

# **Puheteknologian hyödyntäminen rakennusten teknisissä järjestelmissä**

Veli Möttönen & Jouko Pakanen  
VTT Rakennustekniikka

Johannes Peltola, Marko Salmela & Tapio Seppänen  
VTT Elektroniikka



ISBN 951-38-5321-7 (nid.)  
ISSN 1235-0605 (nid.)

ISBN 951-38-5322-5 (URL: <http://www.inf.vtt.fi/pdf/>)  
ISSN 1455-0865 (URL: <http://www.inf.vtt.fi/pdf/>)

Copyright © Valtion teknillinen tutkimuskeskus (VTT) 1998

#### JULKAISIJA – UTGIVARE – PUBLISHER

Valtion teknillinen tutkimuskeskus (VTT), Vuorimiehentie 5, PL 2000, 02044 VTT  
puh. vaihde (90) 4561, telekopio 456 4374

Statens tekniska forskningscentral (VTT), Bergsmansvägen 5, PB 2000, 02044 VTT  
tel. växel (90) 4561, telefax 456 4374

Technical Research Centre of Finland (VTT), Vuorimiehentie 5, P.O.Box 2000, FIN-02044 VTT, Finland  
phone internat. + 358 0 4561, telefax + 358 0 456 4374

VTT Rakennustekniikka, Rakentaminen ja kiinteistönhallinta, Kaitoväylä 1, PL 18021, 90571 OULU  
puh. vaihde (08) 551 2111, faksi (08) 551 2090

VTT Byggnadsteknik, Byggnad och fastighetsförvaltning, Kaitoväylä 1, PB 18021, 90571 OULU  
tel. växel (08) 551 2111, fax (08) 551 2090

VTT Building Technology, Construction and Facility Management, Kaitoväylä 1, P.O.Box 18021,  
FIN-90571 OULU, Finland  
phone internat. + 358 8 551 2111, fax + 358 8 551 2090

VTT Elektroniikka, Elektroniikan piirit ja järjestelmät, Kaitoväylä 1, PL 1100, 90571 OULU  
puh. vaihde (08) 551 2111, faksi (08) 551 2320

VTT Elektronik, Elektroniska kretsar och system, Kaitoväylä 1, PB 1100, 90571 ULEÅBORG  
tel. växel (08) 551 2111, fax (08) 551 2320

VTT Electronics, Electronic Circuits and Systems, Kaitoväylä 1, P.O.Box 1100,  
FIN-90571 OULU, Finland  
phone internat. + 358 8 551 2111, fax + 358 8 551 2320

Toimitus Leena Ukoskoski

Libella Painopalvelu Oy, Espoo 1998

Möttönen, Veli, Pakanen, Jouko, Peltola, Johannes, Salmela, Marko, Seppänen, Tapio. Puheteknologian hyödyntäminen rakennusten teknisissä järjestelmissä [Utilising speech technology in technical systems of buildings]. Espoo 1998, Valtion teknillinen tutkimuskeskus, VTT Tiedotteita – Meddelanden – Research Notes 1918. 73 s. + liitt. 1 s.

**Avainsanat** intelligent buildings, electric appliances, security systems, speech technology, voice communication

## TIIVISTELMÄ

Asukkaan, huoltomiehen tai muun käyttäjän kommunikointi rakennuksen teknisen laitteen tai järjestelmän kanssa tapahtuu yleensä erilaisten alfanumeeristen tai graafisten näyttölaitteiden ja näppäimistöjen avulla. Laitteen valmistuskustannusten minimointi on eräs syy siihen, että nämä käyttöliittymät on toteutettu muuttamalla näppäimellä ja yhden tai kahden rivin alfanumeerisella näytöllä. Tämä yhdistelmä ei useinkaan tarjoa riittäviä vuoropuhelumahdollisuuksia käyttäjän ja tietokoneen välillä.

Puheteknologia tarjoaa menetelmiä ja tuotteita näiden laitteiden ja järjestelmien käyttöliittymien toiminnan parantamiseen. Puheteknologia, joka soveltaa esimerkiksi puheen tallennusta ja toistoa, puhekomentoja, puhujantunnistamista, puheentunnistamista tai puhesynteesiä erilaisten koneiden ja laitteiden rakenteessa ja toiminnassa, onkin nopeasti yleistymässä. Laaja-alaisesta hyödyntämisestä huolimatta puheen käsittelyyn menetelmissä on vielä ratkaisemattomia ongelmia. Esimerkiksi täydellinen vuoropuhelu ihmisen ja koneen välillä ei ole vielä toteutunut.

Elektronisia komponentteja, ohjelmistotuotteita sekä kehitystyökaluja puheteknologian soveltamiseen on tarjolla kaikille puheteknologian osa-alueille. Puheviestintää soveltavien tuotteiden kustannukset sopivat myös hintakriittisille tuotteille, jos puheen käsittelyn vaatimustaso on kohtuullinen. Vaatimustason kasvaessa joudutaan elektronisten komponenttien valinnassa siirtymään analogisista tai signaalinkäsittelyn kannalta hitaista digitaalipiireistä signaaliprosessorisovelluksiin. Kehitystyökalua valittaessa sopiva vaihtoehto voi olla yksinkertainen demo-työkalu, puheen- ja signaalinkäsittelyyn suunniteltu studiolaitteisto tai koko käyttöliittymän suunnitteluun tarkoitettu järjestelmä.

Puheteknologian nykyinen kehitystaso sekä eri valmistajien tuottamat elektroniset komponentit, ohjelmistot ja kehitystyökalut luovat hyvät edellytykset sisällyttää puheviestintää myös rakennusten teknisiin laitteisiin ja järjestelmiin. Sopivia kohteita ovat esimerkiksi kodinkoneet, erilaiset automaatiojärjestelmät sekä turvajärjestelmät. Omat erityisen käyttäjäryhmänsä puheteknologiatuotteille muodostavat vanhukset ja vammaiset. On myös odotettavissa, että tämän teknologian myötä syntyy kokonaan uusia tuotteita palvelemaan rakennusten käyttäjiä.

Möttönen, Veli, Pakanen, Jouko, Peltola, Johannes, Salmela, Marko, Seppänen, Tapio. Puheteknologian hyödyntäminen rakennusten teknisissä järjestelmissä [Utilising speech technology in technical systems of buildings]. Espoo 1998, Technical Research Centre of Finland, VTT Tiedotteita – Meddelanden – Research Notes 1918. 73 p. + app. 1 p.

**Keywords** intelligent buildings, electric appliances, security systems, speech technology, voice communication

## ABSTRACT

Communication between a user and a technical device or system of a building typically takes place by means of various types of alphanumeric or graphic displays and keyboards. To minimize manufacturing cost, these user interfaces are realized using only a few buttons and an alphanumeric display with one or two rows of text. However, this combination often does not provide enough possibilities for interaction between the user and the computer.

Speech technology offers methods and devices for improving the functionality of the interfaces of these devices and systems. Speech technology, which is rapidly becoming more common, applies speech record and playback, voice commands, speaker identification, speech recognition or speech synthesis in the operation of many kinds of machines and equipment. Despite broad use of the technology, speech processing methods still contain unsolved problems. For example, fluent conversations between man and machine have yet to be realized.

Electronic components, software products and development tools are available for all areas of speech technology. The cost of products that apply voice communication is suitable even for products whose price is critical. If requirements of speech processing are critical, the choice of electronic components must be changed from analog or slow digital circuits to signal processor applications. A suitable development tool could be a simple demo tool, studio equipment designed for speech and signal processing, or a system for designing the entire user interface.

The present level of development in commercial speech technology and the electronic components, software and development tools provide good prerequisites for including voice communication in the technical devices and systems found in buildings. Suitable applications include home appliances, various automation systems and security systems. Elderly and disabled people constitute a special group of users. It is also probable that this technology will spawn completely new products for building environment.

## ALKUSANAT

“Puheteknologian hyödyntäminen rakennusten teknisissä järjestelmissä” on tutkimus, jonka rahoittajina ovat toimineet Teknologian tutkimuskeskus (TEKES), VTT sekä yritykset Esmi Oy, Neste Oy, Ouman Oy ja Securitas Tekniikka Oy.

Tutkimusprojektin johtoryhmään ovat kuuluneet Mika Lautanala (TEKES), Veijo Lappalainen (VTT), Esko Paajanen (Esmi Oy), Seppo Peltola (Neste Oy), Timo Peltola (Ouman Oy) sekä Heikki Leppänen (Securitas Tekniikka Oy).

Tutkimus on laadittu VTT Rakennustekniikan ja VTT Elektroniikan yhteistyönä. VTT Rakennustekniikan tutkimus kohdistuu mm. rakennusten sähkö-, automaatio-, tieto- ja teleteknisiin järjestelmiin, laitejärjestelmien kehittyneisiin säätö- ja hallintamenetelmiin, vikadiagnostiikkaan sekä kiinteistöjen ylläpidon tekniikkaan. VTT Elektroniikka on elektroniikkaan erikoistunut t & k -yksikkö, joka tarjoaa palveluja kaikille elektronisia osia sisältävien tuotteiden valmistajille. Asiakaskuntaan kuuluu yrityksiä erityisesti seuraavilta toimialueilta: elektroniikka, tietoliikenne, prosessiautomaatio, koneteollisuus ja instrumentointi. Aikaisempaa erityisosaamista tämän selvityksen tekemiseksi on muodostettu mm. puheentunnistuksen, puheenkoodauksen, käyttöliittymien ja sulautettujen järjestelmien tutkimushankkeissa.

Tutkimuksen käytännön toteutuksesta ovat vastanneet Veli Möttönen (VTT Rakennustekniikka), Johannes Peltola (VTT Elektroniikka), Marko Salmela (VTT Elektroniikka), Tapio Seppänen (VTT Elektroniikka), sekä Jouko Pakanen (VTT Rakennustekniikka), joista viimeksi mainittu on toiminut hankkeen projektipäällikkönä.

# SISÄLLYSLUETTELO

TIIVISTELMÄ .....	3
ABSTRACT .....	4
ALKUSANAT .....	5
KÄSITTEET .....	8
1 JOHDANTO .....	9
2 KÄYTTÖLIITTYMIEN ONGELMAT RAKENNUSTEN TEKNISISSÄ LAITTEISSA JA JÄRJESTELMISSÄ.....	10
2.1 Käyttöliittymän ominaisuuksia.....	10
2.2 Puutteellinen käyttöliittymä.....	10
2.3 Puutteellisten käyttöliittymien syyt .....	11
2.4 Puheteknologian mahdollisuudet.....	11
3 PUHEVIESTINNÄN MENETELMIEN PERIAATTEET .....	13
3.1 Puheviestit .....	13
3.1.1 Puheen tallentamiseen perustuva viestin tuottaminen .....	14
3.1.2 Tekstin muuttaminen puheeksi: puhesynteesi.....	15
3.2 Automaattinen puheen tunnistaminen .....	18
3.2.1 Historiaa .....	18
3.2.2 Puheentunnistusjärjestelmien luokittelu .....	19
3.2.3 Puheentunnistuksen perusteet lyhyesti.....	21
3.2.4 Haastavimmat tutkimuskohteet.....	24
3.3 Puhujantunnistus .....	25
4 PUHETEKNOLOGIAN KAUPALLINEN LAITE- JA OHJELMISTOTARJONTA .....	27
4.1 Elektroniset komponentit ja ohjelmistot.....	27
4.1.1 Puheviestit.....	27
4.1.2 Puheentunnistus .....	33
4.1.3 Puhujantunnistus .....	40
4.1.4 Elektronisten komponenttien ja piiriratkaisujen valinta .....	41
4.1.5 Verkotettu arkkitehtuuriratkaisu .....	43
4.1.6 Komponenttien ja ohjelmistojen hankintakustannuksia .....	44
4.2 Kehitystyökalut ja -ympäristöt .....	45
4.2.1 Työkaluja puheviestejä hyödyntäviin järjestelmiin .....	45
4.2.2 Työkaluja puheentunnistusjärjestelmiin .....	47
4.2.3 Työkaluja puhujantunnistusjärjestelmiin .....	50
4.2.4 Virtuaalisuunnittelu.....	50
5 SOVELLUSESIMERKKEJÄ.....	55
5.1 Puheviestejä hyödyntäviä tuotteita.....	55

5.1.1 Paikannus- ja opastejärjestelmät .....	55
5.1.2 Ajoneuvo- ja kulutuselektronikka.....	56
5.1.3 Matkapuhelimet ja henkilöhakulaitteet .....	57
5.1.4 Terveystieteiden sovellukset .....	57
5.1.5 Turvatekniikka .....	57
5.1.6 Laitteiden käytönopastus.....	58
5.2 Puheentunnistusta hyödyntäviä tuotteita .....	58
5.2.1 Ajoneuvoelektronikka.....	59
5.2.2 Matkapuhelimet .....	60
5.2.3 Kotiautomaatio.....	60
5.2.4 Tiedonkeruulaitteet ja erikoistietokoneet.....	62
5.3 Puhujantunnistusta hyödyntäviä tuotteita.....	63
<b>6 PUHETEKNOLOGIAN SOVELTAMINEN RAKENNUSTEN TEKNISISSÄ LAITTEISSA JA JÄRJESTELMISSÄ.....</b>	<b>64</b>
6.1 Kodinkoneet .....	64
6.2 Automaatio .....	64
6.3 Turvajärjestelmät.....	65
6.4 Vammaiset ja vanhukset.....	65
6.5 Muut puheteknologiasovellukset.....	66
<b>7 CASE - PUHETEKNOLOGIATUOTTEEN SUUNNITTELU .....</b>	<b>67</b>
<b>8 YHTEENVETO .....</b>	<b>70</b>
<b>LÄHDELUETTELO.....</b>	<b>71</b>

## LIITTEET

Liite 1: Yritysten WWW-sivujen osoitteita

# KÄSITTEET

## *Puheenkoodaus*

Puheenkoodaus (*speech coding*) on signaalinkäsittelyyn perustuvaa tekniikkaa, jossa puhesignaaliin sisältyvä, kuulon kannalta olennainen informaatio pyritään pakkaamaan tiiviiseen muotoon siirtoa tai tallennusta varten ilman, että äänen laatu kuitenkaan heikkenee merkittävästi [Karjalainen 1993].

## *Puheenkäsittely*

Puheenkäsittely (*speech processing*) on tutkimuksen ja teknologian haara, jossa pyritään kehittämään keinoja ihmisen ja koneen välistä puhekommunikaatiota varten ja ihmisten välisen puheviestinnän helpottamiseksi teknisin keinoin [Karjalainen 1993].

## *Puhsynteesi*

Puhsynteesi (*speech synthesis*) eli keinotekoinen puheen tuottaminen koostuu menetelmistä tuottaa koneella puhesignaaleja ihmiselle jonkin luonteeltaan luonnollisen kielen informaation perusteella [Karjalainen 1993].

## *Puheentunnistus*

Puheentunnistuksessa (*speech recognition*) pyritään puhsynteesiin nähden päinvastaiseen tulokseen, ihmisen tuottaman puhesignaalin muuntamiseen koneessa esitettävään muotoon esimerkiksi sanoiksi, lauseiksi tai merkityskuviksi [Karjalainen 1993].

## *Puhujantunnistus*

Puheentunnistusteknologian käyttö henkilön identifioimiseksi (*speaker identification*) tai väitetyn henkilöllisyyden varmistamiseksi (*speaker verification*).

## *Puheviesti*

Puheviesteillä tarkoitetaan koneen tai laitteen puhemuodossa käyttäjälle antamia viestejä, opasteita, ohjeita jne.



# 1 JOHDANTO

Puheviestintä on luonnollinen ja eräs käyttökelpoisimpia kommunikointitapoja ihmisten välisessä kanssakäymisessä, sillä puhuminen on mahdollista myös muiden tehtävien ohella. Puheviestinnän etuna onkin se, että se vapauttaa kädet ja näköaistin muita toimintoja varten, mikä monissa yhteyksissä lisää sekä toimintamahdollisuuksia että turvallisuutta. Mm. näiden syiden vuoksi puheen käyttö on erityisen mielenkiinnon kohteena ihmisen ja koneen välistä vuorovaikutusta kehitettäessä.

Puheviestintää hyödyntävää tekniikkaa, puheteknologiaa, on onnistuneesti sovellettu useilla eri aloilla. Esimerkiksi vammaisille ja vanhuksille tarkoitetuissa tuotteissa sovelletaan puheteknologiaa viestinnän ja ympäristön hallinnan apuvälineissä. Parhailtaan tämä teknologia on leviämässä viihde-elektronikan ja kodintekniikan tuotteisiin. Puheteknologian eri muodoissaan odotetaan vaikuttavan voimakkaasti sekä ihmisten väliseen puheviestintään että ihmisen ja koneen vuorovaikutukseen. Tämä uusi tekniikka tullee lähitulevaisuudessa näyttämään merkittävää osaa käyttöliittymien kehittämisessä. Uusien käyttöliittymien täysi hyödyntäminen edellyttää kuitenkin sitä, että käyttäjä voi vapaasti valita hänelle itselleen ja kulloiseenkin tilanteeseen sopivimman kommunikointitavan.

Onkin odotettavissa, että puheteknologiaa sisältävät tuotteet tulevat lisääntymään nopeasti. Menetelmät ja tekniikka ovat jo valmiina moniin käyttötarkoituksiin. Tämän vuoksi eri sovellusalueilla on mietittävä, miten puheviestintää eri muodoissaan kannattaa soveltaa, mitä puheteknologian menetelmiä on kehitetty ja millaisia hyödyntämiskelpoisia teknisiä ja kaupallisia tuotteita on jo nyt tai lähitulevaisuudessa saatavilla. Tämän tutkimuksen tarkoitus on kartoittaa näitä asioita rakennusten teknisten järjestelmien ja laitteiden osalta.

Tutkimus on tilannekatsaus puheteknologiaan ja sen soveltamismahdollisuuksiin rakennusten teknisissä järjestelmissä. Alan nopean kehittymisen vuoksi monet raportissa esitetyt tiedot, erityisesti elektronisten komponenttien ja ohjelmistojen osalta, vanhenevat nopeasti. Luvussa 2 esitetään rakennusten teknisten järjestelmien käyttöliittymien ominaisuuksia ja niissä esiintyviä ongelmia sekä kerrotaan puheteknologian tarjoamista mahdollisuuksista ratkaista näitä ongelmia. Luvussa 3 käsitellään lyhyesti puheen synteesin ja tunnistamisen teoreettista taustaa. Tavoitteena on antaa lukijalle mielikuva siitä, mitä menetelmiä käytetään ja mitä ongelmia vielä on ratkaisematta. Luvussa 4 esitellään puheteknologian soveltamiseen tarkoitettujen elektronisten komponenttien ja ohjelmistojen kaupallista tarjontaa sekä tuotteiden kehittämistä helpottavia kehitystyökaluja ja -ympäristöjä. Luvussa 5 esitellään joitakin puheteknologiaa hyödyntäviä tuotteita, jotka jo ovat tai lähiaikoina tulevat markkinoille. Tavoitteena on antaa esimerkkejä siitä, miten puheteknologiaa nykyisin sovelletaan eri aloilla. Luvussa 6 hahmotellaan puheteknologian erilaisia soveltamismahdollisuuksia rakennusten teknisissä laitteissa ja järjestelmissä. Lopuksi, luvussa 7, kuvataan käytännön esimerkki puheteknologiatuotteen suunnittelusta.

## 2 KÄYTTÖLIITTYMIEN ONGELMAT RAKENNUSTEN TEKNISISSÄ LAITTEISSA JA JÄRJESTELMISSÄ

### 2.1 KÄYTTÖLIITTYMÄN OMINAISUUKSIA

Ihmisen ja tietokoneen välinen kommunikointi on vuoropuhelua, jossa käytetään erilaisia vuorovaikutustapoja ja -laitteita. Tavallisimpia laitteita ovat standardisoidut tietokoneen näppäimistöt näyttöineen ja muine apulaitteineen. Yhä enemmän käyttäjä joutuu kuitenkin kommunikoimaan erilaisten sulautettujen järjestelmien kanssa, joiden vuorovaikutusrajapinta eli käyttöliittymä voi olla edellisistä poikkeava. Asuin- tai toimistorakennus on tyypillinen esimerkki ympäristöstä, joista löytyy molempia edellä mainittuja käyttöliittymäratkaisuja. Rakennusten teknisten järjestelmät ja laitteet saattavat olla sekä standardeja tietokonejärjestelmiä että tiettyyn tarkoitukseen suunniteltuja laitteita.

Laitteen käyttöliittymä ei ole välttämättä toimiva kokonaisuus, vaikka sen tekninen toteutus ja varustelutaso olisivat hyvät, sillä käyttäjän ja tietokoneen fyysinen rajapinta on vain osa käyttöliittymää. Siihen kuuluu myös kaikki se, mitä käyttäjän on ymmärrettävä tietokoneen ja tietyn ohjelman toiminnasta ja toteutuksesta voidakseen saavuttaa haluamansa tavoitteet [Koivunen 1993]. Hyvä tekninen toteutus ja monipuoliset vuorovaikutusmahdollisuudet ovat kuitenkin perusedellytyksiä pyrittäessä tyydyttävään lopputulokseen käyttöliittymän suunnittelussa. Rakennusten teknisillä järjestelmillä ja laitteilla nämäkin ovat usein puutteellisia. Seurauksena on käyttöliittymiä, joista hyvällä suunnittelullakaan ei saada toimivia ratkaisuja.

### 2.2 PUUTTEELLINEN KÄYTTÖLIITTYMÄ

Perusongelmana useissa käyttöliittymissä on se, että ne on varustettu yhden tai kahden rivin alfanumeerisella näyttölaitteella, jonka avulla ei ole mahdollista tulostaa selväkielisiä viestejä. Viestit joudutaan koodaamaan, mikä aiheuttaa ongelmia niiden ymmärtämisessä. Toimintaohjeet, statustiedot, virheilmoitukset ym. jäävät puutteellisiksi.

Toiminnoiltaan ja ominaisuuksiltaan riittävä näyttölaite mahdollistaa useiden tietojen yhtäaikaista esittämisen, mikä havainnollistaa ja helpottaa kommunikointia käyttäjän kanssa. Jos näyttölaite on puutteellinen, joudutaan esimerkiksi käyttäjän opastaminen hoitamaan samaa näyttölaitetta apuna käyttäen. Tällöin itse näyttöltä katoaa tilanne, johon opastusta tarvitaan.

Eräs keskeinen ongelma käyttöliittymissä on se, että monipuolinenkin vuoropuhelu laitteen tai järjestelmän kanssa on mahdollista mutta vaatii oikeiden näppäinyhdistelmien muistamista. Tällöin käyttäjän on opeteltava erilaisia näppäinproseduu-

reja, joilla tiedonvaihto laitteen kanssa käy päinsä. Jos laitetta käytetään harvoin, joudutaan turvautumaan käyttöohjeeseen tai käyttäjän käsikirjaan oikean näppäinyhdistelmän varmistamiseksi. Tämä johtaa helposti kommunikoinnin välttelemiseen ja laitteen tai teknisen järjestelmän ominaisuuksien vajavaiseen hyödyntämiseen. Tavallista on myös se, että käyttöliittymän omat puutteet estävät käyttäjän luontevan opastuksen.

## 2.3 PUUTTEELLISTEN KÄYTTÖLIITTYMIEN SYYT

Hyvän käyttöliittymän suunnittelu vaatii tietoa ihmisen ja tietokoneen välisestä vuorovaikutuksesta ja sen perusmekanismeista. Vuorovaikutustapahtumaa voidaan tarkastella monesta eri näkökulmasta ja alan tutkimus on selkeästi poikkitieteellistä. Käyttöliittymän suunnittelu on siten käytännössä usean eri alan asiantuntijan yhteistyötä. Moni käytännön laitteisiin toteutettu käyttöliittymä on puolestaan yksittäisen suunnittelijan aikaansaannos, ja käyttöliittymän ominaisuudet ovat sen mukaiset.

Rakennusten teknisen järjestelmät ja laitteet eivät ole hankintahinnoiltaan suoraan vertailukelpoisia teollisuuden vastaavien tuotteiden kanssa. Rakennusalan tuotteen hankintahinnalla ja alhaisilla kustannuksilla on suuri merkitys tuotteen menekkiin. Tätä suuntausta on edistänyt myös tyypillinen rakennuttamistapa, jossa rakennuttaja ei ole rakennuksen loppukäyttäjä. Tämä näkyy myös laitteiden varustelutasossa ja laadussa. Tällä menettelyllä on ollut oma vaikutuksensa myös teknisiin laitteisiin ja niiden käyttöliittymäratkaisuihin.

Sulautetut järjestelmät ovat edullisina toteutuksina nopeassa tahdissa syrjäyttäneet vanhat tekniset ratkaisut ja tuoneet tietokonetekniikan uusille sovellusalueille. Mikroprosessorikomponenttien suorituskyky on kasvanut, mikä on mahdollistanut monipuolisten toimintojen sisällyttämisen fyysisesti pienikokoisiin ja halpoihin laitteisiin. Käyttöliittymän kannalta olennaisissa vuorovaikutuslaitteissa kehitys ei ole kuitenkaan ollut samanlaista. Esimerkiksi hyvän käyttöliittymän kannalta olennainen graafinen näyttölaite on useimmiten liian kallis kustannuskriittisiin rakennusteknisiin tuotteisiin. Jos tyydytään muutaman rivin alfanumeerisiin näyttölaitteisiin, hyvää vuorovaikutusta käyttäjän ja tietokoneen välille on vaikea toteuttaa.

## 2.4 PUHETEKNOLOGIAN MAHDOLLISUUDET

Puheteknologia koostuu monista erilaisista puheenkäsittelyn tekniikoista ja menetelmistä, joilla on merkitystä hyvän käyttöliittymän suunnittelussa ja toteutuksessa. Tällaisia osa-alueita ovat mm. puheviestien tallennus ja toisto, puheenkoodaus, puheentunnistaminen ja ymmärtäminen, puhujantunnistaminen sekä puhesynteesi ja -analyysi. Puheteknologian hyödyntämisellä rakennusten teknisissä laitteissa ja järjestelmissä ei ole mahdollista korjata kaikkia käyttöliittymissä ilmeneviä ongelmia. Puheviestintä tuo käyttöliittymään kuitenkin uuden ulottuvuuden, jota ei ole vielä täysimääräisesti hyödynnetty. Vaikka todellisen vuoropuhelun toteuttamiseen käyttäjän ja tietokoneen välillä voi kulua aikaa, puheteknologia eri muodois-

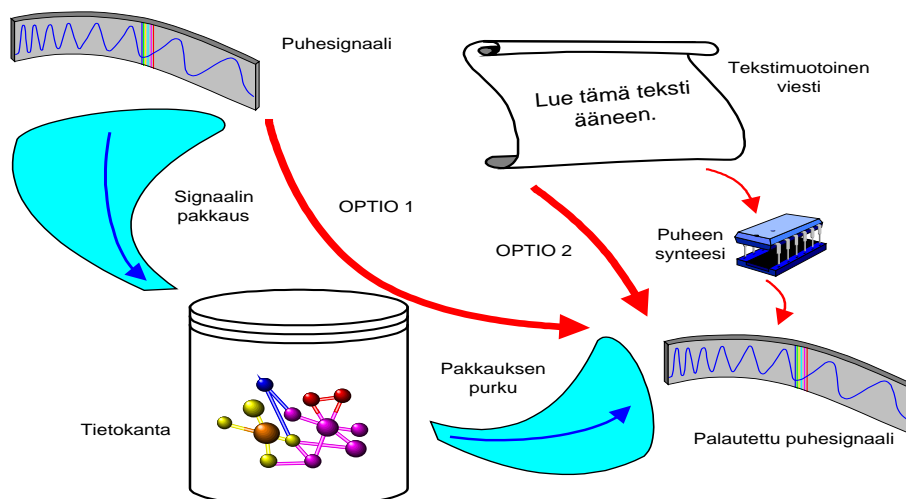
saan tarjoaa jo nyt useita merkittäviä etuja käyttöliittymän suunnitteluun. Eräs merkittävä etu on se, että puheviestinnän sisällyttäminen laitteen toimintaan voidaan toteuttaa kustannuksiltaan edullisesti. Tämä laajentaa sen soveltamismahdollisuuksia. Puheviestinnän avulla laitteista on mahdollista tehdä helppokäyttöisempiä, jolloin ne palvelevat paremmin eri käyttäjäryhmiä, kuten vammaisia tai vanhuksia, sekä soveltuvat käyttökohteisiin, joissa esimerkiksi käsiä tarvitaan muihin tehtäviin. Puheviestintä tekee vuoropuhelusta joustavamman ja mahdollistaa kommunikoinnin jopa ilman käsiä ja näköyhteyttä. Puheteknologiaa hyödyntäen on mahdollista luoda uusia palveluja ja tuotteita rakennusten eri käyttäjäryhmille ja siten helpottaa rakennuksen käyttöä.

## 3 PUHEVIESTINNÄN MENETELMIEN PERIAATTEET

### 3.1 PUHEVIESTIT

Puheviestien avulla on mahdollista antaa esimerkiksi paikantamiseen ja liikkumiseen liittyviä opasteita sekä laitteiden ja järjestelmien käyttöä helpottavia, käyttötilanteeseen sopivia havainnollisia käyttöohjeita. Puheviesti on monissa sovelluskohteissa tehokas ‘tulostusmuoto’. Puheviestien toteutus on useimmiten halvempi ja kestävämpi kuin näyttö, valaistusoloista riippumaton sekä näyttöä turvallisempi ja luotettavampi sellaisissa käyttökohteissa, joissa käyttäjän tulee samanaikaisesti seurata silmillään muita asioita. Puheviestit eivät sovi hyvin tarkoituksiin, joissa viestit ovat pitkiä tai henkilökohtaisia (esim. julkisella paikalla ilman kuulokkeita) tai kun ympäristössä on paljon melua taikka puheviesti voi häiritä toisia henkilöitä.

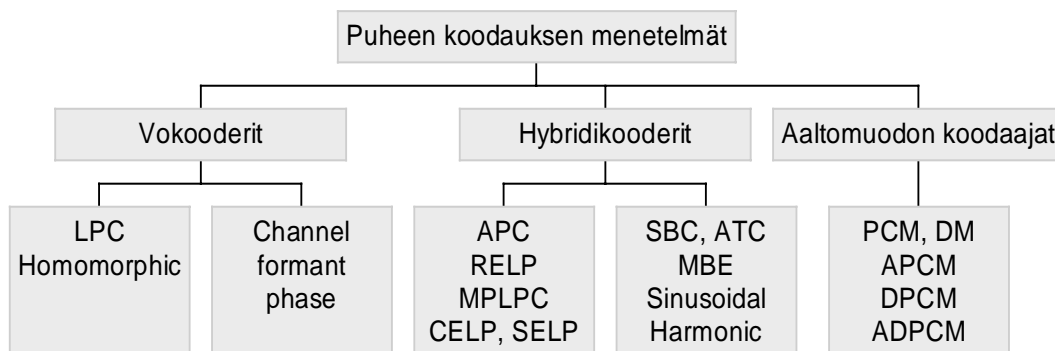
Puheviestit voidaan toteuttaa joko erilaisia puheen tallennus- ja toistovälineitä käyttämällä taikka muuttamalla tekstiä keinotekoiseksi puheeksi puhesynteesin avulla, kuva 1. Puheen tallentamisen ja toiston avulla voidaan toteuttaa edullisesti hyvän äänenlaadun omaavia puheviestejä. Niiden käyttö rajoittuu kuitenkin sovelluksiin, joissa sanomien sisältö voidaan ennakolta määrätä ja sanomat ovat kohtalaisen lyhyitä. Ennen tallentamista puhesignaali tyypillisesti pakataan pienempään tilaan eli kompressoidaan. Kompressointiastetta voidaan kasvattaa käyttämällä hidasta näytteistystaajuutta ja vähäistä bittimäärää kullekin näytteelle äänenlaadun kustannuksella. Puhesynteesin avulla sanomia voidaan tuottaa joustavammin, mutta keinotekoisen puheen äänenlaatu ei vielä yllä tallennetun puheen tasolle.



Kuva 1. Puheen synteesi kompressoidusta äänestä ja suoraan tekstistä.

### 3.1.1 Puheen tallentamiseen perustuva viestin tuottaminen

Digitaalisesti tallennettaessa puhe täytyy ensin kompressoida eli kutistaa sovelluksen kannalta riittävän pieneen muistitilaan. Puheen kompressointiin on kehitetty runsaasti menetelmiä, joista osaa on sovellettu sulautettuihin tuotteisiin, kuten matkapuhelimiin ja robottipuhelimiin. Kompressoinnista puhuttaessa käytetään myös nimitystä koodaus. Puheenkoodausmenetelmät voidaan ryhmitellä kolmeen luokkaan: aaltomuodon säilyttävät menetelmät, puheen ominaisuuksia hyödyntämään suunnitellut ratkaisut ja näiden yhdistelmät, kuva 2 [Kondozi 1994].



Kuva 2. Puheenkoodausmenetelmien luokittelu ja esimerkkejä [Kondozi 1994].

#### Aaltomuotokooderit

Aaltomuodon koodaimet (waveform coders) pyrkivät säilyttämään signaalin yleisen muodon eivätkä ole erityisesti juuri puheen koodaamiseen tarkoitettuja. Suorituskykyä voidaan hyvin arvioida signaali-kohinasuhteella, koska kvantisointi tuottaa suurimmat virheet. Aaltomuodon koodaimet toimivat menestyksekkäästi lähinnä yli 16 kb/s kompressoasteella. Esimerkkejä ovat CCITT:n standardoimat PCM (pulse-code modulation, 64 kb/s) ja ADPCM (adaptive differential PCM, 32 kb/s).

#### Vokooderit

Puheen ominaisuuksia erityisesti hyödyntävät kooderit (vocoders, voice coders) hävittävät puhesignaalin sellaisia piirteitä, jotka eivät ole kriittisiä ihmiskorvalle. Koodauksessa signaalin aaltomuotoa ei pyritä säilyttämään, mistä johtuen signaali-kohinasuhteen käyttäminen suorituskyvyn arviointiin ei ole järkevää. Syntetisoidun puheen laatua arvioidaankin subjektiivisilla kuuntelukokeilla. Vocoder koostuu kahdesta osasta, analysaattorista ja syntetisaattorista. Analysaattori mallintaa puhesignaalin matemaattisesti hyödyntäen ihmisen puheenmuodostuksen tietämystä ja tallentaa muistiin mallin parametrit. Syntetisaattori muodostaa mallista puhesignaalin. Mallinnus ei ole tarkka, vaan sen yhteydessä katoaa jonkin verran tietoa signaalin muodosta. Näitä koodausmenetelmiä käytetään erityisesti sovelluksissa, joissa puhetta on siirrettävä alle 4,8 kb/s nopeudella. Esimerkkinä on Yhdysvaltain hallituksen standardoima LPC-10 (linear predictive coding, 2,4 kb/s), jota käytetään lähinnä ei-kaupallisiin sovelluksiin.

## Hybridikooderit

Hybridikooderit pyrkivät yhdistämään näiden kahden lähestymistavan edut ja ne voidaan jakaa kahteen luokkaan, taajuus- ja aika-alueissa toimiviin ratkaisuihin. Taajuusalueessa toimiva menetelmä laskee peräkkäisille signaalisegmenteille lyhytaikaisen spektrin, jakaa sen alikaistoihin ja koodaa kunkin alikaistan erikseen. Alikaistan koodauksessa käytettävä bittimäärä määritetään adaptiivisesti tarkastellen mm. kaistan energian määrää tai signaali-kohinasuhdetta. Eniten käytetyt menetelmät tässä luokassa ovat suodatinpankkia käyttävä SBC (sub-band coding) ja taajuusmuunnosta käyttävä ATC (adaptive transform coding), joita käytetään nopeuksilla 9,6 - 16 kb/s. MBEV-tekniikalla (multi band excited vocoder) päästään jopa 4,8 kb/s nopeuksiin hyvällä äänenlaadulla.

Aika-alueessa toimivat menetelmät perustuvat aikasarjan lineaariseen ennustamiseen. Tämä pohjautuu siihen, että puheen tilastolliset ominaisuudet voidaan mallintaa hyvin tarkasti lähde-suodatinmallilla, joka olettaa, että puhe syntyy ohjaamalla lineaarista aika-muuttuvaa suodatinta periodisella pulssijonolla soinnillisissa kohdissa ja satunnaisella kohinalla soinnittomissa kohdissa. Periodisen pulssijonon käsite tulee äänihuulten värähtelystä ääntä muodostettaessa ja lineaarinen suodatin tulee ääniväylän hengitysilmaa muokkaavasta vaikutuksesta. APC (adaptive predictive coding) on ensimmäisiä tähän luokkaan kehitettyjä menetelmiä ja sitä käytettiin nopeuksilla <16 kb/s. Muita esimerkkejä ovat RELP, MPLPC, CELP ja SELP. Nopeuteen 4,8 kb/s pääsevä CELP soveltaa vektorikvantisointia ja koodikirjoja mallinnuksen jäännösvirheen koodaamisessa saavuttaakseen lisäkompressiota. Motorola on kehittänyt yhden piirin toteutuksen CELP:n eräälle versiolle VCELP (vector sum excited LPC). VCELP-tekniikkaa käytetään mm. Pohjois-Amerikassa ja Japanissa. Toinen laajaa suosiota saavuttanut esimerkki on GSM-standardin käyttämä RPE-LTP (regular-pulse excited - long-term prediction), jolla päästään nopeuteen 13 kb/s. Tästä on kehitetty hiljattain myös EFR-versio (enhanced full rate), joka koodikirjoja käyttäen parantaa äänenlaatua entisestään kompressiosuhteen pysyessä samana.

### 3.1.2 Tekstin muuttaminen puheeksi: puhesynteesi

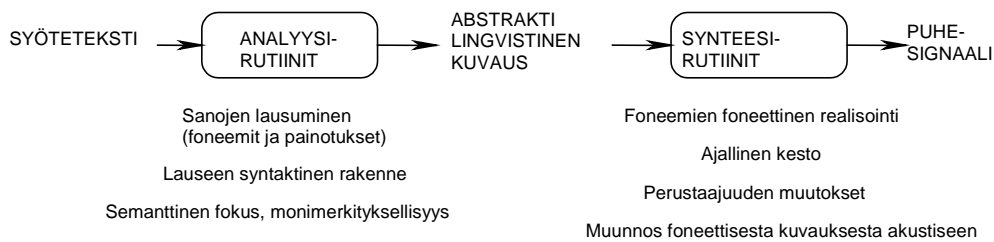
Puhesynteesi eli keinotekoinen puheen tuottaminen koostuu menetelmistä tuottaa koneella puhesignaaleja ihmiselle jonkin luonteeltaan luonnollisen kielen informaation perusteella [Karjalainen 1993]. Tämä luku perustuu lähteeseen [Klatt 1987].

Ensimmäinen puheen synteesin laite kehitettiin Bell Telephone Laboratories -laitoksessa jo 1930-luvun puolivälissä. Laite perustui sähköisiin ja mekaanisiin resonaattoreihin ja summereihin, ja sitä ohjattiin pianon kaltaisen näppäimistön ja polkimien avulla. Myöhemmin 1940-luvulla äänten synteettiseen muodostukseen käytettiin jo taajuussisältöä kuvaavia spektrogrammeja. 1960-luvun alkupuolella kehitettiin lineaarisiin suodattimiin pohjautuva akustinen teoria, minkä tuloksena toteutettiin menestyksekkäitä formantti- ja artikulaatiosyntetisaattoreita. Syntetisaattorit ovat hyödyntäneet 1970-luvulta lähtien elektronisten piirien sijasta tietokoneita tai erityisesti suunniteltuja digitaaliratkaisuja. Tärkeä virstanpylväs on

vuonna 1972 demonstroitu OVE II, jota käyttäen pystyttiin osoittamaan, että tyyppillinen kuulija ei enää erottanut, oliko kyseessä luonnollisesti vai keinotekoisesti tuotettu puheääni. Nykyisin uskotaan, että artikulaatorinen malli eli ihmisiäänen muodostuksen malli tarjoaa parhaan perustan puhesynteesille.

Tekstin automaattinen muuntaminen tekopuheeksi onnistutaan tekemään nykyisin verrattain hyvin muutamilla laboratorioprototyypeillä ja kaupallisilla tuotteilla. Edistymistä ovat edesauttaneet seuraavien alojen tulokset: lingvistiikka, äänikuvioiden luonnehdinta akustis-foneettisilla menetelmillä, havaintopsykologia, puheen tuottamisen matemaattiset mallit, rakenteellinen ohjelmointi ja tietokonetekniikka.

Puhesynteesin vaiheet on esitetty kuvassa 3. Ensimmäisessä vaiheessa analysoidaan lauserakenne ja kunkin sanan foneettinen koostumus. Foneemeja eli äänneitä on suomenkielessä on vajaat 30. Toisessa vaiheessa abstrakti lingvistinen esitystapa muunnetaan puheen aaltomuodoksi. Muuntamisessa huomioidaan myös äänneelle haluttu kesto-aika, perustaajuus ja äänenpainotus.

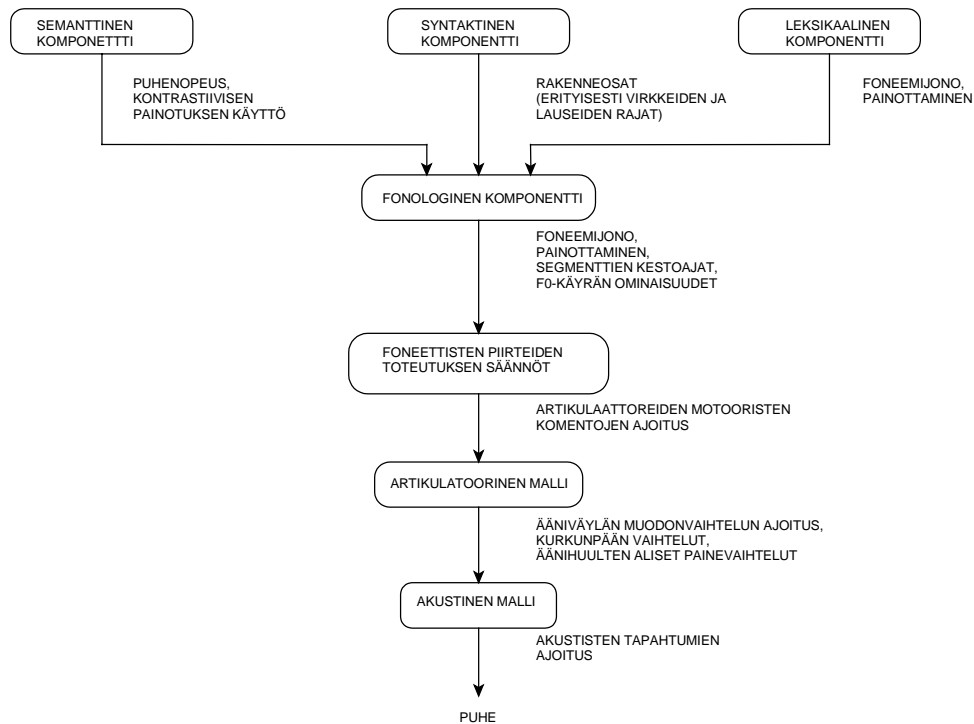


Kuva 3. Puhesynteesin periaatekaavio [Klatt 1987].

Intuitiivisesti saattaisi näyttää siltä, että pienemmissä sovelluksissa nykyisen muistiteknologian ansiosta olisi käytännöllistä tallentaa tietyn kielen sanojen aaltomuotoja muistiin ja sitten yksinkertaisesti koota niitä peräkkäin lauseiden tuottamiseksi. Tämä lähestymistapa tuottaa kuitenkin huonoja tuloksia, koska sanat lausutaan eri tavalla erikseen kuin lauseen yhteydessä. Ensinnäkin, lauseen yhteydessä luonnollisesti sanottuina sanojen pituus on vain puolet siitä, mitä ne ovat erikseen sanottuina. Toisekseen, kieliopillisista ja sisällöllisistä tekijöistä riippuvat lauseen painotuskuvio, rytmi ja sävelkulku muuttuvat luonnottoman kuuloisiksi. Kolmanneksi, sanojen alku- ja loppuosat osittain sulautuvat yhteen luonnollisessa lauseessa, mikä on tärkeää luonnollisuuden ja ymmärrettävyyden takaamiseksi. Ainoa tyydyttävä tapa simuloida näitä vaikutuksia on kulkea läpi kieliopillisen, fonologisen ja foneettisen muunnoksen kautta.

Käymättä tarkemmin lävitse yksityiskohtia kuvassa 4 esitetään yksinkertaistettu kaavio puheen synteesin lingvistisestä menetelmäpuutteesta.





Kuva 4. Puheen synteessin lingvistinen menetelmäpuite [Klatt 1987].

Syntetisoidun puheen laadun arviointiin voidaan käyttää seuraavia kriteerejä:

- ymmärrettävyys (intelligibility)
  - *yksittäiset sanat*. Esimerkiksi loppusointitestissä kuulija valitsee jonkin kuudesta vaihtoehdosta, jotka eroavat toisistaan vain alkutai loppukonsonantissa.
  - *sanat lauseyhteydessä*. Esimerkiksi kuulovammaisten testit, testit puheen häiritsemiseksi kohinalla, merkityksettömien lauseiden testit.
  - *kappaleiden ymmärtäminen*. Esimerkiksi luetun ja kuullun ymmärtämisen testit.
- luonnollisuus (naturalness)
  - on vaikeasti mitattava suure.
  - esimerkiksi eri laitteilla tuotettujen lauseiden vertailu pareittain, joista valitaan parempi
  - luonnollisuutta ei saa sekoittaa ymmärrettävyyteen (vaikka korreloivatkin toisiinsa)
- sopivuus tiettyyn sovellukseen (suitability)
  - tärkein päätöskriteeri.

Tekstistä puheeksi muuntaminen on edelleen aktiivisen tutkimuksen kohteena. Alla luetellaan eräitä tärkeimpiä tutkimusongelmia:

- puheeksi muunnettavan tekstin analysointi
  - *erikoisteksti*. Numerot, erikoiset merkkijonot ja lyhenteet.

- *kielioppi ja merkityssisältö*. Erityisesti puheen synteisiin erikoistunut kieliopin analyysi, merkityssisällön arviointi tekstistä (“She hit the old man with the umbrella.”).
- *äänteiden ja painotusten ennakointi*. Äänneyhdistelmät, hankalien erisnimien lausuminen.
- synteesi
  - *prosodiikka*. Perustaajuuden ohjaus (äänenkorkeus), paremmat ajallisen keston säännöt, mekanismit vaihtelun mallintamiseksi sääntöihin.
  - *phonologia*. Tarkempi lausetason äänten hallinta.
  - *akustis-foneettiikka*. Ymmärrettävyyden parantaminen pienillä äänellisillä vihjeillä.
  - *äänen laatu*. Erityisesti syntetisoidun naisenään laadun parantaminen.

## 3.2 AUTOMAATTINEN PUHEEN TUNNISTAMINEN

Luku 3.2 perustuu lähteeseen [Rabiner 1993].

### 3.2.1 Historiaa

Automaattinen puheen tunnistaminen (ASR, automatic speech recognition) on ollut tutkimuskohteena jo 50 vuotta. Suuresta panostuksesta huolimatta ollaan vielä kaukana tavoitteesta luoda kone, joka ymmärtää kaikkien puhujien keskustelua kaikissa ympäristöissä. Tämän moniteknologisen ongelman mittasuhteista saa käsityksen tiedostamalla, että ratkaisua on etsitty yhdistelemällä yhden tai useamman seuraavan alan tuloksia: signaalin käsittely, fysiikan akustiikka, hahmontunnistus, tietoliikenne- ja informaatioteoria, lingvistiikka, fysiologia, tietokonetekniikka ja tietojenkäsittelytiede sekä psykologia.

Ensimmäiset laajempaa huomiota saavuttaneet tulokset tehtiin Bellin laboratoriossa 1952 ja RCA:n laboratoriossa 1956. Bellissä kehitetty laite kykeni tunnistamaan yhden puhujan lausumat numerot 0 - 9. RCA:n laite puolestaan pystyi tunnistamaan 10 erilaista tavua, jotka esiintyivät 10 monitavuisessa sanassa. Tulokset saavutettiin molemmissa hyödyntämällä vokaalien spektraalista tietoa. Vuonna 1959 University College in England -tutkimuslaitoksessa kyettiin tunnistamaan vokaalien (4) lisäksi myös konsonanteja (9). Tässä tutkimuksessa käytettiin ensimmäistä kertaa tilastollista tietoa foneemien sallituista järjestyksistä sanoissa. Myös MIT Lincolnin laboratoriossa saatiin tuloksia vuonna 1959 vokaalien tunnistamisesta kahden konsonantin välistä; järjestelmä toimi puhujariippumattomasti.

Japanilaiset tulivat mukaan puheentunnistuksen tutkimukseen 1960-luvun alussa saavuttamalla useita mielenkiintoisia tuloksia. Tunnetuimpia julkistuksia esittivät Radio Research Lab in Tokyo (1961), Kyoto University (1962) ja NEC Laboratories (1963), jotka kehittivät erikoiselektronikkaa puheen tunnistukseen. Kolme tärkeintä projektia toteutettiin 1960-luvulla seuraavissa paikoissa: RCA Laborato-

ries Yhdysvalloissa (puheen nopeuden vaihtelu, sanojen alku tai loppu), Neuvostoliitossa (dynamic time warping, jatkuva puhe) ja Carnegie Mellon University (jatkuva puhe). Jälkimmäisin laitos on edelleen maailman johtava jatkuvan puheen tutkimuslaitos.

1970-luvulla saavutettiin useita merkittäviä tuloksia Neuvostoliitossa, Japanissa ja Yhdysvalloissa. Yksittäisten sanojen tunnistaminen kehitettiin käyttökelpoiseksi teknologiaksi. Neuvostoliittolaiset osoittivat, kuinka hahmontunnistusta käytetään puheentunnistamisessa. Japanilaiset kehittivät dynaamisen ohjelmoinnin tekniikkaa. Yhdysvaltalaiset osoittivat, kuinka puheen koodaamisessa käytettyä LPC-tekniikkaa (linear predictive coding) voidaan soveltaa myös puheen tunnistuksessa. IBM:n tutkijat ratkoivat suuren sanavaraston problematiikkaa ja AT&T Bell Labsissa tutkittiin puhujariippumatonta tunnistusta.

1980-luvulla tutkimuksen pääpaino oli edelleen jatkuvan puheen tunnistamisessa (NEC/Japani, JSRU/Englanti, Bell Labs/USA). Vuosikymmenen puolivälissä yleistyivät äänteiden tilastollista mallintamista tekevä kätkeyt Markovin mallit (HMM, Hidden Markov Model) johtavana menetelmänä. Neuroverkkoja opetettiin käyttämään puheen tunnistuksessa. Jatkuvan puheen ja laajan sanavaraston tunnistaminen yleistyi Yhdysvalloissa voimakkaasti DARPA:n (Defence Advanced Research Projects Agency) käynnistettyä laajan tutkimusohjelman. Parhaat tulokset saavutettiin seuraavissa paikoissa: CMU (SPHINX-järjestelmä), BBN (BYBLOS-järjestelmä), Lincoln Labs, SRI, MIT ja AT&T Bell Labs.

1990-luvulla on keskitytty seuraaviin kohteisiin: jatkuva puhe, erittäin suuret sanavarastot (kymmeniä tuhansia), puhujasta riippumattomuus ja sovellukset. Tärkeitä tutkimusaiheita ovat myös ulkoisten audiohäiriöiden suodattaminen puhesignaalista ja dialogin hallinta ihmisen ja tietokoneen välillä.

### **3.2.2 Puheentunnistusjärjestelmien luokittelu**

Perusratkaisultaan puheentunnistusjärjestelmiä on kahdenlaisia: puhekomentojärjestelmiä, jotka pystyvät tunnistamaan yksittäisiä sanoja tai sanayhdistelmiä, sekä laajempaan puheen tunnistukseen pystyviä sanelujärjestelmiä, jotka pystyvät parhaimmillaan tunnistamaan normaalia puhetta. Eräänlaisena välimuotona voidaan pitää järjestelmiä, jotka pystyvät tunnistamaan ennalta määrättyjä avainsanoja normaalista puheesta (keyword spotting). Tällaisen järjestelmän avulla komennot voidaan antaa esim. muodossa 'Muista panna *ulkovalot päälle* jo *kello 18*.'

Puhekomentojärjestelmät mahdollistavat käyttäjän navigoinnin puheen avulla tietokoneohjelmistojen 'valikkoviidakoissa', laitteiden päälle/pois-kytkemisen ennalta ohjelmoitujen kommentojen avulla, puhelinumerovalinnan nimien perusteella ja vastaavankaltaiset yksinkertaiset toiminnot.

Sanelujärjestelmät mahdollistavat edellistä monipuolisemmat toiminnot, kuten tietojen syötön tietokantoihin ja tekstin syötön tekstinkäsittelyjärjestelmiin. Soveltamismahdollisuudet ovat lähes rajattomat. Parhaimmat jatkuvan puheen tunnistukseen tarkoitetut järjestelmät pääsevät 95 %:n tunnistustarkkuuteen. Voice Infor-

mation Associates on arvioinut, että puheentunnistustuotteiden markkinat kasvavat vuoteen 2001 mennessä 810 miljoonaan Yhdysvaltain dollariin [Thyfault 1998]. Joissakin arvioissa markkinoiden arvon on ennustettu vuosisadan vaihteessa olevan jopa 3,8 miljardia USD [Forman 1997]. Lähiaikojen tavoitteena on kehittää tosiaikaiseen kielenkääntämiseen ja keskusteluun pystyviä järjestelmiä.

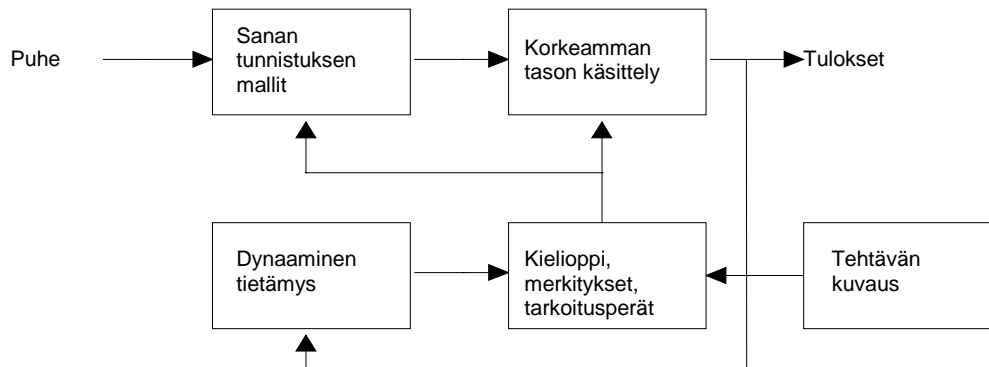
Taulukossa 1 on esitetty erilaisten puheentunnistusjärjestelmien periaatteita sekä niiden etuja ja haittoja.

*Taulukko 1. Erilaisten puheentunnistusjärjestelmien etuja ja haittoja.*

	<b>Puhekomento- järjestelmät</b>	<b>Sanelujärjestelmät</b>	
		<b>Sana kerrallaan</b>	<b>Jatkuva puhe</b>
<b>Puheen syöttö</b>	Käyttäjät lausuvat yksittäisen sanan tai sanaryhmän komennon antaessaan.	Käyttäjien tulee lausua sanat hitaasti ja huolellisesti.	Käyttäjät lausuvat sanat normaalin puheen tavoin.
<b>Edut</b>	Erittäin tarkka	Soveltuu sanelunomaiseen käyttöön. Suuri sanavarasto Kohtuulliset prosessointivaatimukset	Soveltuu sanelumaiseen käyttöön. Suuri sanavarasto Helppo ja luonnollinen käyttää
<b>Haitat</b>	Rajalliset soveltamiskohteet Ei sovi sanelunomaiseen käyttöön Pieni sanavarasto	Rajallinen tarkkuus Hidas Hankalakäyttöinen Luonnoton puhetapa	Rajallinen tarkkuus Erittäin suuret prosessointivaatimukset

### 3.2.3 Puheentunnistuksen perusteet lyhyesti

Puheen tunnistuksen yleinen lohkokaavio on esitetty kuvassa 5.

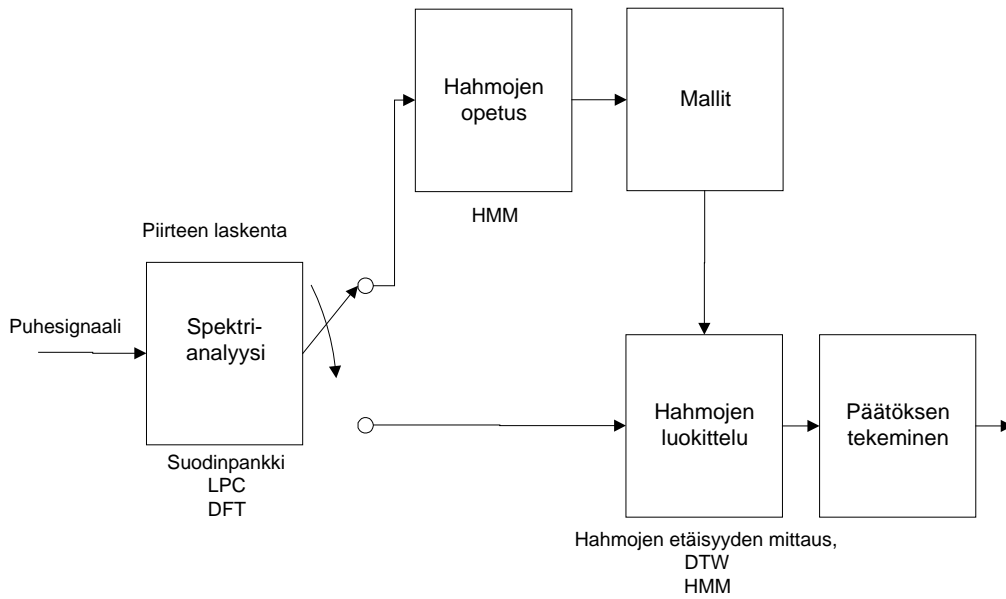


Kuva 5. Puheen tunnistuksen yleinen lohkokaavio [Rabiner 1993].

Ensimmäinen lohko tekee alustavat sanan tunnistukset äänisignaalin perusteella. Näistä muodostetaan muiden lohkojen toimesta sanaketju, joka on mielekäs kieliopin, merkityssisällön ja sovelluksen kannalta. Nämä korkeamman tason käsitteet prosessoidaan laskentamoduulilla, joka käyttää dynaamista tietämyksen esitystapaa. Kuvassa esiintyvää takaisinkytkentää käytetään tämän tietämyksen muutteluun, sillä sen avulla voidaan rajoittaa tehokkaasti lauseiden tulkinta- vaihtoehtoja jatkossa.

Suorituskykyä voidaan kasvattaa lisäämällä järjestelmään dialogin hallinnan moduuli, joka varmentaa käyttäjältä tunnistustuloksen pyytäen tarvittaessa uudelleenlausumisen. Vapaamuotoisen lauseen tunnistus on huomattavasti vaikeampaa kuin määrämuotoisen puheen, kuten komentosanaketjun. Luotettavuutta saavutetaan tässä lisää käyttämällä tilanneherkkää tunnistusta, jossa sanan tunnistushetkellä käytetään kieliopin ja sanaston tarjoamia rajoitteita vähentämään vaihtoehtojen määrää.

Ennen tunnistukseen siirtymistä käyttäjän on kuitenkin opetettava järjestelmä esittelemällä sille kattava joukko ääninäytteitä. Näytteille tehdään spektrianalyysi, ja tuloksia käytetään ääniteiden mallien luomiseen ja myöhemmin tunnistukseen. Kuva 6 havainnollistaa tunnistimen kaksijakoista, opetustilaan ja tunnistustilaan sijoittuvaa toimintaa.

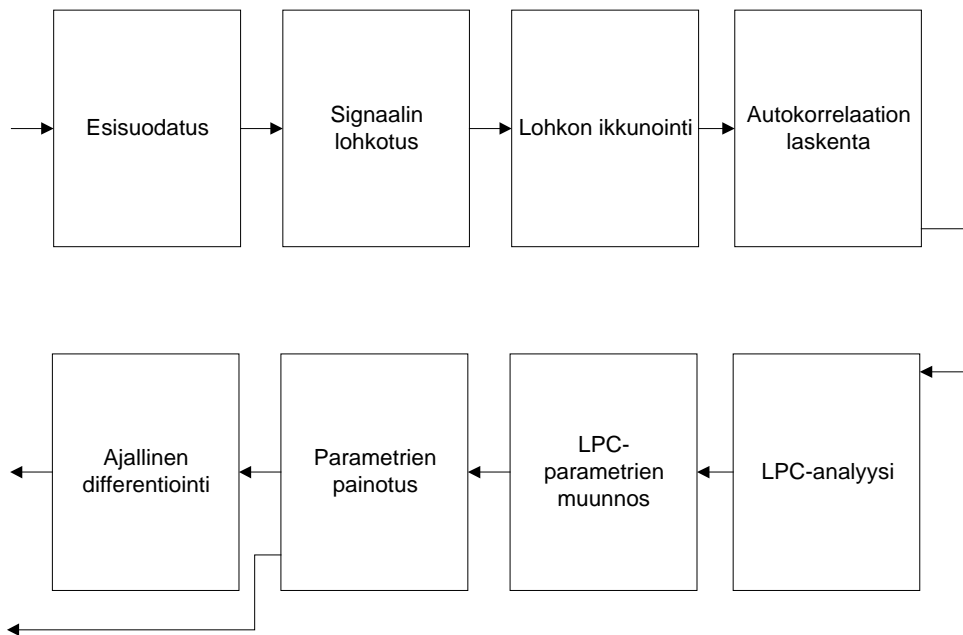


Kuva 6. Sanantunnistimen yksinkertaistettu lohkokkaavio [Rabiner 1993].

Käytännössä järjestelmät voidaan jakaa kahteen ryhmään, sanan tunnistajiin ja foneemin tunnistajiin sen mukaan, millaisia äänteitä niille opetetaan. Ensimmäisen ryhmän järjestelmät pyrkivät oppimaan mallintamalla kokonaisen sanan yhdellä mallilla. Sanat voidaan tämän jälkeen tunnistaa vertaamalla puheesta laskettua luonnehdintaa (mallia) mallitietokantaan.

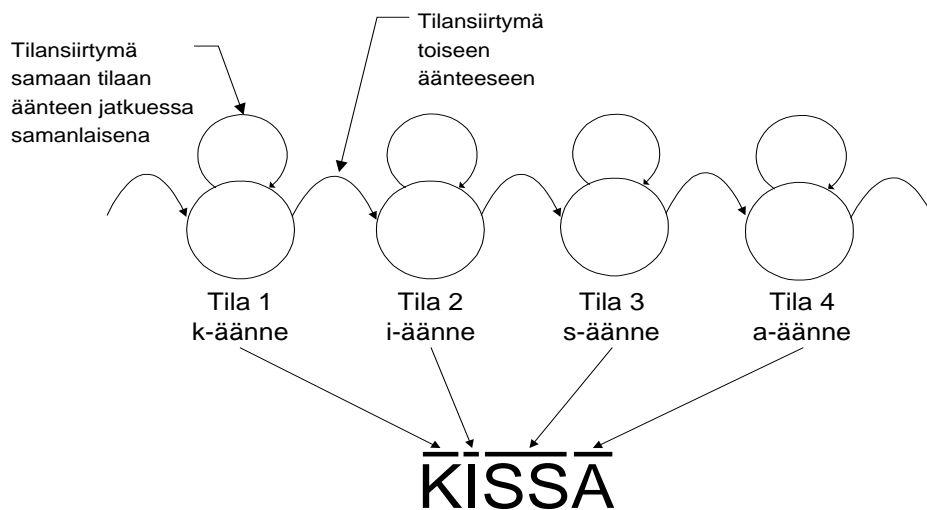
Toisen ryhmän järjestelmät mallintavat kunkin erilaisen äänten eli foneemin yhdellä mallilla. Sanojen tunnistaminen tapahtuu tämän jälkeen tunnistamalla ensin puheen peräkkäiset äänteet ja sitten vertaamalla tunnistettua äänneketjua tietokantaan tallennettujen sanojen äänneketjuihin. Sanaston äänneketjuja ei välttämättä tallenneta etukäteen tietokantaan, vaan ne voidaan muodostaa tekstimuodossa olevista sanoista foneemisääntöjä käyttämällä vasta tunnistushetkellä.

Äänisignaalin opettamiseksi tai tunnistamiseksi näyte tulee ensin muuntaa sellaiseen muotoon, joka ei vie paljon muistitilaa ja joka parhaalla tavalla luonnehtii äänen ominaisuuksia. Yleisin menetelmä on laskea äänisegmentille kepstri (cepstrum), joka yleensä muodostetaan taajuussisältöä kuvaavasta tehotiheyspektristä. Puhujasta ja äänen korkeudesta riippumattomuuden saavuttamiseksi kepstriä täytyy jälkikäsitellä lifterointimenetelmällä. Tuloksena on yhtä äänisegmenttiä luonnehtiva piirrevektori, jossa on yleensä noin 10 elementtiä. Vektoriin lisätään yleensä vielä signaalisegmentin tehollisarvo voimakkuusvaihteluiden mallintamiseksi. Puheen dynamiikan pitkäaikaisempien ominaisuuksien mallintamiseksi lasketaan peräkkäisten segmenttien piirrevektorien väliset erovektorit. Tällä tavoin saadaan segmenttiä lopullisesti kuvaava piirrevektori, jossa on siis noin 25 elementtiä. Kuvassa 7 esitetään piiirteen laskennan vaiheet.



Kuva 7. Äänisegmentin mallintaminen piirrevektorilla [Rabiner 1993].

Tämän hetken käytetyin menetelmä äänten dynaamisten ominaisuuksien mallintamiseksi on tilastollista lähestymistapaa soveltava kätketyt Markovin mallit (HMM, Hidden Markov Model). Syötteenä malliin ovat edellä kuvatut piirrevektorit. Ennen HMM-mallin käyttämistä tunnistuksessa se on opetettava kattavalla ääninäytteistöllä. Kuten edellä todettiin, äänneyksikkönä voi toimia kokonainen sana tai foneemi. Tällöin siis saadaan yksi HMM-malli jokaiselle tunnistettavalle sanalle tai foneemille. HMM-malliin liittyy joukko parametreja, jotka opetusvaiheessa lasketaan ääniaineistosta. Parametrit ovat itse asiassa todennäköisyyksiä, jotka kuvastavat sitä, millä todennäköisyydellä tietyt osäänteet seuraavat toisiaan puheessa. Kuvassa 8 on esimerkkinä puheen tunnistuksessa paljon käytetty left-right-tyyppinen HMM-malli, jota voidaan soveltaa esimerkiksi foneemin aikadynamiikan mallinnukseen. Jokaiseen mallin tiloja yhdistävään nuolella merkittyyyn siirtymään liittyy kaksi todennäköisyysparametria. Opettaminen on aikaa vievä operaatio, ja siihen käytetään iteratiivista estimate-maximize-algoritmia. Varsinaisessa tunnistuksessa käytetään mm. Viterbi-algoritmia, joka on eräänlainen haku-algoritmi.



Kuva 8. Left-right-tyyppinen HMM-malli.

### 3.2.4 Haastavimmat tutkimuskohteet

Erikseen lausuttujen komentosanojen tunnistaminen on nykyteknologialla verrattain luotettavaa, erityisesti mikäli tilannerajoitteilla voidaan pienentää mahdollisuuksien määrää tunnistustilanteessa. Esimerkiksi numeroiden 0 - 9 tunnistaminen onnistuu lähes virheettömästi, mikä on mahdollistanut puheella ohjatun numerovalinnan ja pikavalinnan eräissä matkapuhelimissa. Sen sijaan jatkuvan puheen tunnistaminen on vaikeampaa sanarajojen hämärtyksen vuoksi. Samoin vaadittava laskenta-aika on selvästi suurempi, mikä hankaloittaa reaaliaikaisten sovellusten toteuttamista. Jatkuvan puheen tunnistamisessa käytetään sanamallien lisäksi kielimallia, joka sisältää kielioppiin ja merkitysisältöön liittyviä rajoitteita.

Toinen haaste on erittäin suurten sanavarastojen (kymmeniä tuhansia sanoja) sovellukset. Laajojen sanastojen tapauksessa tunnistin usein sekoittaa sanoja toisiinsa ja laskenta-aika kasvaa.

Jokaisen henkilön äänenmuodostuksessa on vaihtelua siinä määrin, ettei samaa sanaa voi sanoa kahta kertaa täsmälleen samalla tavalla. Tämän vuoksi tilastollisen mallinnuksen käyttäminen tuottaa parhaan suorituskyvyn. Suunniteltaessa tunnistimesta monen käyttäjän järjestelmää ongelma kasvaa edelleen. Puhujariippumaton puheen tunnistaminen on edellytys laajoja käyttäjäkuntia koskeville sovelluksille, kuten tietokoneistetuille puhelinpalveluille.

Ympäristön hälyäänet muuttavat äänisignaalia voimakkaasti ja heikentävät suorituskykyä merkittävästi. Häiriösignaalien suodattamiseen ja puhesignaalin korostamiseen on kehitetty runsaasti menetelmiä ja tutkimus on edelleen aktiivista.



### 3.3 PUHUJANTUNNISTUS

Puhujantunnistus on yksi nopeasti kehittyvistä biometrisistä henkilöntunnistusteknologioista. Tutkimusten mukaan puhujantunnistus on useimmissa turvajärjestelmissä merkittävästi turvallisempi kuin perinteinen salasanasuojaus. Yleensä puheen perusteella tapahtuva henkilöntunnistus liitetään osaksi useampitasoista tunnistusjärjestelmää. Muina tunnistuskeinoina käytetään esimerkiksi salasanaa, kasvonpiirteiden tunnistusta tai sormenjälkiä. Taulukossa 2 vertaillaan biometrinen henkilöntunnistusmenetelmien ominaisuuksia.

*Taulukko 2. Biometrinen henkilöntunnistusmenetelmien ominaisuuksia [Phillips 1997].*

Biometrinen menetelmä	Tunnistuspisteen hinta	Edut	Haitat	Sovelluskohteet
Kasvonpiirteet	1 500 USD	<ul style="list-style-type: none"> <li>• helppo</li> <li>• nopea</li> <li>• yksi halvimmistä menetelmistä</li> </ul>	<ul style="list-style-type: none"> <li>• altis hämäysyrityksille</li> <li>• heikko valaistus voi vaikuttaa tunnistukseen</li> </ul>	<ul style="list-style-type: none"> <li>• yleinen</li> </ul>
Käsiala	1 000 USD	<ul style="list-style-type: none"> <li>• halpa</li> </ul>	<ul style="list-style-type: none"> <li>• fyysinen kunto ja tunteet voivat vaikuttaa</li> </ul>	<ul style="list-style-type: none"> <li>• teollisuus</li> </ul>
Puhe	1 200 USD	<ul style="list-style-type: none"> <li>• halpa</li> <li>• soveltuu hyvin etäkäyttöön</li> </ul>	<ul style="list-style-type: none"> <li>• hidas</li> <li>• fyysinen kunto ja tunteet voivat vaikuttaa</li> </ul>	<ul style="list-style-type: none"> <li>• etäpalvelut ja toiminnot</li> </ul>
Sormenjäljet	1 200 USD	<ul style="list-style-type: none"> <li>• halpa</li> <li>• erittäin turvallinen</li> </ul>	<ul style="list-style-type: none"> <li>• haavat ja lika voi vaikuttaa tunnistukseen</li> </ul>	<ul style="list-style-type: none"> <li>• poliisitoimi ja vastaavat</li> </ul>
Silmän verkkokalvo	5 000 USD	<ul style="list-style-type: none"> <li>• erittäin vaikea hämätä</li> </ul>	<ul style="list-style-type: none"> <li>• intrusiivinen ja epämurkava</li> </ul>	<ul style="list-style-type: none"> <li>• ydinvoimalat</li> <li>• lääketiede</li> <li>• vankeinhoito</li> </ul>

Järjestelmässä puhe muunnetaan ensin digitaalisiksi signaaleiksi, jotka sen jälkeen analysoidaan puheen yksilöivien ominaisuuksien mallintamiseksi. Saatua tulosta verrataan aikaisemmin tallennettuihin malleihin puhujan henkilöllisyyden varmistamiseksi. Puhujantunnistusjärjestelmällä voidaan esimerkiksi varmistua siitä, kuuluuko henkilönumeron lausunut henkilö tosiasiallisesti yrityksen henkilökuntaan.

Puhesignaali on monimutkainen funktio puhujan fysikaalisista ominaisuuksista, kuten äänihuulten ja -kanavan mittasuhteista, ympäristöstä (esimerkiksi taustamelu ja siirtokanava) ja tunnetilasta (esimerkiksi fysikaalinen ja henkinen stressi) [Naik 1990]. Näistä syistä johtuen puhesignaalissa esiintyy suurta vaihtelua puhujien välillä ja jopa samallakin puhujalla eri ajanhetkinä. Tunnistavien ja erottelevien piirteiden valinta on kriittistä tunnistimen suorituskyvylle. Yleisesti ottaen piirteiltä vaaditaan seuraavia ominaisuuksia:

- erottelee puhujat, mutta on epäherkkä puhujan äänen vaihteluille
- ovat helposti mitattavissa puhesignaalista
- ovat stabiileja ajan suhteen
- ovat vaikeita väärentää.

Ihminen käyttää puhujan tunnistamisessa varsinaisen äänen ominaisuuksien lisäksi sellaisia “korkean tason” piirteitä, kuten murre, asiayhteys, puhetyyli ja puhujan tunnetilat. Tutkijat ovat koettaneet löytää näiden piirteiden akustisia vastineita, mutta toistaiseksi heikolla menestyksellä. Myös näiden luonnehdintojen opettaminen tunnistimelle on osoittautunut vaikeaksi. Tästä johtuen käytetään akustisesta signaalista johdettuja “matalan tason” parametreja, kuten äänenkorkeutta, spektrin arvoja, formanttitaajuuksia ja signaalin energiaprofiileja. Näiden suureiden aikadynaamiikkaa verrataan tietokantaan periaatteessa samoilla menetelmillä kuin mitä puheentunnistuksessa käytetään (DTW, HMM).

Puhujan- ja puheentunnistamisessa käytetään samankaltaisia menetelmiä. Tämä on sinänsä ymmärrettävää, koska nykyiset puheentunnistimet ovat hyvin puhujariippuvaisia. Tätä havaintoa hyödyntämällä puhujantunnistin voidaankin opettaa puhujakohtaisilla ääninäytteillä, jolloin tunnistetaan samanaikaisesti sekä puhe että puhuja.

## 4 PUHETEKNOLOGIAN KAUPALLINEN LAITE- JA OHJELMISTOTARJONTA

### 4.1 ELEKTRONISET KOMPONENTIT JA OHJELMISTOT

#### 4.1.1 Puheviestit

##### Äänentallennuspiirit

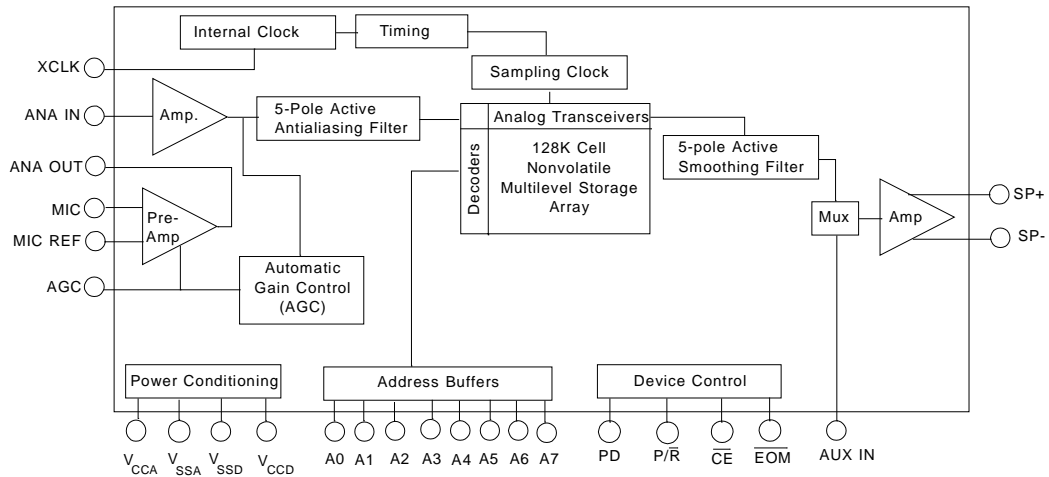
Perinteisiä puheentallennus- ja toistovälineitä ovat mm. nauhurit. Yhä laajempaan käyttöön ovat kuitenkin tulleet elektronisten piirien avulla toteutetut ratkaisut, joissa voidaan käyttää joko analogisia tai digitaalisia äänentallennuspiirejä. Seuraavassa on esitetty esimerkkejä molemmista piiriteknisistä toteutustavoista.

*Information Storage Devices, Inc. (ISD)* valmistaa patentoimallaan ChipCorder®-teknologialla yhdellä mikropiirillä toteutettuja äänentallennus- ja -toistopiirejä. Mikropiirien ensimmäiset versiot esiteltiin jo vuonna 1991 [Goodenough 1996]. Valmistaja lupaa tuotteilleen yli 100 000 luotettavaa tallennuskertaa. Mikropiiri säilyttää tallennetut viestit ilman teholähdettä, mikä on eduksi virtakatkosten ja pariston vaihtotilanteiden varalta varsinkin turva- ja hälytysjärjestelmissä.

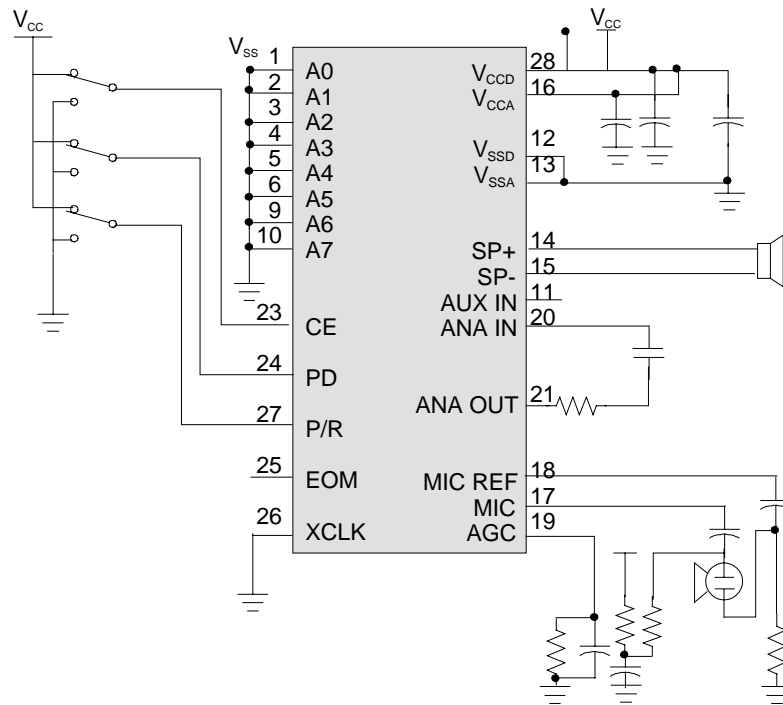
Yrityksen mukaan ChipCorder-teknologian keskeisiä etuja äänen liittämässä tuotteisiin ovat mm.

- Luonnollinen äänentoisto luo tuotteesta korkealaatuisen vaikutelman.
- Yhden mikropiirin toteutus yksinkertaistaa suunnitteluprosessia, mahdollistaa tilaasäästävät tuotteet sekä lyhentää tuotekehitysaikaa.
- Äänen tallennus pysyväismuistiin varmistaa viestien säilymisen myös pariston vaihdon aikana.
- Tallennuksen uusimismahdollisuus pidentää tuotteiden ikää ja lisää näin tuotteiden arvoa.
- Suuri tallennustiheys mahdollistaa kustannuksiltaan tehokkaat tuotteet kustannustietoisille markkinoille.

ChipCorder-teknologia mahdollistaa äänen tallentamisen ilman A/D-muunnosta EEPROM-muistiin. Tämä minimoi tarvittavien ulkoisten piirien määrän. Täydellisten tallennus- ja toistotoimintojen toteuttamiseen tarvitaan lisäksi vain mikrofooni, kaiutin, teholähde ja muutamia passiivisia komponentteja, kuvat 9 ja 10.



Kuva 9. ISD1000A-sarjan piirien lohkokkaavio [ChipCorder® Data Book].



Kuva 10. ISD1016/1020A-piirien sovellusesimerkki [ChipCorder® Data Book].

Koska EEPROM-piirit ovat pysyvämuistia, laitteet eivät tarvitse paristovarmistusta äänitettyjen viestien säilyttämiseen. Tämän vuoksi tehonkulutusta on voitu merkittävästi pienentää, mikä on tärkeää mm. kannettavissa paristokäyttöisissä laitteissa.

ChipCorder-tekniikalla yhteen muistisoluuun voidaan tallentaa tavanomaisen yhden bitin, 1/0, (kaksi jännitetasoa), sijasta 256 eri jännitetasoa, mikä tavanomaisiin digitaalisiin ratkaisuihin verrattuna mahdollistaa muistisolua kohden jopa kahdeksankertaisen tietomäärän tallennuksen.

ISD:n tuotesarjojen ominaisuuksia on esitetty taulukossa 3. Valmistajalta on saatavissa myös useita tuotteiden testaus- ja kehitysjärjestelmiä.

*Taulukko 3. ISD:n analogisten äämentallennus- ja -toistopiirien ominaisuuksia [ChipCorder® Data Book].*

Tuotesarja	Tuotenumero	Tallennusaika	Näytteenotto-taajuus (kHz)	Tyypillinen suodattimen kaistanleveys (kHz)
<b>ISD1000A</b>	ISD1016A	16 s	8.0	3.4
	ISD1020A	20 s	6.4	2.7
<b>ISD1100</b>	ISD1110	10 s	6.4	2.6
	ISD1112	12 s	5.3	2.2
<b>ISD1200</b>	ISD1210	10 s	6.4	2.6
	ISD1212	12 s	5.3	2.2
<b>ISD1400</b>	ISD1416	16 s	8.0	3.3
	ISD1420	20 s	6.4	2.6
<b>ISD2500</b>	ISD2560	60 s	8.0	3.4
	ISD2575	75 s	6.4	2.7
	ISD2590	90 s	5.3	2.3
	ISD25120	120 s	4.0	1.7
	ISD2532	32 s	8.0	3.4
	ISD2540	40 s	6.4	2.7
	ISD2548	48 s	5.3	2.3
	ISD2564	64 s	4.0	1.7

*Oki Semiconductor (OKI)* valmistaa 'solid-state recorder'-piireiksi kutsumiaan äämentallennukseen ja -toistoon tarkoitettuja IC-piirejä. Piireissä käytetään joko ADPCM (Adaptive Differential Pulse Code Modulation) tai SBC (Sub Band Coding) -puheenkoodausmenetelmää. Tuotannossa on sekä pelkästään toistoon että tallennukseen ja toistoon tarkoitettuja piirejä. Piirit sisältävät A/D- ja D/A-muunnitimet ja muut tarvittavat toiminnot, joten kaiuttimen, kaiutinvahvistimen, datan tallentamiseen soveltuvan muistin sekä mikrofonin liittämisen jälkeen järjestelmällä voidaan toistaa ja tallentaa ääntä nauhurin tavoin.

Äänen laatu ja toisaalta äänen vaatima tallennustila riippuvat voimakkaasti käytetystä näytteenottotaajuudesta. Näytteenottotaajuuden kasvattaminen parantaa äänen laatua mutta lisää samalla tarvittavaa muistin määrää ja siten myös kustannuksia. Tuotteita valmistettaessa joudutaan valinta tekemään sovelluksessa tarvittavan äänenlaadun ja kustannusten välillä. ADPCM-menetelmää käytettäessä sovelluksen tarvitsema muistin määrä saadaan laskennallisesti seuraavasta kaavasta:

$$\text{Muistin määrä (bitteinä)} = \text{äänen tallennusaika (s)} \times \text{näytteenottotaajuus (Hz)} \times \text{datan formaatti (3 tai 4 bitin ADPCM)}$$

Esimerkiksi 32 s äänitallenne näytteenottotaajuudella 8 kHz tarvitsee muistitilaa: 32 s x 8 kHz x 4 bittiä = 1 024 000 bittiä eli 128 kB.

OKIn valmistamien piirien ominaisuuksia on esitetty taulukossa 4.

*Taulukko 4. OKIn digitaalisten äänen tallennus- ja toistopiirien ominaisuuksia [Oki Semiconductor].*

Tuote-numero	Tuote-kuvaus	Näytteenotto-taajuus (kHz)	Ulkoisen muisti	Toisto-aika (SA/MCU)	Sisäinen A/D, D/A	Maksimi kuormitus-/varallaolo-virta	Käyttö-lämpötila
<b>MSM5218</b>	Recorder	4 ~ 16	RAM / ROM	variable	10-bit D/A	6 mA	-30 ~ +70 °C
<b>MSM6388</b>	Recorder, Stand alone / MCU	4 ~ 36	Serial Register / DRAM / SRAM	120 sec / variable	12-bit A/D 12-bit D/A	10 mA / 10 µA	-40 ~ +85 °C
<b>MSM6588</b>	Recorder, Stand alone / MCU, Voice activated rec.	4 ~ 16	Serial Register / DRAM / SRAM	170 sec / variable	12-bit A/D 12-bit D/A	15 mA / 10 µA	-40 ~ +85 °C
<b>MSM6588L</b>	Recorder, Stand alone / MCU, Voice activated rec.	4 ~ 16	Serial Register / DRAM / SRAM	170 sec / variable	12-bit A/D 12-bit D/A	15 mA / 15 µA	-40 ~ +85 °C
<b>MSM6688</b>	Recorder, Stand alone / MCU	4 ~ 16	Serial Register / DRAM / SRAM	1024 sec / variable	12-bit A/D 12-bit D/A	30 mA / 10 µA	-40 ~ +70 °C
<b>MSM6688L</b>	Recorder, Stand alone / MCU	4 ~ 16	Serial Register / DRAM / SRAM	1024 sec / variable	12-bit A/D 12-bit D/A	30 mA / 15 µA	-40 ~ +70 °C
<b>MSM6789A</b>	Sub-band coder (SBC) Recorder, Stand alone / MCU	6 / 8	Serial Register / DRAM / SRAM	2048 sec / variable	12-bit A/D 12-bit D/A	35 mA / 10 µA	0 ~ +70 °C
<b>MSM6789L</b>	Sub-band coder (SBC) Recorder, Stand alone / MCU	6 / 8	Serial Register / DRAM / SRAM	2048 sec / variable	12-bit A/D 12-bit D/A	35 mA / 10 µA	0 ~ +70 °C
<b>MSM9841</b>	Recorder w/ Internal FIFO Memory	4 ~ 32	MCU Addressable Memory	variable	14-bit A/D 14-bit D/A	15 mA / 10 µA	-40 ~ +85 °C

*Winbond Electronics Corp.* valmistaa 'voice recorder controller'-piireiksi kutsumiaan IC-piirejä. Puheen koodaukseen käytetään modifioitua ADM-algoritmia. Piiriin sisältyy äänen digitointiin ja muodostukseen tarvittavien A/D- ja D/A-muuntimien lisäksi mm. mikrofoniesivahvistin, AGC (automatic gain control), anti-aliasing/smoothing-suodattimet, kaiutinvahvistin ja tarvittavat ohjauspiirit. Piiriä on helppo käyttää puheen tallennukseen ja toistoon. Tallennusaika on periaatteessa rajoittamaton, sillä se riippuu ainoastaan ulkoisen muistin määrästä. Piiriä voidaan käyttää muistioissa, onnitelukorteissa, herätyskelloissa, leikkikaluissa ja monissa muissa sovelluksissa.

Äänen digitaaliseen tallennukseen ja toistoon soveltuvia piirejä löytyy edellä mainittujen lisäksi myös muilta komponenttivalmistajilta.

### **Tekstistä puheeksi puhesynteesin avulla**

Puhesynteesi on vakiintunutta ja koeteltua tekniikkaa, jota voidaan soveltaa edullisesti monilla sovellusalueilla. Puhesynteesijärjestelmät sovitetaan kielikohtaisesti. Puhesynteesin avulla tuotetun puheen ymmärrettävyys on nykyisin jo hyvä, joskin se kuulostaa vielä yksitoikkoiselta ja konemaiselta. Puheen laatua pyritään kuitenkin jatkuvasti parantamaan. Kehityskohteina ovat mm. puheen luonnollisuuden, puhetyylin, persoonallisuuden sekä aksentin parantaminen ja joustava vaihto. Lisäksi kehitetään tunteiden vaikutuksen mallinnusta sekä puheen tarkoitukseen liittyvää puhesynteesiä.

Puheviestien tuottaminen puhesynteesin avulla tekstistä puheen tallentamisen sijaan mahdollistaa viestien joustavan laatimisen ja helpon ylläpidettävyyden. Tekstiedostojen luonti on helppoa eikä edellytä erikoisolosuhteita tai -välineitä (esim. äänistudiota) sanomien laatimiseksi. Tämän vuoksi puhesynteesijärjestelmät tuntuvat ideaalisilta puheviestijärjestelmien rakentamiseen. Kaupallisissa tuotteissa ne ovat kuitenkin vielä harvinaisia, mikä johtuu ennen kaikkea kahdesta syystä, puheen laadusta ja järjestelmien hinnasta.

Tekstin puheeksi muuntavia järjestelmiä on saatavissa monilta valmistajilta useina kieliversioina; suomenkieli ei kuitenkaan ole tuetuimpien kielten joukossa. Suomalainen *Timehouse Oy* valmistaa suomenkielistä MikroPuhe-puhesyntetisaattoria. Se perustuu mikrofoneemiseen synteesitekniikkaan, jossa puheen muodostamiseen käytetään ihmisäänestä poimittuja ääninäytteitä. MikroPuhe on toteutettu ohjelmallisesti ja se on saatavissa PC-, Apple Macintosh - ja puhelinkäyttöön soveltuvina versioina.

Belgialainen *Lernout & Hauspie Speech Products N.V. (L & H)* on eräs maailman johtavista puheteknologian kehittäjistä. Sen omistajiin kuuluu mm. Microsoft. Yrityksen ydinosuusalueita ovat automaattinen puheentunnistus, tekstin puheeksi muuntavat järjestelmät sekä puheen ja musiikin koodaus. Lisäksi yritys valmistaa yhteistyössä toisten yritysten kanssa tuotteita mm. sanelimiin ja kielenkäännöskoneisiin.

L & H on merkittävä tekstin puheeksi muuntavien puhesynteesiohjelmistojen yhtiön toimittaja. Yrityksen tuotteita on saatavissa useille eri laitealustoille, puhelinjärjestelmiin, signaaliprosessoreihin sekä ohjelmistoina eri käyttöjärjestelmiin. Taulukossa 5 on tuotteiden ominaisuuksia ja taulukossa 6 niiden soveltamiskohteita.

*Taulukko 5. L & H:n tekstin puheeksi muuntavien tuotteiden ominaisuuksia [Lernout & Hauspie 1997].*

<b>Tuote</b>	<b>Prossessorin valmistaja</b>	<b>Keskeiset ominaisuudet</b>	<b>Saatavat kielet</b>	<b>Kehitteillä olevat kielet</b>	<b>Markkina-alue</b>
<b>tts2000/M</b>	<ul style="list-style-type: none"> <li>• Intel</li> </ul>	<ul style="list-style-type: none"> <li>• rajoittamaton sanavarasto</li> <li>• muokattavat sanastotyökalut</li> <li>• äänen-voimakkuuden, nopeuden ja korkeuden säätö</li> <li>• sisiäinen prosodiatyökalu</li> </ul>	englanti (US) saksa hollanti ranska espanja italia korea	japani kiina, (mandaria) arabia englanti (UK)	Tietokoneet & Multimedia
<b>tts2000/T</b>	<ul style="list-style-type: none"> <li>• Intel</li> <li>• RISC-tietokoneet</li> <li>• Texas Instruments</li> <li>• AT&amp;T</li> </ul>	<ul style="list-style-type: none"> <li>• rajoittamaton sanavarasto</li> <li>• muokattavat sanastotyökalut</li> <li>• äänen-voimakkuuden, nopeuden ja korkeuden säätö</li> <li>• monikanavainen</li> <li>• sisäinen prosodiatyökalu</li> </ul>	englanti (US) saksa hollanti ranska espanja italia korea	japani kiina, (mandaria) arabia englanti (UK)	Tietoliikenne.
<b>tts3000/A</b>	<ul style="list-style-type: none"> <li>• Texas Instruments</li> </ul>	<ul style="list-style-type: none"> <li>• rajoittamaton sanavarasto</li> <li>• muokattavat sanastotyökalut</li> <li>• äänen-voimakkuuden, nopeuden ja korkeuden säätö</li> <li>• sisäinen prosodiatyökalu</li> </ul>	englanti (US) saksa hollanti ranska espanja italia korea englanti (UK)	japani kiina, (mandaria) arabia	Ajoneuvo-elektroniikka
<b>tts3000/C</b>	<ul style="list-style-type: none"> <li>• Intel</li> <li>• Texas Instruments</li> <li>• AT&amp;T</li> <li>• Toshiba</li> <li>• Motorola</li> <li>• Zilog</li> </ul>	<ul style="list-style-type: none"> <li>• rajoittamaton sanavarasto</li> <li>• muokattavat sanastotyökalut</li> <li>• äänen-voimakkuuden, nopeuden ja korkeuden säätö</li> <li>• sisäinen prosodiatyökalu</li> </ul>	englanti (US)	saksa hollanti ranska espanja italia korea englanti (UK) japani kiina, (mandaria) arabia	Kulutus-elektroniikka



Taulukko 6. L & H:n tekstin puheeksi muuntavien tuotteiden soveltamiskohteita [Lernout & Hauspie 1997].

Markkinalohko	Proessoripiiri	Asiakkaita ja lopputuotteita
<b>Ajoneuvo-elektronikka</b>	Motorola (M56XXX DSP) Texas Instruments (TMS320C5X DSP)	<ul style="list-style-type: none"> <li>• liikennetiedotusjärjestelmät</li> </ul> Philips Car Systems Blaupunkt GmbH Robert Bosch GmbH International GmbH Sagem
<b>Kulutuselektronikka</b>	<i>Puhesynteesi:</i> Texas Instruments (TSP50C10 tai TMS320C2X) Analog Devices (ADSP21XX)  <i>Analyysi:</i> Motorola (M68EC000)	<ul style="list-style-type: none"> <li>• taskukokoiset kielenkäännöslaitteet, elektroniset kalenterit</li> </ul> Inventec Corporation Kinpo Electronics Inc. Vtech Electronics Holdings, Ltd
<b>Tietokoneet ja multimedia</b>	PC 80486 DX Windows OS äänikortti  (minimivaatimukset)	<ul style="list-style-type: none"> <li>• kielenkäännös</li> <li>• faksin luku</li> <li>• tietokoneen etäkäyttö</li> <li>• kielten opetus</li> </ul> Frank AudioData Globalink I.R.I.S (Image Recognition Integrated Systems) Novell Proteq Technologies Samsung
<b>Tietoliikenne</b>	Texas Instruments (TMS320C3X)  AT&T (DSP3210)	<ul style="list-style-type: none"> <li>• interaktiiviset puhelinvastajat</li> <li>• sähköpostin ja faksin lukija</li> <li>• automaattinen tiedonhaun opastus</li> <li>• Call Centers</li> </ul> AT&T Dialogic Corp. Deutsche Bundespost Telekom

#### 4.1.2 Puheentunnistus

Puheentunnistusmarkkinoilla on sekä puheentunnistussytimien valmistajia että ohjelmistojen ja komponenttien toimittajia, jotka hyödyntävät tuotteissaan jonkin alkuperäisvalmistajan puheentunnistusalgoritmeja. Kaikista tuotteista ei välttämättä käy selville, minkä yrityksen puheentunnistusalgoritmeja tuotteessa käytetään.

Seuraavassa on esitelty joitakin keskeisiä puheentunnistustuotteiden valmistajia sekä heidän tuotteitaan. Valmistajia ja soveltajia on lukuisia ja niiden määrä kasvaa nopeasti. Ohjelmien soveltaminen suomen kielelle on mahdollista, sillä useimpien järjestelmien toiminta perustuu kätkeytyihin Markovin malleihin, jossa sanoja verrataan tilastollisesti tietokantaan tallennettujen sanojen akustisiin jälkiin eikä pyritä tunnistamaan foneemeja ja sanarunkoja.

*Advanced Recognition Technologies, Inc. (ART)* kehittää sekä puheen- että käsi- alantunnistusteknologiaa. Yrityksen missiona on tarjota luonnollisia tiedonsyöttö- ratkaisuja tietoliikennelaitteisiin (kiinteisiin ja langattomiin puhelinjärjestelmiin), tietokoneisiin, kulutuselektronikkaan sekä muille sovellusaloille.

ART:n smARTspeak® on kielestäriippumaton puhekomento- ja -ohjauskäyttöön tarkoitettu ohjelmisto. Se soveltuu mm. matkapuhelimien, puhelinvastaajien, kodintekniikan, pelien, lelujen ym. laitteiden ohjaukseen. SmARTspeak-puheentun- nistusydyntä on käytetty mm. Motorolan GSM-puhelimissa.

Laitteistovaatimuksina ovat:

- CPU: 8-bittinen suoritin (esim. 8031, 8088), 80x86 tai muu pro- sessori
- käyttöjärjestelmä: DOS, Windows 3.xx, Windows 95 tai laitteis- tokohtainen käyttöjärjestelmä
- puheentunnistus: puhujakohtainen, erilliset sanat tai fraasit
- opetus: 1 - 2 näytettä jokaisesta sanasta tai fraasista
- puheensyöttölaite: mikä tahansa mikrofoni
- näytteenottotaajuus: 6 kHz tai suurempi, 8 bit PCM
- ROM: koodin koko vähemmän kuin 20 KB
- RAM: 200 tavua / näyte
- tuetut kielet: mikä tahansa kieli
- asennus: voidaan asentaa mille tahansa prosessorille, mikrosuorit- timelle tai ASIC-piirille.

*Dragon Systems, Inc.* valmistaa tuotteita sekä loppukäyttäjille että laitteisto- ja oh- jelmistovalmistajille. Tuotteita on saatavissa useille kielille sovitettuna puheentun- nistuksen eri osa-alueille: jatkuva ja epäjatkuvu puheentunnistus, puhekomennot, asiakaskohtaiset puhelinsovellukset, sovelluskehittäjän työvälineet jne. Tuotteita on saatavissa myös useille laitealustoille (MS-DOS, Windows, Windows 95, Win- dows NT, Apple Macintosh, MAC OS, DSP ja puhelintekniikka). Dragon System- sin puheentunnistustekniikkaa on hyödynnetty mm. useiden tietotekniikkayritys- ten tuotteissa (IBM, Media Vision, Microsoft ja WordPerfect).

L & H:n puheentunnistustuotteet on suunnattu neljälle eri markkinalohkolle, auto- elektroniikkaan (asr200/A), kuluttajatuotteisiin (asr200/C), tietoliikenteeseen (asr1500/T) sekä tietokoneisiin ja multimediaan (asr1500/M). Puheentunnistusme- netelmät on sovitettu useiden valmistajien prosessoripiireille. Esimerkkejä käyte- tyistä prosessoripiireistä ja sovelluskohteista on esitetty taulukossa 7.

Taulukko 7. L & H:n puheentunnistustuotteiden sovelluskohteita [Lernout & Hauspie 1997].

Markkinalohko	Prosessoriipiiri	Asiakkaita ja lopputuotteita
<b>Autoelektroniikka</b>	Analog Devices (ADSP2101)	<ul style="list-style-type: none"> <li>• navigointijärjestelmät</li> <li>• komento-ohjausjärjestelmät</li> </ul> Americon Inc., Sagem
<b>Kulutus-elektroniikka</b>	Texas Instruments (TMS320C50) Analog Devices (ADSP2101)	<ul style="list-style-type: none"> <li>• Interactive Set-Top Box</li> </ul> SaveSmart
<b>Tietokoneet ja multimedia</b>	PC 80486 DX Windows OS äänikortti (minimi)	<ul style="list-style-type: none"> <li>• kielten opetus</li> <li>• tietokantasovellukset</li> </ul> The Learning Company, NASA-Martin Merietta Services Inc.
<b>Tietoliikenne</b>	Yleiskäyttöinen tietokone + DSP (esim. Texas Instruments TMS320C31)	<ul style="list-style-type: none"> <li>• puhelinvastaajajärjestelmät</li> <li>• puhelinpankki</li> </ul> AT&T, Dialogic Corp., Deutsche Bundespost, Telekom, Unisys

Yritykseltä on saatavilla mm. sanapohjainen tunnistin Windows CE -tuotteisiin sekä TI C20x (Texas Instruments) ja AD21xx (Analog Devices) kiinteän pilkun signaaliprosessorisarjoille. Tunnistin kykenee tunnistamaan 40 sanaa kerralla ja ohjelmistorajapinta antaa mahdollisuuden sanastojen vaihtamiseen tunnistuksen aikana, mikä mahdollistaa kontekstiriippuvaisen puheentunnistuksen. Sanastot voivat olla itse luotuja käyttäjäkohtaisia sanastoja tai valmiiksi puheentunnistimen mukana toimitettuja. Tunnistimen löydettyä sanan ohjelmointirajapinta tarjoaa tunnistinta käyttävälle sovellukselle muutaman parhaan tunnistustuloksen antavaa sanaa, jolloin sovellus voi parantaa tunnistustarkkuutta jättämällä tilanteeseen sopimattomat sanat pois.

*Verbex Voice Systems, Inc.* kehittää ja valmistaa sekä puhujakohtaisia että puhujasta riippumattomia puheentunnistusohjelmistoja ja -tuotteita, jotka pystyvät tunnistamaan luonnollisella tavalla annettuja sanayhdistelmiä, numerosarjoja jne. Ohjelmistot mahdollistavat tietokoneen ohjauksen puhekomentojen avulla. Puheentunnistusohjelmistoja on saatavissa DOS, Unix, Windows 95 ja Windows NT -käyttöjärjestelmiin.

Yritys valmistaa lisäksi sulautettuihin järjestelmiin tarkoitettua VCORE™ (Verbex Voice Systems Continuous Onboard Recognizer Engine) -puheentunnistusyöntä. Tuotteen soveltamiskohteita ovat mm.

- PDA-laitteet, käsipuhelimet

- teollisuuden tiedonkeruupäätteet
- viihde-elektroniikka: pelit ja Internet TV
- ajoneuvonavigaattorit.

VCORE™ on suunniteltu mahdollisimman helpoksi siirtää eri ympäristöihin, joten sitä voi käyttää periaatteessa missä tahansa ympäristössä, joka tarjoaa riittävästi laskentakapasiteettia puheentunnistuksen tarpeisiin. Tyypillisesti resurssien tarve on noin 13 MIPS laskentatehoa ja 512 kB muistia. Lisäksi suoritusympäristön tulee tarjota joitakin ohjelman kehitykseen ja suoritukseen liittyviä elementtejä, kuten äänen luvun, muistin varauksen ja viestintäjärjestelmän.

VCORE™ voi käyttää mukana toimitettavia sanastoja tai käyttäjä voi luoda omia. Suunniteltaessa varsinaista sovellusta voidaan puheentunnistukseen käyttää VCORE™:n tarjoamaa ohjelmointirajapintaa, APIa. API tarjoaa menetelmät kontrolloida tunnistimen toimintaa sekä antaa tunnistetut sanan sovellukselle. Jos VCORE™:ä aiotaan käyttää ympäristössä, johon sitä ei aikaisemmin ole käännetty, on riskinä kääntämisen epäonnistuminen, vaikka järjestelmä onkin suunniteltu sekä ohjelmakoodin että dokumentoinnin osalta mahdollisimman helpoksi siirtää.

*Voice Control Systems, Inc. (VCS)* on yksi johtavista puheentunnistusytimien toimittajista. Yrityksen tuotteita on saatavissa 50 kielelle, mm. suomen kielelle, sovitettuna. Tuotteita käytetään puheenkäsittelysovelluksissa televiestintäalalla, henkilökohtaisissa tietokoneissa sekä auto- ja kulutuselektroniikassa. Yrityksen oman ilmoituksen mukaan lähes 95 %:ssa puheella aktivoitavista sovelluksista käytetään tällä hetkellä VCS:n teknologiaa.

VCS:n teknologiaa on käytössä mm. televerkkojen automaatiassa (1-800-COLLECT), puhelinpankissa (National Westminster Bank, Ltd.), PC:n multimediaäänikortissa (Creative Labs), matkatietojärjestelmässä (Lufthansa), puheohjatussa puhelimessa (AT&T VoiceTouch) sekä puheohjatussa autopuhelimessa (Mercedes-Benz). Yrityksen teknologiaa sisältyy lisäksi monien laitevalmistajien tuotteisiin (mm. Dialogic, Natural MicroSystems, Brite Voice Systems, OKI Telecom, Hughes Network systems, Creative Labs, Oki Semiconductor) [Voice Control Systems 1997]. VCS:n puheentunnistustuotteita on sovitettu sekä omille että muiden valmistajien laitealustoille.

Auto- ja kulutuselektroniikan alueelle yritys näkee olevan puheentunnistusteknologian markkinoita ja soveltamiskohteita mm. puhelimissa, puhelinvastaajissa, videonauhureissa, televisioissa, henkilökohtaisissa muistioissa ja videopeleissä. Tuotevalmistajat voivat hankkia VCS:n puheentunnistusteknologiaa joko lisensoimalla ohjelmiston tai ostamalla puheentunnistustekniikan sisältäviä piirisarjoja.

Taulukossa 8 on esitetty esimerkki käytettävistä sanastoista ja eri puheentunnistustapojen soveltamisesta matkapuhelimiin.

Taulukko 8. Puheentunnistusohjelmiston sanastoesimerkkejä.

Puhujasta riippumaton				Puhujakohtainen
Numerot	Komennot	Kontrollisanat	Yleisnimet	Henkilönimet
Nolla	Soita	Kyllä	Koti	Äiti
Yksi	Yhdistä	Ei	Toimisto	Jussi
...	Lisää	Avusta	Lääkäri	Liisa
Yhdeksän	Poista	Peru	Kauppa	Mauri Makkonen

Muita puheentunnistusohjelmistojen valmistajia on lukuisia, mm. IBM, Kurzweil ja Fonix, joka on uusi tulokas ja jonka puheentunnistusmenetelmä perustuu omaan patentoituun ratkaisuun.

### Puheentunnistukseen kehitettyjä elektronisia piirejä

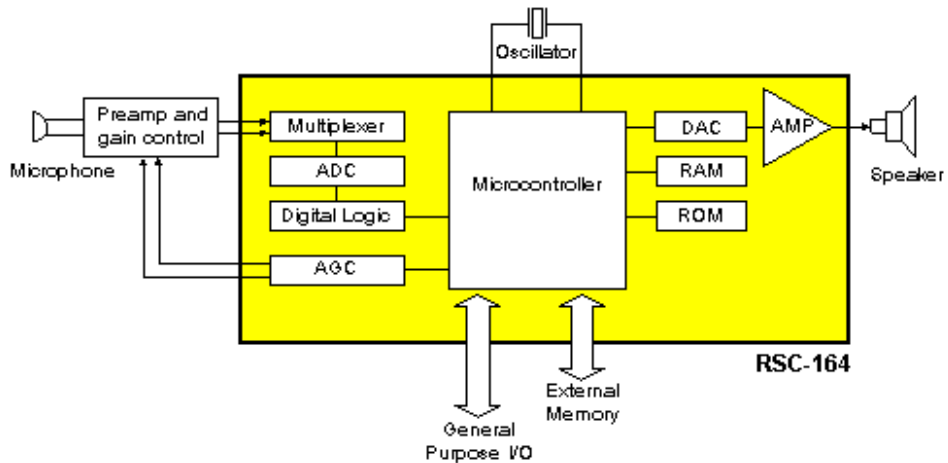
NEC valmistaa PSC2000 (Peripheral Speech Chip) -piiriä, joka on suunniteltu toimimaan yhdessä NEC:n uPD70108 'V20' -mikroprosessorin ja muiden komponenttien kanssa. Piirisarja on tarkoitettu edullisia puheentunnistussovellutuksia varten.

Oki Semiconductor valmistaa puheentunnistuspiiriä MSM6679A-110 (Voice Recognition Processor), jonka ominaisuuksia ovat

- puhujasta riippumaton puheen tunnistus
  - jokaisessa sanastossa jopa 61 sanaa
  - tuki usealle sanastolle
- puhujakohtainen puheen tunnistus
  - jokaisessa sanastossa jopa 25 sanaa
  - tuki usealle sanastolle
- elektroninen äänen tallennus ja toisto
- puhesynteesi.

Samalle piirille on integroitu lisäksi muita aputoimintoja. Piiriä käytetään mm. auto- ja kulutuselektronikassa sekä tietokoneissa.

Sensory Inc. kehittää ja tuotteistaa teknologioita, joiden avulla tuotteiden kommunikointitapoja voidaan parantaa. Yritys myy mm. Interactive Speech™ -piirisarjaa, joka on suunnattu edullisiin kuluttajatuotteisiin, kuten kotielektronikkaan, henkilökohtaisiin turvatuotteisiin, kauko-ohjaimiin, kannettaviin elektronisiin laitteisiin sekä tietoliikennelaitteisiin. Yrityksen ilmoituksen mukaan piirisarjan kehittämisen päämääränä on ollut "to bring life to products through speech and audio technology". Kuvassa 11 on RSC-164-piirin lohkokaavio ja taulukossa 9 piirien keskeisiä ominaisuuksia.



Kuva 11. RSC-164-piirin lohkokaavio [RSC-164 Data Sheet].

Taulukko 9. Interactive Speech™ -piirisarjojen ominaisuuksia [Sensory 1997].

Ominaisuudet	RSC 164	RSC 164i	Voice Password	Voice Direct
<b>Puhujasta riippumaton</b>	kyllä	kyllä	ei	ei
<b>Puhujakohtainen</b>	kyllä	kyllä	ei	kyllä
<b>Puhujantunnistus</b>	kyllä	kyllä	kyllä	ei
<b>Puhesynteesi</b>	kyllä-custom	kyllä-custom	kyllä-standard	kyllä-standard
<b>Puheen tallennus ja toisto</b>	kyllä	ei	ei	ei
<b>Ohjelmitava mikrosuoritin</b>	sisäinen/ ulkoinen	sisäinen/ ulkoinen	ulkoinen	ulkoinen
<b>Asiakaskohtainen sovitus</b>	kyllä	kyllä	ei	ei

Texas Instrumentsin DSP-piirisarjoille on sovitettu useiden puheentunnistus- ja puhesynteesialgoritmien kehittäjien tuotteita, mm. Voice Processing Corporationin ja Lernoutin & Hauspien. Taulukoissa 10 ja 11 on yhteenveto tuotteista.

*Taulukko 10. Texas Instrumentsin TMS320 DSP -piirille sovitettuja puheentunnistus- ja puhesynteesialgoritmeja [Texas Instruments 1997, s. 18-38 - 18-40].*

Algoritmi	Valmis- taja	TMS320- piiri	Ohjelma- -muisti (k word)	Data- muisti (k word)	Proses- sorin suoritus- kyky (MIPS)	Kielet
<b>Yksittäiset sanat, puhujakohtainen tunnistus</b>						
VproPRL™	VPC	'C3x				21
<b>Jatkuva puhe, puhujasta riippumaton tunnistus</b>						
L & H. asr1000/T Automatic Speech Recognition ASR	L & H	'C2x/'C5x	64	28+16- 370/kanava	12	englanti (US), ranska, saksa, espanja, hollanti, italia
L & H. asr1000/T Automatic Speech Recognition ASR	L & H	'C3x	100 <sup>1</sup>	28+16- 370/kanava <sup>2</sup>	8	nglanti (US), ranska, saksa, espanja, hollanti, italia
VProPRL&#153	VPC	'C3x				21
<b>Jatkuva puhe, puhujasta ja kielestä riippumaton tunnistus</b>						
L & H. asr1500/M Automatic Speech Recognition ASR	L & H	'C2x/'C3x/ 'C5x				7
L & H. asr1500/T Automatic Speech Recognition ASR	L & H	'C2x/'C3x/ 'C5x				7

Taulukko 11. Texas Instrumentsin TMS320 DSP -piirille sovitettuja puheentunnistus- ja puhesynteesialgoritmeja(jatkoa) [Texas Instruments 1997 s. 18-38 - 18-40].

Algoritmi	Valmistaja	TMS320-piiri	Ohjelma-muisti (k word)	Data-muisti (k word)	Proessorin suorituskyky (MIPS)	Kielet
<b>Tekstistä puheeksi</b>						
ELOQUENS® 2000	CSELT	'C3x	80	820	2.4	italia
L & H.tts2000/T	L & H	'C31	110	370-560	4	7
L & H.tts3000/A	L & H	'C25	8	220	5	7
L & H.tts3000/C	L & H	'C10	110	100	0.3	7
L & H.tts3000/T	L & H	'C3x		220-400	5	7
<b>Puhujasta riippumaton tunnistus</b>						
AURIS® 1010	CSELT	'C3x	64	256	6.4	italia
AURIS® 2010	CSELT	'C3x	100	512	11.2	italia
FLEXUS® 1000	CSELT	'C3x	100	512	6.4	italia
L & H. asr200/A Automatic Speech Recognition ASR	L & H	'C25/'C5x	7	8+5+4.5/wo rd	10	englanti
L & H. asr200/A Automatic Speech Recognition ASR	L & H	'C3x	12	16+5+4.5/w ord	8	englanti
VProPRL™	VPC	'C3x				21

### 4.1.3 Puhujantunnistus

Belgialainen *Keyware Technologies N.V.* -yhtiö julkisti vuoden 1997 loppupuolella VoiceGuardian-henkilöntunnistusjärjestelmän, joka pohjautuu ääninäytteen perusteella tapahtuvaan äänen ominaispiirteiden vertailuun. Järjestelmä toimii Windows NT -käyttöjärjestelmässä ja voidaan yhdistää internet/intranet-ympäristöön. Valmistajan ilmoituksen mukaan äänitunnistus on turvallinen sekä matkimista että nauhoitettua ääninäytettä vastaan.



Tunnistusjärjestelmää varten tarvitaan ainoastaan mikrofoni sekä Pentium-prosessorilla, 32 MB RAM-muistilla ja Windows 95 tai Windows NT -käyttöjärjestelmällä varustettu PC. Tunnistustilanteessa henkilön lausumaa salasanaa verrataan edeltä käsin tallennettuun ääninäytteeseen. Tallennettu ääninäyte voi sijaita samassa tietokoneessa, tietoverkossa sijaitsevassa toisessa tietokoneessa taikka esimerkiksi henkilön omalla älykortilla.

Ensimmäisiä VoiceGuardian-tunnistusjärjestelmän käyttökohteita ovat erilaiset pankkijärjestelmät sekä tietojen turvalliseen käyttöön ja kulkuoikeuksiin liittyvät sovellukset.

*Sensory Inc:n* valmistama Voice Password™ -mikropiiri on suunnattu edullisiin turvatuotteisiin. Piiri käyttää salasanaan ja ääninäytteeseen perustuvaa neuroverkopohjaista tunnistusteknologiaa (ks. taulukko 9).

*Voice Control Systems Inc.* on julistanut oman puhelinjärjestelmiin (langattomat järjestelmät mukaan lukien) liitettävän SecureDigits™ -puhujantunnistustuotteen. Tuotteen avulla puhujantunnistus voidaan tehdä samanaikaisesti normaalisti lausuttujen numeroiden tunnistuksen kanssa. Järjestelmä on tarkoitettu liitettäväksi ennen kaikkea puhelinpankkijärjestelmien kaltaisiin sovelluksiin.

#### **4.1.4 Elektronisten komponenttien ja piiriratkaisujen valinta**

Puheenkäsittelytekniikka on signaalinkäsittelyä, jossa puhesignaalia prosessoidaan analogisessa ja/tai digitaalisessa muodossa. Aivan yksinkertaisimmissa sovelluksissa riittää pelkkä analoginen signaalin käsittely. Useimmiten puheviestit muutetaan ensin digitaalisiksi, käsitellään digitaalisena, ja vasta lopuksi prosessoitu signaali muutetaan analogiseen muotoon puheääneksi. Komponentit ja piiriratkaisut valitaan puheenkäsittelysovellukseen usean eri tekijän perusteella. Harkittavia perusteita ovat mm. signaalinkäsittelyn tehokkuus, tallennetun ja tuotetun puheen laatu, ohjelmoitavuus, laajennettavuus, valmiin tuotteen tuotantomäärät, komponenttien hintatekijät ja kehitystyökalujen ominaisuudet.

Elektronisten komponenttien ja piiriratkaisujen osalta puhesignaalien digitaaliseen käsittelyyn löytyy lukuisia erilaisia vaihtoehtoja. Yksinkertaisimmissa tapauksissa sovellus ei tarvitse erillistä CPU:ta ja siihen riittää pelkkä puheenkäsittelyyn suunniteltu mikropiiri muutamalla komponentilla lisättyinä. Mikäli kyseessä ei ole reaaliaikainen sovellus tai kohde ei vaadi nopeata signaalinkäsittelyä, konventionaaliset mikroprosessoriratkaisut, esimerkiksi yhden piirin mikrotietokoneet 8- tai 16-bittiset mikrokontrollerit, ovat usein sopiva ratkaisu. Näin varsinkin silloin, kun kokonaisuus koostuu useista erilaisista tiedonkäsittely- ja I/O-toiminnoista ja puheenkäsittely on vain yksi osa koko järjestelmän toimintaa. Edellytyksenä tietenkin on, että CPU ehtii suorittaa kaikki toimintonsa sujuvasti. Jos prosessorin kuormitus kasvaa liiaksi, voidaan puheenkäsittely siirtää mikrokontrollerilta omalle mikropiirilleen, sillä osa puheteknologian piiriratkaisuista on suunniteltu suoraan mikroprosessorien liitäntäpiireiksi. Ne kykenevät suorittamaan tiettyjä signaalinkäsittelytehtäviä itsenäisesti ja vähentävät siten prosessorin kokonaisuormitusta.

Monet puhesignaalien käsittelyalgoritmit vaativat kuitenkin tehokasta tiedonkäsittelykapasiteettia ja vapaasti ohjelmoitavia algoritmeja, esimerkiksi puhesignaali, jota siirretään 16 kb/s nopeudella. Jos tämä signaali halutaan koodata laadukkaasti lopputuloksen aikaansaavalla algoritmilla reaaliaikaisena toteutuksena, se vaatii lähes 40 MIPSin käsittelytehon, taulukko 12. Perinteisellä tekniikalla valmistetut mikrokontrollerit, jotka toimivat 25 MHz nopeudella, kykenevät saavuttamaan vain noin kolmen tai neljän MIPSin tehon [Anon. 1996]. Siksi useimmat puheenkäsittelysovellukset tehdäänkin signaaliprosessorien (DSP) avulla. Niiden arkkitehtuuri poikkeaa konventionaalisesta prosessoritekniikasta. Signaaliprosessorit noudattavat ns. *modifioitua Harvardin arkkitehtuuria*, jossa data- ja ohjelmamuisti on erotettu toisistaan, so. komponentin sisällä molemmilla on oma väylä tiedonsiirtoa varten. Siten datan ja koodin siirto CPU:lle voi tapahtua yhtä aikaa. Lisää tehoa on saatu aikaan myös ns. *multiply and accumulate* (MAC) -käskyillä, joissa esimerkiksi kerto- ja yhteenlaskuoperaatio voidaan suorittaa yhden kellojakson aikana. Myös ns. *pipeline-tekniikka*, jossa käskyn eri osat suoritetaan rinnakkain, tuo lisää tehoa prosessoriin.

Nykyisten signaaliprosessorien suorituskyky on jo noin 400 MIPSiä [Bursky 1998]. Tekniset edellytykset puheen käsittelylle vaativiinkin sovelluksiin ovat siten olemassa. Tämä koskee pian myös ns. hintakriittisiä tuotteita. Puheteknologian komponentit, ohjelmistot ja työkalut kehittyvät voimakkaasti. Lähivuosien aikana puheen käsittelytekniikka rutinoituu ja se hallitaan nykyistä paremmin. Tämä kehitys tulee näkymään paitsi lisääntyvänä puheteknologian tuotesovelluksina myös kustannuksiltaan edullisina alan komponenttien, ohjelmistojen ja työkalujen hintoina.

Elektronisten komponenttien ja piiriratkaisujen ohella puheteknologiasovellus vaatii useimmiten oman ohjelmistonsa itse puheen käsittelyyn. Puheenkäsittelysovelluksiin on tarjolla jo lukuisia kaupallisia ohjelmistoja, joita voidaan hyödyntää osana puheteknologiatuotteita. Edellä taulukoissa 10 ja 11 on muutamia esimerkkejä alan ohjelmistotarjonnasta. Komponenttivalmistajien luetteloista löytyy myös informaatiota kolmansien osapuolten toimittamista puheenkäsittelyohjelmista, joita heidän tuotteissaan voi hyödyntää [esim. Texas Instruments 1997]. Ohjelmistotuotteiden valinta on tehtävä rinnakkain komponentti- ja piiriratkaisujen kanssa.

Taulukko 12. Esimerkkejä eri standardien vaatimista datansiirtonopeuksista, puheenkoodaustekniikoista ja tarvittavista tiedonkäsittelynopeuksista [Texas Instruments 1997, s. 1 - 23].

Standardi	Datan siirtonopeus (kbps)	Puheenkoodaus- tekniikka	Suorituskyky (MIPS)
G.711	64	LOG	0 – 0.5
G.721	32	ADPCM	5
G.728	16	LD CELP	38
GSM	13	RPE-LTP	3.2
TETRA	7.2	ACELP	10
IS-54	7.9	VSELP	18
DOD 4.8	4.8	CELP	10
LPC	2.4	LPC	3

#### 4.1.5 Verkotettu arkkitehtuuriratkaisu

Puheenkäsittelylaitteiston arkkitehtuurit voidaan jakaa karkeasti kahteen luokkaan, keskitetyihin ja hajautettuihin ratkaisuihin. Keskitetyssä ratkaisussa kaikki puheenkäsittelyä suorittava elektroniikka on keskitetty esimerkiksi yhdelle piirikortille, jolloin lopputulos on kompakti 'stand-alone'-järjestelmä. Tämä ratkaisu soveltuu hyvin mm. ääniviestien tallentamiseen, kun viestejä ei tarvitse muuttaa.

Eräs merkittävä teknologiatrendi on koota erillisiä toimilaitteita ja sensoreja tietoverkkojen osaksi. Tällaisessa hajautetussa arkkitehtuurissa puhesovelluksen toteutus tehdään kahdelle tai useammalle piirikortille, jotka yhdistetään tiedonsiirtolinjalla.

Esimerkkejä verkotetuista ratkaisuista:

1. Digitaaliset ääniviestit on tallennettu keskustietokoneeseen (esim. PC), josta ne siirretään linjaa pitkin käyttökohteeseen. Mikäli käyttökohteita on runsaasti, ääniviestien päivitys tapahtuu vaivattomasti kohdistamalla muutokset vain yhteen paikkaan. Tiedonsiirtoyhteys vaatii lisäelektroniikkaa myös terminaalissa.
2. Puheentunnistukseen perustuva ohjaus tai käyttäjän tunnistus. Terminaali sisältää mikrofonin, puheen pakkauselektroniikkaa ja tiedonsiirron vaatiman elektroniikan. Puhe pakataan reaaliajassa ja siirretään tehokkaaseen keskustietokoneeseen analysoitavaksi.

3. Keskusvalvomossa halutaan seurata reaaliajassa terminaalien aktiviteettia esimerkiksi vianvalvonnan ja kulunvalvonnan yhteydessä.

Edellä kuvatuissa konsepteissa korostuu terminaalilaitteiden yksinkertaisuus, mikä voi olla tärkeä näkökohta niiden lukumäärän muodostuessa suureksi. Keskustietokone takaa suuren, skaalautuvan tallennus- ja laskentakapasiteetin. Puhesignaalin pakkaus voidaan suorittaa tehokkaasti reaaliajassa joko dsp-prosessorilla tai erikoispiirillä.

Nykyisellä tiedonsiirtoteknologialla voidaan toteuttaa nopeita langattomia tai langallisia siirtoyhteyksiä tehonkulutuksen silti muodostumatta liialliseksi. Rakenuksen sisäinen WLAN-ratkaisu (wireless local area network) on yksi potentiaalinen vaihtoehto.

Siirtoviiveet eivät oikein suunniteltuina muodosta suorituskyvyn pullonkaulaa. Esimerkiksi 1 Mb/s siirtonopeudella yhden sekunnin pituisen ääniviestin (8000 x 16 bittiä) siirtäminen kestää noin 122 ms, kun mukaan lasketaan vain bittien siirtoaika. Tähän lukuun tosin summautuu toteutusteknologiasta riippuvia lisäviiveitä. Käytännössä puhesignaali voidaan pakata huomattavasti pienempään tilaan, mikä vähentää siirtonopeusvaatimuksia mutta lisää laskentavaatimuksia. Kaupallisesti saatavien ratkaisujen siirtonopeudet vaihtelevat voimakkaasti yltäen aina gigabit-tiin sekunnissa saakka.

#### 4.1.6 Komponenttien ja ohjelmistojen hankintakustannuksia

Nopeasti kehittyvillä aloilla, kuten tietokonetekniikassa ja puheteknologiassa, komponenttien ja ohjelmistojen hinnat muuttuvat nopeasti. Hintoihin vaikuttavat tuotteen ominaisuuksien lisäksi voimakkaasti ostoaikajankohda, hankintamäärät jne. Varsinkin ASIC-piirien hinnoittelusta on erittäin vaikea saada tietoja. Taulukoissa 13 ja 14 on esitetty eri lähteistä koottuja suuntaa antavia hintatietoja.

*Taulukko 13. ISD:n puheen tallennus- ja toistopiirien hintoja.*

Tuote ja valmistaja	Komponentin hinta
ISD33000 Series Information Storage Devices, ISD (matkapuhelimiin, kannettaviin laitteisiin jne.)	16,60 - 8,30 USD 50 000 kpl erissä tallennusajasta, pakkauksesta ja lämpötilavaatimuksista riippuen
ISD1500 Series Information Storage Devices, ISD (leluihin, lahja- ja mainostuotteisiin jne.)	0,90 USD (ISD1506) 0,99 USD (ISD1510) 1,65 USD (ISD1520) 10 000 kpl erissä

Taulukko 14. Puheentunnistuspiirien ja -ohjelmistojen hintoja.

Tuote ja valmistaja	Tuotteen hinta	Demo-järjestelmän hinta
SmARTspeak® SDK Software Developer's Kit Advanced Recognition Technologies, Inc.	495 USD	
HM2007 Speech Recognition Chip HUALON Microelectronic Corp.	n. 16 USD	n. 160 USD Demo board
PSC2000 Peripheral Speech Chip NEC	n. 10 - 20 USD määristä riippuen	
OKI VRP6679 Voice Recognition Processor OKI Semiconductor	n. 20 USD	n. 900 USD Demo board
VCS Continuous Speech Recognition Voice Control systems, Inc.	ohjelmiston lisenssi- maksu vaihtelee määristä riippuen n. 5 - 500 USD	
VCS Isolated Word Speech Recognition Voice Control systems, Inc.	ohjelmiston lisenssi- maksu vaihtelee määristä riippuen n. 1 - 500 USD	

## 4.2 KEHITYSTYÖKALUT JA -YMPÄRISTÖT

### 4.2.1 Työkaluja puheviestejä hyödyntäviin järjestelmiin

Mikäli puheen tallennuskapasiteettiin ja sen hintaan sekä puheen tallentamiseen tarvittavaan studioaikaan ei tarvitsisi kiinnittää huomiota, olisi puheviestijärjestelmien rakentaminen yksinkertaista. Tallennettaisiin vain jokainen tarvittava viesti erikseen järjestelmään. Kustannussyistä tämä ei kuitenkaan yleensä ole mahdollista. Useimmiten joudutaan toimimaan rajallisen tallennuskapasiteetin ja tuotekehitysbudjetin puitteissa. Tämän vuoksi järjestelmiä toteutettaessa joudutaan turvautumaan puheen koodaukseen, sanojen ja sanayhdistelmien ketjutukseen, puhesynteesiin jne. Niinpä järjestelmien toteutusta helpottavat hyvät tuotekehitysympäristöt, joita on saatavissa sekä komponentti- ja ohjelmistovalmistajilta että valmistajasta riippumattomilta toimittajilta.

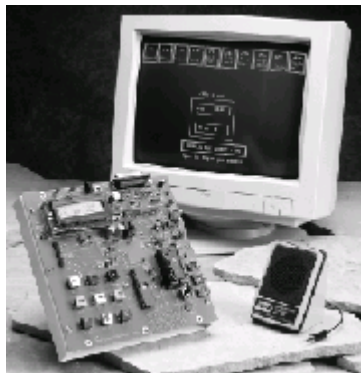
*Information Storage Devices Inc.* valmistaa useampia kehitystyökaluja ChipCorder®-piireihin pohjautuvien tuotteiden kehitystä varten:

ISD-SD200/203 Universal Sound Development and Programming System on PC-pohjainen järjestelmä, joka mahdollistaa äänen editoinnin ja tallennuksen tarkan ohjelmoinnin ISD:n ChipCorder®-tuotteisiin. Editointi, viestien sijoittaminen ja ohjelmointi tehdään Windows-pohjaisen Sound Mapper 200 -sovellusohjelmiston avulla. Järjestelmään kuuluu Main System Module, kahdeksankantainen ohjelmointikotelo tai Handler Interface, Sound Mapper -sovellusohjelmisto ja tarvittavat kaapelit. Järjestelmän avulla voidaan maksimoida viestien tallennus ja ohjel-

moida kahdeksan piiriä samanaikaisesti. ISD-SD203 ohjelmointikotelo tukee ISD33000-sarjan tuotteita ja kehitteillä oleva ISD-SD201-ohjelmointikotelo kaikkia muita tuotteita. Järjestelmän hinnat ovat: ISD-SD200 Main System Module n. 1 250 USD, ISD-SD203P Programming Pods for PDIP n. 1 650 USD ja ISD-SD203E Programming Pods for TSOP n. 1 650 USD.

ISD-ES301 Development System on suunniteltu sekä tuotekehitystyötä että ISD3000-tuotteiden arviointia varten, kuva 12. Järjestelmän ominaisuuksia ovat:

- rinnakkaisporttiliityntä ISD3000-laitteiden SPI-portin ohjausta varten
- ulkoinen liityntä ISD3000:n suoraa käsittelyä varten
- 9-näppäiminen näppäimistö yksinkertaisia ohjaustoimintoja varten
- ulkoinen mikrofoniliitin
- line input ja line output -liittimet
- ulkoinen mikrofoni ja kaiutin
- automaattinen vahvistuksen säätö mikrofoniliitynnässä
- toiston voimakkuuden säätö
- sisäinen paristoteholähde
- ulkoisen teholähteen liitin
- ISD-ES301 sisältää kannan Motorola 68HC705-C4A -mikrosuoritinta varten, joka liittyy ISD33000-laitteisiin, LCD-näyttöön ja näppäimistöön. Motorolan suoritinta varten kirjoitettu ohjelmisto tarjoaa käyttöliittymän ISD33000-laitteiden ohjaukseen.
- ISD-ES301 Main System Modulen hinta on n. 500 USD.



*Kuva 12. ISD-ES301 Development System [ChipCorder® Data Book].*

*Quadravox Inc.* valmistaa PC-pohjaisia järjestelmiä, jotka helpottavat puheen taikka yleisemmin äänen liittämistä tuotteisiin ja järjestelmiin. Kehitysvälineet on tarkoitettu sellaisten tuotteiden kehittämiseen, joissa laitteen antaman herätteen perusteella tuotetaan jokin puhe- tai ääniviesti. Kehitysvälineitä on saatavissa sekä ISD:n että Texas Instrumentsin piirejä varten.

iQbox© Speech Development System tukee kaikkiin nykyisiin ISD ChipCorder®-piirisarjoihin pohjautuvien tuotteiden kehitystä. Järjestelmässä on varauduttu myös uusiin nykyistä pitemmän tallennusajan sisältäviin piirisarjoihin. Järjestelmään sisältyy ohjelmisto, dokumentointi, ohjelmointilaitteisto sekä tarvittavat kaapelit. Tarvittavan PC-laitteiston minimivaatimukset ovat: 2MB kiintolevytilaa, Windows 3.1 tai uudempi ja 16 bitin äänikortti. Noin 300 USD maksavan järjestelmän ominaisuuksia ovat:

- Windows, wav-formaatti (11 025 kHz, 16 bittiä, mono)
- ohjelmisto tukee sanomien ketjutusta ja monimutkaisten, tiiviiden sanomajärjestelmien rakentamista
- signaalin esiprosessointi
- datan tallennus ennalta määrättyihin paikkoihin
- tuottaa automaattisesti hakemistot sanoille ja ketjutuksille
- saatavissa on myös valmiita sanastoja eri käyttötarkoituksia varten.

Qbox Pro Speech Prototyping System tukee Texas Instrumentsin LPC-dekoodausalgoritmeja tuotesarjoihin TSP50C1X ja MSP50C3X sekä piireihin TSP50C20, TSP53C30 ja TMS5220. Tarvittavan PC-laitteiston vaatimukset ovat: Windows 3.1 tai uudempi, 486/66 tai nopeampi CPU, 8 MB RAM, 16 bitin äänikortti (11 025 kHz tallennus ja toisto). Noin 200 USD:n hintaisen järjestelmän ominaisuuksia ovat:

- Windows, wav-formaatti (16 bittiä, mono)
- sisältää ohjelmiston, jolla 11 025 kHz:n näyte voidaan muuttaa 8 ja 10 kHz:n tiedostoiksi
- ei vaadi lisälaitteita, vain äänikortti
- LPC analyysi ja koodaus
- graafinen editointi.

#### **4.2.2 Työkaluja puheentunnistusjärjestelmiin**

Karkeasti luokitellen puheentunnistusjärjestelmien kehitysympäristöt voidaan jakaa kolmeen osaan, Windows 95/NT -pohjaisiin tunnistimiin, sulautettujen järjestelmien DSP- ja yleisprosessoreihin sekä sulautettuihin järjestelmiin tarkoitettuihin puheentunnistuspiireihin. Lähes kaikilta aikaisemmin luetuilta puheentunnistusjärjestelmien toimittajilta löytyy ohjelmistoja ja evaluointikortteja, joilla pyritään helpottamaan tunnistimien käyttöönottoa omissa sovelluksissa. Kehitysympäristöjen hinnat ovat yleensä kohtuullisia, sillä tunnistinkohtaisten kehittimien käyttö lisää sitä tukevan tunnistimen myyntiä. Tyypillisesti hinnat vaihtelevat sadasta dollarista muutamaan tuhanteen dollariin. Kaikki työkalupaketit eivät sisällä mahdollisuuksia ääniaineiston lukuun mikrofoniin tietokoneeseen. Tähän tehtävään sopivat kuitenkin mainiosti normaalit PC-mikroille tarkoitetut äänikortit ja äänenkäsittelyohjelmistot. Esimerkkinä äänikorteista voidaan mainita *Terratecin* tuoteperhe, joka on Suomessa aika hyvin edustettu. Terratecin hinnat ovat mallista riippuen 500 - 3000 mk. Yleensä äänikortin mukana tulee ohjelmisto äänen käsittelyä ja editointia varten, mutta vaativiin erikoistehtäviin, kuten suodatukseen tai

kohinan analysointiin, voidaan käyttää erikseen myytäviä äänenkäsittelyohjelmistoja. Näistä esimerkkinä voidaan mainita *Sonic Foundryn* valmistama *Sound Forge 4.0*.

Windows 95/NT-käyttöjärjestelmiin saataviin puheentunnistimiin löytyy tunnistimen valmistajalta yleensä myös kehitysympäristö, jonka avulla on helppo tehdä kyseisen valmistajan puheentunnistinta käyttäviä sovelluksia eri ohjelmointikielillä. Tähän asti kehitysympäristöjen käyttö on sitonut tuotteen yhteen puheentunnistusjärjestelmän toteutukseen, sillä eri valmistajien kehitysympäristöt eivät ole olleet keskenään yhteensopivia. Puheentunnistimien yleistyessä pyritään kuitenkin standardoimaan tapa, jolla puheentunnistin ja sitä käyttävä sovellus kommunikoi keskenään. Käyttöön ollaan ottamassa kaksikin standardia, Microsoftin suunnittelema Windows-käyttöjärjestelmälle tarkoitettu Microsoft Speech API (SAPI) sekä yleiskäyttöisempi SRAPI, jonka takana ovat mm. Novell Inc., Dragon Systems, IBM, Kurzweil AI, Intel ja Philips Dictation Systems. Molemmat standardit tarjoavat sekä puheentunnistimelle että sitä käyttävälle sovellukselle helppokäyttöiset kommunikointimenetelmät komentosanojen ja sanelun mahdollistamiseksi. Varsinaista puheentunnistusta rajapinnat eivät tarjoa, joten puheentunnistusta käyttävää tuotetta levitettäessä täytyy mukana toimittaa myös puheentunnistin, josta joutuu maksamaan lisensointimaksun tunnistimen valmistajalle. Tilanne muuttunee tulevaisuudessa, koska Windows-käyttöjärjestelmään tullaan liittämään jossain vaiheessa puheentunnistin.

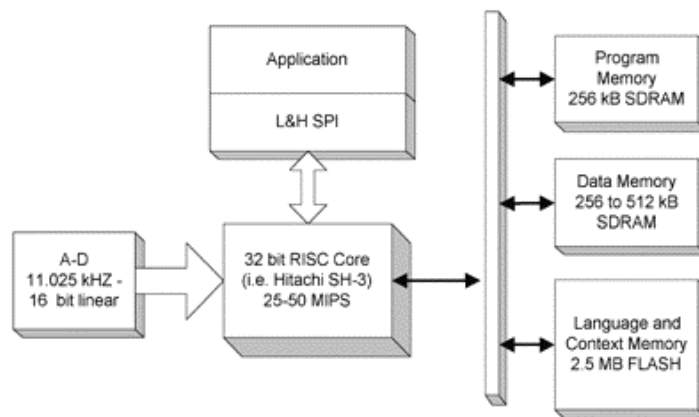
*Dragon Systems* tarjoaa *NaturallySpeaking*-puheentunnistimelleen ohjelmointipakettia, jonka avulla voidaan toteuttaa ko. tunnistinta käyttävä sovellus. Sovellus voi käyttää puheentunnistinta helppokäyttöisen ActiveX-teknologian avulla tai SAPI:n kautta. Tällä hetkellä tuettuja kieliä ovat amerikan- ja brittienglanti, ranska, espanja, saksa ja italia. Tuotekehitysympäristö tarjoaa sovellukselle pohjan dialogeihin ja sanavirheiden korjaamiseen, sanaston hallinnan sekä integroinnin sovelluksen lähdekoodiin, jolloin sovellus voi vastata nopeasti puheentunnistimelta tuleviin komentoihin tai saneluun. Sanaston hallinta mahdollistaa tilannekohtaisten sanastojen sekä kielimallin käytön, jolloin tunnistusmahdollisuuksia rajaamalla saadaan tunnistustulokset luotettavammiksi. Sanastoina voidaan käyttää tunnistimen mukana tullutta sanastoa tai ne voidaan rakentaa puhujariippumattomista foneemimalleista itse. Sanaston hallintaan on suunniteltu graafista käyttöliittymää hyödyntävä helppokäyttöinen työkalu.

ActiveX-teknologiaa käyttämällä on helppo toteuttaa puheentunnistinta hyväksi käyttäviä sovelluksia esim. C++ tai Visual Basic -ohjelmointikielillä. ActiveX-komponentit tarjoavat käyttöliittymät sanastoille, dialogeille ja puhemakroille sekä mahdollistavat myös puheella tapahtuvan palautteen käyttäjälle. Sovelluksen voi toteuttaa myös käyttämällä SAPI-rajapintaa, jolloin toteutus on vaikeampaa ja hitaampaa kuin ActiveX-komponentteja käyttämällä mutta puheentunnistin on huomattavasti helpompi vaihtaa johonkin toiseen SAPIa tukevaan puheentunnistimeen. Levitettäessä suuria määriä sovellusta voidaan *Dragon Systemsin* kanssa sopia lisenssihinta tuotteen mukana toimitettavalle *NaturallySpeaking*-tunnistimelle.



Sulautettujen järjestelmien puheentunnistuksen toteuttamiseen ei toistaiseksi ole yleisesti käytettyjä standardeja, vaan jokainen puheentunnistimen valmistaja toteuttaa oman rajapinnan tarjoamalleen puheentunnistusohjelmistolle tai mikropiirille. Jos Java tulevaisuudessa yleistyy myös sulautetuissa tuotteissa, tilanne saattaa muuttua, sillä SRAPI-rajapinnan on tulevaisuudessa tarkoitus tukea myös Java-ohjelmointikieltä. Sulautettujen järjestelmien suunnittelussa ja toteutuksessa tulevat suurimmat kustannukset yleensä puheentunnistuksessa käytettyjen piirien ja niitä tukevien oheispiirien vaatiman piirilevyn suunnittelusta ja toteutuksesta. Yksinkertaisemmat puheentunnistukseen tarkoitetut erikoispiirit tulevat yleensä halvemmaksi ja helpommaksi toteuttaa kuin signaaliprosessori- tai mikroprosessoripohjaiset tuotteet. Niillä ei kuitenkaan yleensä saavuteta yhtä luotettavaa ja joustavaa järjestelmää kuin prosessoripohjaisilla tuotteilla.

*Lernout & Hauspie* valmistaa kehitysyökaluja erilaisiin sulautettuihin tuotteisiin. ASR1600 for RISC tarjoaa foneemipohjaisen puheentunnistimen RISC-tyyppisille mikroprosessoreille, kuten Hitachi SH-3:lle. Kuvassa 13 nähdään tyyppillisen toteutuksen lohkokaavio.



Kuva 13. Foneemipohjaisen puheentunnistimen lohkokaavio [*Lernout & Hauspie 1998*].

Kuvasta nähdään, että tuotteen suunnittelijan täytyy tarjota puheentunnistimelle ympäristö, jossa tunnistin toimii. Puheentunnistin taas tarjoaa ohjelmointirajapinnan varsinaiselle sovellukselle, jolloin suunnittelijan ei tarvitse huolehtia äänen käsittelystä tai tunnistusprosessista. Kehitysympäristöön kuuluu MS Windowsissa toimiva testiympäristö, jolla voidaan suunnitella varsinainen sovellus puheentunnistuksen osalta sekä testata tunnistinta erilaisilla tunnistustilanteilla ja parametreilla. Lisäksi työkaluihin kuuluu ohjelma sanaston suunnitteluun kontekstin ja foneemikuvausten osalta.

Piiriä MSM6679A-110 VRP varten *Oki Semiconductor* tarjoaa erilaisia evaluation-piirilevyjä sekä PC-pohjaisia ohjelmistoja ääninäytteiden keräämistä ja piirille talletusta varten. Lisäksi on saatavilla ainakin yksi esimerkkisovellus lähdekoodineen. Esimerkissä piirillä on toteutettu äänikomennoilla toimiva puhelinnume-

ron valinta. Mukana tulevaa lähdekoodia voi muuttaa omiin sovelluksiin sopivaksi tai käyttää esimerkikoodina. Ääninäytteiden keräämiseen tarkoitettu ohjelmisto sisältää helppokäyttöisen hiiriohjatun äänieditorin sekä kortin, jolla ääninäytteet voidaan kirjoittaa ROM-piirille. Varsinainen puheentunnistus tapahtuu piirillä olevaa APIa käyttäen. Puheentunnistuspiiriä voi komentaa suoraan samalla piirillä oleva isäntäprosessori tai piiriä voi käyttää sarja- tai rinnakkaisliitännän kautta. OKIn piiri vaatii kohtalaisen paljon oheislogiikkaa toimiakseen ja piirin suunnittelu vastaa hyvin paljon yksinkertaisen sulautetun järjestelmän tekemistä.

*Sensory Inc.* tarjoaa puheentunnistuspiirilleen RSC-164-kehitysympäristöä, johon sisältyvät evaluointikortti sekä tarvittavat ohjelmistot. Ohjelmistot sisältävät koodikirjaston, emulaattorin, simulaattorin, assemblerin ja linkkerin ja tarjoavat täydellisen ohjelmistonkehitysympäristön puheentunnistuspiirille. Piirin ohjelmointi muistuttaa hyvin paljon normaalin assembly-kielisen ohjelmiston suunnittelua. Evaluointikortin avulla on helppo tarkastaa ohjelmiston ja mahdollisten puheentunnistusvirheiden käyttäytyminen ennen varsinaisen sovelluksen toteuttamista. Piiri ei vaadi kovin paljon muuta elektroniikkaa toimiakseen, mutta äänen sisäänoton analogiaosat joudutaan toteuttamaan itse.

### 4.2.3 Työkaluja puhujantunnistusjärjestelmiin

Vuoden 1997 loppupuolella saatiin hyväksytyä SVAPI (The Speaker Verification Programming Interface Standard), jota hyödyntävät monet ohjelmistovalmistajat omissa tuotteissaan. Se helpottaa eri valmistajien tuotteiden liittämistä toisiinsa ja nopeuttanee näin alan kokonaismarkkinoiden kasvua. SVAPIin takana ovat mm. Novell (primary sponsor), CitiCorp, Dialogic, Higgins and Associates, Hughes, IBM, INS, ITT Industries, MIT Lincoln Laboratory, Motorola, SRI/Nuance, T-Netix, Texas Instruments, Veritel, VeriVoice ja US hallitus. API tarjoaa yhteisen rajapinnan puhujantunnistusjärjestelmien ja niitä hyväkseen käyttävien sovellusten välille. Tällä hetkellä kuitenkin lähes kaikilla puhujantunnistusjärjestelmiä tarjoavilla valmistajilla on valmistajakohtainen rajapinta sovelluksia varten.

*Keyword Technologies* tarjoaa VoiceGuardian-puhujantunnistustuotteeseensa ohjelmointirajapinnan sekä tuotekehitystyökaluja, jotka helpottavat tuotteen käyttämistä omista sovelluksissa. Ohjelmistorajapintaa voidaan käyttää kolmella eri tavalla,

- suorina funktiokutsuina
- OLE-tekniikalla
- ActiveX-tekniikalla, mikä mahdollistaa myös verkon yli tapahtuvan autentikoinnin.

### 4.2.4 Virtuaalisuunnittelu

Tuotteiden lisääntyvä monimutkaisuus ja kiristynyt kansainvälinen kilpailu, jossa innovointi ja dynaamisuus ovat keskeisiä edellytyksiä menestymiselle, asettavat uusia haasteita myös rakennusten teknisiin järjestelmiin ja laitteisiin perustuvien

tuotteiden kehittämiseksi. Lisääntyvästä monimutkaisuudesta ja toiminnallisuuden monipuolistumisesta huolimatta tuotteiden tulisi olla entistä helpokäyttöisempiä. Kilpailukyvyyn lisäämiseksi tuotekehitykseltä vaaditaan yhä enemmän kykyä suunnitella, testata ja arvioida tulevan tuotteen ominaisuuksia mahdollisimman pitkälle jo ennen, kuin tuotteesta valmistetaan ensimmäinen fyysinen prototyyppi. Lupaava lähestymistapa vastata näihin haasteisiin on *virtuaalisuunnittelu*.

Virtuaaliprototyyppi (virtuaalimalli, digitaaliprototyyppi) on tuotteen tai tuotekonseptin digitaalinen tietokonemalli, jota simuloimalla pyritään käyttäjälle luomaan mahdollisimman todenmukainen vaikutelma tuotteen ominaisuuksista. Simuloitavia ominaisuuksia voivat olla tuotteen visuaalinen ulkoasu, toiminnallisuus, ääniominaisuudet ja käyttöliittymä. Jopa tuotteen käyttöön liittyviä tunto- ja voimavasteita voidaan simuloida erityisellä käyttöliittämälaitteistolla. Virtuaalisuunnittelu puolestaan tarkoittaa prototyyppiprozessia, jossa virtuaaliprototyyppiä käytetään tuotekonseptien innovointiin, suunnitteluun, testaukseen ja arviointiin. Niitä voidaan käyttää joko korvaamaan tai täydentämään perinteisiä fyysisiä prototyyppiä. Virtuaalisuunnittelu pohjautuu keinotodellisuustekniikoiden sekä kehittyneiden mallinnus-, simulointi- ja käyttöliittämättekniikoiden yhdistettyyn soveltamiseen.

### **Sovelluskohteet ja hyödyt**

Rakennusten teknisten järjestelmien ja laitteiden käyttöliittymiä kehitettäessä virtuaalisuunnittelua voidaan soveltaa ja hyödyntää seuraavasti:

- käyttöliittymän ominaisuuksien, kuten ulkoasun, toiminnallisuuden ja käytettävyyden, varhaiseen havainnollistamiseen ja testaamiseen
- niiden käyttöliittymän ominaisuuksien suunnitteluun ja testaamiseen, joiden suunnittelu perinteisillä menetelmillä on vaikeaa
- erilaisten käyttöliittymäratkaisujen nopeaan ja kustannustehokkaaseen kokeiluun ja arviointiin
- korvaamaan perinteisiä fyysisiä prototyyppiä, joiden valmistus ja muuttaminen on kallista, aikaavievää ja materiaaleja kuluttavaa
- sellaisten käyttöliittymäratkaisujen tutkimiseen, joiden toteuttaminen ei nykyisellä teknologialla ole mahdollista, mutta tarvittavan teknologian uskotaan olevan käytettävissä lähitulevaisuudessa
- yhdistämään erityyppisiä ja -tasoisia simulointimalleja, kuten laitteisto- ja ohjelmistosimulointimalleja, jolloin parannetaan eri suunnittelualojen välistä kommunikointia moniteknologisen tuotteen tuotekehityksessä.

### **Rakennusten teknisten laitteiden tai järjestelmien käyttöliittymien virtuaalisuunnittelu**

Rakennusten automaatio- ja tietotekniikkajärjestelmät ovat nykyisin tyypillisesti moniteknologiatuotteita, joiden tuotekehitykseen liittyy useita suunnittelualueita, kuten ohjelmistosuunnittelua, laitteistosuunnittelua, mekaniikkasuunnittelua ja käyttöliittymäsuunnittelua. Virtuaaliprototyyppien rakentaminen käsittää useita osa-alueita. Mallinnus voi lähteä liikkeelle järjestelmän tai laitteen käyttöliittymän

kolmiulotteisen rakenteen mallintamisesta. Kolmiulotteisiin malleihin liitetään usein myös materiaalitekstuurit, jotka kuvaavat pintarakenteissa käytettäviä materiaaleja. Myös mekaniikan toiminnallisuuden eli kinematiikan mallintaminen on yleensä tarpeen, vähintäänkin tuotteen käyttöliittymään sisältyvien näppäinten ja muiden kontrollien osalta. Käyttäytymisen ja toiminnallisuuden mallintaminen käsittää mm. käyttöliittymän, sulautetun ohjelmiston ja laitteiston sekä muun elektroniikan toimintojen ja käyttäytymisen kuvaamisen suoritettavilla simulointimalleilla. Simuloinnin havainnollistamiseen käytetään seuraavia havainnollistamismuotoja ja -tekniikoita:

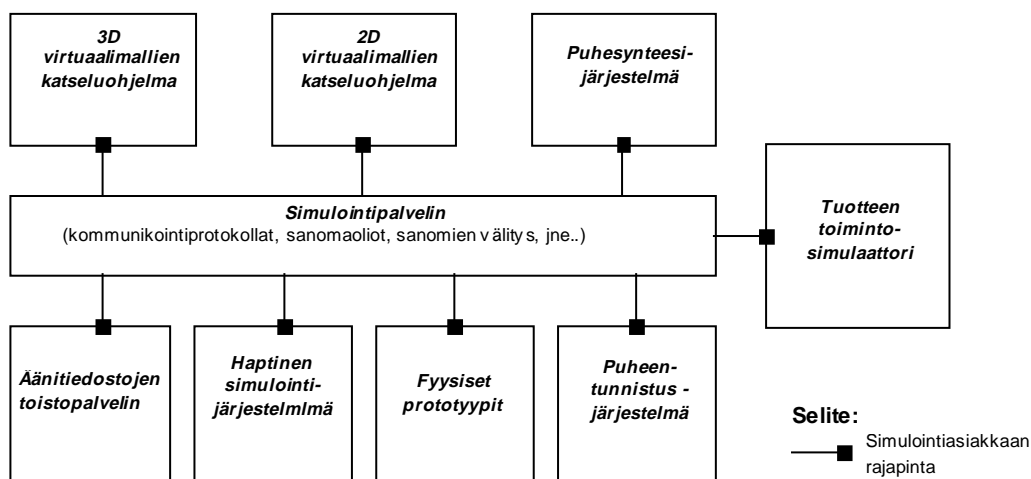
- *Visuaalisessa havainnollistamisessa* (visual rendering) pyritään antamaan mahdollisimman todentuntuinen kolmiulotteinen kuva virtuaaliprototyypistä. Kolmiulotteisten kuvien todentuntuisuutta voidaan lisätä käyttämällä stereokuvajärjestelmiä, joilla kuviin saadaan mukaan syvyysvaikutelma. Tarvittaessa, esimerkiksi suorituskykyyn liittyvistä syistä, 3D-visualisointi voidaan korvata myös perinteisemmällä 2D-visualisoinnilla.
- *Ääniin perustuvassa havainnollistamisessa* (auditory rendering) tuotteen ääniominaisuuksia jäljitellään tuottamalla synteettisiä ääniefektejä. Pidemmälle menevässä akustisessa havainnollistamisessa voidaan luoda virtuaalinen kolmiulotteinen ääniympäristö, jossa kullekin äänilähteelle on määritelty sijainti ja jossa kolmiulotteisen tilan vaikutusta ääniin, esimerkiksi heijastusten ja jälkikaikujen kautta, simuloidaan laskennallisesti. Myös puhesynteesi edustaa tätä havainnollistamisen muotoa.
- *Haptinen havainnollistaminen* (haptic rendering) tarkoittaa voima- ja tunto-vasteen tarjoamista erityisten haptisten käyttöliitäntälaitteiden kautta. Käyttäjä vastaanottaa haptiseen havainnollistamisen liittyvän palautteen pääasiallisesti tunto- ja lihasaistiensa kautta. Voima- ja tuntoaistimukset kytkeytyvät voimakkaasti ihmisen näköaistiin sekä kolmiulotteisten tilojen ja muotojen hahmottamiseen. Tämän vuoksi haptisen palautteen tarjoaminen tehostaa olennaisesti kolmiulotteisten virtuaaliobjektien tarkastelua ja käsittelyä.

Käyttäjän ja virtuaaliprototyypin välinen vuorovaikutus järjestetään erilaisten käyttöliitäntälaitteiden kautta, jotka voidaan luokitella syöte- ja tulostekäyttöliitäntälaitteisiin. Tulostelaitteita ovat erilaiset näyttölaitteet (esim. monitori, silmikkonäytöt, HDM-näyttökypärät), äänisyntetisaattorit kaiuttimineen sekä tunto- tai voimavasteen tarjoavat haptiset laitteet. Syötelaitteita ovat perinteisen näppäimistön ja 2D-hiiren lisäksi 3D-hiiri, käyttäjän tai ympäristön objektien monitorointiin käytettävät seurantalaitteet (tracking devices), haptiset ohjauslaitteet, mikrofonit ja puheentunnistusjärjestelmät.

### **Puheteknologioita soveltavien käyttöliittymien virtuaalisuunnittelu-ympäristö**

Kuvassa 14 on esitetty arkkitehtuurikuvaus virtuaalisuunnittelu ympäristöstä, joka soveltuu puheteknologioita hyödyntävien käyttöliittymien kehitykseen. Ympäristö sisältää seuraavat komponentit:

- Virtual Model 2D/3D Viewer, jota käytetään käyttöliittymän visuaalisen havainnollistamiseen. 3D-visualisointimalli tuotetaan joko 3D-mallinnustyökaluilla tai konvertoimalla se erityisillä konversiotyökaluilla kolmiulotteisista CAD-malleista. Visuaalinen palaute käyttäjälle välitetään normaalin monitorin kautta ja sitä voidaan haluttaessa tehostaa stereokuvajärjestelmällä (kuva 15). 2D-visualisointimalli rakennetaan tuotteen käyttöliittymän bittikarttakuvista.
- Logic Simulator, jolla kuvataan valittu osa tuotteen käyttäytymisestä ja toiminnallisuudesta ja joka toimii muut simulointikomponentit yhdistävänä komponenttina.
- Haptic System, jota käytetään voima- ja tuntovasteen simuloinnissa sekä visuaalisen mallin kanssa tahdistettuna kinemaattiseen simulointiin. Haptisen palautteen antamiseen tarvitaan erityinen tunto- ja voimavasteen tuottava haptinen ohjauslaite (kuva 15).
- Puhetunnistukseen (Speech Recognition System) ja -tuottamiseen (Speech Synthesis System) liittyvät ohjelmistot ja laitteet (äänisyntetisaattorit, kaiuttimet, stereokuulokkeet, mikrofonit, puheanalyysi- ja synteessiohjelmistot, jne), jotka on kytketty toiminnallisuuden simulointimalleihin.
- Virtuaaliprototyypiin voidaan kytkeä myös toimivia laitteita ja ohjelmistoja (Physical Mock-Ups). Tätä voidaan hyödyntää esimerkiksi kehitettäessä olemassa olevasta tuotteesta uuden sukupolven tuotetta. Valmiina olemassa olevan teknologian hyödyntäminen tuotekehityksessä nopeuttaa huomattavasti seuraavan sukupolven prototyypin tuottamista.



Kuva 14. Virtuaalisuunnittelu ympäristön arkkitehtuuri.

## VRP virtuaalisuunnitteluympäristö

VTT Elektronikassa on kehitteillä kuvan 14 arkkitehtuurin sisältämät komponentit tarjoava virtuaalisuunnitteluympäristö (VRP Prototyping System), joka on tarkoitettu erityisesti monimuotoisen käyttöliittymän omaavien elektroniikka- ja tietoliikennetuotteiden tuotekehityksen tukemiseen, kuva 15.



*Kuva 15. Virtuaaliprototyyppien havainnollistamisessa käytettäviä laitteita.*

## 5 SOVELLUSESIMERKKEJÄ

### 5.1 PUHEVIESTEJÄ HYÖDYNTÄVIÄ TUOTTEITA

#### 5.1.1 Paikannus- ja opastejärjestelmät

Puhesanomien käyttöä viestintäkeinona on tutkittu laajasti näkövammaisten ja lukuhäiriöisten apuvälineiden kehittämisen yhteydessä. Ihmisten tehokas liikkuminen edellyttää sopivia suunnistamisen apuvälineitä. Tavallisille ihmisille liikenne-merkit, katuosoitteet jne. toimivat tällaisina opasteina. Lukuhäiriöiset ja näkövammaiset eivät kuitenkaan voi tukeutua niihin. Tätä puutetta korvaamaan on Yhdysvalloissa kehitetty kaksi toimintaperiaatteeltaan poikkeavaa puheopastekonseptia, Verbal Landmarks® ja Talking Signs®. Järjestelmät poikkeavat toisistaan mm. viestin lähetys- ja tallennustapojen perusteella. Verbal Landmarks -järjestelmä käyttää induktiosilmukkaa ja synteettistä puhetta ja Talking Signs vastaavasti infrapunälähetintä ja digitaalisesti tallennettua tai synteettistä puhetta [Bentzen & Mitchell 1993]. Molempien hankkeiden yhteydessä on tehty merkittävää kehitystyötä mm. puheen liittämässä opasteisiin.

The Smith-Kettlewell Eye Research Instituten johdolla vuodesta 1978 alkaen toteutetussa Talking Signs -hankkeessa on tutkittu paitsi puheopasteita, jotka passiivisesti kertovat sijainnin ja antavat ohjeita johonkin paikkaan pääsemiseksi, myös etäopastetekniikkaa, jonka avulla voidaan tuottaa toistuvaa, suunnasta riippuvaa ääniviestiä. Talking Signs -järjestelmä koostuu opasteissa sijaitsevista infrapunälähtimistä sekä kädessä pidettävästä vastaanottimesta, joka sisältää kaiuttimen tai kuulokeliitäntän. Infrapunälähetite on suuntaherkkä (vrt. Verbal Landmarks, jossa käytetään induktiosilmukkaa), mikä lisää henkilön mahdollisuuksia paikantaa itsensä suhteessa opasteeseen. Talking Signs -opasteet ovat lyhyitä, yksinkertaisia ja selkeitä ja ne voidaan toteuttaa tallennetun puheen tai koneen tuottaman synteettisen puheen keinoin. [Bentzen 1995 , Crandall ym. 1995]

Talking Signs -järjestelmän esimerkkikohteita ovat Yhdysvalloissa mm.

- Department of Parking and Traffic, San Francisco
- San Francisco Municipal Railway
- Bay Area Rapid Transit (BART)
- The Washington D.C. Metro.

Yhdysvaltojen ulkopuolelta kohteita löytyy mm. Glaskowsta (Skotlanti), Venetsiasta (Italia) ja Yokohamasta (Japani).

Jatkokehityskohteena Talking Signs -järjestelmässä on puheopasteiden liittäminen erilaisten laitteiden, kuten esimerkiksi pankkiautomaattien, käyttöön. Tällöin sama järjestelmä voi ensin opastaa pankkiautomaatin luo, sen jälkeen kertoa ohjeet automaatin käytöstä ja lopulta opastaa painonappien sijainnin ja niiden toiminnot.

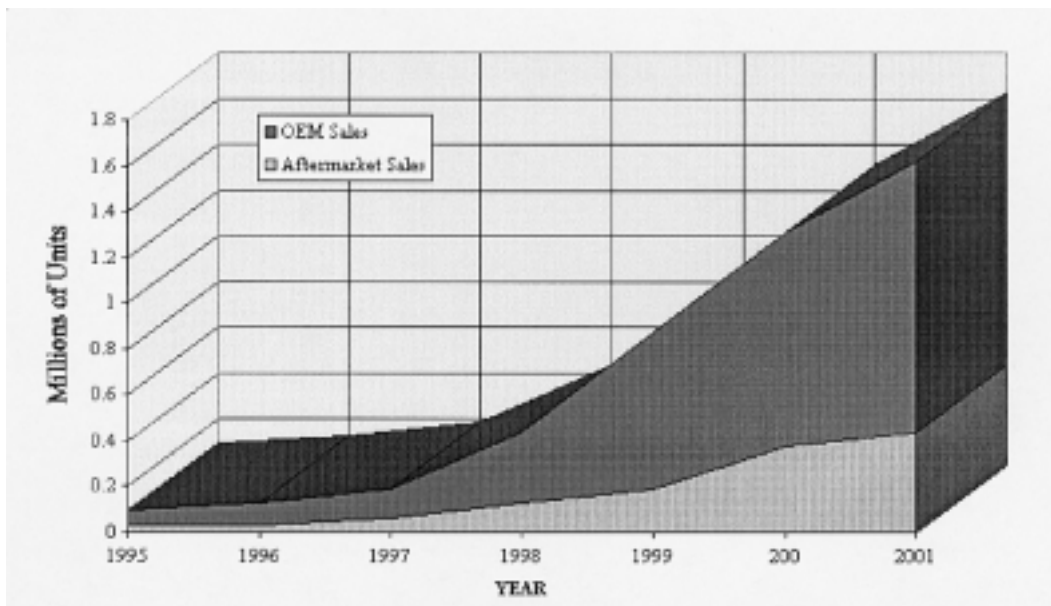
Vastaavan kaltaisen järjestelmän kehittyneemmän version käyttöönottoa suunnitellaan eurooppalaisessa yhteishankkeessa, jossa on osallistujia Belgiasta, Italiasta, Englannista ja Ranskasta. Hankkeessa on tavoitteena kytkeä kaikki opasteet verkon välityksellä keskustietokoneeseen, jolloin opasteiden valvonta ja muuttaminen on vaivatonta. Järjestelmästä tulee lisäksi monikielinen, jolloin käyttäjä voi itse valita, millä kielellä hän haluaa opasteet kuunnella. Lisäksi järjestelmästä tulee reaaliaikainen, sillä tarkoituksena on sen liittäminen mm. matkatietojärjestelmiin, joiden kautta saadaan tietoa junien ja linja-autojen lähtöajoista, laitureista jne. Suunnitteilla on myös liityntä GPS-paikannusjärjestelmään. [Stephens 1995, Strothotte ym. 1995]

Puheviestien käyttöä itsepalveluautomaateissa jne. tutkitaan mm. EU:n TIDE (Telematics for the Integration of Disabled and Elderly people) -projektiin liittyvässä Saturn-projektissa.

### **5.1.2 Ajoneuvo- ja kulutuselektronikka**

Liikennetietojärjestelmät ja ajoneuvonavigaattorit tarjoavat monia soveltamismahdollisuuksia puheviesteille. Autonavigaattoreiden markkinoiden oletetaan kasvavan lähivuosina voimakkaasti, kuva 16 [VODIS]. Grundig tuo vuoden 1998 aikana Euroopan markkinoille suunnatun autoradion, jossa RDS-järjestelmän kautta saatu liikennetiedote voidaan tallentaa automaattisesti muistiin. Tiedote voidaan kuunnella autoon tullessa, jolloin voidaan ennalta varautua kulloiseenkin liikennetilanteeseen. Liikkuva Systems International on tuonut Yhdysvaltain markkinoille Retki MetroGuide -autonavigaattorin, jossa ajo-ohjeet ja matkakohteisiin liittyvät tiedot tulostetaan synteettisen puheen avulla. Vastaavanlaisia tuotteita valmistavat useat muutkin yritykset.





Kuva 16. Autonavigaattorimarkkinoiden kehitysennuste Euroopassa [VODIS].

### 5.1.3 Matkapuhelimet ja henkilöhakulaitteet

Useat matkapuhelimien valmistajat ovat lisänneet tuotteisiinsa lyhyiden puheviestien tallennusmahdollisuuden. Tällöin puhelinta voidaan tarvittaessa käyttää myös käteväenä muistiona. Tällaisia puhelimia on saatavissa useilta laitevalmistajilta.

Henkilöhakulaitteissa puheen tallennus- ja toistopiirejä käytetään muistitoiminnon lisäksi hakukoodien muuttamiseen ennalta tallennetuiksi puheviesteiksi. Tällaisia henkilöhakulaitteita valmistaa mm. Swissphone Telecommunications Inc.

### 5.1.4 Terveydenhoidon sovellukset

Esimerkkinä puheen tai paremminkin äänen tallennuksen ja toiston käytöstä terveydenhoitoalan sovelluksissa voidaan mainita Telescan Medical Systemsin luoma palvelukonsepti, jossa sydänvaivoista kärsivä potilas itse tallentaa tarkoitusta varten kehitetyllä TeleScan 2000 -laitteella 45 sekunnin näytteen sydänäänistään. Näytteen hän lähettää puhelimen avulla palvelukeskukseen, joka näytteen esikäsitteilyn jälkeen toimittaa sen edelleen potilaan lääkärille analysoitavaksi.

### 5.1.5 Turvatekniikka

Puheviestien käyttö tarjoaa uusia keinoja turvallisuuteen liittyvien laitteiden käytön helpottamiseen ja käytettävyyden parantamiseen. Esimerkiksi MobileComm on lisännyt hätäpuhelimeensa toiminnon, jossa puhelinkaapin oven aukaisu lähettää hälytyskeskukseen automaattisesti viestin hälyttävän puhelimen sijainnista, kuva 17. (Esim. 'Hälytys puhelimesta 12, autohallin siivessä A.')

Tämän viestin avulla apu voidaan toimittaa oikeaan paikkaan huolimatta siitä, osaako hätäantanut avun tarvitsija käyttää hätäpuhelinta tai osaako hän kyseisen maan kieltä.



Kuva 17. EC-911-hätäpuhelin [ISD Customer. Part Number 2000996P005].

Kanadalainen TelAlert Inc. valmistaa turvajärjestelmää, jossa turvapuhelin soittaa saamansa hälytystiedon perusteella selväsanaisen hälytysviestin ennalta määrättyihin puhelinnumeroihin. Järjestelmään voidaan liittää erilaisia ilmaisimia, savuilmaisin, liikeilmaisoin, turvaranneke jne.

### 5.1.6 Laitteiden käytönopastus

Ennalta tallennettujen puheviestien käyttö tarjoaa uusia keinoja myös laitteiden käytönopastukseen. Suomessa on käytössä mm. Tomra Oy:n valmistamia pullonpalautusautomaatteja, jotka neuvovat käyttäjille pullojen oikean asettelun.

Hankalissa paikoissa ja käyttöolosuhteissa käytettävissä, kannettavissa tiedonkeruulaitteissa puheviestit voivat täydentää näyttöä varmistamalla syötetyn tiedon. Worthington Data Solutions valmistaa esimerkiksi viivakoodinlukijaa, joka ilmoittaa puheviestillä virheellisistä syöttötiedoista.

## 5.2 PUHEENTUNNISTUSTA HYÖDYNTÄVIÄ TUOTTEITA

Markkinoilla on saatavissa lukuisia kaupallisia puheentunnistusjärjestelmiä. Niiden tehokkuus, käyttötarkoitus ja soveltamismahdollisuudet poikkeavat kuitenkin suuresti toisistaan. Ensimmäiset yleiskäyttöiset puheentunnistusjärjestelmät julkaistiin kaupallisesti vuonna 1990. Yleiskäyttöisen järjestelmän ominaisuuksia ovat mm.

- puhujakohtainen puheentunnistus, joka vaatii käyttäjäkohtaisen opetuksen
- yksittäisten sanojen tunnistaminen lauseista
- sanaston koko 5 000 - 50 000 sanaa.

Puhujasta riippumatonta, yksittäisten sanojen tunnistusta rajoitetulla sanavarastolla on onnistuneesti sovellettu mm. matkapuhelimien käyttöliittymissä. Rajoitetun

sanavaraston puhekomento-ohjausjärjestelmiä on yleisesti saatavissa halvalla ja jopa ilmaiseksi mm. joidenkin PC-pakettien mukana. Suuren sanavaraston sisältävät, puhujasta riippumattomat puheentunnistusjärjestelmät ovat vielä kalliita, ECU 1 500 - ECU 3 700 [Kouroupetroglou & Németh 1995], jopa enemmänkin.

Käytännön sovelluksissa kriittisiä ovat sekä tunnistustarkkuus että puhujien ja ympäristön vaihtelun sieto. Tuskin yksikään järjestelmä toimii hyvin kaikissa olosuhteissa. Useita toimivia suuren sanavaraston omaavia jatkuvan puheen tunnistamiseen soveltuvia prototyypijärjestelmiä on jo esitelty, mutta ne on yleensä tarkoitettu rajatulle sovellusalueelle, kuten terveydenhuollossa röntgenkuvien saneeluun.

Tietoliikenne- ja tietokonetekniikka ovat tällä hetkellä suurimmat puheentunnistuksen soveltamisalueet. Tietoliikenteessä puheentunnistusta käytetään hyvin monenlaisissa sovelluksissa mm. tiedonhaun apuna. Useat teleyhtiöt ovat hankkineet puhelinyhteyksien kautta toimivia, puhujasta riippumattomia puheentunnistusjärjestelmiä. Vaikka ne toimivat pääasiallisesti pienellä sanavarastolla (esim. numerot ja paikannimet), niillä on erikoistarkoituksissa suuret käyttömahdollisuudet.

Seuraavassa on joitakin esimerkkejä tuotteista, joissa puheohjaus on toteutettu yksinkertaisten puhekomentojen avulla, mikä on moniin käyttötarkoituksiin riittävä puheentunnistuksen taso. Jatkuvan puheen tunnistaminen täydellisesti on luonnollisesti hyödyllinen tekstin ja muun tiedon syöttämisessä tietojärjestelmiin (ns. sanelinsovellukset).

### **5.2.1 Ajoneuvoelektronikka**

Puheohjauksen soveltaminen ajoneuvoissa käytettäviin elektronisiin laitteisiin on parhaillaan suuren mielenkiinnon kohteena, koska puhelinten, radion, CD-soittimien jne. käyttö ajon aikana lisää onnettomuusriskiä. Puheohjauksen avulla voitaisiin merkittävästi lisätä turvallisuutta. Melun ja muiden häiriöiden vuoksi auto on kuitenkin vaikea ympäristö puheentunnistuksen käyttöön.

Euroopan komission "Telematics Applications of Common Interest"-ohjelmaan kuuluvan tutkimushankkeen, Voice Operated Driving Information Systemin (VODIS), ensisijaisena tavoitteena on kehittää autokäyttöön soveltuvia puheohjattuja käyttöliittymiä. Päämääräksi on asetettu puheohjatun käyttöliittymän soveltaminen Blaupunkt Berlin RCM202A Navigointijärjestelmään. Käyttöliittymän avulla ohjataan navigointijärjestelmää, GSM-puhelinta, CD-soitinta ja radiota. Kehitysryhmään kuuluu jäseniä sekä Karlsruhen että Carnegie Mellonin yliopistoista. Yhteistyökumppaneina on lukuisia yrityksiä, Robert Bosch GmbH, Multicom Research Inc., RWTH-Aachen, Volkswagen AG, Renault Research Innovation, PSA Peugeot Citroen, Lernout & Hauspie Speech Products, Institute for Perception Research/IPO ja INESC.

Microsoft Corporation julkisti vuoden 1998 alussa Windows® CE 2.0 -käyttöjärjestelmään pohjautuvan Auto PC -järjestelmän, joka sisältää sisäänrakennettuna puheentunnistuksen, kuva 18. Auto PC -konseptin mukaisia laitteita kehittävät lu-

kuisat laitevalmistajat, mm. Alpine, Clarion ja Samsung. Perusversiossaan laitteet sisältävät puheohjauksella toimivan radion ja CD-soittimen. Lisäksi järjestelmään voi kuulua esimerkiksi GPS-vastaanotin, CD-levyjen vaihtaja sekä liityntä matkapuhelimelle.



*Kuva 18. Auto PC Powered by Microsoft® Windows® CE 2.0 [Microsoft 1998].*

## 5.2.2 Matkapuhelimet

Joihinkin käsipuhelimiin on liitetty puheohjattu puhelinnumeron valinta. Käyttäjä voi soittaa haluttuun puhelinnumeroon lausumalla esimerkiksi numeron yhteyteen tallentamansa nimen. Tällainen puhelin on saatavissa esimerkiksi Philipsiltä. Nokia valmistaa vastaavan kaltaista puhelinta Japanin markkinoille.

## 5.2.3 Kotiautomaatio

Yhdysvalloissa on kaupallisesti saatavissa useita puheohjattuja kotiautomaatiojärjestelmiä sekä erillisiä puheohjauspaketteja, joiden avulla puheohjaus voidaan liittää olemassa oleviin automaatiojärjestelmiin.

Automated Voice Systems, Inc.(AVSI) valmistaa kolmea erikokoista kotiautomaatiojärjestelmää, joita voidaan hallita puhekomentojen avulla. Järjestelmien avulla voidaan ohjata kaikkia sähkö- ja elektroniikkalaitteita, esimerkiksi valaistusta, stereoita, televisiota, lämmitys- ja jäädytysjärjestelmiä sekä turvatekniikkaa.

Puheohjaus tunnistaa mitä tahansa kieltä tai murretta. Järjestelmään sisältyy sekä puhujakohtainen että puhujasta riippumaton puheentunnistus. Käyttöönottovaiheessa käyttäjä opettaa järjestelmälle tietyjen sanaryhmien sanat. Järjestelmien ominaisuuksia on esitetty taulukossa 15.

Taulukko 15. AVSI:n kotiautomaatiojärjestelmien ominaisuuksia [AVSI 1997].

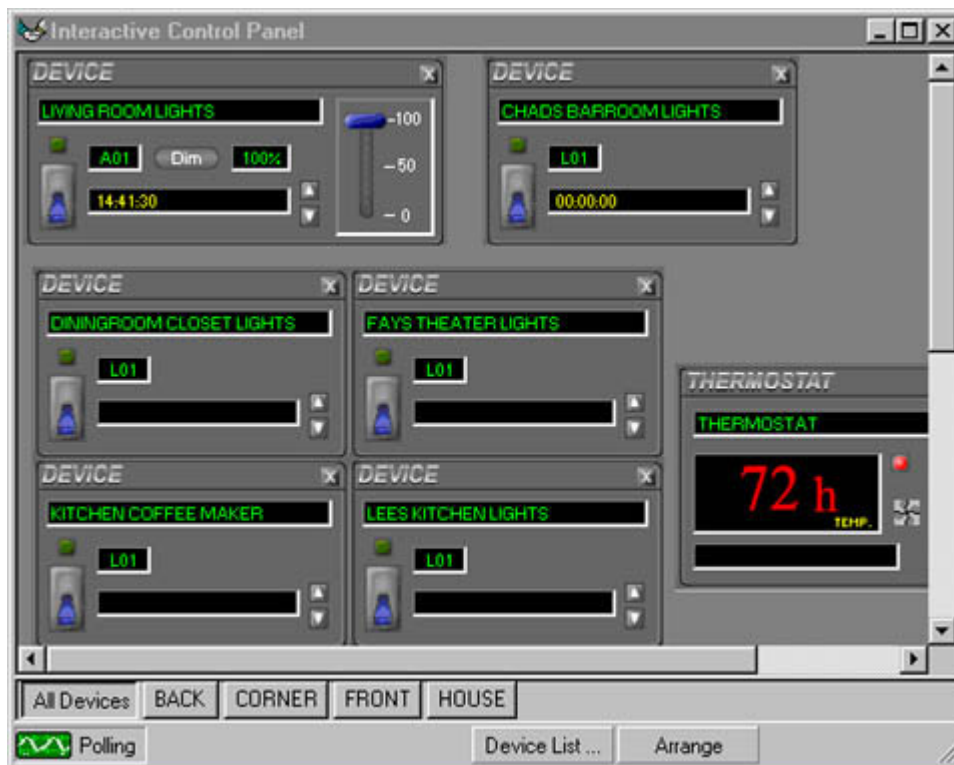
	<b>E.C.U Envi- ronmental Control Unit</b>	<b>Butler-In-A- Box</b>	<b>Mastervoice Series II</b>
<b>Ääniohjausta käyttäviä henkilöitä</b>	1	2	4
<b>Ääniohjattavia laitteita</b>	42	32	42
<b>Kosketusohjattavia laitteita</b>	256	256	256
<b>Ajastimia</b>	16	32	99
<b>Makroja</b>	8	16	32
<b>Puhelinnumeroita</b>	16/henkilö, 15 merk- kiä/numero	16/henkilö, 15 merkkiä/numero	16/henkilö, 15 merkkiä/numero
<b>Laiteluettelossa olevia laitteita</b>	64	64	96
<b>Suositusuhinta</b>	1795 USD	2995 USD	3995 USD

Lisäksi AVSI valmistaa erillistä IntellaVoice -järjestelmää, jonka avulla X-10-pohjainen kotiautomaatiojärjestelmä voi vastata puheviestien avulla. Järjestelmän ohjehinta Yhdysvalloissa on alle 400 USD.

Home Automated Living -yrityksen HAL2000 on monipuolinen kotiautomaatio-ohjelmisto, jonka keskeisenä osana on puhujasta riippumaton puhekomento-ohjaus. Järjestelmä on liitettävissä mm. X-10-pohjaisiin kotiautomaatiolaitteisiin. Laitteistovaatimukset ovat:

- Windows 95
- keskusyksikkö vähintään 486
- 12 MB keskusmuistia
- 60 MB kiintolevytilaa
- PhoneBlaster äänikortti (tai vastaava)
- puhelin ja internet-liityntä.

Kuvassa 19 on esitetty käyttöliittymäesimerkkinä HAL 2000 -järjestelmän interaktiivinen säätöpaneeli.



Kuva 19. HAL 2000 -kotiautomaatiojärjestelmän interaktiivinen säätöpaneeli [HAL 2000].

Applied Future Technologies, Inc. -yrityksen HomeVoice™-ohjelmiston avulla puhujasta riippumaton puheohjaus on liitettävissä useisiin kotiautomaatiojärjestelmiin. Puhekomennot liitetään automaatiojärjestelmään ohjelmoituihin toimintoihin. Puheentunnistukseen voidaan sisällyttää lisäksi puhujantunnistus, jolloin esimerkiksi turvajärjestelmän kytkeminen pois päältä edellyttää komentojen antajaksi määrätyn henkilön. HomeVoice™ pystyy myös vastaamaan puheviestien avulla. Viestit voivat olla joko ennalta talletettuja äänitiedostoja tai puhesyntetisaattorilla tuotettua puhetta. HomeVoice™-järjestelmän puheentunnistus pohjautuu Verbex Voice Systemsin puheentunnistustekniikkaan. Laitteistovaatimukset ovat:

- Windows 95
- keskusyksikkö Pentium 75 MHz
- 16 MB keskusmuistia
- 165 MB kiintolevytilaa
- Sound Blaster yhteensopiva äänikortti
- mikrofoni, kaiuttimet.

Ohjelmiston hinta on noin 240 USD.

#### 5.2.4 Tiedonkeruulaitteet ja erikoistietokoneet

Puheentunnistuksen liittäminen tiedonkeruulaitteisiin helpottaa toimintaa monissa liikkuvuutta ja joustavuutta vaativissa teollisuuden ja kaupan tiedonkeruutarpeis-

sa. Tällaisia tiedonkeruulaitteita on jo saatavissa joiltakin valmistajilta (Verbex Voice Systems, Inc. ja Vocollect, Inc.). Samoin jotkin henkilökohtaiseen käyttöön tarkoitetut elektroniset kalenterit ja muistiot sisältävät puheentunnistustekniikkaa, esimerkiksi Voice Powered Technology International, Inc. -yhtiön valmistamat IQ•VOICE™ Organizer -mallit.

Erikoistarkoituksiin, kuten sotilas- ja huoltotehtäviin, suunnitelluissa ns. päällepuettavissa tietokoneissa puheentunnistus on olennainen osa käyttöliittymää. Tällaisia tietokoneita valmistavat mm. ViA Inc., kuva 20, ja Xybernaut Corporation.



*Kuva 20. ViA Wearable - päällepuettava tietokone [ViA 1998].*

### 5.3 PUHUJANTUNNISTUSTA HYÖDYNTÄVIÄ TUOTTEITA

Puhujantunnistus tai -varmistus on usein liitettyä puheentunnistukseen taikka se on yhtenä osana erilaisia tunnistusmenetelmiä sisältäviä henkilön tunnistusjärjestelmiä.

Puhujantunnistusta kehitetään ennen kaikkea uusien telepalveluiden ja tietoverkkojen käyttöturvallisuuden lisäämiseksi. Euroopassa tutkimusta tehdään mm. COST250 (Speaker recognition in Telephony) ja Esprit CAVE (Caller Verification In Banking and telecommunication) -projekteissa. Tele- ja tietoverkkosoveluksiin kehitettyjä tuotteita tarjoavat esimerkiksi iNTELiTRAK Technologies, Inc., ITT, Keyware Technologies ja Veritel Corporation.

Puhujantunnistusta käytetään myös kulkuoikeuksien tarkastamisessa rajan ylityksissä sekä vankeinhoidon apuna. Esimerkiksi yhdysvaltalaisen BI Incorporated -yrityksen BI Voice Verification -järjestelmää käytetään kotona olevien, mutta rajoitetun ulkonaliikkumisoikeuden omaavien rikollisten valvonnassa.

# 6 PUHETEKNOLOGIAN SOVELTAMINEN RAKENNUSTEN TEKNISISSÄ LAITTEISSA JA JÄRJESTELMISSÄ

## 6.1 KODINKONEET

Kodinkone on tyypillinen laite, johon voidaan sisällyttää entistä enemmän älykkyyttä pienin kustannuksin. Tätä ei voi kuitenkaan täysin hyödyntää ilman monipuolisia ja kalliita käyttöliittymäratkaisuja. Puheteknologia tarjoaa siihen edullisen vaihtoehdon ja mahdollistaa siten kodinkoneen toimintojen ja älykkyyden täysipainoisemman hyödyntämisen.

Puheteknologia helpottaa kodinkoneen käyttöä tekemällä käyttöliittymästä ymmärrettävän. Sumean logiikan soveltamisella kodinkoneiden ohjaukseen pyritään nykyisin samaan lopputulokseen. Kumpikin lähestymistapa ratkaisee ongelman eri tavalla. Sumean logiikan avulla kodinkoneet kykenevät toimimaan oikein käyttäjän antamien epätasmoisten tai osittain puutteellistenkin komentojen mukaan. Puheteknologia tekee puolestaan mahdolliseksi sen, että kodinkone opastaa puheviestien avulla käyttäjää antamaan eri käyttötilanteisiin oikeita ja täsmällisiä komentoja kodinkoneelle. Toisaalta puheteknologiaa voidaan hyödyntää myös siten, että kone tulkitsee puheentunnistuksen avulla käyttäjän antamat, hieman epätasmoiset ohjeet ja valitsee niiden perusteella oikean toimintaohjelman.

Edellä mainittujen puheviestien ja puheentunnistuksen lisäksi kodinkoneiden ohjaukseen soveltuvat myös puhujantunnistus, puhekomennot ja puhesynteesi. Puhujantunnistamisen avulla kodinkone voi ottaa huomioon kunkin käyttäjän omat tottumukset tai tieto- ja taitotason. Puheentunnistuksella varustettu kone kykenee puolestaan vastaamaan käyttäjän esittämiin erityiskysymyksiin ja opastamaan esimerkiksi puhesynteesin avulla käyttäjää erilaisissa käyttö- ja vikatilanteissa.

Sopivia kodinkoneita puheteknologian sovelluksiin ovat mm. viihde-elektronikka, pesukoneet, uunit, liedet ja tietokoneet. Esimerkkejä, joissa puheteknologiaa sovelletaan kodinkoneiden ohjaukseen, on helppo keksiä, sillä monet kodinkoneet sisältävät runsaasti erilaisia ohjelmia ja toimintamuotoja, mutta niiden käyttöliittymä on puutteellinen. Teknisesti ja myös kustannuksiltaan puheteknologia soveltuu itse asiassa kaikenlaisissa kodinkoneissa hyödynnettäväksi, mutta puheviestintä on toistaiseksi verrattain uusi piirre niiden toiminnassa ja vaatii kuluttajien hyväksymisen. Siten on vaikea ennakoida, missä laajuudessa tai missä kodinkoneissa uusi teknologia yleistyy. Sama koskee muitakin puheteknologian sovellusalueita.

## 6.2 AUTOMAATIO

Automaation sovellusalueella on tarjolla jo kaupallisia puheteknologiatuotteita, kuten edellä luvussa 4 todettiin. Rakennusten valaistus, lämmitys ja jäähdytys, ilmastointi, säätölaitteet, valvontajärjestelmät sekä pientalojen kotiautomaatio ovat tämän alueen tyypillisiä kohteita. Tähän ryhmään voitaneen lukea myös ne laitteet



ja järjestelmät, jotka helpottavat LVIS-järjestelmien huoltoa ja ylläpitoa. Huolto- toiminta nopeutuu, jos huoltomiehen ei tarvitse samanaikaisesti käyttää käsiänsä sekä huoltotoimenpiteen suoritukseen että sen kirjaamiseen. Jälkimmäinen voidaan hoitaa esimerkiksi puheentunnistuksen avulla. Vastaavia järjestelmiä on jo käytössä muualla kuin LVIS-alalla.

Puheen tallennus ja toisto sopii laitteisiin, joiden käyttöliittymä muilla tekniikoilla jäisi puutteelliseksi tuotteen vaatimuksiin kuuluvan alhaisen hankintakustannuksen vuoksi. Puhekomennot ja erityissanavarastolla toimiva puheentunnistus soveltuvat laitteiden toimintojen ohjaukseen, jälkimmäinen soveltaen sanojen tarkkailu (word spotting) -menetelmää. Myös puhesynteesi tai minimitapauksessa tallennetut puheviestit ovat tarpeellisia, sillä käyttäjä haluaa varmennuksen antamansa komennon tai ohjeen perille menosta. Puhujantunnistus sopii puolestaan kohteisiin, joissa järjestelmän käyttöoikeus on rajattu tai halutaan varmistaa, etteivät ulkopuoliset henkilöt voi käyttää järjestelmää. Viime mainittu ominaisuus tulee esille esimerkiksi ohjattaessa automaatiotoimintoja rakennuksen ulkopuolelta puhelinverkon avulla.

### 6.3 TURVAJÄRJESTELMÄT

Turvajärjestelmillä tarkoitetaan tässä kulunvalvonta-, rikosilmoitus- ja paloilmotusjärjestelmiä. Niiden ohjaukseen soveltuvat esimerkiksi puheen tallennus ja toisto, puhujantunnistus sekä puhekomennot. Puheen tallennuksen ja toiston avulla rakennuksen käyttäjää voidaan paitsi opastaa ja tiedottaa myös varoittaa vaaratilanteista tai liikkumisesta luvattomalla alueella.

Puhujantunnistus voidaan liittää esimerkiksi osaksi rikosilmoitusjärjestelmää tai kulunvalvontajärjestelmää. Yhdessä puhekomentojen kanssa sitä voisi hyödyntää molemmissa järjestelmissä. Se mahdollistaa kortittoman ja samalla viiveettömän liikkumisen kulkuoikeuksilla rajatussa rakennuksessa. Jos tunnistus halutaan tehdä varmemmaksi kuin puhujantunnistuksella on käytännössä mahdollista, voidaan tämä menetelmä yhdistää jonkin toisen biometrisen tunnistusmenetelmän kanssa.

Paloilmotusjärjestelmissä puheen tallennus ja toisto soveltuu opastustarkoituksiin hälyttävän kohteen luokse tai opastamaan rakennuksen sisällä olijat ulos. Kulkuohje voidaan tallentaa paloilmotusjärjestelmää asennettaessa. Periaatteessa samanlainen toiminto voidaan sisällyttää jopa pieneen palohälyttimeen, sillä toiminnon tekninen toteutus voidaan tehdä kustannuksiltaan hyvin edullisesti.

### 6.4 VAMMAISET JA VANHUKSET

Vammaiset ja vanhukset muodostavat oman erityisryhmänsä, joille puheteknologia voi tarjota uusia erilaisia tuotteita ja palveluja. Lähivuosikymmenien aikana tapahtuva vanhusten määrän voimakas lisääntyminen pakottaa etsimään uusia ratkaisuja vanhusten kotona asumisen ja liikkumisen helpottamiseksi. Puheteknologia voi auttaa vanhuksia esimerkiksi kodinkoneiden, kotiautomaation ja turvajärjestelmien käytössä. Jos laitteiden suunnittelussa sovelletaan puheteknologiaa,

voitaisiin niiden toiminnoissa pienin kustannuksin ottaa huomioon vanhusten kaltaiset käyttäjien erityisryhmät. Vammaisten osalta tilanne on sama. Puheviestintä on hyvä menetelmä ohjata erilaisia laitteita ja apuvälineitä, jos liikuntakyky tai näkökyky on rajoitettu. Vastaavan ryhmän muodostavat lukutaidottomat tai osittain kielitaidottomat henkilöt, jotka ymmärtävät puheviestejä. Kaikilla näillä alueilla voi hyödyntää mm. puheen tallennusta ja toistoa, puhekomentoja, puhujantunnistusta ja puhesynteesiä.

## 6.5 MUUT PUHETEKNOLOGIASOVELLUKSET

Puheteknologian yleistyminen muuttaa entisiä ja luo uusia tuoteratkaisuja. Opatustus- ja informaatiojärjestelmät on eräs alue, jolla puheteknologiaa voidaan hyödyntää rakennuksissa nykyistä enemmän. Puheen tallennusta ja toistotekniikkaa hyödynnetään jo nyt mm. museoiden ja julkisten rakennusten opastusjärjestelmissä. Sen avulla voidaan vähentää työvoiman tarvetta rutiinomaisissa tehtävissä. Samanlaiset mahdollisuudet koskevat myös informaation välittämistä rakennuksen käyttäjille. Tässä voi hyödyntää mm. puhekomentoja ja puheentunnistusta sekä puhesynteesiä.

Useat rakennusten tekniset järjestelmät ja laitteet liittyvät telekommunikaatiojärjestelmiin, joka on puheteknologian voimakkaimpia kasvualueita. Osa puheteknologian sovelluksista rakennuksissa tulee yleistymään telekommunikaation kautta. Laitteiden ja järjestelmien toiminnan ohjaus tai niiden toiminnan tarkistus tapahtuu vaivattomasti hyödyntämällä olemassa olevia langallisia tai langattomia viestintämahdollisuuksia. Telekommunikaatioalueella monet puheteknologian eri menetelmät ovat tavalla tai toisella jo käytössä.

## 7 CASE - PUHETEKNOLOGIATUOTTEEN SUUNNITTELU

### **Puhuva Lämmönsäädin**

Ouman Oy:n Puhuva lämmönsäädin on tuote, joka on tarkoitettu kiinteistöjen lämmityksen ja käyttöveden säätöön. Ouman Oy julkisti uuden tuotensa keväällä 1998. Lämmönsäädin oli aluksi tarkoitus suunnitella perinteistä tekniikkaa soveltaen. Yrityksen osallistuminen puheteknologiaprojektiin muutti kuitenkin aiempia suunnitelmia, ja toimiva puheviestintä otettiin tavoitteeksi osana säätimen käyttöliittymän suunnittelua. Puhuva lämmönsäädin koostuu kahdesta toisiinsa kytkettävästä laitteesta, varsinaisesta säätölaitteesta sekä puheviestejä tallentavasta ja toistavasta yksiköstä, kuva 21. Puheviestiyksikkö on pistokeliittimellä säätölaitteeseen kytkettävä lisävaruste. Sen tarkoituksena on opastaa säätölaitteen käyttäjää.



*Kuva 21. Ouman Oy:n Puhuva lämmönsäädin.*

### **Suunnittelukriteerejä**

Uudelle tuotteelle ja sovellettavalle puheteknologialle asetettiin mm. seuraavia vaatimuksia. Koska kyseessä oli kiinteistöihin sijoitettava LVIS-alan tuote, sen hinta ei saanut muodostua liian korkeaksi kilpaileviin tuotteisiin nähden. Puhe-

viestien äänen laadun olisi oltava kohtuullinen tai hyvä, jotta käyttäjälle syntyisi luottamus laitteen toimintaa kohtaan. Laitteen olisi kyettävä tallentamaan ja toistamaan viestejä neljästä kahdeksaan minuuttia. Puheteknologiatuotteiden valintaan vaikuttivat myös mm. tarvittavien komponenttien lukumäärä ja siten muodostuvan piirilevyn koko, sillä käytettävä tila oli rajallinen. Myös kehitystyökalun saatavuudella ja sen ominaisuuksilla oli merkitystä. Esimerkiksi kehitystyön helppous ja nopeus oli eräs työkalun valintaan vaikuttava tekijä, samoin hinta. Jos puheteknologiakomponenteille löytyi kakkostoimittaja, se katsottiin sille eduksi.

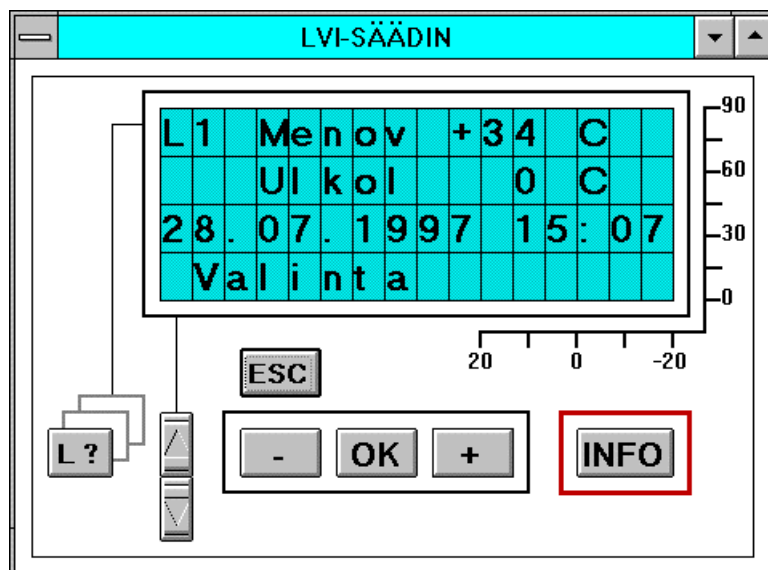
### **Valitut komponentit ja työkalut**

Puheen tallennukseen ja toistoon sopivia puheteknologian tuotteita on runsaasti saatavana. Puheteknologian uutuudesta johtuen komponenteille ei yleensä löydy kakkosvalmistajia. Siksi tätä kriteeriä ei pidetty ehdottomana vaatimuksena. Tallennusajan, äänen laadun sekä työkalujen suhteen useat eri vaihtoehdot tulivat kyseeseen. Jonkin verran valintaa rajoittivat tarvittavien komponenttien lukumäärä ts. piirilevyn koko, tarvittavan kehitystyökalun ominaisuudet sekä komponenttien ja työkalujen hintatekijät. Toteutuksessa päädyttiin ISD:n piiritekniikan valintaan. Toisena vaihtoehtona olleet OKIn puheteknologiatuotteet karsiutuivat lähinnä suuremman piirilevyn tilan tarpeen ja vaativamman kehitystyön takia. Kehitystyökaluksi valittiin Quadravoxin ohjelmistot ja laitteet.

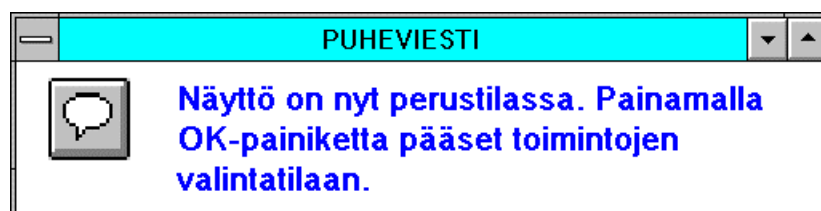
### **Käyttöliittymän ja puheviestiyksikön suunnittelu**

Hintavaatimuksista johtuen säätölaitteen käyttöliittymä varustettiin 4 x 16 merkin alfanumeerisella LCD-näytöllä sekä muutamalla näppäimellä. Laite sisältää kuitenkin monipuolisia toimintoja, joiden hallinta yksinkertaisella käyttöliittymällä on vaativaa. Näytön toimintaohjetekstit joudutaan usein lyhentämään tai itse toiminto keskeyttämään, jotta asiaa koskeva toimintaohjeteksti saadaan näytölle. Puheviestiyksikkö opastaa käyttäjää käyttöliittymän toiminnoissa. Samanaikainen näytön teksti sekä puheviesti tekevätkin kommunikoinnin laitteen kanssa yksinkertaiseksi ja selkeäksi. Käyttöliittymän tekstien sekä puheviestien suunnittelussa hyödynnettiin Visual Basic -ohjelmointiympäristöä, jonka avulla tekstien ja niihin liittyvien puheviestien alustava suunnittelu oli sujuvaa, kuvat 22 ja 23.

Puheviestiyksikkö suunniteltiin aluksi sijoitettavaksi säätölaitteekotelon sisään lisävarusteeksi. Myöhemmin tämä ajatus hylättiin ja yksiköstä tehtiin täysin erillinen lisälaitte. Varaamalla itse säätölaitteeseen kustannuksiltaan huokea liitäntä puheviestiyksikkö voidaan myöhemminkin liittää osaksi kaikkiin samanmerkkisiin säätölaitteisiin. Koska puheviestiyksikkö on erillinen, irrotettava, taskukokoinen laite, se on luonteeltaan apuväline, jota kokenut käyttäjä ei välttämättä tarvitse. Laite voidaan esimerkiksi jättää asiakkaalle väliaikaisesti säätölaitteen toimintoihin tutustumista ja laitteen toimintojen opiskelua varten, mutta tarvittaessa laite voidaan asentaa myös pysyvästi säätölaitteen yhteyteen.



Kuva 22. Ikkuna lämmityksen säätökäyrän asetelusta säätölaitteen käyttöliittymän toimintaa simuloivassa Visual Basic -ohjelmassa. Ikkunaan liittyvä puheviesti on esitetty kuvassa 23.



Kuva 23. Kuvan 22 ikkunaan liittyvä puheviesti simuloitaessa säätölaitteen käyttöliittymää Visual Basic -ohjelmointikielellä.

Valitut komponenttiratkaisut ja työkalut mahdollistivat laitteiston nopean toteutuksen rinnakkain varsinaisen säätölaitteen suunnittelun kanssa. Valmis puheviestisyksikkö oli toteutukseltaan kuta kuinkin suunnitellun kaltainen. Puhemuistin säästämiseksi jotkin lauseet jouduttiin kokoamaan erillisistä sanoista, jolloin puhujan äänen painotus oli joissakin kohdissa luonnollisesta poikkeava. Virhe voidaan kuitenkin helposti korjata uudella äänityksellä. Kehitystyökalussa ilmeni selviä valmistusvikoja, jotka reklamoinnin ja ohjelmistopäivityksen jälkeen kuitenkin poistuivat. Quadravoxin kehitystyökalu vaati toimiakseen PC:n äänikortteineen ja -ohjelmistoineen. Muutaman äänityskokeilun jälkeen näytti ilmeiseltä, että PC-ympäristössä nauhoitettu puhe sisälsi tarpeettomia häiriöitä. Siksi lopullinen äänitys tehtiin studio-olosuhteissa ja valmis äänite ohjelmoitiin työkalujen avulla ISD-piirille. Puhe on tallennettu 8 Khz:n näytteenottotaajuudella, joka mahdollistaa puhelinverkon kaistaleveyden käytön puheviestin toistossa. Äänen laatu sovellukseen nähden vaikutti riittävältä. Äänen laatuun vaikuttavat paitsi nauhoituksen ja toiston kaistaleveys myös äänentoistotekniikka.

## 8 YHTEENVETO

Puheen käyttö ihmisen ja koneen välisessä vuorovaikutuksessa on ollut suuren mielenkiinnon kohteena useiden vuosikymmenien ajan. Automaattisen puheentunnistuksen historia ulottuukin noin viidenkymmenen vuoden taakse. Puheen käyttö tarjoaa uusia mahdollisuuksia laitteiden ja järjestelmien käyttöliittymien toteuttamiseen. Puheteknologian avulla on mm. mahdollista vapauttaa kädet ja näköaisti muita toimintoja varten, mikä lisää käyttäjän turvallisuutta ja toimintamahdollisuuksia. Tästä syystä vammaiset ja vanhukset muodostavat oman erityisen käyttäjäryhmänsä puheteknologiatuotteille.

Käytännössä puheteknologian sovellutukset ovat vielä olleet harvinaisia. Viime vuosien aikana menetelmiä ja teknologiaa on kuitenkin kehitetty sille tasolle, että sen soveltaminen on kaupallisiin tuotteisiin mahdollista. Puheteknologiatuotteiden markkinoiden odotetaan kasvavan voimakkaasti lähimpien vuosien aikana. Suurimmat odotukset ovat tele- ja tietoliikenteessä sekä kulutuselektronikassa. Näiden alojen tuotteissa puheteknologiaa käytetään nykyisellään jo monipuolisesti.

Puheteknologian hyödyntäminen on kaupallisesti pisimmällä puheviestien osalta. Puheviestit voidaan toteuttaa yksinkertaisimmillaan integroidun puheen tallennus- ja toistopiirin avulla. Puheentunnistustekniikka on puolestaan parhaillaan kaupallistumassa ja avaa samalla tietä puhujantunnistuksen sovelluksille. Puheteknologian leviämistä edistää tarvittavien elektronisten komponenttien ja ohjelmistojen sekä tuotekehitystä helpottavien kehitystyökalujen tarjonnan lisääntyminen samoin kuin yleisesti hyväksytyjen ohjelmistorajapintojen aikaansaaminen.

Puheteknologian nykyinen kehitystaso luo hyvät edellytykset puheviestinnän sisällyttämiselle myös rakennusten teknisiin laitteisiin ja järjestelmiin, joissa puheteknologian soveltaminen vielä on harvinaista. Hyvät mahdollisuudet ovat olemassa jopa uusien, käyttäjiä aikaisempaa paremmin palvelevien tuotteiden kehittämiselle.

## LÄHDELUETTELO

Anon. 1996. DSPs wrestle with CPUs in the embedded arena. *Computer Design*, Vol. 35, nro 4, March, s. 75 - 87.

AVSI 1997. Automated Voice Systems, Inc. [online]. [cited 1997-12-30]. Available from Internet: <URL:<http://www.mastervoice.com/>>

Bentzen, B. L. & Mitchell, P. 1993. Audible signage as a wayfinding aid: Comparison of "Verbal Landmarks®" with "Talking Signs®". *Accessible Design for the Blind*, Berlin, MA -1503, September 10, 1993 [cited 1997-12-22]. 16 s. Available from Internet: <URL:<http://www.ski.org/RERC/WCrandall/ACBPAPER.HTM>>

Bentzen, B. L. 1995. Overview of Talking Signs Concept. Teoksessa: *Proceedings of a Symposium on Consumer, User Agency, Researcher, and Commercial Experience with Talking Signs and Related Technologies*, San Francisco, California, June 27, 1995. [cited 1997-12-22] S. 4 - 5. Available from Internet: <URL:<http://www.ski.org/RERC/WCrandall/SYMPOS.HTM>>

Bursky, D. 1998. Superscalar processor delivers 400 MIPS for DSP and control needs. *Electronic Design*, Vol. 46, Number 3, February 9, s. 37 - 42.

ChipCorder® Data Book. Voice Recording and Playback ICs 1997 Edition. Information Storage Devices, Inc. 554 s.

Crandall, W., Bentzen, B. L., Myers, L. & Mitchell, P. 1995. Transit Accessibility Improvement Through Talking Signs® REMOTE Infrared Signage: A Demonstration and Evaluation. The Smith-Kettlewell Eye Research Institute, San Francisco, March 15, 1995 [cited 1997-12-22]. 21 s. Available from Internet: <URL:<http://www.ski.org/RERC/WCrandall/FINAL.HTM>>

Forman, P. P. 1997. IBM, Dragon tools create buzz. *Computer Reseller News*, December 1, 1997 [online]. [cited 1998-04-06]. Available from Internet: <URL:<http://www.techweb.com/se/directlink.cgi?CRN19971201S0042>>

Goodenough, F. 1996. Single IC stores and plays back 4 minutes of voice. *Electronic Design*, July 22, s. 67 - 76.

HAL 2000 [online]. Homea Automated Living [cited 1997-12-10]. Available from Internet: <URL: <http://www.automatedliving.com/>>

ISD Customer Application. Where safety and technology meet. Information Storage Devices. Part Number 2000996P005.

Karjalainen, M. 1993. Puheen- ja signaalinkäsittely. Teoksessa: Hyvönen, E., Karanta, I. & Syrjänen, M. (toim.). *Teokälyn ensyklopedia*. Hämeenlinna: Oy Gaudemus Ab. S. 99 - 103. ISBN 951-662-5559-2

Klatt, D. H. 1987. Review of Text-to-Speech Conversion for English. *Journal of the Acoustic Society of America*, Vol. 82, nro 3, September, s. 737 - 793.

Koivunen, M-R. 1993. Ihmisen ja koneen vuorovaikutus. Teoksessa: Hyvönen, E., Karanta, I. & Syrjänen, M. (toim.). *Teokälyn ensyklopedia*. Hämeenlinna: Oy Gaudeamus Ab. S. 319 - 326. ISBN 951-662-5559-2

Kondozi, A. M. 1994. *Digital Speech – Coding for Low Bit Rate Communications Systems*. John Wiley & Sons Ltd, New York. 456 s. ISBN 0-471-95064-5

Kouroupetroglou, G. & Németh, G. 1995 [online]. Teoksessa: Roe, P.R.W. (ed.) *Telecommunications for all. ECSC-EC-EAEC, Brussels\* Luxembourg 1995*. Belgium, CD-90-95-712-EN-C, 1995 [cited 1997-12-12]. Available from Internet: <URL:<http://www.stakes.fi/cost219/TELLALL96.DOC>>

Lernout & Hauspie 1997 [online]. *Text To Speech Products* [cited 1997-12-18]. Available from Internet: <URL:<http://www.lhs.com/coretechnologies/ttsproducts.htm>>

Lernout & Hauspie 1998 [online]. *Embedded products, ASR Dev Tools, L & H ASR1600 for RISC* [cited 1998-04-22]. Available from Internet: <URL:<http://www.lhs.com/speechtech/embddevtools/asr1600risc.asp>>

Microsoft 1998 [online]. *PC Companions Auto PC* [cited 1998-04-02]. Available from Internet: <URL:<http://www.microsoft.com/windowsce/autopc/about/datasheet.htm>>.

Naik, J. M. 1990. *Speaker Verification: A Tutorial*. *IEEE Communications Magazine*, January, s. 42 - 48.

Phillips, K. 1997. *Biometric identification comparison chart*. *PCWeek*, March 26, 1997. Ziff-Davis Publishing Company.

Rabiner, L. R. & Juang B-H. 1993. *Fundamentals of Speech Recognition*. Prentice Hall, Englewood Cliffs. 496 s. ISBN 0-13-015157-2

RSC-164 Datasheet [online]. *General Purpose Microcontroller Featuring Speech Recognition, Speech & Music synthesis, Speaker Verification and Audio Record/Play*. Sensory, Inc. [cited 1998-04-02]. Available from Internet: <URL:<http://www.sensoryinc.com/datasheet.shtm>>.

Sensory 1997. *Interactive Speech™ Overview* [online]. Sensory Inc. [cited 1997-12-31]. Available from Internet: <URL:<http://www.sensoryinc.com/interactive.shtml>>

Stephens, R. 1995. *Remote signage plans in Europe* [online]. Teoksessa: *Proceedings of a Symposium on Consumer, User Agency, Researcher, and Commercial Experience with Talking Signs and Related Technologies*, San Francisco, Califor-



nia, June 27, 1995. [cited 1997-12-22] S. 4 - 5. Available from Internet:  
<URL:<http://www.ski.org/RERC/WCrandall/SYMPOS.HTM>>

Strothotte, T., Petrie, H., Johnson, V. & Reichert, L. 1995. MoBIC: user needs and preliminary design for a mobility aid for blind and elderly travellers. 2<sup>nd</sup> TIDE Congress, Paris, La Villette, 26 - 28 April 1995.

Texas Instruments, 1997. TMS320 DSP Development Support Reference Guide, Literature Number: SPRU011E, January 1997.

Thyfault, M. E. 1998. Voice Recognition Goes Mainstream - Talking to a computer will become more common, on the phone or at your PC. Information Week Online, News in Review, January 5, 1998 [online]. Available from Internet:  
<URL:<http://www.techweb.com/se/directlink.cgi?IWK19980105S0030>>

ViA 1998 [online]. ViA Wearable, ViA, Inc. 1998 [cited 1998-04-03]. Available from Internet: <URL: <http://www.flexipc.com/Webpages/wfeatures.htm>>

VODIS, Voice Operated Driver Information System. Project Extension [online]. [cited 1998-04-01]. Available from Internet:  
<URL:<http://guagua.echo.lu/langeng/en/le1/vodis/extn.html>>

Voice Control Systems [online], Voice Control Systems Annual Report, March 1997 [cited 1997-12-29]. Available from Internet:  
<URL:<http://www.voicecontrol.com/financetext.html>>

## YRITYSTEN WWW- SIVUJEN OSOITTEITA

<b>Yritys</b>	<b>URL-osoite</b>
<b>Advanced Recognition Technologies, Inc.</b>	<a href="http://www.artcomp.com/">http://www.artcomp.com/</a>
<b>Dragon Systems, Inc.</b>	<a href="http://www.dragonsys.com/">http://www.dragonsys.com/</a>
<b>Information Storage Devices, Inc.</b>	<a href="http://www.isd.com/">http://www.isd.com/</a>
<b>Keyware Technologies N.V.</b>	<a href="http://www.keyware.be/">http://www.keyware.be/</a>
<b>Lernout&amp;Hauspie N.V.</b>	<a href="http://www.lhs.com/">http://www.lhs.com/</a>
<b>NEC</b>	<a href="http://www.nec.com/">http://www.nec.com/</a>
<b>Oki Semiconductor</b>	<a href="http://www.okisemi.com/">http://www.okisemi.com/</a>
<b>Quadravox, Inc.</b>	<a href="http://www.quadravox.com/">http://www.quadravox.com/</a>
<b>Sensory, Inc.</b>	<a href="http://www.sensoryinc.com/">http://www.sensoryinc.com/</a>
<b>Sonic Foundry, Inc.</b>	<a href="http://www.sfoundry.com/">http://www.sfoundry.com/</a>
<b>TerraTec</b>	<a href="http://www.terratec-us.com/">http://www.terratec-us.com/</a>
<b>Texas Instruments, Inc.</b>	<a href="http://www.ti.com/">http://www.ti.com/</a>
<b>Timehouse Oy</b>	<a href="http://www.kolumbus.fi/~topteam/">http://www.kolumbus.fi/~topteam/</a>
<b>Verbex Voice Systems, Inc.</b>	<a href="http://www.verbex.com/">http://www.verbex.com/</a>
<b>Winbond Electronics Corp.</b>	<a href="http://www.winbond.com.tw">http://www.winbond.com.tw</a>
<b>Voice Control Systems</b>	<a href="http://www.voicecontrol.com/">http://www.voicecontrol.com/</a>