

# PUHEEN TUOTTAMIS- JA HAVAITSEMISPROSESSIEN MALLINTAMISESTA

PERTTI HURME

Jyväskylän yliopisto

## 1. Johdannoksi

Tarkastelen puheen "prosessoimista" - puheen tuottamis- ynnä havaitsemisprosesseja - fonetiikan näkökulmasta. Fonetiikan näkökulmalle on ominaista konkreettisuus: foneetikko pyrkii puhe lähtökohtanaan kohti puhujan ja kuulijan aivoja, kohti puhekäyttäjytymisen mallintamista. Laadittuja malleja ei suinkaan tule pitää "lopullisena totuutena": mallit ovat pelkästään ajattelun apuvälineitä, ongelmien havainnollistajia.

Kiellopin omaksumiseen kiinnitetään kielenopetuksessa vanhastaan suurta huomiota. Sen sijaan puheviestinnän lainmukaisuudet ovat yleensä jääneet vähemmälle huomiolle. Kieli on kuitenkin "vain" viestinnän apuväline; kaikki viestintä ei suinkaan ole kielellistä viestintää.

Tässä esityksessä tarkastelen joitakin puheen tuottamis- ja havaitsemisprosessien keskeisiä piirteitä. Keskityn prosessien periferisiin, lähellä ääniaaltoa sijaitseviin osiin. Onko merkityksen muuntamisessa ääneksi (puheen tuottaminen) ja äänen palauttamisessa takaisin merkitykseksi (puheen havaitseminen) kuitenkin mitään ongelmaa? Eikö puheen tuottaminen ole äänteiden liittämistä peräkkäin, ikään kuin helminauhaksi? Eikö puheen havaitseminen vastaavasti etene äänne äännteeltä? Tällainen naiivi käsitys puheen prosessoimisesta on erheellinen, kuten tässä esityksessä pyrin osoittamaan (vrt. Lieberman 1970). Käsitys perustuu staattiseen eli posturaaliseen näkemykseen fonetiikasta (Laver 1970; vrt. Lehtonen 1976a).

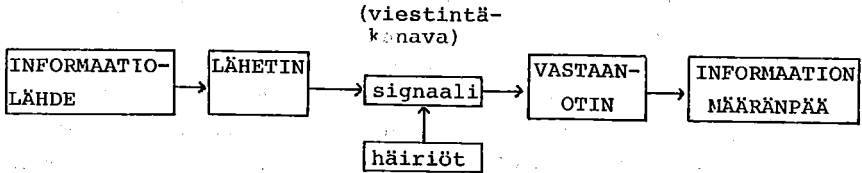
Tarkastelen ensin tiedonvälityksen yleistä mallia. Sitten yritän valaista puheen prosessoimisen todellista kompleksisuutta puheen tuottamis- ja havaitsemisprosessien malleja pohtimalla. Lopuksi selostan yhtä todellisen prosessorin (puheen havaitseminen) simulointiyrittystä, puhetta ymmärtävän koneen kehittämisyrityksiä.

## 2. Tiedonvälityksen yleinen malli

Alunperin Shannonin esittämä tiedonvälityksen yleinen malli (kuvio 1) kuvaa viestin kulkua informaatiolähteestä informaation määränpäästään: lähetin muuntaa lähteen valitseman viestin koodisanomaksi ja lähettää muodostamansa signaalin viestintäkanavaan; vastaanotin ottaa signaalin vastaan, dekoodaa signaalin eli purkaa signaalin koodin sekä välittää viestin mää-

rämpähän; viestintäkanavaan saattaa kohdistua häiriöitä, joiden vuoksi vastaanotettu signaali voi poiketa lähetetystä; saman koodin käyttäminen on viestinnän onnistumisen edellytys. (Vrt. Shannon ja Weaver 1949.)

KUVIO 1: Tiedonvälityksen yleinen malli.



Mallin soveltaminen puheviestinnän kuvaukseen ei ole ongelmattonta: malli ei esimerkiksi paljasta sitä, että puheen tuottamis- ja havaitsemisprosesseissa on paljon yhteistä (vrt. Lashley 1951) - ihminenhan on sekä puhuja että kuulija.

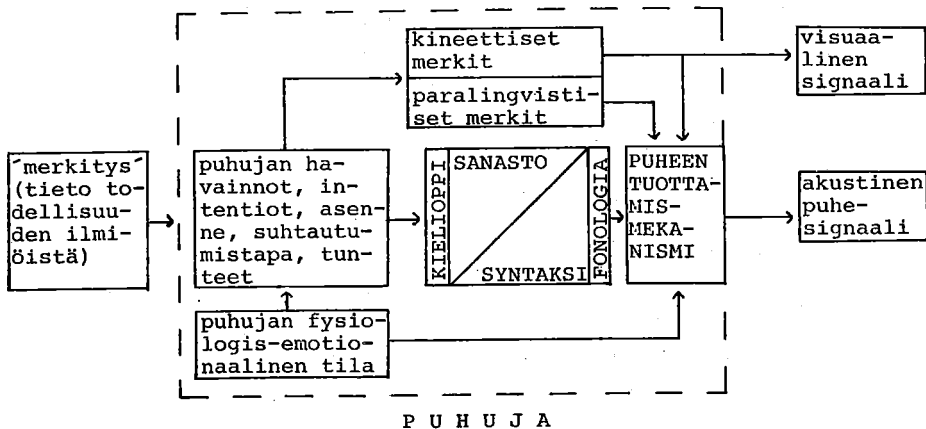
Puheen koodaus puhesignaalin muotoon samoin kuin puhesignaalin dekooodaus ovat varsin monimutkaisia prosesseja (vrt. Lehtonen 1976b: 18-20), joiden kuvaamiseen Shannonin malli on aivan liian yleinen. Seuraavaksi esitänkin muutaman suhteellisen yksityiskohtaisen puheen tuottamis- ja havaitsemisprosessin mallin.

### 3. Puheen tuottamisprosessi

Kuvion 2 malli kuvaa ilmaisuuden syntymistä huomattavasti yksityiskohtaisemmin kuin Shannonin malli, mutta ei kuitenkaan läheskään tyhjentävästi (vrt. Lehtonen 1976b: 19). Lehtosen esittelemä malli kuvaa merkityksen (eli ajatuksen, tiedon) muuntamista akustiseksi puhesignaaliksi sekä myös visuaaliseksi signaaliksi: ilmeiksi, eleiksi ja liikkeiksi.

Lehtosen esittelemän mallin mukaan puheen tuottamismekanismiin vaikuttavat siis monet muutkin kuin kielelliset tekijät: puhujan havainnot, intellitot, asenne, suhtautumistapa, tunteet sekä fysiologiset-emotionaalinen tila. Puhuja myös antaa kuulijalle enemmän kuin pelkän kielellisen muodon: tietoa suhtautumistavastaan, sosiaalisesta ja maantieteellisestä taustastaan, iästään jne.

KUVIO 2: Malli ilmaisun syntymisestä.

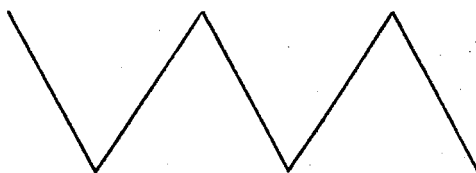


(kaavioon ei ole merkitty eri tasoilla vaikuttavaa palautetta)

Puheen tuottamisprosessin monimutkaisuus tulee selvästi ilmi myös pohdittaessa, miten keskushermosto ohjaa puheen tuottamismekanismien toimintaa. Ehdotetuista ohjausperiaatteista tarkastelen ensin kahta ääripositiota, ketjumallia ja kampamallia (kuvio 3).

KUVIO 3: Puheen tuottamisen ketjumalli ja kampamalli.

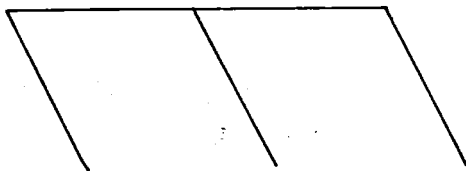
Ketjumalli:



keskushermosto

periferinen hermosto

Kampamalli:



keskushermosto

periferinen hermosto

Ketjumallin eli periferisen palautteen mallin mukaan puhe etenee yksikkö yksiköitä siten, että tuotetusta yksiköstä saatu palaute on seuraavan yksikön tuottamisen edellytys. Ohjausperiaate muistuttaa siis johdannossa esitettyä naiivia käsitystä puheen tuottamisesta. Ketjumallia vastaan on esitetty vakuuttavia argumentteja. Niinpä on voitu todeta, etteivät hermoimpulsit ehdi käytettävissä olevassa ajassa kulkea edestakaisin keskushermoston ja puheen tuottamiseen osallistuvien periferisten ääntöelinten välillä - kyseisten neuronien johtamisnopeus ei ole riittävä.

Kampamalli eli ennaltaohjelmointimalli puolestaan perustuu usean tavun laajuisen puhejakson ennalta suunnitteluun ja suunnitelman toteuttamiseen palautetta odottamatta. Mallia vastaan on esitetty argumentteja, jotka korostavat palautteen merkitystä puheen tuottamisprosessin ohjaamisessa. Niinpä esimerkiksi kynä hampaiden välissä pystytään puhumaan normaalisti, ilman että äänteiden laatu sanottavasti muuttuu, mikä on mahdollista vain mikäli puhekäskeyjä palautteesta saatavan tiedon perusteella muutetaan normaaleista.

Näiden kahden mallin välille sijoittuu sentraalisen palautteen malli (Kent ja Moll 1976). Palaute ei ole periferistä (ääntöelimestä saapuvaa) kuten ketjumallissa, vaan keskushermoston tasolta saapuvaa: mallin mukaan keskushermosto pystyy ennustamaan tiettyjen käskyjen ja käsky-yhdistelmien sensoriset seuraukset. Puheen ohjaus ja tarkkailu tapahtuisi normaalisti keskushermoston tasolla; hitaammin saapuvaa periferistä palautetta käytettäisiin vain ajoittain "puhesuunnitelman" tarkistamiseen ja korjaamiseen.

Puheen tuottamisessa ovat tärkeitä sekä ennalta suunnittelu että aikaisemmin tuotetusta puheesta saatava palaute. Sentraalisen palautteen malli yhdistää onnistuneesti molemmat tekijät.

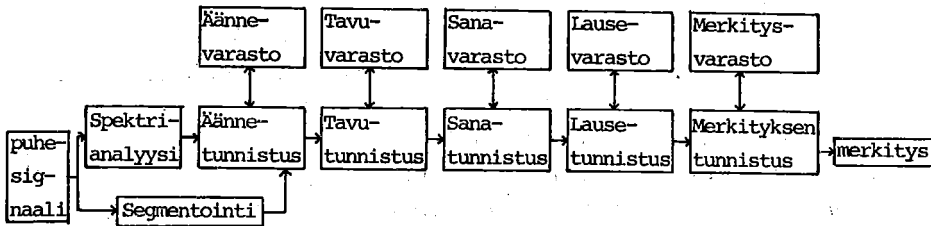
#### 4. Puheen havaitsemisprosessi

Pitääkö sitten naiivi käsitys puheen prosessoimisesta paikkansa havaitsemisprosessin puolella? Tapahtuuko puheen havaitseminen äänne äänneelintä? Tarkastelen aluksi kahta puheen havaitsemisprosessin mallia, seriaalista ja rinnakkaista.

Seriaalisen mallin mukaan (kuvio 4) puheen havaitsemisprosessi etenee seriaalisesti lähellä äänialtoa olevilta tasoilta aina merkitykseen asti. Mallin eri tasoilla siis vastaanotettua signaalia muokataan vähitellen yhä korkeammalle abstraktiotasolle. (Vrt. Karjalainen 1975.)

Seriaalista mallia kohtaan on kuitenkin viime aikoina esitetty kritiikkiä monelta taholta (vrt. Levelt ja Flores d'Arcais 1975). Esimerkiksi automaattisia puheen tunnistimia rakenneltaessa on tultu huomaamaan, ettei puheen segmentointi ole mahdollista pelkästään ääniallosta saatavan tiedon

KUVIO 4: Puheen havaitsemisen seriaalinen malli.



perusteella; lisäksi tarvitaan tietoa "korkeammilta tasoilta" ja tätä tietoa taas ei seriaalisen mallin mukaan ole alemmilla tasoilla käytettävissä.

Seriaaliselle mallille vastakkaisessa rinnakkaisessa mallissa saapuvaa ääniaaltoa ei analysoida peräkkäin taso kerrallaan, vaan erilaiset analyysit (fonologinen, syntaktinen ja semanttinen) työskentelevät samanaikaisesti (vrt. Pisoni ja Sawusch 1975). Rinnakkaisten mallien yleistymisen on osa kehitystä, jossa mallien painopiste on liukunut ääniaallon lähetyviltä korkeammille tasoille:

In the attempts toward a construction of a speech perception theory, we wonder whether more attention should not be given to the possibility of very early semantic and syntactic processing, on the basis of linguistic or even non-linguistic context, of the expectations of the hearer, etc. which could contribute to impose certain restrictions as to the kind of signals which may be accepted as possible candidates for perceptual synthesis, or produce certain selections or modifications on the incoming acoustic events. (Levelt ja Flores d'Arcais 1975: 354.)

Sitaatti korostaa puheen havaitsemisen aktiivista luonnetta. Kielellinen havainto ei ole passiivinen seuraus ulkopuolisesta ärsykkeestä, vaan ihmisen oman toiminnan tulos. Puheen havaitsemisen aktiivinen luonne on osoitettu muun muassa seuraavanlaisten kokeiden avulla: Ilmaisusta poistettiin ääninauhaa leikkelemällä kaikki akustiset vihjeet siihen sisältyvästä [s] -äännteestä. Syntyneeseen taukoon sijoitettiin yskäisy. Kuuntelukokeessa koehenkilöt eivät havainneet [s] -äännteen puuttuvan ilmaisusta, eivät edes siinä tapauksessa, että tiesivät miten oli menetelty. (Warren 1970: 392-393.)

Puheen havaitseminen ei siis tapahdu niinkään sen perusteella mitä toinen viestintäosapuoli tekee kuin sen perusteella mitä tapahtuu omassa aivoissa: "Luultavasti kuullija rakentaa omaa lausettaan rinnan puhujan kanssa käyttämällä tulossa olevan ennakoimiseen paitsi itse fyysikaalisen puhe-

signaalin välittämiä vihjeitä myös oman kielioppinsa ja kielellisen kokemuk-  
sensa sisältämää redundanssia sekä itse puhetilanteeseen sisältyvää ennustet-  
tavuutta." (Lehtonen 1976b: 21.)

Puheen havaitsemisprosessiin vaikuttaa siis monia tekijöitä akustisen  
puhesignaalin lisäksi. Seuraavaan luetteloon on pyritty poimimaan olennai-  
simmat: (1) äänialto eli akustinen puhesignaali, (2) kielen rakenteen (koo-  
din) sisäistynyt tuntemus, (3) puhetilanne, (4) nonverbaaliset vihjeet, (5)  
adaptoituminen viestintäosapuolen äänenkäyttö- ja ääntämisominaisuuksiin ja  
(6) suhtautuminen viestintäosapuoleen (sympatia/antipatia). Akustisen puhe-  
signaalin osuus puheen havaitsemisprosessissa on usein melko pieni muihin,  
niin sanottuihin korkeamman tason tekijöihin verrattuna.

Aidinkielistä puhetta kuunneltaessa merkitys ymmärretään yleensä välit-  
tömästi ja näennäisen vaivattomasti. Vieraskielistä puhetta kuunneltaessa  
sen sijaan korkeamman tason tiedon puutteellisuus pakottaa keskittymään akus-  
tisen puhesignaalin kuuntelemiseen. Oppimisen edistyessä korkeamman tason  
tiedon osuus prosessissa kasvaa akustisen tiedon osuuteen verrattuna. (Vrt.  
Zabrocki 1975.)

## 5. Mallien testaamisesta

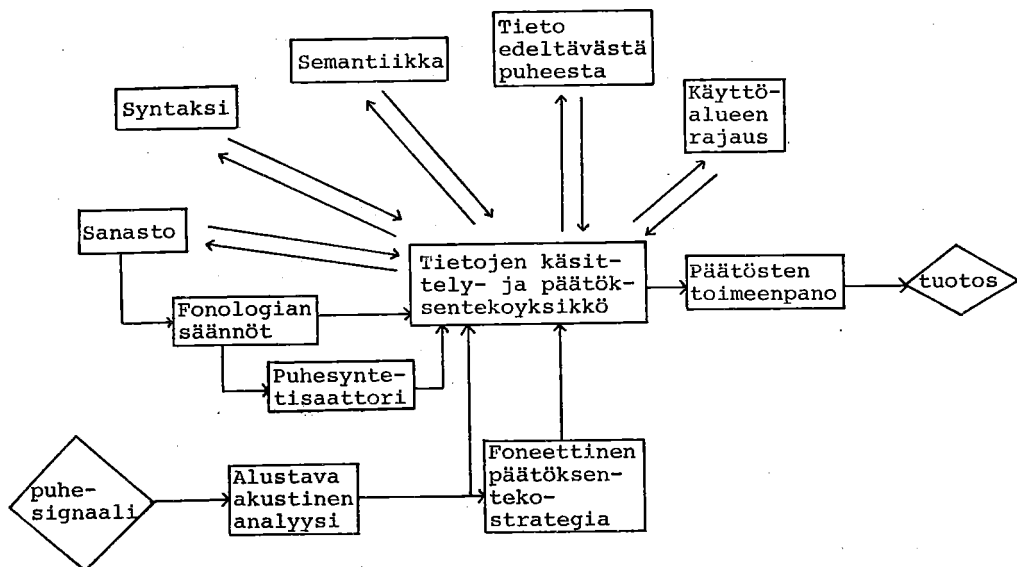
Edellä käsitellyillä puheen tuottamis- ja havaitsemisprosessien malleil-  
la on merkitystä pyrittäessä simuloimaan todellisia prosesseja koneiden avul-  
la. Puhesyntetisaattorien (puhuvien koneiden) ja automaattisten puheen tun-  
nistimien (puhetta ymmärtävien koneiden) suunnittelun ja rakentamisen yhtey-  
dessä on mahdollista testata laadittuja teoreettisia malleja käytännössä.  
Käsittelen seuraavaksi niitä periaatteita, joiden mukaan toistaiseksi menes-  
tyksekkäimmät puhetta ymmärtävät koneet on rakennettu.

Kuvio 5 kokoaa puhetta ymmärtävän koneen olennaisia piirteitä (vrt.  
Klatt 1975).

Klattin laatima malli - itse asiassa yhdistelmämalli useasta puheen ym-  
märtämiseen tähtäävästä koneesta - on aktiivinen ja rinnakkainen. Malli ot-  
taa myös huomioon suhteellisen monta puheen havaitsemisprosessiin vaikutta-  
vaa tekijää.

Klattin malli suorittaa ääniaallon alustavan analyysin sekä karkean  
foneettisen luokittelun. Kaikkien toimenpiteitten tulokset tallennetaan kes-  
keiseen tietojenkäsittely- ja päätöksentekoyksikköön, jonka ominaisuudet  
käytännössä määräytyvät tietokoneen ominaisuuksien mukaan. Korkeamman tason  
analyyttisyyttä edustavat syntetisaattorit (esimerkiksi tilanne-, merkitys-, lauseenmuo-  
dostus-, sanastonmuodostus- ja ääntämisyksiköt) saavat tietoja keskusyksi-  
köstä ja antavat sinne oman tulkintansa puhesignaalista.

KUVIO 5: Puhetta ymmärtävän koneen malli.



Klattin mallissa korostuvat juuri korkeamman tason tekijät. Näiden tekijöiden huomioon ottaminen on itse asiassa välttämätöntä: seriaaliseen malliin (siis puhe-signaalin "jalostamiseen") perustuvan puheen automaattisen tunnistimen rakentaminen ei ole ollut menestyksestä. Sen sijaan Klattin kuvaamat yritykset ovat olleet suhteellisen lupaavia, vaikkakin vielä ollaan varsin kaukana yleisen, kaikenlaisiin tarkoituksiin sopivan puhetta ymmärtävän koneen kehittämisestä.

## 6. Lopuksi

Yllä olen pyrkinyt osoittamaan, että puheen prosessoiminen näennäisesti yksinkertaisuudestaan huolimatta on hyvin monimutkainen tapahtuma: staattiset mallit on hylättävä sekä puheen tuottamis- että havaitsemisprosessessa mallinnettaessa. Puheen tuottaminen ei ilmeisesti tapahdu äänne äänneeltä, vaan prosessiin kuuluu pitkäjänteistä suunnittelua. Vastaavasti puheen havaitsemisprosessi etenee laajasti ennakoiden, ei äänne äänneeltä. Akustisen puhe-signaalin (sekä visuaalisen signaalin) syntyy vaikuttavat monet te-

kijät. Monet tekijät vaikuttavat myös signaalin purkamiseen puheen havaitsemisprosessissa. Puheen prosessoimista voisi yleisesti luonnehtia aktiiviseksi ja dynaamiseksi tapahtumaksi.

Kuitenkin on ilmeistä, että puhetta prosessoidaan useammalla kuin yhdellä tavalla. Puhetta kuunneltaessa huomio voi kiinnittyä joko puhesignaalin sisältämään merkitykseen tai puhesignaalin muotoon. Van Lancker (1975) erottaa puheen prosessoimisessa jatkumon, jonka päissä ovat toisaalta "automaattinen puhe" ja toisaalta "uutta luova puhe" (automatic speech, novel speech). Automaattinen puhe on mekaanista, memoroitua (vrt. Itkonen 1976), "yliopittua": esimerkiksi huudahdukset, kiroukset ja ulkoapitut runot ovat automaattista puhetta. Uutta luova puhe puolestaan on kognitiivista ja symbolista: eläimet eivät pysty tuottamaan uutta luovaa puhetta, mutta kylläkin automaattista puhetta ainakin jossain määrin (vrt. papukajajat!). Jatku- mo automaattisesta luovaan puheeseen tulisi ottaa huomioon puheen tuottamis- ja havaitsemisprosesseja mallinnettaessa.

Lopuksi pari sanaa puheen prosessoimisen mallien suhteesta vieraiden kielten oppimiseen. Staattisten käsitysten hylkääminen johtaa dynaamiseen näkemykseen puheen prosessoimisesta. Tässä näkemyksessä korostuvat muun muassa - puheen tuottamisprosessista puhuttaessa - ääntöelimien saapuvien puhekäskeyjen ajastukseen liittyvät seikat. Niinpä yksityisten äänteiden opettamiselle ja harjoittamiselle ei panna suurtakaan painoa.

Foneettinen tutkimus voi auttaa vieraan kielen opiskelijaa yrittämällä selvittää, mitkä puhesignaalin ominaisuudet ovat tärkeitä syntyperäisen puhujan kannalta. Kaikki puhesignaalin sisältämä tieto ei nimittäin ole yhtä tärkeitä: todennäköisesti kuulija vain silloin tällöin ikään kuin tarkistaa kuulemansa puhesignaalista. Kielenoppijan olisi tärkeä tietää, mikä puhesignaali on sellaista "avaininformaatiota", jota syntyperäinen puhuja käyttäisi hyväkseen.

Esimerkin avaininformaatiosta tarjoavat englannin "soinnittomat" ja "soinnilliset" klusiilit. Nämä äänteet eivät englannissa eroa toisistaan soinnin mukanaolon tai puuttumisen suhteen vaan soinnin alkamishetken (ajastuksen) suhteen. Itse asiassa englannin [b, d, g] -äänteet voivat olla kokonaan soinnittomia. (Vrt. Suomi 1976.) Suomalaiselle ei siis tulisi opettaa, että esimerkiksi englannin [b] on äännettävä aina soinnillisena ja [p] aina soinnittomana. Suomalainen kielenoppija tulisi opettaa kuulemaan (ja myöhemmin tuottamaan) uudenlainen avaininformaatio: soinnin alkamishetken merkitys englannin "soinnittomien" ja "soinnillisten" klusiilien erottamisessa toisistaan. Siitä avaininformaatiosta, jota syntyperäiset



puhujat käyttävät hyväkseen puhesignaalin koodia purkaessaan, tiedetään toistaiseksi melko vähän.

### Kirjallisuutta

- Cohen, A., ja S. Nooteboom (toim) 1975. Structure and Process in Speech Perception. Springer.
- Jeffress, L. (toim.) 1951. Cerebral Mechanisms in Behavior. Wiley.
- Itkonen, T. 1976. Syntaktisten vaikutusyhteyksien luonteesta, Virittäjä 1/1976, 52-81.
- Karjalainen, M. 1975. Kompleksisten informaatioprosessien systeemikuvauksista ja ajattelun mallintamisesta, Kybernetiikan seminaari: seminaariesitelmät. Tampereen Teknillinen Korkeakoulu.
- Kent, R., ja K. Moll 1976. Articulatory Timing in Selected Consonant Sequences, Brain and Language 2, 104-123.
- Klatt, D. 1975. Speech Understanding Systems and Speech Perception Theory, teoksessa Cohen ja Nooteboom (toim.) 1975, ~~344-344~~ 344-344.
- Lashley, K. 1951. The Problem of Serial Order in Behavior, teoksessa Jeffress (toim.) 1951, 112-136.
- Laver, J. 1970. The Production of Speech, teoksessa Lyons (toim.) 1970, 53-75.
- Lehtonen, J. 1976a. Puhetilanne ja fonetiikan normit, teoksessa Nummenmaa (toim.) 1976, 46-51.
- Lehtonen, J. 1976b. Kieli, puhe ja viestintä, teoksessa Sallinen-Kuparinen ja Lehtonen (toim.) 1976, 17-30.
- Levelt, W., ja G. Flores d'Arcais 1975. Some Psycholinguists' Reactions to the Symposium on Dynamic Aspects of Speech Perception, teoksessa Cohen ja Nooteboom (toim.) 1975, ~~345-351~~ 345-351.
- Lieberman, P. 1970. Towards a Unified Phonetic Theory, Linguistic Inquiry 1, 307-322.
- Lyons, J. (toim.) 1970. New Horizons in Linguistics. Penguin.
- Nummenmaa, L. (toim.) 1976. Kieli, konteksti ja tilanne - AFInLA:n syys-symposiumi 1975. Suomen sovelletun kielitieteen yhdistyksen julkaisu- ja 14. Helsinki: AFInLA.
- Pisoni, D., ja J. Sawusch 1975. Some Stages of Processing in Speech Perception, teoksessa Cohen ja Nooteboom (toim.) 1975, 16-35.

- Sallinen-Kuparinen, A., ja J. Lehtonen (toim.) 1976. Puheopin näkökulmia.  
Jyväskylän yliopiston Suomen kielen ja viestinnän laitoksen julkaisuja  
13. Jyväskylä.
- Shannon, C., ja W. Weaver 1949. The Mathematical Theory of Communication.  
University of Illinois Press.
- Suomi, K. 1976. English Voiceless and Voiced Stops as Produced by Native  
and Finnish Speakers. Jyväskylä:Contrastive Studies, 2. Jyväskylä  
Department of English.
- Van Lancker, D. 1975. Heterogeneity in Language and Speech: Neurolinguistic  
Studies. Working Papers in Phonetics (UCLA), 29.
- Warren, R. 1970. Perceptual Restoration of Missing Speech Sounds, Science  
167, 392-393.
- Zabrocki, L. 1975. Kybernetische Modelle der Sprachlichen Kommunikation.  
Polska Akademia Nauk: Komitet Neofilologiczny.