyttöliittymäsuunnittelu tapahtuu kolmella tasolla: dialogikontrolli, keskustelumalli ja puhetulosteiden suunnittelu.
- Toimivia puhepohjaisia keskustelujärjestelmiä on kehitetty tutkimusprojekteissa mutta myös kaupallisiin tarkoituksiin.
- Puhepohjaiset keskustelujärjestelmät tulevat yleistymään edelleen. Haasteita ovat yksinkertaisten ja älykkäiden järjestelmien erottaminen, yleispätevien käyttöliittymäperiaatteiden muodostaminen ja tutkimustyön ja kaupallisten sovellusten yhteensovittaminen.

3. Tietokoneen inhimillistäminen

Tietokoneen inhimillistäminen eli antropomorfismi tarkoittaa sitä, että käyttäjä kokee tietokoneella olevan joitain inhimillisiä piirteitä, joita sillä ei ole (kuten tunteet). Käyttäjä saattaa rationaalisesti ymmärtää, ettei tietokoneella ole inhimillisiä piirteitä, mutta toimia kuitenkin vastaisen kokemuksen perusteella [Reeves & Nass 1996, 3-15]. Inhimillistäminen on tuttua lemmikkieläinten piiristä: esimerkiksi puhuvan papukaijan kanssa keskustelea niin kuin se ymmärtäisi mitä se sanoo, vaikei voikaan olla varma, ymmärtääkö se oikeasti. Riittää, että vaikuttaa siltä, että se ymmärtää. [Hefton 2003].

Käyttöliittymiä suunniteltaessa on tärkeää huomioida inhimillistämisen vaikutukset. Reeves & Nass [1996] ovat tehneet tutkimuksia monista ihmisten välisistä käyttäytymiskaavoista, jotka ulottuvat myös ihmisen ja tietokoneen (ja television yms. uuden median) väliseen vuorovaikutukseen. Tietokoneesta voidaan tietoisesti suunnitella inhimillisen oloinen ja hyväksikäyttää ihmisten käyttäytymiskaavoja. Toisaalta kun tiedetään käyttäytymiskaavat, voidaan välttää inhimillistämistä esimerkiksi tietoisesti suunnitteleamalla tietokone konemaiseksi.

Tässä luvussa käsittelen tietokoneen inhimillistämistä. Ensin käsittelen inhimillistämisen hyötyjä ja haittoja. Tämä käsittely ei olisi perusteltua ilman katsausta siihen, miksi inhimillistämistä oikeastaan tapahtuu. Luettelen myös muutaman tavan vaikuttaa inhimillistämiseen. Viimeisenä tapana mainitsen puhekielisyyden, jota kokeiltiin työn empiirisessä osiossa. Koetta ja tuloksia käsitellään luvuissa kahdeksan ja yhdeksän.

3.1. Inhimillistämisen hyödyt ja haitat

Väitän tietokoneen inhimillistämisessä olevan kyse odotusten toteutumisen vastavuoroisuusperiaatteesta. Tietokoneen käyttämiseen vaadittiin aikanaan kymmenien tai satojen kytkinten vääntämistä, jolloin oli helppo käsittää, miten tietokone toimii, mihin se pystyy; mikä tietokone on. Vaikei koneella olisi edes tekoälyä, toimii vastavuoroisuusperiaate silloin, kun kone *vaikuttaa* inhimilliseltä. Jos tietokone tekee jotain, mikä voidaan ymmärtää inhimilliseksi toiminnaksi, eikä vuorovaikutustilanne paljasta, miten tietokone sen tekee, syntyy inhimillisyyden vaikutelma. Kuten Reeves & Nass [1996, 12] toteavat, ihminen on pitkässä evoluutiossa tottunut, että asiat ovat niin kuin aistii. Käytännössä tällä vastavuoroisuusperiaatteella on merkitystä lähinnä kun tietokone toimii niin, että käyttäjä luulee sen pystyvän sellaisiin asioihin, joihin se ei pysty. Esimerkiksi Gorin ym. [1996] huomasivat, että avoin kysymys ("How may I help you?", suomeksi "Kuinka voin auttaa") johtaa laajaan vaihteluun siinä, mitä käyttäjät sanovat.

Mahdollisuus keskustella tietokoneen kanssa on kiinnostanut tutkijoita ennen kuin tietokoneita on ollutkaan – nähtävästi suurin motivaatio monellekin tutkimukselle on ollut aito innostus mahdollisimman luontevan kielen puhepohjaisen keskustelujärjestelmän kehittämiseksi. Esimerkiksi Kotelly [2003] tukee avoimesti järjestelmän inhimillistämistä,

perustellen sitä laajalla kokemuksella käyttäjistä, jotka pitävät järjestelmää *sosiaalisena toimijana*. Ehkä puhepohjaisten keskustelujärjestelmien kohdalla kyse on siitä, että käyttäjät inhimillistävät järjestelmää joka tapauksessa – sehän puhuu! – joten on suotavaa, että se puhuu inhimillisesti. Kuten Boyce [2000] sanoo: ”Jos tulee mahdolliseksi rakentaa tietokoneita, jotka vastaavat luontevaan luonnolliseen kieleen ja palautuvat virheistä sujuvasti, on todennäköistä että käyttäjien käsitys tietokoneesta sosiaalisena toimijana muuttuu.”

Olen samaa mieltä Boycen [2000] ja Kotellyn [2003] kanssa siitä, että inhimillistäminen sopii usein puhepohjaisille keskustelujärjestelmille, mutta ei suinkaan aina. Kotelly vertaakin puhepohjaisia keskustelujärjestelmiä IVR (Interactive Voice Response, vuorovaikutteinen äänivastaus, yleensä syöte annetaan puhelimen näppäimiä painamalla) –järjestelmiin, joille hän ei suosittele inhimillistämistä. Mielestäni näiden kahden väliin mahtuu monia järjestelmiä, jotka käyttävät puheentunnistusta, mutta joita käyttäjä ei pidä sosiaalisina toimijoina. Varsinkin jos järjestelmä on vahvasti aloitteellinen, käyttäjälle on selvää, että hän on vuorovaikutuksessa koneen kanssa ja tarkoitus on pelkästään välittää tietoa ripeästi eikä keskustella – inhimilliset piirteet tuskin ainakaan lyhentävät vuorovaikutuksen kestoa. Toisaalta esimerkiksi myös puheohjattu sähköpostijärjestelmä (kuten Postimies [Turunen & Hakulinen 2000b]) on sellainen järjestelmä, jolle inhimillistäminen ei välttämättä sovi, vaikka se onkin puhepohjainen keskustelujärjestelmä. Järjestelmässä ovat pääosassa sähköpostiviestit, joita luetaan samalla synteessillä kuin muukin vuorovaikutus, jolloin inhimillistäminen voisi johtaa sellaiseen hämmentävään vaikutelmaan, että joku toinen sosiaalinen toimija lukee viestejä. Käyttäjä voi vain arvailla, osaako tietokone käsitellä myös sähköpostien sisältöä – sehän voisi vaikka levittää salaisuuksia tai käydä ivalliseksi.

Mielestäni jako älykkäisiin ja yksinkertaisiin järjestelmiin antaa hyvän vihjeen siitä, kannattaako inhimillistämistä tukea. Esimerkiksi matkatoimistojärjestelmä on yksinkertainen järjestelmä, jos sillä voi suorittaa vain matkan varauksen, mutta jos järjestelmä osaa esimerkiksi suositella jotain kohdetta (esimerkiksi tarjousten ja asiakkaan mieltymysten perusteella), on järjestelmä älykäs. [Rosenfeld ym. 2001]

Yksinkertaisille järjestelmille sopii komentokäyttöliittymät ja IVR-tyylinen valintavuorovaikutus, mutta myös luonnollinen kieli yksinkertaisessa muodossaan. Käyttäjät yksinkertaistavatkin kyselyn sanamuotoa oma-aloitteisesti, kun kone osoittaa ymmärtämättömyyttä. Näin kävi ainakin käytettävyytsteissä, jotka teimme Interact-projektin puitteissa (lisää Interactista luvussa kahdeksan). Tällöin on turhaa suunnitella järjestelmään inhimillisiä piirteitä. Parhaassakin tapauksessa ne vain nostavat käyttäjien odotuksia järjestelmän kyvyistä. Pahimmassa tapauksessa käyttäjät eivät tiedä ollenkaan, kuinka järjestelmälle voi puhua.

Jos järjestelmä on älykäs (eli älykkään oloinen), voi inhimillistämisestä olla hyötyä. Kotelly [2003] väittää, että inhimilliset ominaisuudet koetaan ystävällisiksi, kohteliaiksi ja

luonteviksi. Uskon, että näin onkin älykkäiden järjestelmien kohdalla. Vastauksia lyhyessä muodossa antava järjestelmä voi vaikuttaa töksäyttelevältä, mikäli se muuten toimii ihmisen kaltaisesti. Kun joitain asioita on tottunut hoitamaan pelkästään ihmisten kanssa, on luontevaa käyttää ihmisten välistä kommunikaatiota muistuttavaa kieltä.

Ongelmallisinta tässä jaottelussa onkin jaottelu itse. Milloin järjestelmä on oikeastaan älykäs? Rosenfeld ym. [2001] luettelevat kuusi esimerkkiä käyttötarkoituksesta, johon sopii yksinkertainen järjestelmä:

- Kodinkoneiden säätäminen ja käyttö.
- Toimistokoneiden säätäminen ja käyttö.
- Julkisen tiedon hakeminen.
- Yksityisen tiedon hakeminen ja muokkaaminen.
- Erilaisten kommunikointivälineiden hallinta (esimerkiksi sähköposti ja faksi).
- Sekalaisten kuluttajasovellusten hallinta (esimerkiksi kartan seuraaminen, lomakkeiden täyttäminen ja Internet-sivujen selailu).

Mielestäni tällainen listaaminen ei ole täsmällistä, eivätkä Rosenfeld ym. [2001] niin väitäkään. Mielestäni jaottelu tulee tehdä sillä perusteella, kokeeko käyttäjä järjestelmän älykkääksi vai yksinkertaiseksi. Aiemmin mainittu matkatoimistojärjestelmä voi ollakin käyttäjän mielestä älykäs järjestelmä, vaikka sillä pystyisi pelkästään varaamaan matkoja, esimerkiksi jos käyttäjä luulee, että vain ihminen kykenee varaamaan matkoja keskustelun perusteella. Toki tällöinkin matkojen varaaminen on toimintona yksinkertainen, eikä inhimillistämistä välttämättä kannata tukea. Toisaalta monet kodinkoneet ja toimistokoneet sijaitsevat fyysisesti niin lähellä toisiaan, että yksinkertaiset konekohtaiset käyttöliittymät eivät riitä erottamaan, mitä konetta käyttäjä kutsuu. Tällaiseen tilanteeseen ratkaisuksi sopisi multimodaalisuuden lisäksi kaksivaiheinen käyttöliittymä, jossa käyttäjä valitsisi laitteen älykkään (tai yksinkertaisen) agentin kanssa ja käyttäisi sitten kyseisen laitteen yksinkertaisia toimintoja.

Inhimillistämisen tukemisen kannattavuus riippuu myös kulttuurista. Monet esimerkit puhepohjaisista keskustelujärjestelmistä ovat Yhdysvalloista, jossa on varsin erilainen palvelukulttuuri kuin Suomessa. Esimerkiksi matkoja varaava järjestelmä kuulostaisi suomalaisista varmasti oudolta, jos se puhuisi amerikkalaisen maireasti, kun suomalaiset ovat tottuneet, että ihmisetkin puhuvat samassa tilanteessa kuivakkaammin.

3.2. Inhimillistämiseen vaikuttaminen

Esittelen tässä kappaleessa kolme tapaa vaikuttaa siihen, inhimillistääkö käyttäjä tietokonetta. Ensimmäinen tapa on käyttää ihmisten välisestä vuorovaikutuksesta tuttuja modaaliteetteja. Puhe on yksi tällainen modaaliteetti, ja siihen liittyen seuraavat tavat

kuvaavat sitä, miten tietokone puhuu. Käyttääkö se yksikön ensimmäistä persoonaa? Puhuuko tietokone puhekieltä? Tapoja olisi lisääkin, esimerkiksi ystävällisyys (kohteliaisuus), tunteiden ilmaiseminen ja tyylin vaihtelu käyttäjän ja tilanteen mukaan. Näistä tavoista on kuitenkin vähän kirjallisuutta eivätkä ne liity suoraan tähän työhön.

3.2.1. Modaliteetit

Perinteisesti tietokoneen kanssa ollaan vuorovaikutuksessa eri modaliteeteilla kuin ihmisten kanssa. Käyttäjä viestii näppäilemällä ja hiirtä liikuttamalla, tietokone viestii rajatulle kaksiulotteiselle ruudulle ilmestyvillä kuvilla ja merkkiäänillä.

Toisinaan ”uusiksi” vuorovaikutustekniikoiksi kutsutut modaliteetit ovatkin tuttuja ihmisten välisestä vuorovaikutuksesta. Jos tietokone viestiikin liikkumalla, siihen yhdistetään käyttäytymispiirteitä, jotka ovat tuttuja eläinten (siis myös ihmisten) kanssa toimimisesta. Esimerkiksi aibo-robotikoirat ovat usein omistajilleen yhtä inhimillisiä kuin lemmikkikoirat muutenkin [Kahney 2001]. Puhe on myös tuttu ihmisten välisestä vuorovaikutuksesta, joten sen käyttäminen ihmisen ja tietokoneen vuorovaikutuksessa edesauttaa koneen inhimillistämistä.

3.2.2. Minä-muoto

Kun tietokone käyttää kieltä (kirjoitettua, puhuttua, viitottua) viestimiseen, suunnittelijan pitää valita, missä persoonamuodossa tietokone puhuu. Esimerkiksi Shneiderman [1987] suosittelee, etteivät tietokoneen viestit sisältäisi pronomineja, siis mieluummin ”aloittaaksesi luennon, paina return” kuin ”Minä aloitan luennon kun sinä painat returnia”. Onkin selvää, että mikäli ohjelmassa ei ole selvää agenttia, joka koetaan itsenäiseksi toimijaksi, on harhaanjohtavaa puhua ensimmäisessä persoonassa – ketä ”minä” tarkoittaa? Toisaalta agentti voi olla ärsyttävä, kuten kaikkietävä Clippy [Microsoft 2003]. Kolmanneksi käyttäjä saattaa olettaa tietokoneen osaavan enemmän kuin se osaa.

AT&T:n puhepohjaisen puhelureititinkeskustelujärjestelmän kehittäjät huomasivat kuitenkin, että perinteinen inhimillistämiskielteisyys ei ole yleistettävissä kaikenlaisiin järjestelmiin. Vertailevassa tutkimuksessa selvisi, että käyttäjät suosivat järjestelmää, joka puhui yksikön ensimmäisessä persoonassa (”How may I help you?” suom. ”Kuinka voin auttaa?”) verrattuna toiseen persoonaan (”What would you like?” suom. ”Mitä haluaisit?”). Toisessa tutkimuksessa havaittiin, että inhimillistäminen on hyväksyttävää ja jopa hyödyllistä, kun se mallinnetaan oikeiden ihmisten mukaisesti. [Boyce 2000] Kun järjestelmän tavoitteena oli joka tapauksessa sallia luonnollisen kielen kyselyt, eivät käyttäjien vapaamuotoiset puhunnokset olleet ongelma.

3.2.3. Puhekieli

Puhekielen ja kirjakielen ero on hyvin selvä suomessa. On harvinaista, että ilmoitamme jonkin asian täysin samassa muodossa puhuttuna ja kirjoitettuna. Vertaapa näitä versioita päivittäisen asian ilmaisemisesta:

- Tiedätkö sinä, missä minun avaimeni ovat? (a)
- Tiedäksä missä mun avaimet on? (b)
- Tiiäkkö sie missä miun avvaimet o? (c)

Vaikka esimerkki a on hyvää suomea, harvat puhuvat näin (kokeile sanoa ääneen). Esimerkki a on kirjakieltä, jota käytetään lähinnä kirjoittamiseen – tosin joissain muodollisissa tilanteissa puhe on ainakin osittain lähempänä kirjakieltä kuin vapaammissa tilanteissa [Vierikko 1972]. Esimerkit b ja c ovat puhekieltä – suomessa puhekieltä esiintyy monina murteina, joissa on vivahde-eroja, ja esimerkkejä voisi helposti listata lisää. Esimerkki b on mielestäni lähempänä yleiskieltä. Televisiolla on suomen kielessä erityisen vahva asema puhutun yleiskielen muodostumiselle, ja yleiskieli onkin lähellä televisiossa puhuttavaa helsinkä.

Saukkosen [1972] kokeellinen vertaileva tutkimus osoitti, että puhekieli on 20 % väljempää (sisältää enemmän sanoja saman asian sanomiseksi) kuin kirjakieli – yllä olevat esimerkit eivät olekaan oikein edustavia todellisesta puhekielestä, sillä ne eivät sisällä täytesanoja, tauotuksia, lauseenvastikkeita, takeltelua tms. puhekielen ominaisuuksia. Näissä esimerkeissä näkyy pelkästään se ero, mitä on puhe- ja kirjakielen sanastoissa.

Puhekieli on todellista käytännön suomea, vaikka puhekielen ja kirjakielen vertailussa onkin helppo osoittaa puhekielen heikkoudet kirjakieleen verrattuna [Hakulinen 1989, 11]. Puhekielen ja kirjakielen ero kapenee tietyissä tilanteissa, joissa on yleistä käyttää hyväksyttävää kirjakielen muotoa. Tällaisia tilanteita ovat ainakin lyhyet ilmaisut ("Anna avaimet") ja kohteliaisuus (esimerkiksi teitittely: "Oletteko Te huomanneet täällä avaimia?"). Lisäksi puhekielen lyhyet ilmaisut käyttävät usein kirjakieleessä hyväksytyjä sanoja, vaikkei ilmaisu muodostaisikaan täydellistä lausetta. Riippuu myös puhujasta, kuinka lähellä tai kaukana toisistaan puhe- ja kirjakieli ovat [Saukkonen 1972].

Puhe- ja kirjakielen ero on intuitiivisesti erilainen englannissa (ja muissa kielissä) kuin suomessa. Englannin kieltä voi puhua kirjakielen kielioppisääntöjä noudattaen vaikuttamatta hassulta – hieman teennäiseltä ehkä. Toki englannissa on murteita kuten kaikissa kielissä, mutta puhepohjaiset keskustelujärjestelmät ovat usein avoimia järjestelmiä, joita käyttävät eri murreryhmien edustajat. Siksi puhepohjaisen keskustelujärjestelmän käyttämä kieli on yleensä yleiskieltä.

Suomenkielisiä puhepohjaisia keskustelujärjestelmiä suunniteltaessa yksi tapa tukea inhimillistämistä on suunnitella järjestelmä puhekieliseksi. Puhekielen käyttö onkin yleistä neuvontapalvelun ihmisille, kuten näkyy Räsänen [2001] tekemistä transkripteistä. Suomenkielissä järjestelmissä ei ole kuitenkaan tiedettävästi kokeiltu puhekieltä.

Englanninkielisissä järjestelmissä käytetään usein lyhyitä tokaisuja, jotka eivät kuulu kirjakieleen. Kotelly [2003] esittää, että lyhyet tokaisut ovat suotavia puhepohjaisille keskustelujärjestelmille. ”OK” tai ”Got it” ilmaisee, että järjestelmä ymmärtää käyttäjää. ”Oh!” keskellä pitkää luetteloä tuo käyttäjän keskittymisen takaisin tehtävään, mikäli hän keskittyy johonkin muuhun. Näiden ilmaisujen käyttö on perusteltua kahdesta syystä: tieto annetaan lyhyessä muodossa ja järjestelmän inhimillisyys koetaan kohteliaaksi, ystävälliseksi ja luontevaksi. Tämä on esimerkki englannin puhekielen onnistuneesta käytöstä puhepohjaisissa keskustelujärjestelmissä, joka ei sinällään silti tarkoita, että kaikkien puhepohjaisten keskustelujärjestelmien tulisi käyttää puhekieltä.

Luku 3 lyhyesti

- Käyttäjät inhimillistävät tietokoneita, vaikka tietäisivät että kyseessä on kone. Inhimillistämisen kannalta on tärkeämpää, vaikuttaako tietokone inhimilliseltä.
- Suunnittelulla voidaan vaikuttaa inhimillistämiseen joko sitä tukien tai heikentäen. Suunnittelusta riippumatta tilanteessa, missä järjestelmä ei ymmärrä käyttäjää, tietokone ei vaikuta inhimilliseltä.
- Inhimillistämisen tukeminen on kannattavaa silloin, kun käyttäjät osaavat odottaa inhimillisyyttä. Toisaalta mikäli järjestelmän kyvyt eivät riitä inhimilliseen vuorovaikutukseen, inhimillistäminen on haitallista kun käyttäjät saavat epärealistisen käsityksen järjestelmän kyvyistä.
- Puhepohjainen keskustelujärjestelmä tukee väistämättä inhimillistämistä, sillä perinteisesti keskustelu on tuttua ihmisten välisestä vuorovaikutuksesta. Lisäksi inhimillistämisen tukemiseksi voidaan käyttää esimerkiksi puhekieltä ja ensimmäistä persoonapronominia järjestelmän puheenvuoroissa.

4. Puhekäyttöliittymien arviointimenetelmiä

On suotavaa, että puhepohjainen keskustelujärjestelmä toteuttaa puhutun kielen mahdollisimman luontevana eli tilanteeseen sopivana. Luonnollisen kielen puhepohjaisen keskustelujärjestelmän tulee siis sekä ymmärtää että puhua luontevaa kieltä. Luontevuus ei kuitenkaan tarkoita sitä, että järjestelmän tulisi käyttää puhekieltä [Doran ym. 2001]. Puhepohjaisen keskustelujärjestelmän tapauksessa luonteva kieli soveltaa kielen normeja tilanteessa, jossa viestivälineenä on puhe ja ihmisen (tai ihmisten) viestikumppanina on kone (tai koneita).

Edellisestä päätellen puhepohjaisen keskustelujärjestelmän tulisi siis automaattisesti tietää, millaista kieltä sen täytyy soveltaa – tilannehan on aina sama: ihminen, kone ja puhe. Asia ei ole aivan näin yksinkertainen, sillä kielen normit ovat periaatteessa henkilökohtaisia. Ihmiset kuuluvat kuitenkin moneen eri sidosryhmään – pienempiin ja laajempiin. Toisin sanoen vaikka teinijengiläiset saattavatkin keskenään puhua varsin erityylistä kieltä kuin mummojengi, tarpeen tullen löytyy yhteinen kieli. Siis luontevan kielen puhepohjaisen keskustelujärjestelmän toteuttamiseksi riittää, että järjestelmä toteuttaa yleiskielen normit kyseisessä tilanteessa. Toki on parempi, jos kohderyhmän voi ottaa suunnittelussa huomioon [Kotelly 2003].

Mistä sitten puhepohjaisen keskustelujärjestelmän kehittäjä tietää, mitä normeja yleiskielellä on tilanteessa, jossa ihminen ja kone keskustelevat puheen välityksellä? Ensinnäkin kehittäjä on itsekin kielenkäyttäjä ja tietää vaistomaisesti, miten koneelle pitäisi puhua ja miten odottaa koneen puhuvan takaisin. Kehittäjän käsitykset yleiskielen normeista saattavat kuitenkin olla vääristyneet, varsinkin kun puhepohjaiset keskustelujärjestelmät ovat vielä niin harvinaisia, että niiden kehittäjä on epäluonnollisen kokenut käyttäjä. Siispä jos pyritään tätä parempaan luontevuuteen, täytyy selvittää muiden kielenkäyttäjien mielipide asiasta. Tarvitaan siis subjektiivisia arvioita puhekäyttöliittymästä.

Puhekäyttöliittymien arviointiin on kehitetty monenlaisia menetelmiä. Arvioinnilla onkin erilaisia päämääriä. Arviointimenetelmiin tutustuessaan huomaa, että ne pyrkivät usein itse järjestelmän kannalta tarpeettomankin hienojakoiseen objektiivisuuteen. Esimerkiksi PARADISE [Walker ym. 1997] on tällainen menetelmä (lisää PARADISEsta kappaleessa 4.1.). Tämä johtuu luultavasti siitä, että puhepohjaiset keskustelujärjestelmät edustavat esimerkiksi graafisiin käyttöliittymiin verrattuna uutta, keskeneräistä teknologiaa. Näin onkin perusteltua tähdätä arvioinnilla tehokkuuden parantamiseen, sillä puhepohjaiset keskustelujärjestelmät eivät voi nykyteknologialla olla kaupallisesti kannattavia kiinnittämättä huomiota tehokkuuteen. Toisaalta juuri kaupallisille järjestelmille olisi tärkeää huomioida käyttäjien subjektiiviset arviot järjestelmästä.

Teknologiakeskeinen suorituksen arviointi kertoo aivan liian vähän käyttäjien todellisista mieltymyksistä [Van Haaren ym. 1998].

Esittelen tässä luvussa muutamia yleisimmin käytettyjä puhekäyttöliittymien arviointimenetelmiä. Tarkoitus ei ole pelkästään esitellä menetelmiä, vaan samalla perustella uuden menetelmän soveltaminen, jolla saadaan subjektiivisia arvioita puhepohjaisista keskustelujärjestelmistä toiminnan tehokkuutta laajemmasta näkökulmasta.

4.1. PARADISE

PARADISE (PARAdigm for DIalogue System Evaluation) [Walker ym. 1997] on kattava puhepohjaisten keskustelujärjestelmien arviointimenetelmä. PARADISE ei kuitenkaan ole täysin nimeänsä ("paradise" on suomeksi paratiisi) vastaava. Se vaatii testaajalta kohtuullisen dialogikorpuksen keräämisen järjestelmällä, annotoinnin ja analysoinnin – käytännössä sekä menetelmän vahvuus että heikkous on sen kattavuus. PARADISE tuottaa mittavat tulokset, jotka kattavat kaikki puhepohjaisen keskustelujärjestelmän arvioitavat osa-alueet, mutta niiden tuottaminen on työlästä ja vie siis resursseja muulta kehitystyöltä. Kritiikkiä esittää myös Hjalmarsson [2002], joka tuli PARADISEa käyttäessään tulokseen, että yleisen puhepohjaisten keskustelujärjestelmien arviointimenetelmän luomiseksi tarvitaan lisäkehitystä, sillä menetelmälle keskeinen tehtävän määrittely ja kyselylomakkeiden suunnittelu on jäänyt vähäiselle huomiolle.

4.2. Paperiprotot, Wizard of Oz ja System in a Loop

Kehitystyössä on usein tärkeää saada käyttäjäpalautetta suunnitteilla olevasta järjestelmästä ennen toimivan järjestelmän rakentamista. Puhepohjaiset keskustelujärjestelmät eivät ole poikkeuksellisen kalliita rakentaa muihin kattaviin sovelluksiin verrattuna, mutta esimerkiksi graafista käyttöliittymää voi kohtuullisella luotettavuudella tutkia niin sanotuilla paperiprotoilla (kuva tai kuvaus järjestelmästä) [MacKenzie 2003]. Paperiproto ei kuitenkaan anna käyttäjälle yhtä todenmukaista kuvaa järjestelmästä, joka käyttää vuorovaikutukseen muita tulostusmodaliteetteja kuin kaksikulotteinen kuva (kuten puhe). Tällöin prototyypin kuvaus voidaan antaa muuten kuin paperilla – puhekäyttöliittymän kuvaus on todenmukaisinta antaa ääninäytteenä.

Wizard of Oz on empiirinen menetelmä, jossa käyttäjälle uskotellaan, että tämä on vuorovaikutuksessa tietokoneen kanssa. Todellisuudessa testaaja simuloi tietokonetta eli kertoo tietokoneelle miten vastata käyttäjälle. Tutkijan ei tarvitse välttämättä edes rakentaa simulaatiojärjestelmää, vaan hän voi käyttää työkaluja, kuten SUEDE [Klemmer ym. 2000] puhekäyttöliittymän simulointiin. Wizard of Oz –menetelmä ei kuitenkaan sovellu kaikenlaisille järjestelmille. Esimerkiksi Salonen [2002] koki menetelmän ongelmalliseksi puheohjatun sähköpostin arviointiin, sillä aidon sähköpostimateriaalin

käyttäminen olisi ollut eettisesti väärin ja toisaalta muunlaiset testijärjestelyt ovat keinotekoisia, mikä näkyy tuloksissa.

System in a Loop on Wizard of Ozin uudempi, laajennettu versio. Kun järjestelmästä on toteutettu joitakin (oletettavasti yksinkertaisia) osa-alueita, ihminen simuloi vain toteuttamattomia osa-alueita. Verrattuna Wizard of Oziin, tästä hyöttyy siksi, ettei järjestelmää simuloiva ihminen vaikuta omalla toiminnallaan tuloksiin yhtä paljon. [McTear 2002]

Tutkijan vaikutus tuloksiin onkin huomionarvoinen asia, sillä missään prototyypitestauksessa ei päästä puolueettomiin tuloksiin – ei paperiprotoissa, Wizard of Ozissa eikä System in a Loopissa. Vaikka valmis järjestelmä annettaisiin käyttäjien testattavaksi, riippuu testin tulokset siitä, mitä käyttäjien pyydetään testata. Ainoa tapa saada luotettavaa tietoa, kuinka käyttäjät oikeasti käyttävät järjestelmää, on iteratiivinen kehitys, eli että järjestelmä annetaan vapaasti käytettäväksi ja kehitetään järjestelmää käyttökokemusten perusteella. Tällöin on huomioitava, että käytön olosuhteet (saatavuus, hinta jne.) vastaavat suunniteltua loppukäyttötilannetta tai ne huomioidaan muuten kehitystä suunniteltaessa.

Edellä mainitut menetelmät ovat työkaluja jonkin tiedon selvittämiseen, mutta ei ole ennalta määrättyä, mitä tietoa näiden menetelmien käyttämisellä halutaan selvittää. Voidaan selvittää esimerkiksi suoritumisen objektiivista tasoa, kuten nopeutta. Näin teki MacKenzie [2003]. Toisaalta prototyypitestauksella voidaan kerätä vapaamuotoisia havaintoja siitä, kuinka ihmiset toimivat järjestelmän kanssa. Näin tekivät Mäkelä ym. [2001]. Itse sovelsin prototyypitestausta puhepohjaisten keskustelujärjestelmien subjektiiviseen arviointiin, jota käsittelen luvuissa kahdeksan ja yhdeksän.

4.3. Käytettävyydestä

Käytettävyyden määrittäminen ei ole yksinkertaista, mutta yleensä hyväksytään Gouldin ja Lewisin [1983] esittämät neljä käytettävyySPIIRRETTÄ:

1. Oppimisen helppous.
2. Hyödyllisyys.
3. Helppokäyttöisyys.
4. Käytön miellyttävyys.

Käytettävyydestä ja käytettävyydestä on julkaistu lukuisia kirjoja, kuten Barnum [2002], Dumas & Redish [1993] ja Weiss [2002]. Käytettävyydestä on yleensä keinotekoinen tilanne, johon kutsutaan (mahdollisuuksien mukaan todellista käyttäjäryhmää edustavia) testikäyttäjiä. Testikäyttäjille annetaan jotain tehtäviä, joita heidän pitää suorittaa. Testaaja tarkkailee testikäyttäjiä joko samassa tilassa, erillisessä tarkkailuhuoneessa tai nauhoitusten perusteella.

Vaikka käytettävyyden määritelmään sisällytetään käytön miellyttävyys, käytettävyydestit keräävät yleensä objektiivista (objektiivisuuteen tähtäävää) kvantitatiivista aineistoa (esimerkiksi virheiden määrä, virheiden korjaamiseen kuluva aika) tai objektiivista (objektiivisuuteen tähtäävää) kvalitatiivista aineistoa (esimerkiksi suoriutumisen eli virheiden löytäminen ja suosituimman toimintatavan löytäminen). Weiss [2002] väittää jopa, että kaikki käytettävyydesteissä kerätty aineisto on kvalitatiivista. Joka tapauksessa, objektiivisena testinä käytettävyydesti ei anna kovinkaan luotettavaa kuvaa käyttäjän mieltymyksistä. Mieltymyksiä mittaamiseen käytettävyydesti on sopimatonta myös usein pienten koehenkilömäärien ja keinotekoisien tilanteiden vuoksi [Dicks 2002]. Käyttäjän mieltymysten selvittämiseksi tarvitaan subjektiivista onnistumista mittaavia subjektiivisia arvioita, joita toki voidaan selvittää käytettävyydestien yhteydessä, esimerkiksi antamalla kyselylomake testien päätteeksi.

4.4. Heuristinen arviointi

Heuristinen arviointi on vapaamuotoinen käytettävyyden arviointimenetelmä. Käytettävyydesteissä yleensä valitaan testikäyttäjiä, jotka eivät välttämättä (joskus jopa toivottavasti) ole käytettävyyden asiantuntijoita ja tarkkaillaan kuinka he suoriutuvat tehtävistä. Heuristiseen arviointiin kutsutaan muutama käytettävyyden asiantuntija, jotka arvioivat sovellusta tarkistuslistan (eng. checklist) perusteella [Nielsen & Molich 1990].

Tunnetuin, heuristiikan keksijän Jakob Nielsenin kehittämä heuristinen tarkistuslista sisältää nykyään kymmenen kohtaa [Nielsen 2003].

- Järjestelmän tilan näkyvyys.
- Järjestelmän ja todellisuuden yhteensopivuus.
- Käyttäjän hallinta ja vapaus.
- Yhtäpitävyys ja standardit.
- Virheiden estäminen.
- Tunnistaminen muistamisen sijaan.
- Joustavuus ja käytön tehokkuus.
- Esteettinen ja minimalistinen suunnittelu.
- Auta käyttäjiä tunnistamaan, diagnosoimaan ja palautumaan virheistä.
- Apu ja dokumentointi.

Tämä tarkistuslista ei kerro kovinkaan paljoa (jos mitään) käyttäjien mieltymyksistä – eikä itse asiassa edes käytettävyydestä. Dicks [2002] esittää kritiikkiä käytettävyydestien väärinymmärryksistä. Yksi tällainen väärinymmärrys on, että monet käytettävyyden tutkijat eivät erota toisistaan käsitteitä ”käytettävyyden” ja ”helppokäyttöisyys”. Mielestäni Dicksin [2002] kohdistamaton kritiikki ”käytettävyyden” ja ”helppokäyttöisyyden” usein sekoittamisesta sopii Nielsenin [2003] heuristiseen tarkistuslistaan.

Muitakin tarkistuslistoja on olemassa. Usein tarkistuslistat ovat sovellusympäristöstä riippuvia, kuten Mzourekin [2003] lista, joka on tarkoitettu NetBeans –sovellusten arviointiin ja Balanced Score Card Institutun [2003] www-sivujen arviointiin tarkoitettu lista. Puhekäyttöliittymille ei ole laadittu varsinaisia tarkistuslistoja, mutta Weinschenkin & Barkerin [2000, 225-266] laatimaa suunnitteluohjelistaa voitaisiin periaatteessa käyttää myös tarkistuslistana. Lista jakautuu 11 kohtaan.

- Virheiden hallinta.
- Palautteen antaminen.
- Varmistukset.
- Odotusten huomioiminen.
- Näppäimistöt ja motoriset toiminnot.
- Sosiaaliset- ja ympäristötekijät.
- Komentokäyttöliittymät.
- Jatkuva puheentunnistus.
- Keskustelu ja puhetulosteet eli järjestelmän puheenvuorot.
- Valikot.
- Muut kuin puheäänet.

Weinschenkin & Barkerin [2000, 225-266] lista ei ole yhtenäinen, eli jotkut kohdat koskevat tietynlaisia järjestelmiä ja toiset toisenlaisia järjestelmiä. Listaa ei olekaan laadittu tarkistuslistaksi, vaan jokaisesta kohdasta on selitetty, kuinka se pitäisi huomioida käyttöliittymäsuunnittelussa. Mutta vaikka listasta laatisikin yhtenäisen tarkistuslistan, ei se mittaisi käyttäjien mielikuvia järjestelmästä vaan järjestelmän toiminnallisuutta, sillä kaikki listan kohdat liittyvät toiminnallisuuden suunnitteluun ja toteutukseen.

Heuristisen arvioinnin perustelu on yleensä halpa kustannus. Nielsen ja Phillips [1993] tulivat vertailussaan päätelmään, että käytettävyydestä on kalliimpaa mutta analyttisesti tarkempaa kuin heuristinen arviointi. Heuristisen arvioinnin kustannuksia alentaa se, että arvioinnin voi suorittaa ilman prototyyppiä (katso osio 2.3.2 arvioinnista ilman prototyyppiä). Vaikka yksittäinen arvioija löytääkin sovelluksesta vain 20-51 % ongelmallisista kohdista, kolme itsenäistä arvioijaa löytävät hyvin suuren osan ongelmakohtista ja enemmän kuin viisi arvioijaa toistavat samoja tuloksia nostaen menetelmän kustannuksia [Nielsen & Molich 1990].

Heuristinen arviointi on menetelmänä mielenkiintoinen ja hyvin perusteltu, sillä kiireisessä IT-bisneksessä on harvinainen tilanne, että on aikaa ja resursseja mittavaan käytettävyydestäukseen – heuristinen arviointikin on parempi kuin ei mitään [Nielsen & Molich 1990].

Yleisesti ottaen tarkistuslista, joka on laadittu käyttöliittymäsuunnittelun lähtökohdista, on vääjäämättä käyttöliittymää eikä miellyttävyyttä arvioiva menetelmä. Vaikkei se olisikaan objektiivinen, se kuitenkin tähtää objektiivisten asioiden mittaamiseen subjektiivisilla menetelmillä.

4.5. Käyttäjakeskeinen koe

Käyttäjakeskeiset kokeet eli subjektiivisen tyytyväisyyden selvittäminen on ollut perinteisesti kiinnostuksen kohde lähinnä markkinoinnin piirissä. Tietojenkäsittelytieteen piirissä järjestelmiä on arvioitu lähinnä edellä esitellyillä menetelmillä. Epäilemättä kaupallisille puhepohjaisille keskustelujärjestelmille on sovellettu käyttäjakeskeisiä kokeita useassakin tapauksessa, mutta nämä sovellukset on tehty kaupallisissa, ei tieteellisissä, intresseissä. Niinpä tällaisten mittausten tuloksia ei ole julkaistu. Poikkeustapaus on ARISE-projekti [Blasband 1998], jossa rakennettiin yleiseurooppalaista juna-aikataulujärjestelmää. Hollannin rautatiet otti projektissa rakennetun VIOS-järjestelmän käyttöön ja van Haaren ym. [1998] arvioivat käytössä ollutta järjestelmää vertaillen kahta eri menetelmää (teknologia- ja käyttäjakeskeinen). Käyttäjakeskeisen kokeen he tekivät SERVQUAL-asteikon (katso luku 6) hyvin omaperäisellä tulkinnalla, jossa käyttäjiltä kysyttiin seitsemän kysymystä joilla mitattiin kolmea ulottuvuutta.

Teknistä ulottuvuutta mitattiin kysymyksellä:

1. Ymmärsikö tietokone kysymyksesi ensi soitolla?

Toiminnallista ulottuvuutta mitattiin kysymyksillä:

2. Kuinka koit keskustelun puhuvan tietokoneen kanssa?
3. Kuinka koit keskustelun pituuden puhuvan tietokoneen kanssa?
4. Kuinka koit puhuvan tietokoneen puheäänen?
5. Kun soitat junaneuvontaan ja sinua palvelee ihminen, onko palvelu huonompi/sama/parempi, kuin tietokoneen antaessa tiedon?
6. Kun soitat OVR:lle (junaneuvonta), valitsisitko puhuvan tietokoneen?

Odotusten ulottuvuutta mitattiin kysymyksellä:

7. Täyttikö puhuva tietokone odotuksesi verrattuna henkilökohtaisiin palveluneuvojiin?

Menetelmien vertailu osoittautui hyödylliseksi. Teknologiakeskeinen testi tulkittiin ensin osoittavan 94 % tyytyväisyyttä ja käyttäjakeskeinen testi osoitti 66 % tyytyväisyyttä. Teknologiakeskeisen testin tuloksia tulkittiin uudelleen ja yhteenvedossa päädyttiin 66 % tyytyväisyyteen. Ero johtuu siitä, että teknologiakeskeisen testin tuloksia tulkittiin optimistisesti. Tilanteissa, joissa käyttäjä ei reagoinut virheeseen, tämän tulkittiin johtuvan testitilanteesta, jossa virheellisellä tiedolla ei ole todellista merkitystä koehenkilölle, joka ei siis välitä meneekö jokin väärin. Käyttäjakeskeisessä testissä ilmeni kolme muuta syytä

virheisiin reagoimattomuuteen: käyttäjä ei tiedä, että voi korjata virheitä, käyttäjä ei tiedä kuinka korjata virheitä tai käyttäjä ei huomannut virhettä. Käyttäjät eivät siis olleetkaan kaikissa niissä tilanteissa tyytyväisiä, joissa teknologiakeskeisen testin perustella näin pääteltiin. Van Haaren ym. [1998] esittävätkin, että käyttäjäkeskeisiä koemenetelmiä kehitettäisiin (erityisesti ”palvelun onnistumisasteen” mittari), sillä teknisesti täydellinenkin järjestelmä on epäonnistunut mikäli käyttäjät kokevat sen kielteisesti.

Mielestäni van Haarenin ym. [1998] tutkimuksen suurin ongelma on siinä, että se mittaa vain osaa käyttäjien kokemuksista. Ongelmallinen on myös heidän omaperäinen tulkintansa SERVQUAL-asteikosta, jota he eivät perustele mitenkään joten on mahdotonta tietää miksi tällaiseen tulkintaan on päädytty. Luultavasti kysymysten määrää on haluttu vähentää ja samalla sisällyttää myös muita tyytyväisyyteen vaikuttavia tekijöitä kuin palvelun laatu. Tuon selityksen hyväksyisin perusteluksi asteikon muokkaamiseen, mutta silti jäisi epäselväksi miksi asteikkoa muokattiin juuri näin.

Luku 4 lyhyesti

- Inhimillistämisen tukemisen onnistumista ja muita inhimillistämisen vaikutuksia täytyy tutkia subjektiivisilla menetelmillä.
- Yleisesti käytetyt puhekäyttöliittymien arviointimenetelmät ovat objektiivisia, siihen tähtääviä tai liian työläitä kun tarvitaan pelkkä subjektiivinen arvio.
- Prototyypitestausten tai käytettävyystudkimuksen yhteyteen voidaan lisätä subjektiivinen arvio.
- Heuristiset tarkistuslistat ja aiemmin puhepohjaisille keskustelujärjestelmille sovelletut subjektiiviset arviot mittaavat liian pientä alaa käyttäjien kokemuksista.

5. Palvelu

”Palvelu on teko, toiminta tai suoritus, jossa asiakkaalle tarjotaan jotain aineetonta, joka tuotetaan ja kulutetaan samanaikaisesti ja joka tuottaa asiakkaalle lisäarvoa; ajansäästöä, helppoutta, mukavuutta, viihdettä tai terveyttä.” Näin kuuluu Zeithamlin & Bitnerin [1996, 5] määritelmä palveluille. Esimerkiksi Ylikoski [2000, 17-32] tarkentaa määritelmää eri näkökulmista, mutta Zeithamlin & Bitnerin määritelmä on tälle työlle riittävän tarkka.

Palvelu on aineetonta. Miten sitä siis voidaan mitata? Joitakin palvelun ominaisuuksia voidaan mitata, esimerkiksi kestoa: kuinka kauan hieronta kestää? Tällainen mittaus voi usein osoittautua huonoksi ilmaisimeksi, kun toinen hieroja poistaa vaivat jo lyhyessä ajassa ja toinen hieroo pitkään ilman tuloksia.

Palvelua itseään ei olekaan tarkoituksellista mitata, vaan sen tuottamaa lisäarvoa. Palvelun lisäarvon mittari on palvelun laatu.

Palveluiden markkinointi on kehittynyt markkinoinnin tieteenhaaran osaksi 1970-luvulta lähtien [Berry & Parasuraman 1993]. Palvelun laatu on edelleen kasvavan mielenkiinnon kohde [Ylikoski 2000 13-16]. Palveluja tarjoaville organisaatioille on tärkeä kilpailuvaltti, että asiakkaat ovat tyytyväisiä palveluun. Vaikka palvelun laatu ei yksin takaa tyytyväisiä asiakkaita, on se hyvin tärkeä tekijä. Kuten Ylikoski [2000, 149] sanoo, ”hyvään laatuun on helppo olla tyytyväinen.”

Palveluiden markkinoinnin piirissä on kehitetty erilaisia tapoja mitata palvelun laatua. Mittaus ei välttämättä aina perustukaan ennalta laadittuihin malleihin, vaan tutkimusmenetelmä voidaan kehittää yksilöllisesti tapauskohtaisesti [Lotti 1998]. Laajamittaisiin tutkimuksiin perustuva SERVQUAL-asteikko on kuitenkin kehitetty niin, että sillä pitäisi voida mitata minkä tahansa palvelun laatua. Asteikko onkin herättänyt paljon mielenkiintoa ja se on yleisessä käytössä.

SERVQUAL-asteikossa on mielenkiintoista juuri se, kuinka sen väitetään voivan mitata kaikenlaisten palveluiden laatua. Yksi tämän työn tärkeimpiä kysymyksiä onkin, voiko asteikolla mitata myös puhepohjaisen keskustelujärjestelmän tuottaman palvelun laatua.

Tässä luvussa käsitellään palvelua. Ensin määritellään palvelun laatu. Seuraavaksi käsitellään palvelun laadun mittaamista. Sitten tutustutaan laatu mittaavaan SERVQUAL-asteikkoon. Luvun päättää kappale, jossa pohditaan SERVQUAL-asteikon sopivuutta puhekäyttöliittymien ja erityisesti puhepohjaisten keskustelujärjestelmien arviointiin. Kappale toimii samalla yhteenvedona tästä ja edellisestä luvusta.

5.1. Palvelun laatu

Laatua on vaikea määritellä. Eri ihmiset arvostavat eri asioita ja yrityksillä saattaa olla toisiinsa verrattuna ristiriitaisia palvelutavoitteita. Tämä johtuu siitä, että palvelun laadun määrittelee asiakas, ja tämä on ainoa oikea lähtökohta laatua tarkasteltaessa. [Ylikoski

2000]. Useat tutkijat ja kuluttajat ovatkin yhtä mieltä siitä, että palvelun laatu tarkoittaa odotusten ja kokemusten erotusta [Kotler 2001; Parasuraman ym. 1985] – laatu on siis subjektiivinen kokemus, jonka kukin asiakas määrittelee henkilökohtaisesti. Jos palvelu on odotettua huonompi, palvelun laatu on matala. Jos palvelu on odotetun kaltainen, palvelun laatu on riittävä. Mikäli palvelun laatu ylittää odotukset, palvelun laatu on korkea. [Kotler 2001]

Palvelun laatu voidaan tiivistää kolmeen osatekijään: tekninen laatu, toiminnallinen laatu sekä imago ja odotukset. Tekninen laatu eli lopputuloksen laatu ja toiminnallinen laatu eli prosessin laatu vaikuttavat kumpikin omalta osaltaan koettuun kokonaislaatuun. Hyvä imago suodattaa satunnaisia virheitä ja huono imago vahvistaa mahdollisia huonoja kokemuksia entisestään. [Ylikoski 2000]

Odotukset määräävät hyväksyttävän teknisen ja toiminnallisen laadun rajat. Odotuksiin vaikuttaa tarpeet, hinta, aikaisemmat kokemukset, kokemukset kilpailijoista, mainosten lupaukset, muiden ihmisten mielipiteet, asiakkaan oma panostus palveluun ja tilanne, jossa palvelua tarvitaan [Ylikoski 2000]. Onkin siis tärkeää löytää sopiva tasapaino miten paljon esimerkiksi mainoksissa luvataan: jos luvataan liian vähän, ei kukaan halua käyttää palvelua, mutta jos luvataan liikaa, palvelun laatu jää matalaksi eivätkä asiakkaat käytä palvelua uudestaan.

5.2. Palvelun laadun mittaaminen

Palvelun laadun määrittely on vaikeaa ja niin on sen mittaaminenkin. Laadun määrittelyn mukaisesti on kuitenkin lähdettävä kliseestä ”asiakas on oikeassa”. Esimerkiksi kaikki Lotin [1998] esittämät tutkimuskeinot palvelun laadun selvittämiseksi perustuvat asiakkaalta kysymiseen. Toki palvelun laatua voidaan selvittää myös Mystery Shopping –menetelmällä [Miettinen 1997; Observa 2003], jossa tutkija tekeytyy asiakkaaksi ja arvioi sitten saamaansa palvelua. Toisaalta tavallaan sekin perustuu asiakkaalta kysymiseen – siinä vain tutkija tekeytyy asiakkaaksi ja käyttää roolia laadun selvittämiseen eli hän kysyy itseltään. Mystery Shopping ei kuitenkaan sovellu automaatin tuottaman palvelun arviointiin, tai ainakin asiantuntija-arviot voi suorittaa avoimesti ilman asiakkaaksi tekeytymistä, sillä automaatti ei erota tutkijoita ja oikeita asiakkaita.

Asiakkaalta kysyminen käy helposti, mikäli tietää mitä kysyy. Yksi vaihtoehto tämän selvittämiseen on asiakkaalta kysyminen. Tämä tarkoittaa lähinnä ryhmäkeskustelua, johon kutsutaan avainasiakkaita. Keskustelemalla selvitetään, mitä asioita asiakkaat pitävät ko. palvelussa tärkeinä. Näissä tärkeissä asioissa onnistumista voidaan sitten kysyä laajemmalla asiakasjoukolta. [Lotti 1998; Ylikoski 2000]

Ryhmäkeskustelua ei kuitenkaan välttämättä tarvita. Palvelun laatu muodostuu palvelusta riippumatta tietyistä ulottuvuuksista (kuten *Palveluvarmuus* ja *Luotettavuus*) [Parasuraman ym. 1985]. Asiakkailta voi kysyä suoraan palvelun onnistumista näissä

ulottuvuuksissa. Tälle ajatukselle perustuu SERVQUAL-asteikko [Parasuraman ym. 1988]. Seuraava luku käsittelee tarkemmin SERVQUAL-asteikkoa.

Luku 5 lyhyesti

- Palvelu on aineetonta lisäarvon tuottamista. Myös koneet voivat tuottaa palvelua, vaikka perinteisesti palvelun tuottajat ovatkin ihmisiä.
- Palvelua arvioidaan mittaamalla palvelun laatua.
- Palvelun laatu mitataan selvittämällä ensin asiakkaille tärkeät ominaisuudet. Laajemmassa tutkimuksessa selvitetään, miten asiakkaat kokevat kyseisen palvelun onnistuneen näissä ominaisuuksissa.
- Palvelun laadusta on olemassa myös malli, jonka mukaan palvelun laatu koostuu aina tietyistä ulottuvuuksista. Tämän mallin mukaan ominaisuuksia ei tarvitse erikseen selvittää. Tälle mallille perustuu SERVQUAL-asteikko, jota esitellään tarkemmin seuraavassa luvussa.

6. SERVQUAL-asteikko

SERVQUAL on 22 alkioita sisältävä asteikko, jolla mitataan palvelun laatuun liittyvien odotusten ja havaintojen suhdetta. Alkiot ovat palvelun laadun ulottuvuuksia koskevia väitteitä (katso alkiot liite 1). [Parasuraman ym. 1988, 1991, 1993, 1994a ja 1994b].

SERVQUALista on kolme versiota, yhden, kahden ja kolmen sarakkeen versiot. Yhden sarakkeen versiossa arvioidaan sitä, vastaavatko havainnot palvelun laadusta odotuksia (kuhunkin väitteeseen liittyen). Kahden sarakkeen versiossa arvioidaan ensin, kuinka havainnot vastaavat vähimmäisvaatimuksia ja sitten, kuinka havainnot vastaavat hyvän palvelun tasoa. Kolmen sarakkeen versiossa arvioidaan erikseen, mikä on vähimmäisvaatimuksen taso, hyvän palvelun taso ja havaitun palvelun taso (kuhunkin väitteeseen liittyen). [Parasuraman ym.1994b] Kaikkia alkioita arvioidaan numeerisella skaalalla (esimerkiksi 1...7).

Kuhunkin sarakkeeseen vastataan itsenäisesti, vaikka kahden ja kolmen sarakkeen versioissa onkin selvää, että vähimmäisvaatimuksen ei pitäisi olla hyvän palvelun tason yläpuolella. Itse asiassa havaittuakin palvelua arvioidaan odotusten perusteella, joten on perusteltua, että vastaaja sovittaa havainnot odotusten arviointiin. Siis odotukset ja havainnot tulisi kysyä joko peräkkäin tai tarjota vastaajalle mahdollisuus nähdä aiemmin tehdyt arviot. Sarakkeisiin tulisi kuitenkin vastata siten itsenäisesti, ettei vastaajaa ainakaan neuvota ottamaan toisia sarakkeita huomioon, vaan arviot tulisi suorittaa vaikutelmien perusteella, jolloin toisten sarakkeiden huomiointikin on vaistonvaraista.

Tässä luvussa käsitellään SERVQUAL-asteikkoa. Ensin tutustutaan laadun ulottuvuuksiin, jonka jälkeen käsitellään SERVQUAL-asteikon sisältöä. Seuraavaksi tarkastellaan SERVQUAL-asteikon saamaa kritiikkiä ja suosiota. Sitten käsitellään SERVQUAL-tutkimuksen tulosten analysointia ja esittämistä. Lopuksi pohditaan, mitä SERVQUAL voi tarjota puhepohjaisten keskustelujärjestelmien arviointiin

6.1. Laadun ulottuvuudet

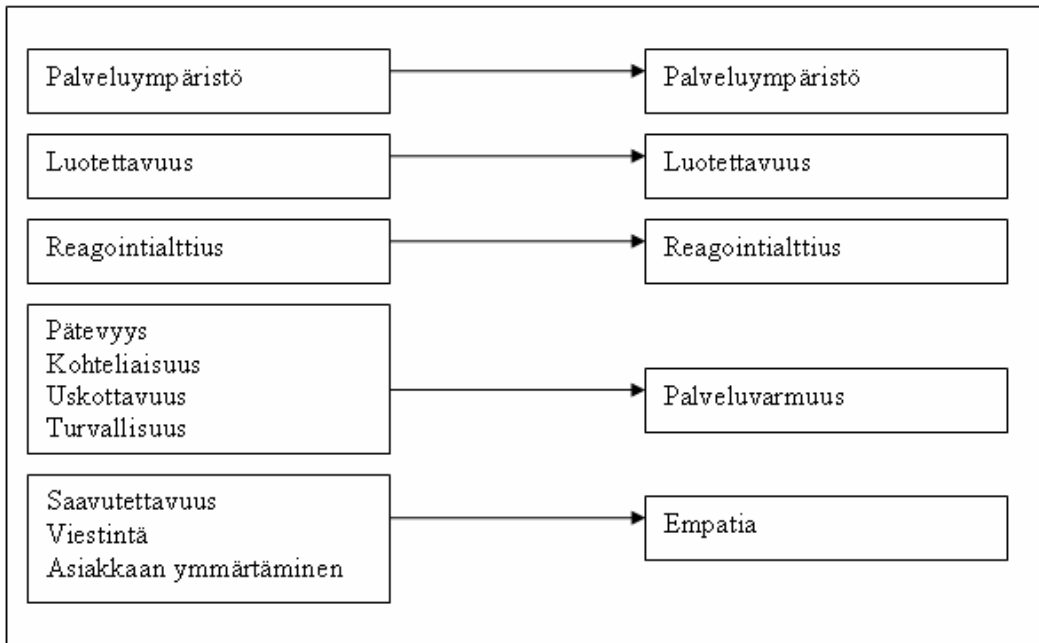
SERVQUAL-asteikko perustuu ajatukselle, että kaikilla palveluilla on jotkin yhteiset laadun ulottuvuudet. Parasuraman, Zeithaml ja Berry [Parasuraman ym. 1985] laativat listan (mallin) ulottuvuuksista useiden johtajien ja avainasiakkaiden haastaatteluihin perustuen ja muokkasivat sitä laaja-alaisten kokemusten perusteella [Zeithaml ym. 1990].

Alkuperäisessä mallissa oli kymmenen ulottuvuutta [Parasuraman ym. 1985] ja muokatussa mallissa on viisi ulottuvuutta [Zeithaml ym. 1990]. Seuraavassa on lueteltu nämä ulottuvuudet selityksineen. Ulottuvuudet on suomentanut Ylikoski [2000, 133] ja Kuuselan [1998, 129] suomenkielisiä selityksiä on käytetty mahdollisuuksien mukaan.

- *Palveluympäristö* tarkoittaa fyysistä palvelua: palvelutilojen ulkoasua, varustetasoa, henkilökunnan ulkoasua ja viestintämateriaaleja
- *Luotettavuus* tarkoittaa kykyä selviytyä luvatuista palveluista luotettavasti ja täsmällisesti
- *Reagointialttius* tarkoittaa halua auttaa asiakkaita ripeästi
- *Pätevyys* tarkoittaa palvelutapahtumassa tarvittavien tietojen ja taitojen hallintaa
- *Kohteliaisuus* tarkoittaa käytöstapoja, asennoitumista, huomaavaisuutta ja ystävällisyyttä
- *Uskottavuus* tarkoittaa luottamusta, rehellisyyttä ja asiakkaan etujen mukaista toimintaa
- *Turvallisuus* tarkoittaa palvelutapahtuman vaarattomuutta ja riskittömyyttä
- *Saavutettavuus* tarkoittaa yhteydenoton helppoutta
- *Viestintä* tarkoittaa asiakkaan kuuntelemista, ymmärtämistä ja kielenkäyttöä
- *Asiakkaan ymmärtäminen* tarkoittaa asiakkaan tarpeiden tiedostamista

Muokatussa mallissa ulottuvuiksi yhdisteltiin niin, että seitsemän alkuperäistä ulottuvuutta korvattiin kahdella uudella [Zeithaml ym. 1990]. Kuvasta 1 selviää, kuinka mallia muokattiin.

- *Palveluvarmuus* tarkoittaa henkilökunnan taitavuutta ja kohteliaisuutta sekä heidän kykyään ilmaista luottamusta.
- *Empatia* tarkoittaa välittämistä eli yksilöllistä huomiota jota yritys tarjoaa asiakkailleen.



Kuva 1: Palvelun laadun ulottuvuudet [Parasuraman ym. 1985] [Zeithaml ym. 1990], suomennokset Ylikoski [2000, 133].

6.2. Suosio, kritiikki ja siihen vastaaminen

SERVQUAL-asteikkoa on sovellettu monille perinteisille palvelualoille, kuten rengaskauppaan [Carman 1990], hammashoitopalveluihin [Carman 1990], majoitukseen [Saleh & Ryan 1992], matkailuun [Fick & Ritchie 1991], autopalveluihin [Bouman & van der Wiele 1992], opetukseen [McElwee & Redman 1993], kirjanpitoon [Freeman & Dart 1993], arkkitehtipalveluihin [Baker & Lamb 1993], sairaaloihin [Babakus & Mangold 1992; Mangold & Babakus, 1991], pankkeihin [Kwon & Lee 1994; Wong & Perry 1991], paikallishallintoon [Scott & Shieff 1993] ja kirjastoihin [Nitecki 2003]. Sitä on sovellettu myös uuden teknologian palveluihin kuten verkkosivuihin [Negash ym. 2002] ja yleensä sähköiseen kaupankäyntiin [van Iwaarden & van der Wiele 2003]. Sitä esitellään markkinoinnin oppi- ja käsikirjoissa [Kuusela 1998; Ylikoski 2000, 133]. Voidaankin sanoa, että SERVQUAL-asteikko on laajalti käytetty.

SERVQUALia kohtaan esitettiin rakentavaa kritiikkiä ja Parasuraman ym. kehittivät asteikkoa moneen otteeseen [Parasuraman ym. 1991, 1993, 1994a ja 1994b]. Esimerkiksi Cronin & Taylor [1992 ja 1994] kritisoivat odotusten mittaamisen luotettavuutta palvelun laadun mittarina. Teas [1993 ja 1994] kritisoivat myös odotusten mittaamista ja havaintoihin vertaamista, todeten ettei laskeneet havainnot tarkoita että laatu olisi parantunut, vaikka odotusten ja havaintojen erotusta tarkastelemalla näin voisikin päätellä. Parasuraman ym. [1994b] vastasivat tähän kritiikkiin luomalla SERVQUAL-asteikosta kolmen sarakkeen mallin, joka lisää luotettavuutta ja parantaa tarkastelumahdollisuuksia.

Kaikkeen kritiikkiin ei pystytty vastaamaan. Parasuraman ym. [1994b] myöntävät tämän todeten, että asteikon käyttäminen kaikessa laajuudessaan saattaa olla ongelmallista, koska odotusten toisistaan erottaminen voi olla vastaajille hankalaa. Parasuraman ym. [1994b] myöntävät myös, että SERVQUAL ei ole tehokkain palvelun laadun mittari psykometristen mittausten perusteella. Toisaalta kun he väittävät, että kolmen sarakkeen SERVQUAL-asteikko on hyödyllisin mittari selvitetessä yksittäisiä palvelun epäkohtia, he eivät tarjoa analyysikeinoja näiden epäkohtien löytämiseen SERVQUALin tuottamasta aineistosta. Lopulta Parasuraman ym. [1994b] esittävät, että SERVQUAL-mittausta kannattaa käyttää pitkittäistutkimuksena, aiempaan laatuun vertailevana.

Kritiikkiä on kuitenkin esitetty lisää. Buttle [1996] jakaa SERVQUAL-asteikon kritiikin teoreettisiin ja toimenpiteellisiin puutteisiin. Teoreettisia puutteita on neljä:

- Paradigman vastaväitteet: SERVQUAL perustuu odotusten täyttymisen mittaamiseen mielipiteiden mittaamisen sijasta.
- Puutteellisesti perusteltu laadun määrittely odotusten ja havaintojen suhteena.
- Prosessikeskeisyys palvelun lopputuloksen arvioinnin kustannuksella.
- Ulottuvuuksien ongelmat: viisi ulottuvuutta eivät ole yleismaailmallisia, malli on liian yksiselitteinen.

Toimenpiteellisiä puutteita on seitsemän:

- Odotukset: termi on moniselitteinen eikä kata kaikkia palvelun laatuun vaikuttavia laatukriteereitä eikä SERVQUAL mittaa absoluuttisia odotuksia.
- Alkiot: neljä tai viisi alkiota eivät riitä kuvaamaan yhtä ulottuvuutta.
- Asiakkaiden muuttuvat käsitykset hyvästä palvelusta.
- Polaarisuus: käänteisesti tarkasteltavat alkiot aiheuttavat hämmennystä (uudemmat SERVQUAL -versiot eivät pidä sisällään käänteisesti tarkasteltavia alkioita).
- Asteikon väärä käyttö.
- Kahtiajakoisuus: Kahteen kertaan toistettu arviointi tylsistyyttää ja hämmentää.
- Vaihtelun hukkaaminen: SERVQUALin yleistarkastelu selvittää vain hieman vaihtelua alkioista toiseen.

Kritiikki on monelta osin aiheellista. Toisaalta esimerkiksi polaarisuuden kritiikki on kohdistettu pelkästään SERVQUALin varhaisiin versioihin. Mielestäni ongelmallisimmat

kohdat ovat prosessikeskeisyys, ulottuvuuksien ongelmat, kahtiajakoisuus ja vaihtelun hukkaaminen.

Buttlen [1996] kritiikki ei sovi kaikilta osin kolmen sarakkeen SERVQUAL-asteikolle. Kolmen sarakkeen mallista selviää paitsi odotusten ja havaintojen erotus myös mielipiteet (*havainnot*-sarake). Myös odotusten jako kahteen eri sarakkeeseen parantaa asteikon validiteettia. Odotusten ja havaintojen erotusmalli (kuilumalli) on perusteltu, toisin kuin Buttle [1996] väittää, esimerkiksi Parasuramanin ym. [1994a & 1994b] toimesta. Alkioiden pieneen määrään ja asiakkaiden muuttuviin käsityksiin kohdistuva kritiikki on aiheellista, mikäli tarkoitus on tosiaan määrittellä absoluuttinen palvelun laadun taso, mutta käytännössä tarkoitus on yleensä löytää kohtia, joista asiakkaat pitävät palvelussa ja joista he eivät pidä. Asteikon väärä käyttö liittyy puolestaan asteikon reliabiliteetin arviointiin eikä varsinaiseen SERVQUAL-tutkimukseen. Tämä osoittaa teoreettisen taustan puutteellisuutta, muttei ole keskeinen ongelma, sillä reliabiliteetin suurimmat aukot liittyvät ulottuvuuksien ongelmiin joita voidaan parantaa.

Mielestäni suurin ongelma SERVQUAL-asteikkoon liittyen on se, että SERVQUAL on kehitetty vain asteikkona eikä asteikkoa soveltavana metodina. SERVQUALin soveltamisesta käytäntöön puhutaan lyhyesti, asia mainiten mutta sitä varsinaisesti käsittelemättä [Parasuraman ym. 1988, 1991, 1993, 1994a ja 1994b; Zeithaml ym. 1990]. Samoin analyysimenetelmät, joita Parasuraman ym. [1994b] esittävät, ovat riittämättömät.

Mikäli SERVQUAL-tutkimus suoritetaan integroituna osana laajempaa tyytyväisyystutkimusta, asteikon prosessikeskeisyys kompensoituu muilla mittauksilla. Kahtiajakoisuuden negatiivisia vaikutuksia voidaan lieventää selvittämällä vastaajille, mistä menetelmässä on kyse selittämällä odotusten ja havaintojen erotuksen merkitys. Limittäisten ulottuvuuksien ongelmat eivät ole keskeisiä, kun vastaajille annetaan mahdollisuus keskustella omista havainnoistaan ja odotuksistaan niissä kohdissa, jotka he itse kokevat tärkeiksi. Tällöin ei tarvitse noudattaa kirjaimellisesti ennalta sovittuja ulottuvuuksia. Vastaajien yksilöllistä vaihtelua ei ole pakko hukata analyysissä, kun ottaa käyttöön uusia analyysimenetelmiä.

Käsittelen SERVQUAL-asteikon soveltamista puhepohjaisiin keskustelujärjestelmiin luvussa 7. Esittämäni metodi on helppo integroida osaksi laajempaa tyytyväisyystutkimusta, ja empiiriseen osuuteen (luvut 8 ja 9) sisältyykin pieni tyytyväisyystutkimus. Metodi sisältää myös vastaajien ohjeistuksen ja mahdollisuuden keskusteluun.

Aineiston analysointia käsittelen kappaleessa 6.3, jossa esittelen myös tyytyväisyystutkimuksista sovelletun analyysimenetelmän joka ei hukkaa vaihtelua. Tämä menetelmä esittää tulokset tärkeys-laatu-taulukossa.

6.3. Aineiston analysointi

SERVQUAL-mittauksen aineistoa ei pitäisi käyttää arvosanamaisesti vaan palvelukuilujen (eli mistä huonot tulokset johtuvat, katso Parasuraman ym. [1985] tai Kuusela [2000]) selvittämiseen sekä pitkän tähtäimen vertailuun: jos palveluvaikutelmat ovatkin korkeammalla hammaslääkärin vastaanotolla kuin pizzeriassa, ei pizzerialan pidä päätellä, että ihmiset käyvät mieluummin hammaslääkärissä kuin heidän pizzerialassaan. Toisaalta palveluvaikutelmien keskiarvo ei luultavasti ole edes sama asia kuin yleisarvosana palvelun laadusta – palvelun laadun ulottuvuudet ovat eriarvoisia, yleensä palveluympäristöä pidetään vähiten tärkeänä [Zeithaml ym. 1990].

Tässä kappaleessa esitellään kaksi tapaa analysoida SERVQUAL-aineistoa. Ensin esittelen lyhyesti tavan selvittää odotusten ja vaikutelmien erotus, joka on mielestäni yksinään riittämätön kriteeri. Laajemmin esittelen tyytyväisyystutkimuksista sovelletun tärkeys-laatu-taulukon, joka on oma panokseni SERVQUALin kehittämiseen paremmaksi arviointityökaluksi.

6.3.1. MSA ja MSS

Kuten SERVQUAL-tutkimuksen piirissä määritellään, palvelun laatu on odotusten ja vaikutelmien erotus. Koska odotuksia on kahdenlaisia (vähimmäisvaatimus ja hyvä taso), laatuakin täytyy mitata kahdella eri mitalla. MSA (Measure of Service Adequacy, palvelun riittävyuden mitta) muodostuu vähimmäisvaatimuksen ja vaikutelmien erotuksista ja MSS (Measure of Service Superiority, palvelun ylemmyyden mitta) muodostuu hyvän tason ja vaikutelmien erotuksesta. [Parasuraman ym. 1994b] Parasuraman ym. [1994b] esittävät esimerkkitutkimuksiensa tulosten esittelyssä MSA- ja MSS-mittojen keskiarvon kullekin ulottuvuudelle. SERVQUAL-asteikon lähtökohta onkin, että kutakin ulottuvuutta mittaavat alkioit voidaan yhdistää ja yleistää ulottuvuuden mitaksi.

6.3.2. Laadun tarkastelu tärkeys-laatu -taulukossa

SERVQUAL-mittauksen aineistosta kannattaa selvittää, mitä asioita asiakkaat pitävät tärkeinä mutta heikosti toteutettuina ja kehittää niitä – samalla voi selvittää mitä asioita asiakkaat eivät pidä niin tärkeinä ja tarkastaa voisiko näihin asioihin satsaamisen sijaan resursseja siirtää tärkeämpiin. Esimerkiksi bussiaikataulupalveluihin sovelletusta SERVQUAL-tutkimuksesta saatetaan huomata, että asiakkaat pitävät luotettavuutta tärkeänä ja empatiaa vähemmän tärkeänä ja että luotettavuudessa ei asiakkaiden mielestä olla vähimmäisvaatimusten tasolla. Tällaisessa tilanteessa bussiaikataulupalvelu voisi esimerkiksi kiinnittää palveluhenkilökunnan koulutuksessa entistä enemmän huomiota oikeiden aikataulutietojen löytymiseen sen sijaan, että järjestettäisiin empatiaa parantavia koulutustilaisuuksia.

Vaikka palvelun laatu ei olekaan ainoa asia, joka vaikuttaa asiakastyytyväisyyteen, laatu yleensä on varmasti tärkein asiakastyytyväisyyden tekijä [Kotler 2001]. Onkin

selkeää esittää palvelun laadun tutkimuksen tulokset asiakastyytyväisyystutkimuksen tulosten tapaisessa taulukossa, vaikka ainakaan Parasuraman ym. eivät näin tee SERVQUALia käsittelevissä artikkeleissaan [1988, 1991, 1993, 1994a, 1994b] eikä palvelun laatua käsittelevässä kirjassaan [Zeithaml ym. 1990], joissa he esittelevät kyllä palvelun laadun tutkimusten tuloksia muuten.

Kun asiakastyytyväisyyttä mitataan suorilla kysymyksillä, joilla saadaan yksiselitteinen vastaus tärkeyteen ja tyytyväisyyteen, on mahdollista tehdä kuvan 2 tyylinen taulukko. Siitä voi tehdä suoria päätelmiä, mutta usein on selkeämpää esittää tulokset kuvan 3 tyyllisenä taulukkona. Kuvan 3 tyylinen taulukko on ainoa vaihtoehto silloin, kun tyytyväisyydestä ja tärkeydestä ei tiedetä tarkkaa arviota vaan pelkästään suuntaa antava arvio (alhainen/korkea). Kuvan 3 tyylinen taulukko on silti hyvä vaihtoehto, vaikka tarkemmat arviot olisi saatavilla – se on helppolukuinen ja jatkotoimenpiteet on yksinkertaista laatia sen perusteella, mitkä ominaisuudet sijoittuvat mihinkin ruutuun. Dutkan [1995] esittämät ja Ylikosken [2000] suomentamat tulkinnat taulukon ruuduista ovat tietenkin vain suuntaa-antavia eikä niitä ole edes tarkoitettu kirjaimellisesti noudatettaviksi.

		Tyytyväisyys					
		1 (matala)	2	3	4	5 (korkea)	Yhteensä
Tärkeys	1 (matala)	0	0	32	44	77	153
	2	0	1	14	30	67	112
	3	1	3	27	20	37	88
	4	1	2	5	5	14	27
	5 (korkea)	2	1	3	4	10	20

Kuva 2: Asiakastyytyväisyystutkimuksen tulosten tarkka tarkastelu. Tyytyväisyyttä ja tärkeyttä on kysytty asteikolla 1...5, kuvassa vastaukset yhden ominaisuuden osalta. Numero tarkoittaa näin vastanneiden summaa [Dutka 1995]

	Alhainen tyytyväisyys	Korkea tyytyväisyys
Alhainen tärkeys	Parannustoimenpiteet eivät kiireellisiä	Tarpeettomia vahvuuksia – voidaan heikentää
Korkea tärkeys	Huomiota tarvitsevat seikat	Organisaation vahvuudet

Kuva 3: Asiakastyytyväisyystutkimuksen tulosten tarkastelu [Dutka 1995] suomennokset Ylikoski [2000]

Kolmen sarakkeen SERVQUAL-mittauksen tärkeimpiä anteja on niin sanotun sietovyöhykkeen määrittelyminen. Sietovyöhyke tarkoittaa vähimmäisvaatimusten ja hyvän tason vaihteluväliä [Parasuraman ym. 1994b]. Kuitenkin juuri tämän sietovyöhykkeen hyödyntäminen mittausaineiston analysoinnissa on täysin unohtunut. Parasuraman ym. [1994b] käyttävät kyllä sekä vähimmäisvaatimuksia että hyvän tason arvioita MSA- ja MSS-arvojen määrittelyyn, mutta sietovyöhykkeen he vain esittävät kaaviossa havaintoarvioiden kanssa esittämättä mitään keinoja, joilla sietovyöhykkeen selvittämisestä olisi hyötyä jatkotoimenpiteiden suunnittelussa. Lisäksi sietovyöhyke, jonka he esittävät, on keskiarvo kaikkien vastaajien sietovyöhykkeestä.

On intuitiivisesti selvää, että vastaajien odotukset ovat yksilöllisiä. Toisille on tärkeämpää, että palvelu osoittaa palveluhalua, toisille on tärkeämpää, että palveluvarmuus on korkea. Koska vastaajat ilmoittavat havaintonsa samalla asteikolla kuin odotuksensakin, on oletettavaa, että he vaistonvaraisesti ilmoittavat, kuinka havainnot sijoittuvat heidän omiin odotuksiin nähden. Siksi havaintoarvojen sijoittumista sietovyöhykkeeseen nähden pitäisi mitata yksilöllisesti.

Havaintoarvojen sijoittumista sietovyöhykkeeseen nähden on luontevaa mitata kolmiportaisella asteikolla, sillä havaintoarvot sijoittuvat joko sietovyöhykkeen alle, sisälle tai sen ylle. Sietovyöhykettä näin hyväksikäyttäen saadaan kuvaan 3 verrattuna tyytyväisyys/laatu -asteikolle yksi aste lisää, *Riittävä laatu*, eli ne ominaisuudet, joissa havainnot sijoittuvat sietovyöhykkeen sisälle. Kuvassa 4 näkyy tämä lisäys. Lisätyt sanalliset tulkinnat kuvaavat mahdollisia toimintaehdotuksia, mutta niiden täsmällinen sisältö ei ole keskeinen, vaan ylipäätään tarkastelun soveltaminen.

	Alhainen laatu	Riittävä laatu	Korkea laatu
Alhainen tärkeys	Parannustoimenpiteet eivät kiireellisiä	Muutoksiin ei tarvetta	Tarpeettomia vahvuuksia – voidaan heikentää
Korkea tärkeys	Huomiota tarvitsevat seikat	Ei tarvitse muuttaa; paranna mikäli ei muita huomiota tarvitsevia seikkoja	Palvelun vahvuudet

Kuva 4: Kolmen sarakkeen SERVQUAL-tutkimuksen tulosten tarkastelu tärkeys- laatu -asteikoilla, tyytyväisyysasteikolle lisätty ”riittävä laatu” verrattuna kuvaan 3

Laadun määrittely SERVQUALin perusteella tälle asteikolle on suoraviivaista. Arvio, joka alittaa vähimmäisvaatimuksen, kertoo alhaisesta laadusta ja arvio, joka ylittää hyvän

palvelun arvon kertoo korkeasta laadusta. Mikäli arvio jää näiden arvojen välille, laatu on riittävä.

Laatu määritellään henkilö- ja ominaisuuskohtaisten sietovyöhykkeiden perusteella. Toisin sanottuna laadun laskemiseen ei käytetä mitään keskiarvoja – ei ulottuvuuden keskiarvoa, vastaajan keskiarvoa eikä vastaajien keskiarvoa. Laatu määräytyy sen perusteella, sijoittuuko vastaajan arvio (havainto) kustakin SERVQUAL-asteikon alkiosta hänen itsensä samalle alkiolle määrittelemien vähimmäisvaatimuksen ja hyvän tason alle, väliin vai ylle.

Tärkeys määritellään sen perusteella, onko odotusten (vähimmäisvaatimus ja hyvän tason arvio) keskiarvo yli keskitason (asteikolla 1-7 keskitaso on 4).

6.4. Puhepohjaisten keskustelujärjestelmien arviointi SERVQUAL-asteikolla

SERVQUAL mittaa palvelun laatua. Se ei mittaa tyytyväisyyttä, vaikka laadulla ja tyytyväisyydellä saattaakin olla korrelaatio.

Palvelun laatu on odotusten ja havaintojen erotus. Siksi SERVQUAL ei mittaa suoriutumista, vaikka suoriutuminen saattaakin parantaa havaittavaa palvelun laatua.

Nykyiset puhepohjaisten keskustelujärjestelmien arviointimenetelmät eivät käytä hyväkseen sitä mahdollista tilannetta, että järjestelmä täyttää palvelun kriteerit ja sitä voidaan arvioida palveluna muiden palveluiden tapaan. Lähtökohta sinänsä on tietojenkäsittelytieteelle ymmärrettävä, tarkoitus on tehdä mahdollisimman ”hyvä” (en nyt puutu siihen, mitä ”hyvä” mahdollisesti tarkoittaa) tekninen sovellus. Kun arvioidaan teknistä sovellusta, on normaalia ottaa tekijän näkökulma – arvioida eri vaihtoehtojen objektiivista paremmuutta.

Puhepohjainen keskustelujärjestelmä on tekninen sovellus, mutta järjestelmät ovat myös tuotteita. Kappaleessa 2.3 esitellyt järjestelmät kuvaavat hyvin sitä, kuinka puhepohjaiset keskustelujärjestelmät ovat usein palvelutuotteita. Onkin hankala kuvitella tilannetta, jossa ihmisen ja tietokoneen välille muodostuu keskustelu ilman, että tietokone suorittaa ihmiselle jotain palvelua.

Puhepohjaisten keskustelujärjestelmien palvelun laatua voi arvioida SERVQUAL-asteikolla. Mittauksella saadaan käyttäjien arvio järjestelmästä palvelun tuottajana. Mikäli asteikolla vertaillaan kahta järjestelmää, jotka suoriutuvat samoista tehtävistä, voidaan tuloksista päätellä, onko järjestelmien välillä eroja palvelun laadun ulottuvuuksissa – kumpi on parempi. Tuloksista selviää myös, mitä seikkoja käyttäjät pitävät tärkeänä ja kuinka kyseinen järjestelmä suoriutuu näissä seikoissa.

Luku 6 lyhyesti

- SERVQUAL on asteikko, jolla selvitetään asiakkaiden subjektiiviset käsitykset palvelun laadusta kaikilla palvelun laadun ulottuvuuksilla.
- SERVQUAL-tutkimuksen tuloksia voidaan tarkastella taulukossa tärkeys-laatu-asteikolla.
- SERVQUAL-tutkimus voidaan soveltaa puhepohjaisten keskustelujärjestelmien arviointiin, sillä yleensä kaikki puhepohjaiset keskustelujärjestelmät ovat palveluita.
- Palvelun laadun arviointi tuo puhepohjaisten keskustelujärjestelmien arviointiin subjektiivisen arvion järjestelmästä tuotteena.

7. SERVQUAL-tutkimuksen puhepohjaisiin keskustelujärjestelmiin soveltaminen

Tässä luvussa esittelen soveltamani metodit, joilla SERVQUAL-tutkimusta voidaan käyttää puhepohjaisten keskustelujärjestelmien arviointiin. Esittämäni kaksi menetelmää poikkeavat perinteisistä palvelun laadun tutkimuksista lähinnä siltä osin, että palvelutilannetta voidaan rekonstruoida erilaisin prototyypein (paperiproto, Wizard of Oz tai keskeneräinen järjestelmä). Tällöin prosessi voidaan aloittaa odotusten selvittämällä ennen kuin vastaajilla (eli koehenkilöillä) on käsitys tutkittavasta palvelusta. Kolmas menetelmä vastaa pitkälti perinteistä palvelun laadun mittausta. SERVQUAL-tutkimus voidaan soveltaa puhepohjaisten keskustelujärjestelmien arviointiin siis kolmella tavalla:

- Pyydetään vastaajien arviot järjestelmän kuvauksesta.
- Annetaan vastaajien kokeilla prototyyppiä ja pyydetään arviot.
- Järjestelmällä on oikeita käyttäjiä, joilta arviot pyydetään.

Esittelen tässä kappaleessa nämä kolme tapaa sanallisesti ja prosessikuvausten muodossa. Tavot eivät kuvauksena eroa toisistaan paljon, mutta niiden toisistaan erottaminen on tärkeää. Tavot poikkeavat toisistaan lähtökohtien suhteen huomattavasti. Itse asiassa tavot sopisivat varmasti muidenkin palveluiden ja palvelujärjestelmien arviointiin, mutta puhun tässä vain puhepohjaisista keskustelujärjestelmistä sillä esittelen tavot puhepohjaisten keskustelujärjestelmien erikoisominaisuudet huomioiden.

Kaikissa metodeissa käytetään kahta lomaketta, odotuslomaketta (liite 1) ja havaintolomaketta (liite 2). Nämä lomakkeet perustuvat kolmen sarakkeen SERVQUAL-asteikkoon. Odotuslomake sisältää sarakkeet *vähimmäisvaatimus* ja *hyvä taso*. Havaintolomake sisältää sarakkeen *havainnot*. Fyysisesti odotukset ja havainnot voivat kyllä olla samalla paperilla, mutta erillisistä lomakkeista puhuminen selkiyttää prosessikuvausta. Mikäli odotukset ja havainnot sijoitetaan fyysisesti yhdelle lomakkeelle, täytyy kiinnittää huomiota, ettei vastaaja saa myöhemmin täytettävistä lomakkeen kohdista tietoja, joita hänelle ei haluta antaa siinä vaiheessa.

Olen suomentanut odotuslomakkeen (liite 1) ja havaintolomakkeen (liite 2) väitteet Parasuramanin ym. [1991] esimerkkien pohjalta. On kuitenkin tärkeää huomata, ettei väitteitä voi suomentaa suoraan. Esimerkiksi Parasuraman ym. [1991] väitteissä puhutaan ”henkilökunnasta” kun tarkoitetaan palvelun tuottajaa. Puhepohjaisen keskustelujärjestelmän tapauksessa tämä pitää siis suomentaa yksinkertaisesti ”palveluksi” tai ”järjestelmäksi”. Suomentamissani väitteissä ei myöskään esiinny palvelun visuaalisen miellyttävyyden arviota toisin kuin Parasuramanin ym. [1991] esimerkeissä. Tämä johtuu siitä, että järjestelmä, jonka palvelun laatua arvioin, on puhelinpohjainen. Näin ollen visuaalisuutta tärkeämpää palveluympäristön

miellyttävyyden arvioimiseksi on arvioida äänen miellyttävyyttä. Puhepohjainen keskustelujärjestelmä voi olla myös esimerkiksi verkkosivuilla, jolloin on perusteltua arvioida myös palvelun visuaalista miellyttävyyttä. Lomakkeita (liitteet 1 ja 2) ei siis pidä käyttää muokkaamatta muuhun kuin puhelinpohjaisten bussiaikataulujärjestelmien arviointiin

Kaikissa metodeissa kerätään myös tyytyväisyyden arvio. Tyytyväisyyttä voi arvioida monilla eri tavoilla – tärkeintä on, että tyytyväisyyttä arvioidaan jotenkin. Tyytyväisyydestä voidaan esimerkiksi pyytää yleisarvosana ja kysyä järjestelmän käyttöhalukkuutta. Prosessikuvauksissa tyytyväisyyden arvio on sijoitettu tutkimuksen viimeiseksi kohdaksi. Periaatteessa se voisi olla ennen *havainnot* -lomakkeen täyttämistä, mutta tällöin palvelun laadun arviointi keskeytyisi. Vapaassa käytössä ollutta järjestelmää arvioivan tutkimuksen yhteydessä tyytyväisyyden arviointi voidaan suorittaa ennen palvelun laadun arviointia, mutta silti on sujuvampaa sijoittaa se loppuun, sillä tällöin vastaajat antavat sanalliset kommentit palvelun laadun arvioinnin huomioiden eivätkä esimerkiksi toista siinä olevia kohtia tai sekoita käsitteitä.

7.1. Kuvausta arvioiva SERVQUAL-metodi

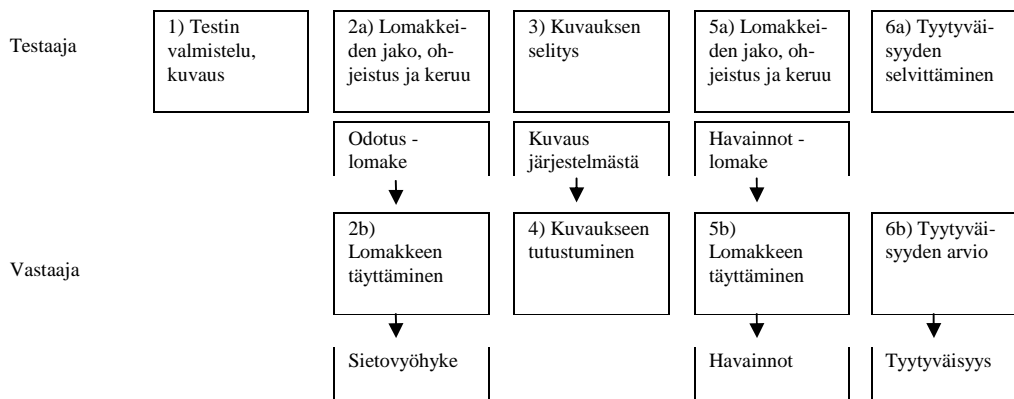
Esittelen tässä kappaleessa kuvausta arvioivan metodin SERVQUAL-asteikon käyttämiseen. Metodi soveltuu palveluille, jotka ovat kehityksen alkuvaiheessa tai jostain muusta syystä ei ole saatavilla järjestelmää vastaajien kokeiltavaksi. Arvioiden saamiseksi ei tarvita minkäänlaista prototyyppiä järjestelmästä, vaan riittää, että on suunnitelma. Suunnitelma itse ei anna vastaajille selkeää kuvaa järjestelmän toiminnasta, elleivät vastaajat ole asiantuntijoita. Paremman kuvan järjestelmän toiminnasta saa kuvauksesta, joka voi olla esimerkiksi paperille kirjoitettu esimerkkidialogi (tai käsikirjoituksen tapainen teksti, joka sisältää myös taukojen kestot ja äänenpainotukset ja muut mahdollisesti käsityksen saamista auttavat merkinnät), äänite tai videonauha. Toinen vaihtoehto on luetella (paperilla tai nauhalla) järjestelmän puhetulosteita ja kertoa missä tilanteessa järjestelmä sanoo näin. Tällöin vastaajan on kuitenkin vaikea arvioida muuta kuin puhetulosteiden suunnittelua.

Joka tapauksessa kuvausta arvioiva metodi ei ole yhtä luotettava kuin metodit, joissa vastaajat saavat käyttää järjestelmää. Esimerkiksi Interact-bussiaikataulujärjestelmää Interact-projektin [Jokinen ym. 2002] puitteissa arvioitaessa huomattiin, etteivät käyttäjät kokeneet pitkiä odotusaikoja kovin häiritsevinä. Syy tähän oli luultavasti käyttäjien keskittyminen tehtävän suoritukseen – kognitiivinen kuormitus oli niin suuri, ettei pitkästä tauosta ollut haittaa vaan käyttäjä sain koota ajatuksensa. Kuvausta lukiessa tai kuunnellessa kognitiivinen kuormitus on tuskin yhtä suuri kuin järjestelmää käyttäessä.

Kuvassa 5 metodi on esitetty prosessikuvauksena. Metodi on kolmiosainen:

- Kerätään koehenkilöiden palveluodotukset
- Annetaan kuvaus palvelusta

- Kerätään koehenkilöiden mielipiteet palvelun laadusta kuvauksen perusteella



Kuva 5: Prosessikuvaus kuvausta arvioivasta SERVQUAL-metodista

Odotukset kerätään siis ennen esimerkkien antamista. Näin siksi, että kokemukset vaikuttavat odotuksiin [Ylikoski 2000, 123-125]. Tätä vaikutusta ei haluta, varsinkaan jos halutaan vertailla eri versioita palvelusta. Odotukset saattavat vääristyä, vaikkei palvelusta olisikaan kuin yksi versio, joten varmuuden vuoksi odotukset kerätään ennen esimerkkien antamista.

Kokeen aluksi koehenkilöiltä kerätään palveluodotukset. Vaikka odotukset kerätään ennen palveluun tutustumista, on tärkeää, että kaikilla koehenkilöillä on sama käsitys minkä palvelun odotuksia kerätään. Esimerkiksi bussiaikataulupalveluun liittyvät odotukset ovat mahdollisesti erilaisia kuin esimerkiksi taksin puhelinneuvontaan liittyvät. Siksi testaajan täytyy antaa ohjeistus siitä, mihin vastataan. Tällöin on tärkeää huomioida myös se, onko kyseessä täysin uusi palvelu vai vanhan palvelun uudenlainen toteutus. Mikäli kyseessä on vain uudenlainen toteutus, voidaan vastaajia ohjeistaa vastaamaan odotuksiin vanhan palvelun perusteella. Tällöin odotukset perustuvat kokemuksiin eikä vain pelkkään arvailuun siitä, mikä mahdollisesti olisi tärkeää.

Testaajan on hyvä selittää vastaajille myös miten vastataan, sillä odotusten (vähimmäisvaatimus ja hyvä taso) ero ei välttämättä ole vastaajalle selvä. Esimerkiksi luvussa kahdeksan esittelemässäni testissä vastaajille ei riittänyt, että pyydettiin ympyröimään ensin vähimmäisvaatimuksen taso ja sitten hyvä taso. Toisaalta jotkut vastaajat ihmettelivät, kuinka vähimmäisvaatimuksia voisi arvioida ennen kuin kuvaus järjestelmästä on annettu. Tällöin riitti, että toisti palvelun luonteen (puhelinpohjainen bussiaikataulupalvelu) ja vakuutti, että odotusten arviot tehdään yleisesti kyseisenlaisista palveluista.

Palveluista annettavat esimerkit kannattaa suunnitella huolella päämäärästä riippuen. Mikäli halutaan saada arvio koko järjestelmästä, esimerkkejä tulee antaa paljon, kattaen

järjestelmän eri osa-alueet. Mikäli halutaan vertailla kahta versiota järjestelmästä, riittää, että esimerkit sisältävät versioiden eron siinä laajuudessa kuin se vaikuttaa järjestelmään – toisaalta kun halutaan vertailla versioiden eroa todenmukaisessa järjestelmässä, täytyy vastaajilla olla käsitys kontekstista (eli miten laadukas palvelu muuten on). Mikäli versioiden eron arviointi on itseisarvoinen tutkimuspäämäärä, esimerkit kannattaa suunnitella siten, että havainnot sijoittuisivat mahdollisimman hyvin sietovyöhykkeen sisälle. Jos versioiden eroa halutaan arvioida todenmukaisessa järjestelmässä, esimerkkien tulisi antaa todenmukainen kuva järjestelmän toiminnasta.

Kuvauksen järjestelmästä voi antaa monella tavalla. Yksinkertaisin tapa lienee kirjoittaa esimerkkejä käyttäjän ja järjestelmän vuorovaikutuksesta ja antaa nämä esimerkit koehenkilöiden luettavaksi paperilla tai vaikka verkkosivuilla. Samat esimerkit voidaan nauhoittaa ja antaa nämä nauhat koehenkilöiden kuunneltavaksi – tällöin koehenkilöt saavat paremman käsityksen järjestelmästä, kun kuulevat sen puheäänän ja vuorovaikutuksen eri vaiheiden kesto sekä vuorovaikutuksen kokonaiskesto realisoituu.

Kun esimerkit on annettu, kerätään koehenkilöiden havainnot havaintolomakkeella. Mikäli havainnot kerätään ryhmässä, testaja odottaa, että kaikki vastaajat ovat valmiita ennen kuin antaa luvan keskustella, sillä muiden vastaajien mielipiteet saattavat muuten vaikuttaa vastauksiin.

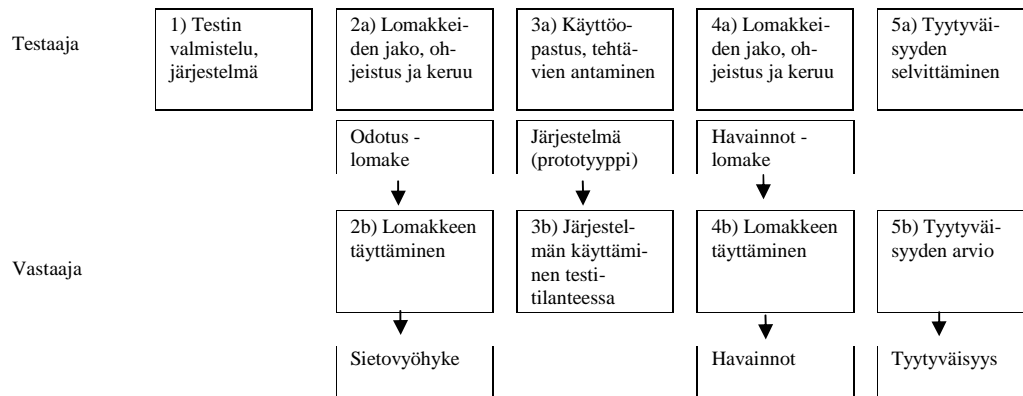
Lopuksi kannattaa tarjota koehenkilöille tilaisuus kommentoida kokemuksiaan kahdestakin syystä. Toisaalta näin päästään usein niiden syiden juurille, mistä pettymykset tai positiiviset yllätykset johtuvat. Lisäksi koehenkilö kokee, että juuri hänen mielipiteistään ollaan kiinnostuneita – tällöin hän on motivoitunut tulemaan muihinkin testeihin ja saman testin seurantaan tutkimukseen.

7.2. Koekäytettävää järjestelmää arvioiva SERVQUAL-metodi

Todenmukaisin tapa antaa esimerkki järjestelmästä on antaa valmis (tai siltä vaikuttava) järjestelmä koekäyttäjille kokeiltavaksi, joko vapaasti tai joihinkin tilanteisiin perustuen. Tämä esimerkin antaminen voidaan yhdistää johonkin muuhun tutkimusmenetelmään, kuten Wizard of Oz, System in a Loop, käytettävyyystestaus tai PARADISE. Metodi soveltuu järjestelmille, joista on toimiva prototyyppi tai kattava suunnitelma Wizard of Oz –tutkimusta varten, muttei vielä vapaassa käytössä olevia toteutusta.

Metodi on esitetty prosessikuvauksena kuvassa 6. Metodi on kolmiosainen:

- Kerätään koehenkilöiden palveluodotukset
- Annetaan koehenkilöiden kokeilla palvelua
- Kerätään koehenkilöiden mielipiteet palvelun laadusta kokeilun perusteella



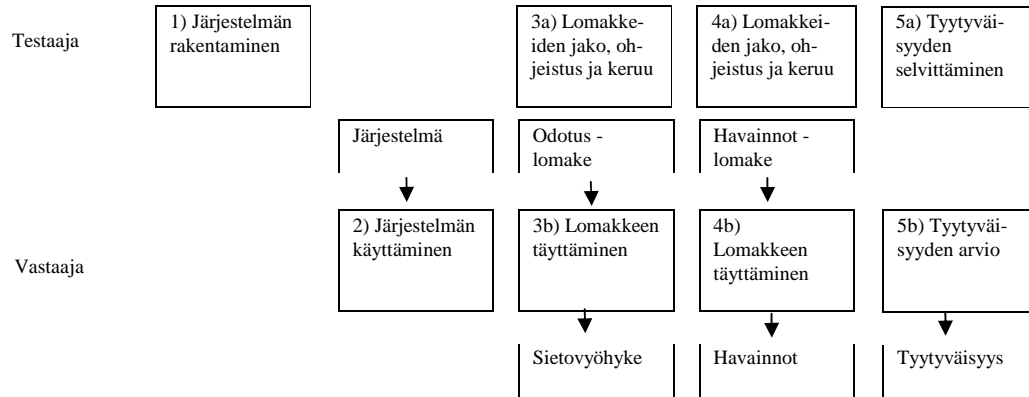
Kuva 6: Prosessikuvaus koekäytettävää järjestelmää arvioivasta SERVQUAL-metodista

Tämä tapa poikkeaa kuvausta arvioivasta metodista vain siltä osin, että vastaaja saa aidon tai siltä vaikuttavan järjestelmän kokeiltavakseen. Mikäli järjestelmän käyttämiseen tarvitaan opastusta, tämä tulee antaa vain siinä laajuudessa, joka on mahdollista myös todellisessa käyttötilanteessa. Mikäli koekäyttö tapahtuu joidenkin tilanteiden pohjalta tai se perustuu tehtäviin, jotka tulee suorittaa, on tilanteiden ja tehtävien suunnittelu yhtä tärkeää kuin kuvauksen suunnittelu kuvausta arvioivan metodin puitteissa. Esimerkiksi helpoista tehtävistä käyttäjä selviytyy helpommin ja saattaa projisoida onnistumisen tunteen järjestelmän ansioksi palvelun laatua arvioidessaan.

7.3. Vapaassa käytössä ollut järjestelmä arvioiva SERVQUAL-metodi

Vapaasti käytettävä järjestelmä ei tarkoita välttämättä samaa kuin valmis järjestelmä tai edes avoimesti saatavilla oleva järjestelmä. Tämä metodi voidaan toteuttaa myös prototyypeillä tai Wizard of Oz -tutkimuksen yhteydessä. Erona koekäytössä olleen järjestelmän arviointiin tässä metodissa järjestelmä annetaan vastaajille vapaaseen käyttöön. Mikäli järjestelmässä on huomattavia puutteita tai Wizard of Oz -tutkimuksen resurssit ovat rajalliset, voidaan järjestelmä antaa valitun joukon käyttöön. Tällöin erona koekäytössä olleen järjestelmän arviointiin on se, että järjestelmä annetaan vastaajien käyttöön heidän omilla ehdoillaan – vaikka he eivät sitten käyttäisi järjestelmää ollenkaan, on varmasti mielenkiintoista kuulla, mistä se johtui.

Mikäli järjestelmä on ollut vapaassa käytössä, ei sen arviointi poikkea muiden palveluiden arvioinnista. Tällöin odotukset täytyy kerätä käytön jälkeen, ellei ole mahdollisuutta tiedustella käyttäjien arvioita ennen järjestelmän lanseerausta. Odotuksia kerättäessä voidaan olla kertomatta tarkasteltavan järjestelmän nimeä, ainakin mikäli sillä on selkeästi saman alan kilpailijoita. Kuvassa 7 näkyy tämän metodin prosessikuvaus.



Kuva 7: Vapaassa käytössä ollutta järjestelmää arvioivan SERVQUAL-metodin prosessikuvaus

Kuvassa 7 ensimmäinen prosessi on järjestelmän rakentaminen. Järjestelmän rakentaja ei tietenkään ole välttämättä sama kuin testaaja, mutta heidän yhteistyönsä on arvion kannalta niin oleellinen, että voidaan ajatella vastaajan kannalta testaajan olevan järjestelmän rakentajan edustaja.

Luku 7 lyhyesti

- SERVQUAL-asteikkoa voidaan soveltaa puhepohjaisten keskustelujärjestelmien arviointiin kolmella tavalla järjestelmän valmiudesta riippuen.
 - Pyydetään käyttäjien arviot järjestelmän kuvauksesta.
 - Annetaan käyttäjien kokeilla prototyyppejä ja pyydetään arviot.
 - Järjestelmällä on oikeita käyttäjiä, joilta arviot pyydetään.
- Kahdessa ensiksi mainitussa tavassa odotukset voidaan kerätä ennen kokemuksia järjestelmästä, jolloin järjestelmä itse vaikuttaa mahdollisimman vähän vastaajien odotuksiin
- Viimeksi mainittu tapa antaa vastaajille parhaan kuvan järjestelmästä, sillä he ovat saaneet käyttää sitä omilla ehdoillaan

8. Interact-bussiaikataulujärjestelmän versioita vertaileva koe

Olin kesällä 2002 harjoittelijana Tampereen yliopiston tietojenkäsittelytieteiden laitoksen TAUCHI-yksikön puhekäyttöliittymiä tutkivassa SPI-ryhmässä (Speech-based and Pervasive Interaction Group). Toimenkuvani oli tehdä käytettävyydestejä USIX-Interact-projektissa rakennetulle Interact-bussiaikataulujärjestelmälle [Jokinen ym. 2002]. Muokkasin myös järjestelmän puhetulosteita, eli sitä, mitä järjestelmä käyttäjälle sanoo ja miten asiat sanotaan. Tulosteita hioessani aloin miettiä, mikä oikeastaan onkaan luontevin tapa puhua automaatille?

Projektissa oli kerätty myös materiaalia Helsingin kaupungin liikenneneuvonnasta. Materiaalia kuunnellessani kävi selväksi, että liikenneneuvoja puhuu puhekieltä. Heräsi kysymys: miksi tulosteet, joita suunnittelemme, ovat kaikki kirjakielisiä? Halusin tehdä vertailevan testin kirjakielisen ja puhekielisen automaatin välillä. Näin sain konkreettisen lähtökohdan luonnollisen kielen puhepohjaisten keskustelujärjestelmien tutkimiseen ja mielenkiintoisen, ennen tutkimattoman aiheen testeille.

Koska lähtökohtani oli vakiintuneen käytännön (kirjakieliset tulosteet) kyseenalaistaminen, oli mielestäni luontevaa kyseenalaistaa myös vallitsevat tutkimusmenetelmät. Ajatukseni oli, että vallitsevat tutkimusmenetelmät eivät ole tarkoitettu vakiintuneen käytännön arvioimiseen, sillä menetelmiä laadittaessa käytäntö on jo vakiintunut.

Uuden tutkimusmenetelmän tarpeellisuus kävi selväksi myös ohjaajan kanssa käydyissä keskusteluissa, sillä puhekäyttöliittymille ei ole kehitetty mitään laajalle levinnyttä käyttäjien omiin mielipiteisiin perustuvaa prototyyppiä arvioivaa menetelmää.

Etsin arviointimenetelmän, joka perustuu tuotenäkökulmaan, jossa arvioinnin perusteeksi riittää yleiskäsitys järjestelmästä ja joka mittaa käyttäjien mielipiteitä. Tämä menetelmä on SERVQUAL (luku 6). Käytin kuvausta arvioivaa SERVQUAL-metodia. Kuvaukset annoin äänitteinä, jotka olin äänittänyt vuorovaikutuksesta aidosti toimivan järjestelmän kanssa.

Tässä luvussa käsittelen järjestämäni koetta. Kerron ensin järjestelmästä ja sen kahden eri version eroista. Sitten tarkastelen testien järjestämistä käytännössä: kerron mukana olleista koehenkilöistä ja koejärjestelyistä.

8.1. Järjestelmän kuvaus

Interact-bussiaikataulujärjestelmä on rakennettu USIX-Interact -projektissa [Jokinen ym. 2002]. Projektissa tutkittiin puhepohjaisia luonnollisen kielen keskustelujärjestelmiä ja kehitettiin suomenkielisen puhepohjaisen keskustelujärjestelmän luomisen edellytyksiä.

Järjestelmän oli tarkoitus kertoa aikataulutietoja monipuolisesti luonnollisen kielen kyselyihin vastaten. Kaikkia suunniteltuja toiminnallisuuksia ei kuitenkaan toteutettu eikä luonnollisen kielen ymmärrys ollut tavoitellulla tasolla. Järjestelmä toimii kuitenkin

osittain. Se osaa kertoa aikataulutietoja käyttäjän esittämien rajoitetun luonnollisen kielen kyselyiden perusteella. Käytännössä järjestelmällä on seuraavat neljä toimintoa:

- Linjojen numeroiden luetteleminen lähtö- ja tulopaikan perusteella.
- Seuraavan lähtöajan kertominen lähtö- ja tulopaikan perusteella. Paikat voivat olla esitetty aiemmassa kyselyssä (linjojen numeroiden selvittäminen).
- Seuraavaksi kulkevan bussin saapumisajan kertominen. Lähtö- ja tulopaikka pitää olla järjestelmän tiedossa.
- Päätepusäkkien nimien kertominen.

Järjestelmän tuntemat paikat ovat Helsingin kaupunginosia. Rajoitettu luonnollinen kieli tarkoittaa sitä, että järjestelmän kieliopissa on muutamia vaihtoehtoisia kysymyslauseita saman asian selvittämiseen. Esimerkiksi linjojen numeroita keskustasta Arabiaan voi kysyä seuraavasti muodostetuilla lauseilla:

- "Miten pääsee/pääsisin/pääsemme/pääsisimme keskustasta Arabiaan?"
- "Mikä bussi ajaa/menee keskustasta Arabiaan?"
- "(Lähtöpaikka) keskusta. (Tulopaikka) Arabia."

Projektin tavoitteet olivat korkealla ja havaitsin lopullisessa bussiaikataulujärjestelmässä vielä monia virheitä. Testasin järjestelmää käyttäjillä kahteenkin otteeseen, ensin projektin puitteissa ja toisen kerran tämän työn suunnitteluvaiheessa erään kurssin (Sosiaalinen suunnittelu) puitteissa. Molemmassa testissä kävi ilmi, ettei järjestelmää voisi käyttää ilman harjoittelua. Koska harjoittelu vaikuttaa tuloksiin (toisille koekäyttäjille riittää, että yksi keskustelu menee onnistuneesti läpi, kun toiset eivät halua aloittaa koetta ennen kuin tietävät miten toimia kaikissa odotettavissa olevissa poikkeustilanteissa), päätin vertailla puhe- ja kirjakielisiä järjestelmiä SERVQUAL-asteikolla pelkkien nauhoitettujen esimerkkidialogien kuulemisen perusteella. Näin testitilanne ei muodostunut liian pitkäksi, pystyin testaamaan monta koehenkilöä kerralla, koehenkilöiden kokemukset järjestelmästä olivat samanlaiset ja pystyin kokeilemaan kuvausta arvioivaa SERVQUAL-metodia käytännössä.

Interact-bussiaikataulujärjestelmä käyttää puhetulosteiden antamiseen Mikropuheen suomenkielistä puhesynteesiä [Mikropuhe 2003]. Edellä mainituissa testeissä havaittiin, että synteesin puheääni ei miellytä kaikkia ja joillain on jopa vaikeuksia saada selvää kaikista sanoista.

8.2. Versioiden erot

Muokkasin järjestelmästä kaksi versiota. Järjestelmien ero oli puhetulosteiden sisällössä eli siinä, mitä sanoja järjestelmä käyttää asioiden sanomiseen. Toinen versio oli kirjakielinen. Kirjakielinen järjestelmä oli lähellä UNIX-Interact -projektin [Jokinen ym. 2002] puitteissa testattua versiota. Toinen versio oli puhekielinen. Puhekielisyys pyrittiin toteuttamaan kuitenkin hienovaraisesti mitään voimakkaita murteita käyttämättä.

Versioiden eroksi valittiin nimenomaan puhe- ja kirjakielisyys, sillä halusin tutkia inhimillistämisen vaikutusta siihen, miten järjestelmä koetaan. Aikatauluneuvonnan puhetyyli oli selvästi puhekielinen. Tietokoneet puolestaan käyttävät yleensä kirjakieltä. Niinpä tietokoneen puhetyylin vaihtaminen puhekieleen oli suoraviivainen tapa tukea inhimillistämistä. Myös ”Sosiaalinen suunnittelu” -kurssin puitteissa tekemissäni testeissä selvisi, että puhekielen ja kirjakielen sanojen ero huomataan paremmin kuin esimerkiksi persoonamuodon vaihtelu.

Koehenkilöt olivat Tampereelta ja järjestelmä Helsingistä. Kummankin kaupungin murteissa persoonapronominin ”sinä” puhekielinen vastike on ”sä”. Niinpä tilanteessa, jossa järjestelmän kirjakielinen versio sanoo ”Pääset keskustasta...”, puhekielinen versio sanoo ”Sä pääset keskustasta...”. Pienempi, vaikeammin huomattava ero on avausreplikkissä, jossa kirjakielinen versio sanoo ”...Kuinka voin auttaa?” ja puhekielinen versio sanoo ”...Kuin voin auttaa?”. Tyypillinen ero puhe- ja kirjakielelle on lukujen lausuminen. Tämä ero toteutuu näissä versioissa linjojen numeroita listattaessa. Esimerkiksi kun kirjakielinen versio sanoo ”...busseilla kuusikymmentäneljä, ...”, puhekielinen versio sanoo ”...kuusnelosella, ...”. Versioiden täsmällinen ero selviää esimerkkidialogeista, jotka ovat liitteenä (liite 3).

Suunnittelin dialogit erään järjestelmää testanneen koehenkilön puhettavan perusteella. Koehenkilön tyyli saada tietoa järjestelmältä oli mielestäni tavanomainen. Hänen esimerkkejään käytettiin dialogien pohjana lähinnä siksi, että käyttäjän puhekielen ilmaisut olisivat luontevia. Toisin sanoen, kun itse suunnittelee miten esimerkkikäyttäjä pyytäisi tietoa järjestelmältä, väkisinkin tiedostaa järjestelmän sanakirjan vajavuudet. Toisaalta ajatuksena oli lisätä esimerkkien aitoutta. Käyttäjä puhuu puhekieltä esimerkkidialogeissa.

Esimerkkidialogit nauhoitettiin vapaaehtoisen esimerkkikäyttäjän ja aidosti toimivan järjestelmän vuorovaikutuksesta. Esimerkkikäyttäjäksi pyydettiin nainen, sillä testaajana koehenkilöiden kanssa toimi mies – näin koehenkilöt tiesivät, että esimerkkikäyttäjä on joku muu kuin testaaja. Myös tässä ajatuksena oli lisätä esimerkkien aitoutta.

Dialogeja oli yhteensä neljä: kaksi tilannetta, joista molemmista puhe- ja kirjakielinen versio. Tilanteita ei tarvittu tätä enempää, sillä aikataulupalvelu ei sisällä monia toimintoja. Päätepyysäkkien listaamistoiminto jäi pois esimerkkitalanteista, sillä se vaikutti lisätoiminnolta joka ei ole keskeinen järjestelmän kannalta.

Vertailemani versiot poikkeavat toisistaan vain vähän. Ne käyttävät hiukan eri sanoja saman asian sanomiseen. Tällöin vertailussa ei olekaan perustelua mitata järjestelmän objektiivista suoriutumista, sillä se on identtinen. Versioiden ero on siinä, miten käyttäjä kokee järjestelmän. Käyttäjän kokemuksia voi mitata ainoastaan käyttäjältä kysymällä.

Esimerkkidialogien nauhoituksessa oli ongelma, jonka vuoksi toiseen tilanteeseen jäi virheellinen ilmoitus: Bussin kerrotaan olevan perillä yli kolme tuntia lähdön jälkeen. Kukaan ei kuitenkaan maininnut virhettä, enkä usko että sitä huomattiin, sillä koehenkilöitä pyydettiin nimenomaan olemaan kiinnittämättä huomiota yksityiskohtiin vaan muodostamaan mielikuva palvelun laadusta.

8.3. Koehenkilöt

Koehenkilöitä eli vastaajia oli yhteensä 11. Yhden vastaukset jätin huomioimatta, sillä hän arvioi ilmeisten väärinkäsitysten vuoksi vähimmäisvaatimukset korkeammalle kuin hyvän tason arviot. Molemmissa ryhmissä oli siis viisi vastaajaa – toisessa kolme naista, toisessa kolme miestä. Kaikki vastaajat olivat alle 35-vuotiaita, yksi oli alle 25-vuotias. Kaikki olivat Tampereen yliopistolla tutkijoina ja/tai opiskelijoina – pääaineita ei kysytty, mutta mukana oli ainakin tietojenkäsittelytiedettä, matematiikkaa, kauppatieteitä ja psykologiaa pääaineenaan lukevia/lukeneita.

8.4. Koejärjestely

Koe alkoi jo siinä vaiheessa, kun kutsuin vapaaehtoisia kokeeseen. Kutsuin koehenkilöitä kolmessa eri yhteydessä: Ensin lähetin sähköpostiviestin opiskelijajärjestöjen sähköpostilistoille, josta sain lopulta kaksi koehenkilöä. Eräältä markkinoinnin laitoksen luennolta sain lisää vapaaehtoisia ja loput vapaaehtoiset löytyivät tietojenkäsittelytieteiden laitokselta.

Kutsujen (paitsi sähköpostikutsun) yhteydessä annoin vapaaehtoisten täytettäväksi taustatietolomakkeen. Taustatietolomakkeen tarkoitus oli muodostaa ryhmistä mahdollisimman samankaltaisia. Lomakkeella kysyttiin koehenkilöiden kokeneisuutta teknologian, julkisen liikenteen ja neuvontapalvelujen käyttämisestä (katso liite 4). Kääntöpuolella oli kalenteri, johon vapaaehtoisia pyydettiin merkitsemään heille sopivat ajat.

Ryhmät kutsuin niin, että mahdollisimman moni pääsisi paikalle yhtä aikaa. Näin siksi, että lopun vapaassa kommentointivaiheessa saataisiin aikaan keskustelua – kuitenkin koehenkilöitä ei ollut liikaa. Kutsutun ryhmän koko vaihteli kahdesta viiteen, paikalle saapuneiden määrä vaihteli yhdestä neljään. Ennen seuraavaan vaiheeseen siirtymistä sähköpostilla kutsutut täyttivät taustatiedot.

Koehenkilöille jaettiin odotuslomakkeet (liite 1). Odotuslomakkeesta selitettiin ”palvelun” ja ”palvelun tarjoajan” ero, sarakkeiden merkitys, skaala ja heitä pyydettiin vastaamaan kaikkiin 22 kohtaan yleisesti puhelinpohjaisiin bussiaikataulupalveluihin

liittyvien odotusten perusteella. Yleensä sarakkeiden merkitys ei selvinnyt vain toteamalla ja kaikkiin koehenkilöiden kysymyksiin lomaketta koskien pyrittiin vastaamaan, niin että kaikki ymmärsivät lomakkeiden täyttämisen samalla tavalla.

Seuraavaksi kuunneltiin esimerkki-dialogit kannettavalta CD-soittimelta. Dialogeista kerrottiin etukäteen, että niissä esiintyy käyttäjä, joka haluaa aikataulutietoja ja järjestelmä, joka vastaa käyttäjän kysymyksiin. Dialogien määrä (kaksi) mainittiin ja koehenkilöitä pyydettiin olemaan kiinnittämättä huomiota yksityiskohtiin. Kerrottiin, ettei yksityiskohtien muistamista kysytä, vaan tarkoitus on muodostaa käsitys automaatin tarjoaman palvelun laadusta.

Kun esimerkki-dialogit päättyivät, koehenkilöt saivat täytettäväkseen havaintolomakkeen (liite 2). Heille kerrottiin, ettei kaikkiin kohtiin voi vastata annettujen esimerkkien perusteella, joten he voivat jättää ne kohdat tyhjiksi. Esimerkkinä näistä kohdista mainittiin kohta 4, jossa kysyttiin palveluun liittyvän materiaalin visuaalista miellyttävyyttä.

Kunkin koehenkilön täytettyä havaintolomakkeen, hänelle annettiin jälkikyselylomake (liite 5). Jälkikyselylomakkeen tarkoitus oli selvittää, kuinka hyvin SERVQUAL-mittaus sopii yhteen tyytyväisyyden kanssa. Yleisarvosanan lisäksi selvitettiin, kuinka valmiita koehenkilöt olisivat käyttämään järjestelmää oikeasti.

Kaikkien täytettyä jälkikyselylomakkeen annettiin mahdollisuus suulliseen kommentointiin ja keskusteluun. Kommentteja oli mahdollista myös kirjoittaa, mikäli vastasi muita nopeammin. Kommenttien kirjoittaminen sallittiin, vaikka keskustelu olisikin ollut jo käynnissä, sillä joillekin on luontevampaa antaa kommentit yksityisesti.

Luku 8 lyhyesti

- Interact-bussiaikatauluautomaatti on kielellistä vuorovaikutusta tutkivassa projektissa rakennettu prototyyppi. Järjestelmä osaa kertoa linjojen numerot kaupunginosien perusteella ja lähtö- ja tuloajat linjan perusteella.
- Interact-bussiaikatauluautomaatista muokattiin puhe- ja kirjakieliset versiot ja nauhoitettiin esimerkkidialogit. Näitä esimerkkidialogeja soitettiin kahdelle viiden koehenkilön vastaajaryhmälle.
- Testissä tutkittiin versioiden eroa, SERVQUAL-asteikon käyttöä puhepohjaisten keskustelujärjestelmien kuvausten arvioinnissa sekä yleistä tyytyväisyyttä Interact-bussiaikatauluautomaattiin kuvausten perusteella.
- Testi järjestettiin kuvausta arvioivan SERVQUAL-metodin mukaisesti.

9. Tulokset ja keskustelu

Tutkin kahta eri asiaa samalla kokeella: mitä eroa on puhe- ja kirjakielisiä tulosteita käytävällä puhepohjaisella keskustelujärjestelmällä (sekä palvelun laadun että yleisen tyytyväisyyden suhteen) ja kuinka SERVQUAL-asteikkoa voidaan soveltaa näiden tulosten saamiseksi. Tulosten esittely jakaantuu näihin kahteen osa-alueeseen.

Tässä luvussa kerrotaan kokeen tuloksista. Ensin vertaillaan testattavina olleita versioita tulosten perusteella. Erot eivät olleet suuria, joten puhekielisyys ei ainakaan paranna palvelun laatua. Sitten kerrotaan SERVQUAL-asteikon soveltuvuudesta tähän vertailutehtävään – metodissa löytyi monta parannettavaa kohtaa, joten osuus sisältää kritiikkiä. Lopuksi kerrotaan kokeenjärjestäjän näkökulmasta yleisiä havaintoja SERVQUAL-asteikon sovittamisesta kyseiseen vertailutehtävään.

9.1. Versioiden arviointi ja vertailu

Vähäisen koehenkilömäärän takia varsinaisia tilastollisesti merkittäviä tuloksia ei ollut odotettavissa. Versioiden arviointi on siis suuntaa-antava. Kehitystyössä tämä onkin riittävä tarkkuus – saadaan selville, kumpi versio on parempi, missä ongelmat ovat ja mitkä tekijät tekee järjestelmästä miellyttävän (ja miellyttävämmän).

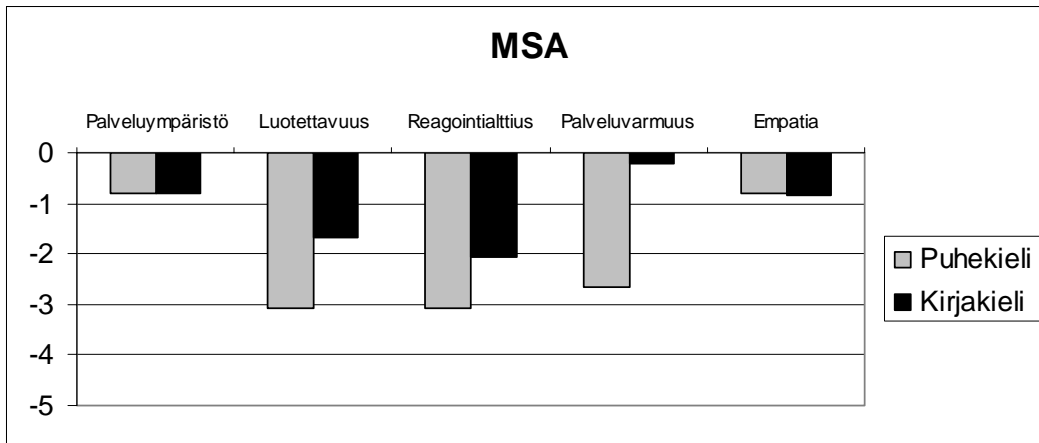
Tulosten analysoinnissa jouduttiin jättämään pois joitain SERVQUAL-asteikon alkioita. Kokeessa ei annettu kaikkea palveluun liittyvää materiaalia (kuten aukioloaikoja). Tämän takia viiteen alkioon (4, 5, 8, 13 ja 19) ei voinut antaa arviota havainnoista. Myös kohtiin 3 ja 17 jättivät jotkut vastaamatta. Kaikki nämä seitsemän alkioita jätettiin kokonaan tulosten analysoinnista pois eli analysoitiin vain niitä alkioita joihin kaikki vastasivat. Näin ollen ulottuvuuteen *Palveluympäristö* jäi kaksi alkioita, ulottuvuuksiin *Luotettavuus*, *Reagointialttius* ja *Palveluvarmuus* jäi kolme alkioita ja ulottuvuuteen *Empatia* jäi neljä alkioita.

Kerron ensin tuloksista MSA- ja MSS-mittareilla analysoituna. Seuraavaksi esitän tulokset tärkeys-laatu –taulukossa. Sitten kerron vapaamuotoisista kommentteista. Testien tuloksia analysoitaessa kävikin selväksi, että pelkät numeeriset arviot kertovat vain mielipiteistä eivätkä mielipiteiden syistä. Kerron lopuksi vielä suuntaviivoja muista kuin SERVQUAL-asteikon tuloksista: jälkikyselyn vastauksista ja koehenkilöiden antamista kommentteista.

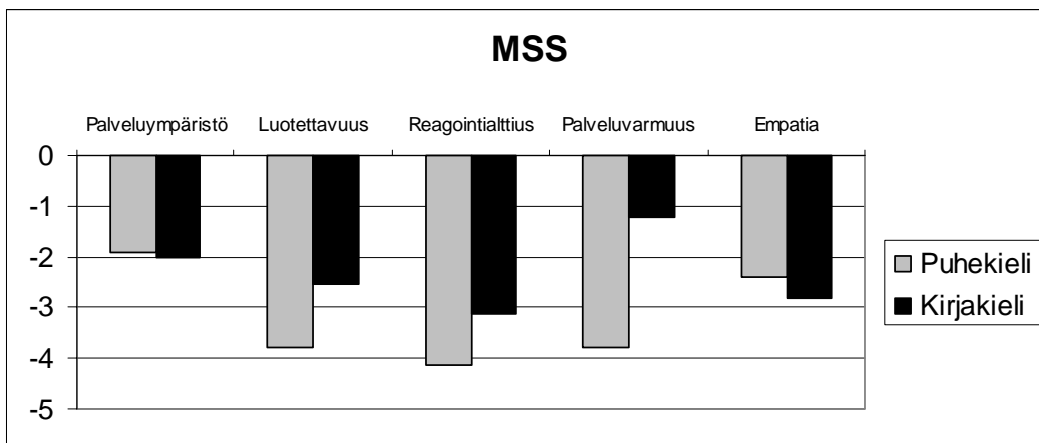
9.1.1. MSA- ja MSS-mittarit

Kuvissa 8 ja 9 näkyy MSA- ja MSS-mittareiden (esitelty osiossa 6.3.1) tulokset. Tulokset on esitetty alaspäin suuntautuvina palkkeina, sillä yleensä tuloksina saadaan sekä positiivisia että negatiivisia arvoja. Tällä kertaa saatiin vain negatiivisia arvoja, sillä havainnot-sarakkeen kaikkien vastaajien keskiarvot olivat vähimmäisvaatimukset-sarakkeen

keskiarvoja alempana kaikilla palvelun laadun ulottuvuuksilla. Mahdolliset arvot ovat vaihteluvälillä -6...6, kun arviointiskaalan vaihteluväli oli 1...7.



Kuva 8: MSA-mittauksen arvot



Kuva 9: MSS-mittauksen arvot

Tärkein havainto näistä mittauksista on se, että molemmat versiot koettiin riittämättömiksi palvelun laadultaan. Havaintojen keskiarvo ei yltänyt vähimmäisvaatimusten keskiarvon tasolle millään ulottuvuudella. Lähimpänä on *Palveluvarmuus*-ulottuvuuden havaintojen keskiarvo kirjakieltä käyttäneellä järjestelmällä, joka on vain 0,2 yksikköä vähimmäisvaatimusten alapuolella. Tästä voidaan päätellä, ettei kirjakielen ja puhekielen ero ole tämän järjestelmän kannalta merkittävä, sillä joka tapauksessa palvelun laatu koettiin riittämättömäksi.

Joitain eroja on kuitenkin havaittavissa. Puhekielinen järjestelmä koettiin vähemmän riittämättömäksi ulottuvuuksilla *Empatia* ja *Palveluympäristö*. Suurempi ero oli kirjakielen hyväksi ulottuvuuksilla *Luotettavuus*, *Reagointialttius* ja etenkin *Palveluvarmuus*. *Palveluvarmuus*-ulottuvuudella olikin ainoa tilastollisesti merkittävä ero MSA- ($F_{1,28} = 13,1$; $p < ,005$) ja MSS-mittareilla ($F_{1,28} = 20,6$; $p < ,0001$). Tämä ero johtui kuitenkin odotusten

eroista (vähimmäisvaatimusten ero $F_{1,28} = 24,9$; $p < ,0001$ ja hyvän tason ero $F_{1,28} = 32,3$; $p < ,0001$) eikä suinkaan havaintojen erosta ($F_{1,28} = 1,8$; $p > ,05$). Näyttää siltä, että palvelun laadun odotukset ovat henkilökohtaisia ja ryhmien tasapuolinen muodostus oli tältä kannalta epäonnistunut.

9.1.2. Vertailu tärkeys-laatu -asteikoilla

Kuvassa 10 näkyy, kuinka vastaajien arviot järjestelmästä asettuvat tärkeys-laatu -asteikolle laadun ulottuvuuksien suhteen osiossa 6.3.2 esitellyn lajittelun mukaan. Ruutujen vasemmalle puolelle sijoitetut luvut ovat puhekielisiä esimerkkejä kuulleen ryhmän lukumääriä ja oikealle puolelle sijoitetut luvut ovat kirjakielisiä esimerkkejä kuulleen ryhmän lukumääriä. Seitsemän alkia jätettiin kokonaan tulosten analysoinnista pois, jolloin ulottuvuuteen *Palveluympäristö* jäi kaksi alkia, ulottuvuuksiin *Luotettavuus*, *Reagointialttius* ja *Palveluvarmuus* jäi kolme alkia ja ulottuvuuteen *Empatia* jäi neljä alkia. Kun vastaajia oli kummassakin ryhmässä viisi, yhteensä ulottuvuuksilla on (edellä mainitussa järjestyksessä) 10, 15, 15, 15 ja 20 arviota ryhmää kohti. Taulukot esitetään erikseen kustakin viidestä palvelun laadun ulottuvuudesta. Taulukot voitaisiin toki yhdistää tai esittää jokaisesta 22 kohdasta erikseen, mutta yhdistäminen hukkaisi ulottuvuuksien väliset erot ja erottaminen taas on hankalaa näin pienellä koehenkilömäärällä. Painotusten visualisoimiseksi vastaajien lukumäärät on kuvattu numeroiden lisäksi mustilla pisteillä ja kohdissa, joissa oli suurin absoluuttinen vaihtelu, on tummennettu tausta.

Palveluympäristö (yht. 10) Puhekieli - Kirjakieli	Alhainen laatu	Riittävä laatu	Korkea laatu
Alhainen tärkeys	2 – 4	3 – 2	1 – 0
Korkea tärkeys	3 – 3	1 – 1	0 – 0

Luotettavuus (yht. 15) Puhekieli - Kirjakieli	Alhainen laatu	Riittävä laatu	Korkea laatu
Alhainen tärkeys	0 – 0	0 – 1	0 – 0
Korkea tärkeys	13 – 11	2 – 3	0 – 0

Reagointialttius (yht. 15) Puhekieli - Kirjakieli	Alhainen laatu	Riittävä laatu	Korkea laatu
Alhainen tärkeys	1 – 3	0 – 1	0 – 0
Korkea tärkeys	13 – 9	0 – 2	1 – 0

Palveluvarmuus (yht. 15) Puhekieli - Kirjakieli	Alhainen laatu	Riittävä laatu	Korkea laatu
Alhainen tärkeys	0 – 3	0 – 2	0 – 1
Korkea tärkeys	14 – 5	1 – 3	0 – 1

Empatia (yht. 20) Puhekieli - Kirjakieli	Alhainen laatu	Riittävä laatu	Korkea laatu
Alhainen tärkeys	0 – 2	3 – 3	1 – 0
Korkea tärkeys	11 – 10	5 – 4	0 – 1

Kuva 10: Versioiden vertailu tärkeys-laatu -taulukossa

Kuvasta 10 näkyy, ettei versioissa ollut suuria eroja. Ulottuvuudet eivät ole absoluuttisesti vertailukelpoisia, mutta pienet erot ovat kuitenkin vähemmän luotettavia. Onkin perusteltua tarkastella vain ruutuja, joissa on huomattavia eroja. Vain kahdessa ruudussa 30:sta vastaajien lukumäärä on toiselle versiolle yli 3 enemmän. Ulottuvuudesta

riippumatta ainoastaan näissä ruuduissa suhteellinenkin ero oli yli 20 prosenttiyksikköä. Nämä kaksi ruutua ovat *Reagointialttius*-ulottuvuuden ruutu ”Alhainen laatu – Korkea tärkeys” ja ennen kaikkea *Palveluvarmuus*-ulottuvuuden ruutu ”Alhainen laatu – Korkea tärkeys”.

Ensimmäisessä näistä on neljän vastaajan ero, mutta se, että kirjakielisiä esimerkkejä kuunnelleet pitivät tätä ulottuvuutta vähemmän tärkeänä, vaikuttaa satunnaiselta. Vastaajia oli kuitenkin hyvin vähän. Palveluvarmuus-ulottuvuudella oleva ero vaikuttavaa vähemmän satunnaiselta edellisen kappaleenkin merkittävyydestulosten valossa. Kirjakielisiä esimerkkejä kuunnelleista kuusi piti tämän ulottuvuuden tärkeyttä alhaisena, kun taas puhekielisiä esimerkkejä kuunnelleista kaikki pitivät tärkeyttä korkeana. Kun toisella ryhmällä oli odotuksetkin alhaalla, on luonnollista, että siinä ryhmässä on myös enemmän tyytyväisiä.

Reeves & Nass ovat eräässä tutkimuksessaan [1996, 75-88] havainneet, että käyttäjät suosivat järjestelmää, joka puhuu samalla tyyllillä kuin he itse. Esimerkkidialogien käyttäjä puhui puhekieltä molempien versioiden kanssa, joten puhekielisen version suosiminen olisi tämän perusteella ollut loogista. Kun puhekielinen versio ei ollut selvästi suositumpi, voidaan tästä päätellä ettei puhekielisyys ainakaan paranna arvioita järjestelmästä.

9.1.3. Vapaamuotoiset kommentit

Vapaamuotoisia kommentteja annettiin sekä kirjallisena että suullisena. Vapaamuotoiset kommentit ovat mielenkiintoisia, sillä niiden antaminen on täysin oma-aloitteista. Näin ollen vapaamuotoiset kommentit kärjistetyksi osoittavat, mitkä kohdat palvelusta jäivät parhaiten vastaajien mieleen.

Kommenttien äänensävy oli yleensä positiivinen. Mahdollisesti vastaajat pitivät testaajaa järjestelmän rakentajana ja olivat siksi kohteliaita. Monet kehuivatkin ensin sanoen, että ohjelma on mukava ja idea on hyvä, mutta luettelivat sitten virheitä.

Selvästi yleisin kommentti koski järjestelmän hitautta eli lähinnä pitkiä prosessointiaikoja, jolloin vuorovaikutukseen tuli tauko. Järjestelmän hakiessa aikatauluja pisimmillään hiljaisuus kesti jopa yli 20 sekuntia. Järjestelmän versioista haluttiinkin arvioita sen todellisen toiminnan puitteissa. Toisaalta kognitiivinen kuormitus oli pienempi esimerkkejä kuunneltaessa joten tauot saattoivat tuntua kärjistetyinkin pitkiltä.

Osa ehdotti parannustoimenpiteitä (eli lähinnä prosessointiääniä, verraten henkilökohtaiseen palveluun jossa kuulee kuinka palveluneuvoja selailee papereita) ja osa totesi suoraan, että järjestelmällä kesti kauan hakea tietoja. Joku sanoikin, ettei olisi itse odottanut niin kauaa, vaan sulkenut puhelimen luullen että järjestelmä on mennyt epäkuuntoon. Havaitsin, että puhetyylillä oli vastaajille vähän merkitystä verrattuna prosessointitaukojen keston. Mikäli olisin muokannut esimerkkejä lyhentämällä prosessointitaukoja, en olisi voinut havaita tätä.

Muita mainittuja ongelmia oli puhesynteesin epäselvä puhetyyli ja vain kysymyksen ensimmäisen osan käsitteleminen. Järjestelmä käyttää Mikropuhe-synteesiä, joka on suomenkielinen puhesynteesi [Mikropuhe 2003]. Jo ennen kokeita oli tiedossa, ettei Mikropuheen kuunteleminen ole helppoa ellei siihen ole tottunut. Myös moniosaisien kysymysten käsittelemättömyys oli tiedossa ennen kokeita. Tämä ongelma tuli ehkä liiankin kärjistetyksi esille, sillä esimerkkidialogit oli suunniteltu siten, että käyttäjä kysyy toistuvasti moniosaisia kysymyksiä. Moniosaisien kysymysten käyttö perustui järjestelmää kokeilleen käyttäjän puhetyyliin. Tällä pyrittiin todenmukaisiin esimerkkidialogeihin. Kaikki järjestelmää kokeilleet eivät käyttäneet moniosaisia kysymyksiä, joten siinä mielessä esimerkkidialogit olivat todenmukaisia, mutta yksipuolisia.

Versioiden vertailun kannalta oli mielenkiintoista kuulla, kommentoivatko vastaajat järjestelmän puhetyyliä. Kirjakielisen järjestelmän esimerkkejä kuulleista kukaan ei kommentoinut puhetyyliä. Yhdessä puhekielisen järjestelmän esimerkkejä kuulleessa ryhmässä keskusteltiin hieman puhetyylistä. Eräs vastaaja oli sitä mieltä, että tietyt puhekieliset ilmaisut kuten sinuttelu käyttäen sanaa ”sä” ovat epäkohteliaita. Muut vastaajat olivat joksinkin samaa mieltä. Tämä onkin tärkeä huomioda inhimillistämisen tukemista suunniteltaessa, sillä inhimillistämistä ei pitäisi tukea muiden ominaisuuksien kustannuksella. Puhekieltäkin on monenlaista, eikä ”sä”-sinuttelun käyttäminen ole kohtelias puhetyyli.

9.1.4. Yleinen tyytyväisyys

Vaikka juuri kukaan ei pitänyt bussiaikatauluautomaattia hyvänä, kaikki olivat valmiita käyttämään sitä, mikäli se olisi ilmainen. Molemmissa ryhmissä yleisarvosanan keskiarvo oli 2.6 (asteikolla 1...7). Silti vain yksi koehenkilö vastasi, ettei voisi kuvitella käyttävänsä esimerkin palveluautomaattia. Yksikään koehenkilö ei vastannut, ettei soittaisi automaatille, jos se olisi ilmaista. Jotkut olivat jopa vastanneet, että voisivat maksaa kyseisestä palvelusta paikallispuhelun hinnan. Tämä kuvaa hyvin sitä, kuinka tyytyväisyyteen vaikuttavat muutkin tekijät kuin palvelun laatu. Mikäli palvelu on ilmainen mutta hyödyllinen, huonokin laatu voi riittää käyttäjälle.

Asiaan saattaa vaikuttaa sekin, että puhekäyttöliittymät ovat harvinaisia Suomessa, joten sitä halutaan kokeilla. Palvelua voidaan kokeilla, vaikka oltaisi varmoja, onko se nopeampi tapa selvittää aikataulut kuin aikataulukirja, jos palvelu on muuten kiinnostava. Tällaisessa tapauksessa on tietysti tärkeää, kuinka palvelu toimii ensi yrittämällä.

Puhepohjaisten keskustelujärjestelmien arvioinnin kannalta SERVQUAL-tutkimusta yhteydessä yleisen tyytyväisyyden selvittäminen tuo lisäarvoa tutkimukseen. Yleisen tyytyväisyyden selvittäessä saa käsityksen siitä, kuinka rajuja muutoksia järjestelmään tarvitsee tehdä, tai onko järjestelmälle nähtävissä käyttöä. Yleisen tyytyväisyyden selvittäminen ei kuitenkaan korvaa SERVQUAL-tutkimusta, sillä siitä ei selviä, mitä

palvelun laadun ulottuvuuksia käyttäjät pitävät tärkeinä ja mitä ulottuvuuksia he pitävät hyvin toteutettuina.

9.2. Metodin kritisointi

Tuloksia tarkastellessa metodin suurin ongelma on se, että ryhmillä oli erilaiset odotukset, vaikka odotukset mitattiin ennen esimerkkien soittamista. Ryhmistä pyrittiin muodostamaan samanlaiset ennakkokyselylomakkeen avulla. Lomakkeen kysymykset oli valittu siten, että ne kattaisivat oleelliset kysymykset bussiaikataulupalvelun käyttöön liittyen: teknologian käyttötottumukset, joukkoliikenteen käyttötottumukset ja aikataulupalvelujen käyttötottumukset.

Lähtökohta oli selvästi väärä. Vaikka valitut kysymykset ovatkin oleellisia palvelun käyttövalmiutta ajatellen, ei se tarkoita, että niiden ja palvelun laadun odotusten välillä olisi minkään suuntaista korrelaatiota. Kun odotukset on mahdollista kerätä ennen ryhmien muodostamista, olisikin parempi vaihtoehto muodostaa ryhmät palvelun laadun odotusten perusteella, niin että odotukset ovat mahdollisimman yhtäpitävät kaikissa ryhmissä, esimerkiksi niin, ettei kummassakaan odotussarakkeessa ole tilastollisesti merkittäviä eroja ryhmien välillä. Toinen ilmeinen vaihtoehto ryhmien välisen eron kokonaan poistamiseksi on ryhmien poistaminen, eli version vertailu ryhmän sisäisesti. Tällainen testiasettelu kuitenkin korostaa versioiden välisiä eroja – näin kävi ainakin edellisessä luvussa mainitsemissani ”Sosiaalinen suunnittelu” -kurssin testeissä.

Toinen ongelma oli siinä, että joitain SERVQUAL-asteikon alkioita jouduttiin jättämään pois. Viiteen alkioon (4, 5, 8, 13 ja 19) ei voinut antaa arviota havainnoista, sillä niihin liittyvää materiaalia ei tarjottu kokeessa. Jotkut jättivät vastaamatta myös kohtiin 3 ja 17 – tämä olisi voitu luultavasti välttää, jos edellä mainitut viisi alkiota olisi lueteltu koehenkilöille.

Kaikkiin alkioihin voisi vastata, mikäli palvelusta annettaisiin hieman enemmän tietoa kuin mitä esimerkkidialogeista käy selville. Lisäksi tarvittavat tiedot ovat jotain visuaalista lisämateriaalia, miten nopeaksi palvelua kehuaan (on luvattu), kauanko palvelua joutuu odottamaan ja mihin aikoihin palvelu on saatavilla.

Lisätiedot voisi antaa esimerkiksi laatimalla palvelusta esittelylipukkeen tai mainoslehtisen, jossa kävisi ilmi nopeuden lupaus ja aukioloajat. Odotusajan voisi ilmoittaa implisiittisesti sisällyttämällä esimerkkidialogien alkuun odotusääniä. Lisätiedot voi antaa tietysti myös muilla tavoilla. Ongelmalliset alkiot voi tarvittaessa jättää pois kyselylomakkeilta, mutta niiden pois jättäminen heikentää menetelmän vahvaa tuotenäkökulmaa.

Ylipäätään SERVQUAL-asteikko osoittautui käsitteellisesti hankalaksi. Odotusten ja havaintojen ero ei ollut kaikille itsestään selvä ja varsinkin vähimmäisvaatimuksen ja hyvän tason ero vaati tarkempaa selittämistä kuin pelkkä asian mainitseminen. Mielestäni kolmen sarakkeen SERVQUAL-version käyttö on kuitenkin perusteltua, sillä se antaa

analyttisempiä tuloksia kuin kahden tai yhden sarakkeen versiot. Asteikko on kuitenkin ymmärrettävä, kyse on vain ohjeiden laatimisen tärkeydestä. Näissä testeissä ei ollut ohjeita kirjoitettuna lomakkeelle eikä täyttämistä annettu esimerkkiä; molemmat ovat tapoja joilla asteikko olisi muuttunut ymmärrettävämmäksi. Tässä testissä asteikon ymmärsi kuitenkin kymmenen vastaajaa yhdestätoista pelkällä kysymyksiin vastaamisella, joten asia on tärkeä huomioida, muttei kriittinen.

Luku 9 lyhyesti

- Vastaajat kokivat molempia versioita arvioineissa ryhmissä järjestelmän palvelun laadun alhaiseksi eli riittämättömäksi.
- Versioissa ei ollut suuria eroja. Puhekielinen versio oli vain hieman parempi kahdella ulottuvuudella. Kirjakielinen versio oli parempi kolmella muulla ulottuvuudella. Tähän vaikutti myös ryhmien epätasapainoiset odotukset, mutta suullisten kommenttien perusteella puhekielisyys ei ainakaan paranna palvelun laatua. Versioiden erot eivät olleet tilastollisesti merkittäviä.
- Vastaajat olivat valmiita käyttämään järjestelmää, vaikka se olisi huono. Ainakin ilmaisena palvelulle olisi testien perusteella kysyntää.
- Odotusten ja havaintojen eroja voisi selittää vielä tarkemmin. Vähimmäisvaatimukset ja hyvä taso eivät ole käsitteenä yksiselitteisiä. Vastaajille voisi myös selittää, miksi odotuksia kerätään erikseen.
- Joitain alkioita jouduttiin jättämään pois analyysistä. Alkiot voi joko jättää pois kokonaan tai luoda testissä puitteet kaikkiin alkioihin vastaamiseksi.

10. Yhteenveto

Tässä työssä tarkasteltiin puhepohjaisia keskustelujärjestelmiä, jotka ovat yleistymässä. Puheen käyttäminen ihmisen ja tietokoneen vuorovaikutuksessa on perusteltu hyödyllä ja luontevuudella. Teknologian kehittyessä erityisesti puheentunnistuksen osalta tulevaisuudessa puhepohjaiset keskustelujärjestelmät voivat olla hyvinkin arkipäiväisiä, mutta teknologisen kehityksen lisäksi käyttöliittymäsuunnittelussa tarvitaan uudenlaisia näkökulmia, ennen kuin tämä on mahdollista. Yleispätevät käyttöliittymäperiaatteet vastaavat tähän tarpeeseen osittain – yksinkertaisia tehtäviä suorittavia koneita voidaan ohjata yksinkertaisella keskustelulla, mikäli yleispätevät käyttöliittymäperiaatteet vähentävät oppimisen rasitusta. Tämän lisäksi etenkin älykkäiden, inhimillisten puhepohjaisten keskustelujärjestelmien kohdalla tarvitaan uusia arviointimenetelmiä.

Inhimillisten puhepohjaisten keskustelujärjestelmien etu on luontevuus. Ajatuksena on, ettei käyttäjän tarvitse opetella järjestelmän käyttämistä, sillä inhimillistämistä hyväksikäyttäen käyttäjä ohjataan vaistonvaraiseen vuorovaikutukseen. Ongelmana inhimillistämässä on lähinnä se, että vaistonvaraisen vuorovaikutuksen käsittely ei ole mahdollista nykyisillä teknologioilla. Puhepohjaisten keskustelujärjestelmien yleistyessä tiettyjen järjestelmien kannattaa olla inhimillisiä ja puhua luontevaa kieltä. Kielen luontevuus on subjektiivinen käsite, joten sitä tulee tutkia subjektiivisilla menetelmillä.

Mikäli järjestelmän kanssa käytävä vuorovaikutus tapahtuu vaistojen avulla (eli käyttäjän ei tarvitse tietoisesti pitää mielessään koneen rajoituksia), on käyttöliittymän arviointikin tapahduttava eri tasolla. Mikäli järjestelmää voidaan pitää palveluna, inhimillistämisen onnistumista ja järjestelmän tehtävään sopivuutta pitää tutkia palveluna ihmisten tuottaman palvelun tavoin. Lisäksi vaikei inhimillisiä puhepohjaisia keskustelujärjestelmiä olisi vielä toteutettu, voidaan puhepohjaisten keskustelujärjestelmien uusista arviointimenetelmistä hyötyä kun saadaan uusia näkökulmia arvioihin.

Palveluita tutkitaan subjektiivisilla mittareilla. Palvelun laatua mitataan odotusten ja havaintojen erolla. SERVQUAL-asteikko perustuu palvelun laadun mittaamiseen viidellä palvelun laadun ulottuvuudella, jotka ovat kaikille palveluille yhteiset.

Sovelsin tässä työssä SERVQUAL-asteikon puhepohjaisten keskustelujärjestelmien arviointiin. Huomioin sovellustyössä myös asteikkoa kohtaan osoitetun kritiikin, vaikei kaikkeen kritiikkiin voidakaan vastata. Laadin SERVQUAL-asteikon soveltamisesta menetit kolmelle eri tavalle arvioida puhepohjaista keskustelujärjestelmää, jotka esitin myös prosessikuvauksena. Valittava metodi riippuu järjestelmän valmiudesta.

Kokeilin soveltamaani SERVQUAL-asteikkoa puhepohjaisen keskustelujärjestelmän inhimillistämisen arviointiin. Kymmenen koehenkilöä arvioi bussiaikataulujärjestelmää ääninäytteiden perusteella. Puolet kuuli esimerkkejä, joissa järjestelmä käyttää puhekieltä

ja puolet esimerkkejä, joissa järjestelmä käyttää kirjakieltä. Vertailussa ei löytynyt suuria eroja – ilmeisesti puhekielisyys ei ainakaan yksinään paranna puhepohjaisen keskustelujärjestelmän palvelun laatua, ainakaan kun palvelu ei ole teknisesti valmis.

Subjektiiivisiä puhepohjaisten keskustelujärjestelmien arviointeja täytyy kehittää edelleen. SERVQUALin soveltamista prototyyppien tutkimiseen tulee hioa, lähinnä parantamalla ohjeistusta ja antamalla enemmän tietoa palvelun kontekstista. Toisaalta arviointimenetelmää tulee suunnata ottamaan huomioon myös muut tyytyväisyyteen vaikuttavat tekijät kuin palvelun laatu. Tätä työtä voidaan käyttää lähtökohtana kattavan subjektiivisen puhepohjaisten keskustelujärjestelmien arviointimenetelmän kehittämiseksi.

Viitteet

- [AVIOS 2003] AVIOS, Applied Voice Input / Output Society verkkosivut, <http://www.avios.com>, haettu 28.4.2003.
- [Babakus & Mangold 1992] Babakus, E. and Mangold, W.G., Adapting the SERVQUAL scale to hospital services: an empirical investigation, *Health Services Research*, Vol. 26 No. 2, February 1992.
- [Baker & Lamb 1993] Baker, J.A. and Lamb, C.W. Jr, Managing architectural design service quality, *Journal of Professional Services Marketing*, Vol. 10 No. 1.
- [Balanced Scorecard Institute 2003] Balanced Scorecard Institute, Checklist for Internet Web Page Quality verkkosivut, <http://www.balancedscorecard.org/bscit/validations/public.htm>, haettu 22.5.2003.
- [Barnum 2002] Barnum, Carol M., *Usability Testing and Research*, Longman, 2002.
- [Berry & Parasuraman 1993] Berry, Leonard L. & Parasuraman, A., Building a new academic field – the case of services marketing, *Journal of Retailing*, Vol. 69, n. 1, Spring 1993.
- [Blasband 1998] Blasband, Marc, The ARISE Project verkkoartikkeli, <http://www.compuleer.nl/Nimes.html>, haettu 28.4.2003.
- [Bohus 2003] Bohus, Dan, spoken dialog dialogue systems links verkkosivut, <http://www-2.cs.cmu.edu/~dbohus/SDS/>, haettu 28.4.2003.
- [Bouman & van der Wiele 1992] Bouman, M. & van der Wiele, T., Measuring service quality in the car service industry: building and testing an instrument, *International Journal of Service Industry Management*, Vol. 3 No. 4 1992.
- [Boyce 2000] Boyce, Susan J., Natural spoken dialogue systems for telephony applications, *Comm. ACM* Vol. 43, No. 9, September 2000.
- [Brown ym. 1993] Brown, Tom J., Churchill, Gilbert A. Jr. & Peter, J. Paul, Research note: Improving the measurement of service quality, *Journal of Retailing*, Vol. 69, n. 1, Spring 1993.
- [Buttle 1996] Buttle, Francis, SERVQUAL: review, critique, research agenda, *European Journal of Marketing*, Vol. 30, No. 1, 1996.
- [Carman 1990] Carman, J.M., Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions, *Journal of Retailing*, Vol. 66 No. 1, Spring 1990.
- [Microsoft 2003] Microsoft Clippy verkkosivut, <http://www.microsoft.com/office/clippy/>, haettu 4.5.2003.
- [Cox ym. 2000] Cox, Richard V., Kamm, Candace A., Rabiner, Lawrence R., Schroeter, Juergen & Wilpon, Jay G., Speech and Language Processing for Next-Millennium Communications Services, *Proceedings of the IEEE*, Vol. 88, No. 8, August 2000.

- [Cronin & Taylor 1992] Cronin, Joseph J. Jr. & Taylor, Steven A., Measuring Service Quality: A Re-examination and Extension, *Journal of Marketing*, Vol. 56, No. 3, July 1992.
- [Cronin & Taylor 1994] Cronin, Joseph J. Jr. & Taylor, Steven A., SERVPERF versus SERVQUAL: Reconciling Performance-Based and Perceptions-Minus-Expectations Measurement of Service Quality, *Journal of Marketing*, Vol. 58, No. 1, January 1994.
- [DARPA Communicator 2003] MITRE, DARPA Communicator verkkosivut, <http://fofoca.mitre.org/>, haettu 25.4.2003.
- [Dicks 2002] Dicks, R.S. Mis-usability: On the uses and misuses of usability testing, *Proceedings of the 20th annual international conference on Computer documentation*, ACM 2002.
- [Doran ym. 2001] Doran, Christine, Aberdeen, John, Damianos, Laurie & Hirschman, Lynette, Comparing Several Aspects of Human-Computer and Human-Human Dialogues, *Proceedings of 2nd SIGdial Workshop on Discourse and Dialogue*, 2001.
- [Dumas & Redish 1993] Dumas, Joseph S., Redish, Janice C., *A Practical Guide to Usability Testing*, Ablex publishing corporation, 1993.
- [Dutka 1995] Dutka, Alan, *AMA Handbook for Customer Satisfaction*, NTC Business Books, 1995.
- [Fick & Ritchie 1991] Fick, G.R. & Ritchie, J.R.B., Measuring service quality in the travel and tourism industry, *Journal of Travel Research*, Vol. 30 No. 2, Autumn 1991.
- [Fonecta 2003a] Fonecta Ltd News verkkosivut, <http://www.fonecta.com/release270502.htm>, haettu 22.5.2003.
- [Fonecta 2003b] 020200 Kysymyksiä ja vastauksia verkkosivut, <http://www.fonecta.fi/020200faq.php>, haettu 22.5.2003.
- [Freeman & Dart 1993] Freeman, K.D. and Dart, J., Measuring the perceived quality of professional business services, *Journal of Professional Services Marketing*, Vol. 9 No. 1, 1993.
- [Galaxy 2003] Galaxy verkkosivut, <http://www.sls.lcs.mit.edu/sls/technologies/galaxy.shtml>, haettu 12.5.2003.
- [Gorin ym. 1996] Gorin, A.L., Parker, B.A., Sachs R.M. and Wilpon, J.G., How May I Help You? *Proceedings of IVTTA*, 1996.
- [Gould & Lewis 1983] Gould, John D. & Lewis, Clayton, Designing for usability: Key principles and what designers think, *Proceedings of CHI*, ACM 1983.
- [van Haaren ym. 1998] van Haaren, Lise, Blasband, Marc, Gerritsen, Marinel, van Schijndel, Marcha, Evaluating Quality of Spoken Dialogue Systems: Comparing a Technology-focused and a User-focused Approach, *Proceedings of First International Conference on Language Resources & Evaluation*, 1998.

- [Hakulinen 1989] Hakulinen, Auli, Puheen ja kirjoituksen pragmatiikkaa, *Äidinkielen opettajain liiton vuosikirja XXXVI Kielen käyttö ja käyttäjä*, Toim. Kauppinen, Anneli & Keravuori, Kyllikki, ÄOL, Helsinki 1989.
- [Hefton 2003] Hefton, Donna, To Be or Not to Be: The Question of Anthropomorphism, verkkoartikkeli, saatavilla <http://www.netpets.com/birds/reference/info/anthro.html>, haettu 9.5.2003.
- [Helin 2003] Helin, Leena, suullinen tiedonanto 22.5.2003
- [Hintikka 2003] Hintikka, Kari A., Koneille uusia käyttöliittymiä, Helsingin sanomat 8.2.2003, saatavilla <http://www.helsinginsanomat.fi/arkisto/juttu.asp?id=20030208ER8>, haettu 25.4.2003.
- [Hjalmarsson 2002] Hjalmarsson, Anna, Evaluating AdApt, a multi-modal conversational, dialogue system using PARADISE, Master's Thesis in Speech Technology, KTH, Stockholm, Sweden 2002. Saatavilla http://www.speech.kth.se/ctt/publications/papers/exjobb_hjalmarsson.pdf, haettu 15.5.2003.
- [Interact 2003] USIX-Interact projektin verkkosivut <http://www.mlab.uiah.fi/interact/>, haettu 25.4.2003.
- [van Iwaarden & van der Wiele 2002] van Iwaarden, J. & van der Wiele, T., A study on the applicability of SERVQUAL dimensions for web sites, ERIM report series Research in Management, saatavilla <http://www.eur.nl/WebDOC/doc/erim/erimrs20020708132000.pdf>, haettu 15.5.2003.
- [ISCA 2003] ISCA, International Speech Communication Association verkkosivut, <http://www.isca-speech.org>, haettu 28.4.2003.
- [Jokinen ym. 2002] Jokinen, Kristiina, Kerminen, Antti, Kaipainen, Mauri, Jauhiainen, Tommi, Wilcock, Graham, Turunen, Markku, Hakulinen, Jaakko, Kuusisto, Jukka and Lagus, Krista: Adaptive Dialogue Systems - Interaction with Interact, *Proceedings of 3rd SIGdial Workshop on Discourse and Dialogue 2002*, saatavilla myös verkossa <http://www.cs.uta.fi/hci/spi/reports/SIGDialFinal.pdf>, haettu 25.4.2003.
- [Kahney 2001] Kahney, Leander, Puppy Love for a Robot, Wired 22.2.2001, saatavilla myös verkossa <http://www.wired.com/news/culture/0,1284,41680,00.html>, haettu 4.5.2003.
- [Kamm ym. 2003] Kamm, Candace, Walker, Marilyn & Rabiner, Lawrence, The Role of Speech Processing in Human-Computer Intelligent Communication verkkoartikkeli, <http://ciips.ee.uwa.edu.au/~roberto/research/speech/local/nsfhcs/talks/rabiner.html>, haettu 15.5.2003.

- [Klemmer ym. 2000] Klemmer, Scott R., Sinha, Anoop K., Chen, Jack, Landay, James A., Aboobaker, Nadeem, Wang, Annie, SUEDE: A Wizard of Oz prototyping tool for speech user interfaces, *Proceedings of UIST*, ACM 2000.
- [Kotelly 2003] Kotelly, Blade, *The Art and Business of Speech Recognition – Creating the Noble Voice*, Addison-Wesley 2003.
- [Kotler 2001] Kotler, P., Principles of marketing, Prentice-Hall, 2001.
- [Kwon & Lee 1994] Kwon, W. & Lee, T.J., Measuring service quality in Singapore retail banking, *Singapore Management Review*, Vol. 16 No. 2, July 1994.
- [Larsson 2003] Larsson, Staffan, Dialogue System Related Links verkkosivut, http://www.ling.gu.se/~sl/dialogue_links.html, haettu 28.4.2003.
- [Lotti 1998] Lotti, Leila: *Markkinointitutkimuksen käsikirja*, WSOY, 1998.
- [MacKenzie 2003] MacKenzie, I. Scott, Using Paper Mockups for Evaluating Soft Keyboard Layouts, julkaisematon käsikirjoitus, saatavilla verkossa <http://www.cs.uta.fi/~scott/mmm/SimpleExperimentPaper.pdf>, haettu 27.4.2003.
- [Mangold & Babakus 1991] Mangold, G.W. and Babakus, E., Service quality: the front-stage perspective vs the back stage perspective, *Journal of Services Marketing*, Vol. 5 No. 4, Autumn 1991.
- [McElwee & Redman 1993] McElwee, G. and Redman, T., Upward appraisal in practice: an illustrative example using the QUALED scale, *Education and Training*, Vol. 35 No. 2, December, 1993.
- [McTear 2002] McTear, Michael F., Spoken dialogue technology: Enabling the conversational user interface, *ACM Computing Surveys*, Vol.34, No. 1, March 2002.
- [Mikropuhe 2003] Mikropuhe verkkosivut <http://www.mikropuhe.com/mikropuhe.asp>, haettu 23.5.2003.
- [Mäkelä ym. 2001] Mäkelä, Kaj, Salonen, Esa-Pekka, Turunen, Markku, Hakulinen, Jaakko and Raisamo, Roope, Conducting a Wizard of Oz Experiment on a Ubiquitous Computing System Doorman, *Proceedings of the International Workshop on Information Presentation and Natural Multimodal Dialogue*, 2001.
- [Negash ym. 2002] Negash, S., Ryan, T. and Igarria, M., Quality and effectiveness in Web-based customer support systems, *Information & Management*, julkaisematon käsikirjoitus, saatavilla verkossa <http://www.sciencedirect.com/>, haettu 15.5.2003.
- [Nielsen & Molich 1990] Nielsen, Jakob & Molich, Rolf, Heuristic evaluation of user interfaces, *CHI '90 Proceedings*, ACM 1990.
- [Nielsen & Phillips 1993] Nielsen, J. & Phillips, Victoria L., Estimating the relative usability of two interfaces: Heuristic, formal, and empirical methods compared, *Proceedings of INTERCHI '93*, ACM 1993.
- [Nielsen 2003] Ten Usability Heuristics verkkosivut, http://www.useit.com/papers/heuristic/heuristic_list.html, haettu 2.5.2003.

- [Nitecki 2003] Nitecki, D.A., SERVQUAL: Measuring Service Quality in Academic Libraries verkkootikkeli, <http://www.arl.org/newsltr/191/servqual.html>, haettu 15.5.2003.
- [Näkövammaisten keskusliitto 2003] Pistekirjoitus verkkosivut, <http://www.nkl.fi/yleistie/pistekir.htm>, haettu 19.5.2003.
- [Miettinen 1997] Miettinen, Anssi, Haamuasiakas haistaa huonon palvelun, Helsingin Sanomat 24.8.1997, saatavilla myös verkossa <http://www.helsinginsanomat.fi/uutisarkisto/19970824/talo/970824ta01.html>, haettu 22.4.2003.
- [Mzourek 2003] Mzourek, Jirka, The UI Checklist verkkosivut, <http://ui.netbeans.org/docs/checklist/>, haettu 22.5.2003.
- [Observa 2003] Gallup Observan verkkosivut <http://www.mdc.fi/observa/> -> Peruspalvelut, haettu 22.4.2003.
- [OVIS 2003] Strik, Helmer, OVIS - Openbaar Vervoer Informatie Systeem verkkosivut, <http://lands.let.kun.nl/TSPublic/strik/ovis.html>, haettu 28.4.2003.
- [OVR 2003] OVR - Openbaar Vervoer Reisinformatie verkkosivut (hollanniksi), <http://www.ovr.nl/>, haettu 28.4.2003.
- [Parasuraman ym. 1985] Parasuraman, A., Zeithaml, Valarie A. & Berry, Leonard L., A Conceptual model of service quality and its implications for future research, *Journal of Marketing*, Vol. 49, No. 4, Fall 1985.
- [Parasuraman ym. 1988] Parasuraman, A., Zeithaml, Valarie A. & Berry, Leonard L., SERVQUAL: A Multiple-item scale for measuring consumer perceptions of service quality, *Journal of Retailing*, Vol. 64, No. 1, Spring 1988.
- [Parasuraman ym. 1991] Parasuraman, A., Berry, Leonard L. & Zeithaml, Valarie A., Refinement and reassessment of the SERVQUAL scale, *Journal of Retailing*, Vol. 67, No. 4, Winter 1991.
- [Parasuraman ym. 1993] Parasuraman, A., Berry, Leonard L. & Zeithaml, Valarie A., Research note: More on improving service quality measurement, *Journal of Retailing*, Vol. 69, n. 1, Spring 1993.
- [Parasuraman ym. 1994a] Parasuraman, A., Zeithaml, Valarie A. & Berry, Leonard L., Reassessment of expectations as a comparison standard in measuring service quality: Implications for further research, *Journal of Marketing*, Vol. 58, No. 1, January 1994.
- [Parasuraman ym. 1994b] Parasuraman, A., Zeithaml, Valarie A. & Berry, Leonard L., Alternative scales for measuring service quality: A comparative assessment based on psychometric and diagnostic criteria, *Journal of Retailing*, Vol. 70, No. 3, 1994.
- [Phonetic 2003a] Phonetic Systems News Releases verkkosivut, <http://www.phoneticsystems.com/news2001/PR12-06-01.htm>, haettu 22.5.2003.
- [Phonetic 2003b] Phonetic Systems News Releases verkkosivut, <http://www.phoneticsystems.com/news2002/PR11-19-02.htm>, haettu 22.5.2003.

- [Rantanen tiedonanto] Rantanen, Leo, sähköpostitiedonanto 25.4.2003.
- [Reeves & Nass 1996] Reeves, Byron & Nass, Clifford, *The Media Equation*, Cambridge University Press, 1996.
- [Rosenfeld ym. 2001] Rosenfeld, Ronald, Olsen, Dan & Rudnicky, Alex, Universal speech interfaces, *ACM Interactions*, Vol. 8, No. 6, ACM Nov./Dec. 2001.
- [Räsänen 2001] Räsänen, Pajari, HKL Liikenneneuvonta, Litterointi, USIX-Interact – projektin sisäiset verkkosivut, saatavilla (ei julkisesti) <http://www.mlab.uiah.fi/interact/internal/index.htm>, haettu 28.5.2003.
- [Saleh & Ryan 1992] Saleh, F. & Ryan, C., Analysing service quality in the hospitality industry using the SERVQUAL model, *Services Industries Journal*, Vol. 11 No. 3 1992.
- [Salonen 2002] Salonen, Esa-Pekka, Hei kuka puhuu? Toimivan puhekäyttöliittymän suunnittelu ja toteutus, Pro gradu –tutkielma, Tampereen yliopisto, Tietojenkäsittelytieteiden laitos 2002, saatavilla http://www.cs.uta.fi/~eps/gradu/eps_gradu.html, haettu 22.5.2003.
- [Saukkonen 1972] Saukkonen, Pauli, Kokeellisia havaintoja puhekielen ja kirjakielen tyylieroista, *Puhekieli ja ilmaisu*, toim. Vierikko, Esko, WSOY, 1972.
- [Scott & Shieff 1993] Scott, D. & Shieff, D., Service quality components and group criteria in local government, *International Journal of Service Industry Management*, Vol. 4 No. 4, 1993.
- [Seneff ym. 1998] Seneff, S., Hurley, E., Lau, R., Pao, C., Schmid, P. and Zue, V., GALAXY-II: A Reference Architecture for Conversational System Development, *Proceedings of ICSLP 98*, 1998.
- [Shneiderman 1987] Shneiderman, Ben, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, Addison-Wesley, 1987.
- [Soliman 1992] Soliman, A.A., Assessing the quality of health care, *Health Care Marketing*, Vol. 10 No:s 1-2, 1992.
- [Speechworks 2003] Speechworks Industry Solutions verkkosivut, <http://www.speechworks.com/industry/index.cfm>, haettu 4.5.2003.
- [Teas 1993] Teas, R. Kenneth: Expectations, Performance Evaluation, and Consumers' Perceptions of Quality, *Journal of Marketing*, Vol. 57, No. 4, October 1993.
- [Teas 1994] Teas, R. Kenneth: Expectations as a comparison standard in measuring service quality: An assessment of a reassessment, *Journal of Marketing*, Vol. 58, No. 1, January 1994.
- [Turunen & Hakulinen 2000a] Turunen, Markku and Hakulinen, Jaakko, Jaspis - A Framework for Multilingual Adaptive Speech Applications, *Proceedings of 6th International Conference of Spoken Language Processing (ICSLP 2000)*, 2000.
- [Turunen & Hakulinen 2000b] Turunen, Markku and Hakulinen, Jaakko, Mailman - a Multilingual Speech-only E-mail Client based on an Adaptive Speech Application

- Framework, *Proceedings of Workshop on Multi-Lingual Speech Communication (MSC 2000)*, 2000.
- [Walker ym. 1997] Walker, M.A., Litman, D.J., Kamm, C.A. and Abella, A., PARADISE: A framework for evaluating spoken dialogue agents, *Proceedings of the 35th Annual Meeting of the Association of Computational Linguistics*, ACL 97, 1997.
- [Weinschenk & Barker 2000] Weinschenk, Susan & Barker, Dean T., *Designing Effective Speech Interfaces*, John Wiley & Sons Ltd. 2000
- [Weiss 2002] Weiss, Scott, *Handheld Usability*, John Wiley & Sons Ltd. 2002.
- [Wong & Perry 1991] Wong, S.M. & Perry, C., Customer service strategies in financial retailing, *International Journal of Bank Marketing*, Vol. 9 No. 3, 1991.
- [Vierikko 1972] Vierikko, Esko, Nuorten poliitikkojen kielestä, *Puhekieli ja ilmaisu*, toim. Vierikko, Esko, WSOY, 1972.
- [Yankelovich, 1996] Nicole Yankelovich, How Do Users Know What to Say?, *ACM Interactions*, Vol. 3, No. 6 November/December 1996.
- [Ylikoski 2000] Ylikoski, Tuire, *Unohtuiko asiakas?* Toinen uudistettu painos, Otava, 2000.
- [Zeithaml & Bitner 1996] Zeithaml, V. & Bitner, M., *Services Marketing*, McGraw-Hill, 1996.
- [Zeithaml ym. 1990] Zeithaml, Valarie A., Parasuraman, A. & Berry, Leonard L., *Delivering quality service. Balancing Customer Perceptions and Expectations*, Free Press, 1990.

Liitteet

Liite 1: Odotuslomake

	Minun vähimmäisvaatimus kuvatulle seikalle	Taso, jolla kuvattu seikka pitää olla, että palvelu on mielestäni hyvä
Väitteet:	matala... ..korkea	matala... ..korkea
1. Äänestä päätellen bussiaikataulupalvelulla on modernit laitteet.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
2. Bussiaikataulupalvelun puheääntä on miellyttävä kuunnella.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
3. Bussiaikataulupalveluun kuuluvat muut äänet ovat myös asiaan kuuluvia ja tekevät toiminnasta miellyttävää.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
4. Bussiaikataulupalveluun liittyvän materiaalin (esitteet yms.) ulkoasu on visuaalisesti miellyttävää.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
5. Bussiaikataulupalvelu on niin nopea kuin sen luvataan olevan.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
6. Kun asiakkaalla on ongelma, bussiaikataulupalvelun tarjoaja osoittaa haluavansa ratkaista sen.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
7. Bussiaikataulupalvelu palvelee oikein ensi yrittämällä.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
8. Bussiaikataulupalvelut on saatavilla niinä aikoina kuin luvataan.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
9. Bussiaikataulupalvelun tarjoaja haluaa ennalta ehkäistä virhetilanteita.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
10. Bussiaikataulupalvelu ilmaisee tarkasti kuinka nopeasti se palvelee sinua.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
11. Bussiaikataulupalvelu palvelee ripeästi.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
12. Bussiaikataulupalvelu on halukas auttamaan.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
13. Bussiaikataulupalvelu ei ole koskaan liian kiireinen vastaamaan asiakkaan pyyntöihin.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
14. Bussiaikataulupalvelu herättää asiakkaissa luottamusta käyttäytymisellään.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
15. Bussiaikataulupalvelun asiakkaat voivat toimia turvallisin mielin.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
16. Bussiaikataulupalvelu on kaikissa tilanteissa kohtelias.	1 2 3 4 5 6 7	1 2 3 4 5 6 7

17. Bussiaikataulupalvelulla on hallussaan tarvittava tieto kaikkiin käyttäjien kysymyksiin vastaamiseksi.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
18. Bussiaikataulupalvelun tarjoaja huomioi asiakkaan yksilöllisesti.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
19. Bussiaikataulupalvelu toimii kaikille asiakkaille sopivaan aikaan.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
20. Bussiaikataulupalvelu huomioi asiakkaan henkilökohtaisesti.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
21. Bussiaikataulupalvelun tarjoaja tahtoo asiakkaan parasta.	1 2 3 4 5 6 7	1 2 3 4 5 6 7
22. Bussiaikataulupalvelu ymmärtää asiakkaiden erikoistarpeet.	1 2 3 4 5 6 7	1 2 3 4 5 6 7

Liite 2: Havaintolomake

	Vaikutelma HKL- bussiaikataulu- automaatin suoriutumisesta kuvatussa seikassa
Väitteet:	matala... ...korkea
1. Äänestä päätellen bussiaikataulupalvelulla on modernit laitteet.	1 2 3 4 5 6 7
2. Bussiaikataulupalvelun puheääntä on miellyttävä kuunnella.	1 2 3 4 5 6 7
3. Bussiaikataulupalveluun kuuluvat muut äänet ovat myös asiaan kuuluvia ja tekevät toiminnasta miellyttävää.	1 2 3 4 5 6 7
4. Bussiaikataulupalveluun liittyvän materiaalin (esitteet yms.) ulkoasu on visuaalisesti miellyttävää.	1 2 3 4 5 6 7
5. Bussiaikataulupalvelu on niin nopea kuin sen luvataan olevan.	1 2 3 4 5 6 7
6. Kun asiakkaalla on ongelma, HKL osoittaa haluavansa ratkaista sen.	1 2 3 4 5 6 7
7. Bussiaikataulupalvelu palvelee oikein ensi yrittämällä.	1 2 3 4 5 6 7
8. Bussiaikataulupalvelut on saatavilla niinä aikoina kuin luvataan.	1 2 3 4 5 6 7
9. HKL haluaa ennalta ehkäistä virhetilanteita.	1 2 3 4 5 6 7
10. Bussiaikataulupalvelu ilmaisee tarkasti kuinka nopeasti se palvelee sinua.	1 2 3 4 5 6 7
11. Bussiaikataulupalvelu palvelee ripeästi.	1 2 3 4 5 6 7
12. Bussiaikataulupalvelu on halukas auttamaan.	1 2 3 4 5 6 7
13. Bussiaikataulupalvelu ei ole koskaan liian kiireinen vastaamaan asiakkaan pyyntöihin.	1 2 3 4 5 6 7
14. Bussiaikataulupalvelu herättää asiakkaissa luottamusta käyttäytymisellään.	1 2 3 4 5 6 7
15. Bussiaikataulupalvelun asiakkaat voivat toimia turvallisin mielin.	1 2 3 4 5 6 7
16. Bussiaikataulupalvelu on kaikissa tilanteissa kohtelias.	1 2 3 4 5 6 7
17. Bussiaikataulupalvelulla on hallussaan tarvittava tieto kaikkiin käyttäjien kysymyksiin vastaamiseksi.	1 2 3 4 5 6 7
18. HKL huomioi asiakkaan yksilöllisesti.	1 2 3 4 5 6 7
19. Bussiaikataulupalvelu toimii kaikille asiakkaille sopivaan aikaan.	1 2 3 4 5 6 7
20. Bussiaikataulupalvelu huomioi asiakkaan henkilökohtaisesti.	1 2 3 4 5 6 7
21. HKL tahtoo asiakkaan parasta.	1 2 3 4 5 6 7
22. Bussiaikataulupalvelu ymmärtää asiakkaiden erikoistarpeet.	1 2 3 4 5 6 7

Liite 3: Esimerkkidialogit

Äänitiedostot saatavilla toistaiseksi verkossa osoitteessa www.cs.uta.fi/~mhartikainen/nauhat . Esimerkkidialogit kirjallisessa asussa alla. Kursivoidut tekstiosuudet ovat järjestelmän puheenvuoroja. Normaalit tekstiosuudet ovat käyttäjän puheenvuoroja. Alaviivat sanassa ”_seuraava_” tarkoittavat korostettua osuutta.

"Maunulasta Kumpulaan", kirjakielinen versio

HKL aikataulupalvelu. Kuinka voin auttaa?

- mikä bussi lähtee seuraavaksi maunulasta kumpulaan ja monelta se on perillä kumpulassa
pääset maunulasta kumpulaan bussilla 52

- mikä on _seuraava_ bussi maunulasta kumpulaan ja monelta se on perillä kumpulassa
Bussi 52 maunulasta kumpulaan lähtee kello 13.51

- monelta se on perillä kumpulassa
Bussi 52 maunulasta kumpulaan on perillä klo 19.19

"Maunulasta Kumpulaan", puhekielinen versio

HKL aikataulupalvelu. Kuin voin auttaa?

- mikä bussi lähtee seuraavaksi maunulasta kumpulaan ja monelta se on perillä kumpulassa
Sä pääset maunulasta kumpulaan viiskakkosella

- mikä on _seuraava_ bussi maunulasta kumpulaan ja monelta se on perillä kumpulassa
Viiskakkonen maunulasta kumpulaan lähtee kello 14.03

- monelta se on perillä kumpulassa
Viiskakkonen maunulasta kumpulaan on perillä klo 19.19

"Keskustasta Patolaan", kirjakielinen versio

HKL aikataulupalvelu. Kuinka voin auttaa?

- no terve, oon rautatieasemalla helsingin keskustassa ja mun pitäis päästä käymään patolassa eli mikähän bussi sinne menis seuraavaksi

En valitettavasti ymmärrä

- eli pitäs päästä patolaan eli mikä numero ja mikä- mihin aikaan menee seuraava bussi
Mistä olitkaan lähdössä?

-Keskustasta

Pääset keskustasta patolaan busseilla kuusikymmentäneljä, kuusikymmentäneljä än, kuusikymmentäyhdeksän ja seitsemänkymmentäkaksi

- no mihin aikaan lähtee seuraava bussi
Bussi kuusikymmentäneljä keskustasta patolaan lähtee klo 18.05

"Keskustasta Patolaan", puhekielinen versio

HKL aikataulupalvelu. Kuin voin auttaa?

- no terve, oon rautatieasemalla helsingin keskustassa ja mun pitäis päästä käymään patolassa eli mikähän bussi sinne menis seuraavaksi

Anteeks, nyt en ymmärrä

- eli pitäs päästä patolaan eli mikä numero ja mikä- mihin aikaan menee seuraava bussi
Mistä sä olit lähössä?

-Keskustasta

Sä pääset keskustasta patolaan kuusnelosella, kuusneljä ännällä, kuusysillä ja seiskakakkosella

- no mihin aikaan lähtee seuraava bussi
Kuusysi keskustasta patolaan lähtee klo 18.15

Liite 4: Taustatietolomake**Taustatietokysely**

Kyselyn tietoja tarvitaan ryhmien muodostamiseen. Henkilöllisyyttä ei yhdistetä muihin tietoihin eikä paljasteta julkaistavassa aineistossa.

nimi _____

s-posti tai puhelin (miten saa parhaiten yhteyden) _____

Ikä

- alle 25
- 25-35
- 35-50
- yli 50

Kuinka usein käytät tietokonetta?

- Erittäin harvoin jos ollenkaan
- Silloin tällöin
- Usein

Miten paljon sinulla on kokemusta puhekäyttöliittymistä?

- En ole koskaan itse kokeillut
- Olen kokeillut, mutta vain muutaman kerran
- Käytän silloin tällöin
- Käytän usein

Matkustatko paikallisbussilla?

- Erittäin harvoin jos ollenkaan
- Silloin tällöin
- Usein

Käyttäisitkö julkista liikennettä toisella paikkakunnalla, esim. Helsingissä vieraillessasi?

- Ei, menen taksilla tai omalla autolla
- Yleensä ei tarvitse, kun on joku muu vaihtoehto, mutta tarpeen tullen käyttäisin
- Käytän pääsääntöisesti julkista liikennettä toisilla paikkakunnilla

Miten selvität aikataulutiedot? (voi olla useita)

- Painetut aikataulut (sis. pysäkillä olevat, irtolehdet, aikataulukirjat)
- WWW-sivuilta
- Soitan neuvontapalveluun
- muuten, miten? _____

Liite 5: Jälkikyselylomake

Jälkikysely

Voisitko kuvitella käyttäväsi esimerkin palveluautomaattia?

- En käyttäisi
- Ehkä kokeilisin tai käyttäisin korkeintaan silloin tällöin
- Kyllä, voisin käyttää tarpeen tullen useinkin

Paljonko maksaisit puhelusta?

- En soittaisi vaikka olisi ilmainen
- Soittaisin vain jos se olisi ilmainen
- Voisin maksaa paikallisverkkomaksun hinnan
- Maksaisin korkeintaan _____ (kirjoita hinta itse)

Yleisarvosana (ympyröi) 1 2 3 4 5 6 7

Muita kommentteja: