

Tutkimusmenetelmät puheen tunnistuksen mallien synnyttäjinä

PEKKA NIEMI ja OLLI AALTONEN

1. Puheen havaitsemisen kolme mallia

Puheen havaitsemisesta tuli keskeinen tutkimuskohde fonetiikassa, kun »pattern-playback»-puhesynteesi kehitettiin Haskinsin laboratorioissa 1950-luvun alkupuolella. Keinotekoisen puheen avulla oli mahdollista tutkia puheen akustiikan ja äänteiden tunnistamisen välistä suhdetta tarkemmin kuin luonnolliseen puheeseen perustuvilla menetelmillä. Pian voitiinkin osoittaa, että englannin vokaalien tunnistamiseen riittää kaksi alinta formanttia (Delattre, Liberman, Cooper ja Gerstman 1952) ja että klusiilien tärkein akustinen vihje on toiseksi alimman formantin siirtymä (Delattre, Liberman ja Cooper 1955).

Tunnistusmetodiin perustuva tutkimus sai uuden ulottuvuuden, kun koehenkilöiden havaittiin erottelevan tarkemmin eri foneemiluokkiin kuuluvat ärsykkeet kuin samaan foneemiluokkaan kuuluvat (Liberman, Harris, Hoffman ja Griffith 1957). Tätä ilmiötä kutsutaan kategoriseksi havaitsemiseksi. Sen on väitetty koskevan vain puheen, erityisesti klusiilien havaitsemista. Osoittautui kuitenkin pian, että synteettisten vokaalien erottelu on enemmän jatkuvaa kuin kategorista (Fry, Abramson, Eimas ja Liberman 1962). Siksi kategorinen havaitseminen alettiin liittää puheen tuottamiseen. Toisin sanoen: artikulatorisesti kategoriset äänteet havaitaan kategorisesti, artikulatorisesti jatkuvat äänteet taas jatkuvina. Tämä tulkinta kategorisen havaitsemisen alkuperästä on kantavana periaatteena puheen havaitsemisen motorisen teorian eri versioissa (Liberman, Cooper, Shankweiler ja Studdert-Kennedy 1967). Lisäksi kategoriseen havaitsemiseen liitetään erilaisia foneemin rajailmiöitä (Repp 1982), jotka puolestaan ovat johtaneet tämän käsitteen korostumiseen puheen havaitsemisen malleissa. Foneeminrajalla tarkoitetaan sellaista kohtaa moniulotteisessa akustisessa foneemiavaruudessa, jossa koehenkilöt tunnistavat ärsykkeen yhtä todennäköisesti eri foneemiluokkiin. Foneeminraja käsitetään mentaaliseksi, kuulijan fonologista järjestelmää heijastavaksi edustumaksi. Tietoisien kielellisen havainnon oletetaan perustuvan juuri näihin rajoihin pikemmin kuin kielellisiä kategorioita edustaviin prototyyppeihin. Näitä teo-

PEKKA NIEMI ja OLLI AALTONEN

reettisiä tarkastelutapoja voidaan nimittää *foneettis-fenomenologisiksi*, koska ne korostavat ärsykkeen foneettisia piirteitä ja tunnistusmenetelmää.

Edellä kuvatut puheen havaitsemisen erityispiirteitä painottavat mallit ovat riidoin niiden mallien kanssa, jotka kiinnittävät päähuomion havaitsemisen sensorisiin tapahtumiin (Fujisaki 1979, Schouten 1980). Tämä auditiiviseen todellisuuteen perustuva tarkastelutapa tulkitsee sekä kategorisen kuulemisen että foneeminrajat psykoakustisiksi ilmiöiksi olettamalla kummankin taustalle yhteisen psykoakustisen tekijän (Pastore, Ahroon, Baffuto, Friedman, Puleo ja Fink 1977). Kategorisen havaitsemisen selitetään johtuvan käytettävästä ABX-menetelmästä, ja itse asiassa myös ei-kielelliset ärsykkeet voidaan sopivassa koejärjestelyssä havaita kategorisesti (MacMillan 1983, Massaro ja Cohen 1983). Tässä *auditiivis-sensorisessa* tulkinnassa foneeminrajat ovat heijastumia puheärsykkeiden akustisen monimuotoisuuden tuottamista psykoakustisista rajoista, jotka liittyvät tavalla tai toisella kuulon herkkyyden muutoksiin.

Erityisesti 1970-luvulla vahvistui ajattelutapa, jonka mukaan puheen havaitsemisessa on eri vaiheita. Vaikutteet tulivat uudesta kognitiivisesta psykologiasta ja perusoletuksena on, että puheen havaitseminen on ajassa etenevä tapahtumasarja, jonka ääripäinä ovat varhainen auditiivinen analyysi ja myöhempi foneettinen päätöksenteko (Sawusch ja Nusbaum 1983, Wood 1974). Auditiivinen analyysi on automaattista ja tiedostamatonta akustisen informaation käsittelyä, joka on luonteeltaan jatkuvaa. Kielelliset ärsykkeet käsitellään samalla tavoin kuin ei-kielelliset ärsykkeet. Foneettinen päätöksenteko puolestaan on tietoista ja kategorista. Näin ollen *monivaihemallit* sisältävät piirteitä sekä tunnistusmenetelmään pohjautuvasta, foneettisesta ajattelutavasta että psykoakustisiin mittauksiin perustuvasta, auditiivisesta tutkimusotteesta.

Seuraavassa perustelemme näkemystä, että kaikki kolme mallia seuraavat tietyistä menetelmän valinnoista.

2. Tutkimusmenetelmät mallien synnyttäjinä

Puheen havaitsemisen perusmallit voidaan äärimmilleen pelkistää kuvata kahden tekijän avulla. Nämä ovat puheärsykkeen *kooditus* ja sen *tunnistaminen*. Mallien yksinkertaisuus tuntuu hämmästyttävältä, kun otetaan huomioon pitkälle kehitetyt tavat kuvata ja analysoida itse puheärsyke. Teesimme onkin, että mallit ovat syntyneet ja pysyvät hengissä lähinnä koska kutakin niistä tukee juuri sille sopiva tutkimusmetodiikka. Mikä vielä pahempaa, käytetyt menetelmät pystyvät tuottamaan vain vahvistavaa näyttöä; niiden avulla ei kyetä osoittamaan itse mallia vääräksi. Näin muodostuu tutkijain esiteoreettisten oletusten seurauksena puheen havaitsemisen koulukuntia, jotka vain väkivoimalla pystyvät keskustelemaan keskenään. Ehdotamme, ettei mikään näistä

Tutkimusmenetelmät puheen tunnistuksen mallien synnyttäjinä

ajattelutavoista ole sellaisenaan väärä vaan ainoastaan yksipuolinen. Kukin niistä nimittäin pystyy vaihtelevalla menestyksellä selittämään puheen havaitsemista. Tarkkaavainen lukija tosin huomaa, että lopussa ilmaiseimme oman mieltymyksemme integroiviin vaihemalleihin.

3. Tunnistusmenetelmä ja yhden tekijän mallit

Suosittu tutkimusmenetelmä on ollut puheärsykkeen tunnistaminen, mikä voi luonnollisesti tapahtua oikein tai väärin. Valitsemalla ärsykejä sekä foneemialueen keskiöstä että läheltä foneeminrajaa saadaan hyvännäköisiä käyriä, jotka enimmäkseen viittaavat siihen, että tunnistus on tietyn kriittisen kohdan jälkeen kutakuinkin täydellinen. Tämantapaiset havainnot johtavat aiheellisesti »kategorisen kuulemisen» teorioihin.

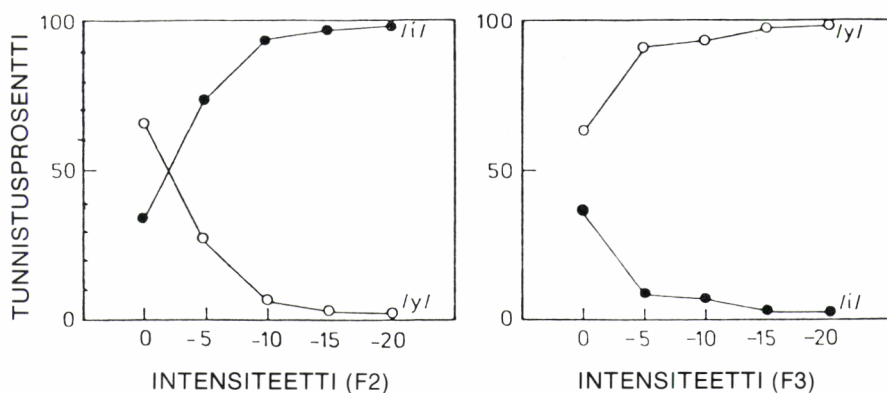
Tunnistuksen eräänlainen lopullisuus jää kuitenkin helposti huomaamatta tässä tutkimuksessa. Kaikki siihen johtaneet prosessit ovat jo tapahtuneet niin, ettei tutkijalla ole niistä mitään tietoa. Tätä puutetta voidaan korvata antamalla yksityiskohtainen akustinen kuvaus ärsykkeestä. Kun tähän lisätään kuvaus ärsykkeen edustuksesta sisäkorvan kalvossa, päästään jo tarkkaan ennusteseen tunnistuksesta (ks. esim. Suomi 1984). On luonnollista, että tällaiset tulokset tulkitaan perifeeristen aistiprosessien valossa, koska ärsykkeiden kuvaus on ankkuroitu juuri niihin. Tälläkään kerralla ei siltä saada tietoa niistä prosesseista, jotka lopulta johtavat tunnistukseen. Audiitiivis-sensorinen ja foneettis-fenomenologinen malli voidaan siis näyttää toisiaan täydentäviksi, eikä niiden välillä ole varsinaista ristiriitaa. Kumpikaan ei pysty tuottamaan havainnoja, joiden valossa toinen malli osoittautuisi kokonaan vääräksi; mallit elävät omaa elämäänsä kohtaamatta juuri toisiaan. Ne eivät myöskään pysty valaisemaan puheen tunnistuksen prosessiluonnetta. Edelleen vastaamatta jää siis kysymys, kuinka ärsykkeestä kerääntyy tietoa ja mitkä tämän informaation piirteet ovat tärkeitä tunnistukselle.

4. Reaktiolatenssit: prosessin epäsuora kuvaus

Lasta ei saa heittää pesuveden mukana. Prosessikuvaukseen tulee siis kuulua myös kaikki se tieto, jonka tunnistusmenetelmä antaa. Niinpä seuraavassa esiteltävät kolme vaihtoehtoista menetelmää todella sisältävät tunnistuksen sen tavanomaisessa muodossa. Lisäksi on kiinnostavaa tietää, kuinka nopeasti tunnistus tapahtuu. Kumpikin tieto on helposti saatavissa esittämällä koehenkilölle foneemeja ja mittaamalla hänen reaktioaikansa siihen, kun hän painaa sitä nappia, jolla on »sama nimi» kuin esitetyllä foneemilla. täten myös virhetunnistuksille saadaan reaktioaika. Tällä ns. mentaaliseen kronometriaan pe-

PEKKA NIEMI ja OLLI AALTONEN

rustuvalla tutkimuslogiikalla on satavuotinen perinne kokeellisessa psykologiassa. Menettelyn ydin on siinä, että muuttelemalla sopivasti koetilannetta voidaan tehdä epäsuoria päätelmiä itse prosessin kulusta. On nimittäin loogisesti pakko ajatella, että esimerkiksi ärsykkeen voimakkuus vaikuttaa ainakin sen kooditukseen eli siihen, kuinka tieto ärsykkeestä kerääntyy aivokuoreen. Yhtä pakottava oletus on, että vastausvaihtoehtojen lisääminen vaikuttaa tunnistuksen vaikeustasoon. Saavutettu etu on, että samalla kokeella pystytään vaikuttamaan sekä kooditusprosessiin että tunnistukseen. Tulosten tarkastelussa kumpaakin näistä voidaan analysoida esimerkiksi tavanomaisessa varianssi-analysissä. Täten kyetään myös kysymään, kumman mallin painovoima on suurempi, auditiivis-sensorisen vai foneettis-fenomenologisen. Havainnollistamme asiaa hiukan kuvioilla 1 ja 2 (Aaltonen ja Niemi 1984).

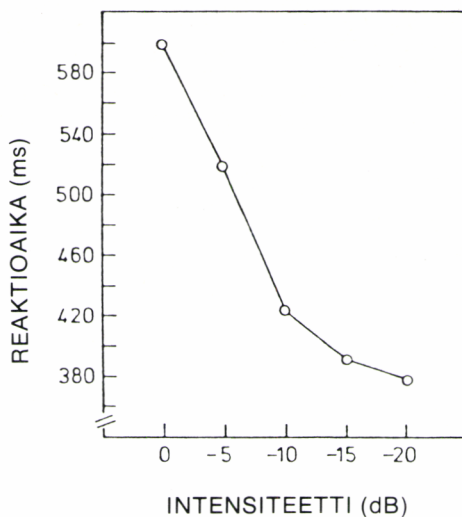


KUVIO 1. Tunnistuksen muuttuminen /i/—/y/-jatkumossa, kun yhden formantin amplitudia vähennetään. (Abskissan yksikkö dBA.)

Reaktiolatenssien etu on se, että ne ovat perinnäistä tunnistusta herkempi mitta. Nopeushan viittaa helppoon tunnistukseen, hitaus taas kertoo jonkinlaisesta vaikeudesta. Kummassakin tapauksessa voi oikeiden tunnistusten prosenttinen määrä silti olla sama. Kuviot 1 ja 2 havainnollistavat asiaa. Jos niitä tarkastellaan erikseen, syntyy todella kaksi erilaista päätelmää. Kuvion 1 tunnistustulokset viittaavat kategoriseen kuulemiseen: epävarmuusalue on kapea ja jo 5—10 dBA:n muutokset F2:ssa tai F3:ssa tuovat lähes virheettömän tunnistuksen, kun lähtökohtana on täydellinen epävarmuus eli foneeminraja. Kuvion 2 reaktioajat sen sijaan tukevat jatkuvan erottelun ideaa, koska vielä 10 dBA:n muutoksen jälkeenkin ärsykkeen laatu paranee niin paljon, että reaktiolatenssi lyhenee merkitsevät 40 ms.

Kun tällaiset havainnot yhdistetään, on vaikea löytää perusteita puhtaan kategorisen kuulemisen mukaiselle tulokinnalle. Näyttää siis siltä, että myös

Tutkimusmenetelmät puheen tunnistuksen mallien synnyttäjinä



KUVIO 2. Reaktiolatenssin muuttuminen /i/—y/-jatkumossa, kun F2:n tai F3:n amplitudia vähennetään. (Tiedot yhdistetty.)

koodituksen aikaisilla tapahtumilla on oma tärkeä osuutensa puheärsyksen tietoisessa havaitsemisessa.

Reaktioaikamenetelmä vaikuttaa lupaavalta. Se sisältää kaikki tunnistusmenetelmän hyvät puolet ja antaa lisäksi tietoa tunnistuksen sujumisesta reaaliajassa. Siihen liittyy kuitenkin yksi puute. Ihmiset eivät nimittäin toimi aina kuten tutkija haluaa, vaikka heitä siihen lempeästi kehoitettaisiinkin.

5. Erilaiset koehenkilöt: »speed-accuracy trade-off»

Tavallinen havainto fonetiikan laboratoriossa on, että ihmisillä on taipumus tehdä annettu tehtävä eri tavoin. Joku toimii rauhallisesti eikä juuri tee virheitä. Toinen taas hosuu; hän on kyllä nopea, mutta hänen suorituksensa laatu on heikko. Vaikka kokeen tekijöiden omat havainnot usein kertovatkin tällaisesta vaihtelusta, sitä on jokseenkin mahdotonta analysoida perinnäisellä tunnistusmenetelmällä. Pelkkä reaktioaika ei myöskään kerro koko totuutta. Henkilöhän voi saada nopean reaktioajan jonkin foneemin kohdalla siitä syystä, että hän tuskin malttaa kuunnella ärsykettä loppuun ennen kuin jo vastaa.

Taulukossa 1 on havainnollistettu erilaisia mahdollisuuksia, jotka väistämättä tulevat vastaan käytännön koetilanteessa. Siinä hypoteettiset koehenkilömme saavat samat keskimääräiset reaktioajat, mutta heidän virhemääränsä vaihtelevat. Huomion arvoista on nyt, että heidän suorituksensa tulkitaan eri tavoin. Koehenkilön A tunnistusprosentti pysyy samana kaikissa tilanteissa

(95 %). Reaktioaikojen hidastuminen tilanteesta 1 tilanteeseen 3 voidaan siis tulkita täysin foneettisesti ilman psykologisia viittauksia esimerkiksi koehenkilön tilaan. Koehenkilön B kohdalla tilanne on päinvastainen. Nopeus johtuu selvästi siitä, että hän hutilo. Kun virheet ja reaktioajat asetetaan vastakkain, käy ilmi, että kysymyksessä on puhdas nopeuden ja tarkkuuden välinen vaihtosuhte. Selitys onkin nyt psykologinen. Koehenkilö C puolestaan on tutkijan unelma. Tilanne 3 on hänelle vaikein kummankin mittarin valossa. Voidaan sanoa, että foneettinen tilanne säätelee erittäin voimallisesti hänen suoritustaan. Koehenkilö D on jossain määrin pulmallinen. On mahdotonta tietää, onko hän ollenkaan toiminut täydellä kapasiteetillaan, koska yhtään lipsahdusta ei ole sattunut. Hänen kohdallaan tulkinnan tulee olla erityisen varovainen.

TAULUKKO 1. Neljän erilaisen hypoteettisen koehenkilön suoritus. Keskimääräiset reaktioajat ovat samat, virheiden määrä sen sijaan vaihtelee. Kunkin henkilön suoritus tulkitaan eri tavoin.

	Koetilanne		
	1	2	3
A. RT (ms)	400	500	600
virheet %	5	5	5
B. RT (ms)	400	500	600
virheet %	9	5	1
C. RT (ms)	400	500	600
virheet %	1	5	9
D. RT (ms)	400	500	600
virheet %	0	0	0

Suorituksen nopeuden ja laadun yhteisessä analyysissä yhdistetään parhaalla mahdollisella tavalla tavanomainen tunnistusmenetelmä ja sen kehittyneempi sukulainen latenssimenetelmä. Tällaisilla kokeilla pystytään epäsuorasti pureutumaan puheärsyksen tunnistuksen prosessiluonteeseen sekä myös jossain määrin valaisemaan koehenkilön psyykkisen tilan vaikutusta itse suoritukseen. Epäsuorat päättelyt eivät kuitenkaan riitä, jos puheen havaitsemisen prosessiluonteesta ollaan vakavasti kiinnostuneita. Ne eivät myöskään riitä, jos kilpailevat pääteoriat halutaan alistaa testiin, joka kohtelee kumpaakin tasapuolisesti. Tämä mahdollisuus avautuu, kun prosessia pystytään seuraamaan reaaliajassa eli »on-line» ärsyksen esittämisestä aina tunnistusreaktioon asti.

Tutkimusmenetelmät puheen tunnistuksen mallien synnyttäjinä

6. Synteesi: puheen havaitsemisen reaaliaikainen kirjaus

Puheen havaitsemisen »on-line»-tutkimukseen on kaksi periaatteellista ratkaisua. Voidaan tutkia havaitsijan autonomisen hermoston reaktioita, jotka vaihtelevat ymmärtämisponnistelujen aiheuttaman kuormituksen tahdissa. Tällaisia mittoja ovat esimerkiksi sydämen syketaajuus, sähkögalvaaninen ihoreaktio ja — luultavasti parhaana näistä mittareista — pupillin koon vaihtelu. Toinen mahdollisuus on mitata puheen vastaanottamisen tuottamia sähköisiä muutoksia aivokuoressa. Tässä on kysymyksessä aivosähkökäyrän (EEG) versio eli ns. tapahtumapotentialiaali (event-related potential, ERP). Toisin sanoen kirjataan erillisen tapahtuman aivokuoressa aikaansaama reaktio. Tällainen tapahtuma voi olla esimerkiksi foneemin kuuleminen.

Valinta näiden vaihtoehtojen välillä ei ole loppujen lopuksi vaikea. Autonomisen hermoston toiminta on verraten epäspesifistä, ja mikä tärkeintä, sen latenssit ovat vähintään 0,5—1 s. Tarkkaan »on-line»-mittaukseen on siis hankala päästä. Sen sijaan ERP-latenssit ovat erittäin pienet. Tämä johtuu siitä, että mittaus voidaan kohdistaa suoraan sinne, missä prosessi tapahtuu, eli aivokuoreen. Niinpä foneeminerottelun varhaiset vaiheet on helppo kirjata otsalohkosta.

Esimerkkimme koskee edelleen suomen kielen /i/—/y/-jatkumoa (Aaltonen, Niemi, Nyrke ja Tuhkanen 1985). Ärsykeitä oli kolme: /i/, /y/ sekä näiden välinen äänne, rajaärsyke, joka tunnistettiin jommaksikummaksi 0,50:n todennäköisyydellä. Koehenkilön tehtävä oli laskea harvinaisten ärsykkeiden määrä 300:n sarjasta, jossa niitä oli 45. Loput 255 koekerroista koostuivat ns. standardiärsykkeen esittämisestä. Kokeessa haluttiin saada selville kaksi asiaa. Ensiksi pyrittiin selvittämään sitä, miten ärsykkeiden välinen ero näkyy varhaisessa ns. poikkeavuusnegatiivisuudessa (mismatch negativity, MMN, ks. Näätänen 1982). Tämä alustava erottelu on kutakuinkin automaattinen, ja sen oletetaan tapahtuvan samalla tavoin sekä puhe- että merkityksettömillä ärsykkeillä. Toiseksi tutkittiin sitä vähimmäisaikaa, jonka kuluessa tietoinen päätöksenteko on jo tapahtunut. Tämän ajan ilmaisee vanhaan tapaan mitattu P3-komponentti. Tutkimuksesta on helppo nähdä yhteys puheen havaitsemisen päämalleihin. MMN mittaa asiaa auditiivis-sensorisen mallin kannalta ja P3 puolestaan foneettis-fenomenologisen mallin kannalta.

Taulukossa 2 esitetään MMN:n ja P3:n latenssit, kun itse kukin kolmesta äännetyypistä oli harvinaisena ärsykkeenä, joiden lukumäärä laskettiin. Nyrkisääntö on, että lyhyt latenssi merkitsee foneeminerottelun helppoutta. Ilmenee kaksi kiinnostavaa seikkaa. Ensiksi: varhainen automaattinen erottelu on terävimmillään, kun verrattavien foneemien fysikaalinen ero on mahdollisimman suuri (kaksi alinta tilannetta taulukossa 2). Tämä havainto on kutakuin-

TAULUKKO 2. MMN:n ja P3:n latenssit eri ärsykepareilla. MMN on mitattu otsalohkosta (Fz) ja P3 päälaen lohkosta (Pz).

Ärsykkeet		Latenssit (ms)	
Standardi	Harvinainen	MMN	P3
rajaärs.	/i/	227	388
/i/	rajaärs.	220	415
rajaärs.	/y/	235	413
/y/	rajaärs.	233	410
/y/	/i/	212	369
/i/	/y/	217	416

kin sitä, mitä voidaan ennustaa auditiivis-sensorisen mallin pohjalta. Tilanne muuttuu kuitenkin toiseksi, kun katsotaan myöhempää P3-komponenttia, joka syntyy tietoisesta tunnistamisesta. Nyt erottuu helpoimmin /i/ siitä riippumatta, kuinka etäällä fyysikaalisesti parin toinen jäsen on (ensimmäinen ja toiseksi alin tilanne taulukossa 2). Paras selitys saattaisi tässä olla fenomenologinen viittaus /i/-foneemin erottuvuuteen, koska se on ääritapaus suppeiden vokaalien jatkumossa.

7. Kokonaisuus on enemmän kuin osien summa

Viesti on selvä, vaikka kokonaiskuva saattaakin olla yksityiskohdiltaan epätarkka. Menetelmän valinnat näyttävät todella ohjanneen puheen havaitsemista koskevaa teorianmuodostusta. Kukin menetelmä tutkii todellisuutta omalta kapealta näkökulmaltaan ja vain hypoteesin mukaiset tulokset julkaistaan. Päähypoteesit ovat kuitenkin niin etäällä toisistaan, ettei »kriittistä koetta» ole kyetty rakentamaan. Nähdäksemme ainoa realistinen menetelmä tämän tavoitteen saavuttamiseksi on puheen havaitsemisen reaaliaikainen kirjaus. Tällöin vältytään siltä, että prosessista irrotetaan vain jokin osa tutkittavaksi. Ongelmaksi jää vieläkin, onko »on-line»-menetelmä kyllin herkkä poimimaan esiin todellisuudessa tapahtuvia foneeminerottelun prosesseja. Sen on täytettävä ainakin kaksi kriteeriä. Ensiksi: koehenkilön tehtävän tulee sisältää tavanomainen tunnistus sellaisenaan, jotta suora vertailtavuus foneettis-fenomenologiseen tutkimusotteeseen säilyisi. Toiseksi: »on-line»-menetelmän tulee olla foneeminerottelun alkuvaiheissa vähintään yhtä herkkä kuin parhaat niistä koeasetelmista, joilla on tuotettu tietoa auditiivis-sensorisen mallin tueksi. Alustavat tulokset viittaavat siihen, että psykofysiologisen aivotutkimuksen

Tutkimusmenetelmät puheen tunnistuksen mallien synnyttäjinä

uudet menetelmät täyttävät molemmat ehdot. Näin päästään rakentamaan todellisia prosessin kuvauksia, joissa foneeminerottelun eri vaiheet tulevat tutkimuksen kohteiksi samojen ehtojen alaisina.

LÄHTEET

- AALTONEN, OLLI — NIEMI, PEKKA 1984: On the processing of simple linguistic stimuli. — Suomen soveltavan kielitieteen yhdistyksen (AFinLA) julkaisuja 37 s. 101—111.
- AALTONEN, OLLI — NIEMI, PEKKA — NYRKE, TIMO — TUHKANEN, MARJA 1985: Mismatch response of event-related brain potential in auditory vowel discrimination. — Abstracts of the papers presented in the Finnisch-Soviet Symposium »EEG and neural activity in psychological research», Helsinki, May 20—25.
- DELATTRE, PIERRE — LIBERMAN, ALVIN — COOPER, FRANKLIN — GERSTMAN, LOUIS 1952: An experimental study of the acoustic determinants of vowel color. — *Word* 8 s. 195—210.
- DELATTRE, PIERRE — LIBERMAN, ALVIN — COOPER, FRANKLIN 1955: Acoustic loci and transitional cues for consonants. — *Journal of the Acoustical Society of America* 27 s. 769—773.
- FRY, DENNIS — ABRAMSON, ARTHUR — EIMAS, PETER — LIBERMAN, ALVIN 1962: The identification and discrimination of synthetic vowels. — *Language and Speech* 5 s. 171—189.
- FUJISAKI, HIROYA 1979: Some remarks on recent issues in speech perception research. — *Proceedings of the Ninth International Congress of Phonetic Sciences* 1 s. 93—99. Copenhagen.
- LIBERMAN, ALVIN — HARRIS, KATHERINE — HOFFMAN, HOWARD — GRIFFITH, BELVER 1957: The discrimination of speech sounds within and across phoneme boundaries. — *Journal of Experimental Psychology* 53 s. 358—368.
- LIBERMAN, ALVIN — COOPER, FRANKLIN — SHANKWEILER, DONALD — STUDDERT-KENNEDY, MICHAEL 1967: Perception of the speech code. — *Psychological Review* 74 s. 431—461.
- MACMILLAN, NEIL 1983: A psychological interpretation of a »categorical perception» experiment by Hary and Massaro. — *Perception & Psychophysics* 34 s. 494—498.
- MASSARO, DOMINIC — COHEN, MICHAEL 1983: Categorical or continuous speech perception: a new test. — *Speech Communication* 2 s. 15—35.
- NÄÄTÄNEN, RISTO 1982: Processing negativity: an evoked-potential reflection of selective attention. — *Psychological Bulletin* 92 s. 605—640.
- PASTORE, R. — AHROON, W. — BAFFUTO, K. — FRIEDMAN, C. — PULEO, J — FINK, E. 1977: Common-factor model of categorical perception. — *Journal of Experimental Psychology: Human Perception and Performance* 3 s. 686—696.
- REPP, BRUNO 1982: Categorical perception: issues, methods, findings. — *Haskins Laboratories Status Report on Speech Research SR-70* s. 99—183.
- SAWUSCH, JAMES — NUSBAUM, HOWARD 1983: Auditory and phonetic processes in place perception for stops. — *PP* 34 s. 560—568.
- SCHOUTEN M. E. H. 1980: The case against a speech mode of perception. — *Acta Psychologica* 44 s. 71—98.
- SUOMI, KARI 1984: On talker and phoneme information conveyed by vowels. — *Speech Communication* 3 s. 199—209.
- WOOD, CHARLES 1974: Parallel processing of auditory and phonetic information in speech perception. — *Perception & Psychophysics* 15 s. 501—508.

Research methods as creators of models of speech perception

PEKKA NIEMI and OLLI AALTONEN

Speech perception became an important research topic along with the invention of »pattern-playback» speech synthesis in the 1950's. It then became possible to manipulate the acoustic components of speech with an accuracy previously unattainable. The experiments as such did not undergo any real change but rather continued to be based on the traditional identification procedure. Two sets of theoretical models emerged from subsequent research. Firstly, the identification procedure gave rise to phonetic-phenomenological models which emphasize psychological factors such as categorical perception. Secondly, the increased possibilities for controlling the physical stimulation led to an auditory-sensory interpretation of speech perception. This approach underlines the importance of psychoacoustic factors which may be derived directly from the functioning of the auditory system.

It is suggested in the present paper that research on speech perception has suffered from the above state of affairs. The sole dependent measure has been identification which, in turn, has been interpreted in either of the two mutually exclusive frameworks. In other words, theory formulation has, in a certain sense, remained second to experimental method.

While both types of model admittedly contain some truth, it has been practically impossible to test them empirically in such a way that each one would be given a fair chance within the same experiment. It is suggested that recent multi-stage models better serve this purpose. They rest on the reasonable assumption that speech is perceived in real time and, hence, research should tap the perceptual process at various points. The unsatisfactory nature of the identification methods is now obvious: it deals only with the end product of the process and ignores the rest. A forerunner for the multi-stage process approach were response latency measures which are also known as reaction time. Having a well-

established position in cognitive psychology, they could be easily adapted to speech perception research. Moreover, they can be essentially improved by explicitly taking the role played by the subject's speed-accuracy trade-off in responding into account. Unfortunately, response latency method does not quite solve the anomalous situation because it only permits indirect inferences about the perceptual process. For example, if stimulus intensity produces interesting effects, it may be concluded that auditory-sensory factors are heavily involved in the identification. It should be noted, however, that these factors were not directly observed. The inference was based on the logical assumption that stimulus intensity must affect the early sensory processes to a great extent.

As said above, response latency is not a genuine on-line measure monitoring of the perceptual process in real time. The authors argue that true progress can be made only if sensory-auditory events are registered simultaneously with the late psychological processes immediately preceding the conscious identification. To meet this goal it is necessary that the respective cortical events be measured. One potentially useful method is the so-called mismatch negativity of the event-related cortical potentials which has been developed by Risto Näätänen and his associates at the University of Helsinki. The basic task consists of a traditional phonetic discrimination expressed as an identification of a target phoneme. In addition to that, brain potentials are systematically registered from the termination of the stimulus until the response is given. Preliminary evidence suggests that auditory-sensory processes can be differentiated from phonetic-phenomenological processes within the same experiment in this way. It may thus be concluded that the two sets of models are not contradictory, as they appear to be if only the identification method is employed.