Vokaalien psykoakustisen laadun määrittämisestä

Algoritmisen menetelmän kuvaus ja tuloksia suomen monoftongeista[1]

KARI SUOMI

1. Kohti kuulohavainnon kannalta realistisempaa vokaalien kuvausta

Perinnäistä vokaalien analyysimenetelmää, äänispektrografiaa, ja siihen liittyvää vokaalien kuvaustapaa on viime aikoina alettu arvostella yhä enemmän. Kritiikkiin voidaan osoittaa ainakin seuraavat neljä toisiinsa kytkeytyvää syytä. Ensiksikin itse tutkimusväline, spektrografi, on — huolimatta eri valmistajien laitteisiinsa aikojen kuluessa tekemistä teknisistä parannuksista — ajastaan jäljessä sikäli, että se ei ota huomioon mm. psykoakustiikassa viime vuosikymmeninä saavutettuja tutkimustuloksia, jotka koskevat akustiselle signaalille ihmisen kuuloradassa tapahtuvia muutoksia. Siksi spektrogrammit ovat kuulohavainnon kannalta enemmän tai vähemmän epärealistisia (tässä suhteessa peruskonstruktioon ehdotetuista parannuksista ks. esim. Carlson ja Granström 1982). Klatt (1982) tiivistää oman näkemyksensä samasta asiasta seuraavasti: »As far as speech perception research is concerned, it is not inconceivable that the sound spectrograph has had an overall detrimental influence over the last 40 years by emphasizing aspects of speech spectra that are probably not direct perceptual cues (and in some cases may not even be resolved by the ear).»

Toiseksi: kun vokaalispektreistä mitataan vain pari kolme parametria, nim. alimpien formanttien taajuudet, heitetään samalla hukkaan suuri määrä informaatiota, jolla on potentiaalista merkitystä vokaalien havaitsemisessa. Tieto eri akustisten muuttujien vaikutuksesta kuulohavaintoon on vielä monelta osin vajavaa, ja tällöin on tietenkin järkevämpää olla hukkaamatta tietoa heti analyysin alkuvaiheessa. Bladon (1982) muistuttaa aiheellisesti siitä usein unohdetusta näkökohdasta, että manipuloitaessa spektrin huippuja (formantteja) myös spektrin kokonaismuoto muuttuu ja päinvastoin, joten

---

[1] Haluan kiittää Olli Aaltosta lukuisista vokaalien havaitsemista koskeneista keskustelutuokioista, Pekka Porria aineiston tilastollisesta käsittelystä ja luokitteluohjelmien tekemisestä, kaikkia koehenkilöitäni aikansa uhraamisesta ja Turun yliopistoa taloudellisesta tuesta tässä artikkelissa selostetuille tutkimuksille.

vokaalien havaitsemisesta saatu kokeellinen tieto tukee teorioita usein yhtä hyvin, pidettiinpä näissä havainnon kannalta keskeisinä formantteja tai spektrin muotoa. Kun siis ratkaisu teorioiden välillä täytyy tehdä muilla perustein, kallistuu vaaka mm. informaation vähäisemmän redusoitumisen ansiosta spektrin koko muodon huomioon ottavan kuvauksen puolelle. Formanttikuvauksen puolella vaakakupissa tosin painaa parametrien pieni määrä, parametrien suhde perinnäisiin artikulatorisiin vokaalin kuvauksen ulotteisiin ja edellisistä seuraava eräänlainen havainnollisuus, mutta näidenkin arvo määräytyy viime kädessä kuvauksen todenmukaisuudesta.

Kolmanneksi: ei ole suinkaan kiistatonta, että vokaalien sointivärin havainto perustuu juuri formantteihin, ts. formanttien perseptuaalinen relevanssi on kyseenalainen. Vaikka usein ei voidakaan tehdä ratkaisua koko spektrin muotoon ja formantteihin perustuvien esitysten välillä, on tilanteita, joissa jälkimmäiset ovat selvästi alakynnessä. Niinpä formanttianalyysissä ovat tunnetusti hankalia tapaukset, joissa vokaalilla on selvästi havaittava sointiväri mutta ei sitä vastaavia, teorian mukaisia formanttihuippuja. On myös tutkimusparadigmoja, jotka mahdollistavat ratkaisun tekemisen teorioiden välillä. Bladon (1982) ja Lindblom (Bladon ja Lindblom 1981) ovat osoittaneet, että subjektiivisesti koetut etäisyydet vokaalien välillä eivät ole ennustettavissa formanttien taajuuksien perusteella. Formanttikuvauksen perusteella esim. etäisyys [i]—[ä] näyttää noin kaksi kertaa suuremmalta kuin etäisyys [i]—[e], mutta subjektiivisesti se ei ole likikään niin suuri. Sen sijaan koko spektrin muotoon perustuvat etäisyysmitat korreloivat subjektiivisten arvioiden kanssa hyvin (vrt. tuonnempana kohdassa 4.5 laskettuihin psykoakustisiin etäisyyksiin).

Lopuksi voidaan vielä mainita se, että silloinkin kun analysoitavissa spektreissä on selviä energiahuippuja, formanttien mittaukseen liittyy monia ratkaisemattomia ongelmia. Aina ei edes ole täysin selvää, onko formantti käsitettävä ääniväylän siirtofunktion teoreettiseksi maksimiksi taajuusalueella vai akustisen spektrin konkreettiseksi energiahuipuksi (ks. esim. Fant 1970, Papçun 1980), ja tämän käsitteellisen sekaannuksen lisäksi akustisen spektrin huippukohtienkin määrittelyssä on monia vaikeuksia (joita ovat seikkaperäisesti käsitelleet mm. Iivonen 1979 ja Karjalainen 1982a). Ongelmat ratkaistaan usein vetoamalla aprioriseen tietoon vokaalin formanttien todennäköisistä arvoista joko aiempien mittausten tai teoreettisten laskelmien perusteella (ks. esim. Ladefoged 1967: 86, Pols ym. 1973). Tällöin akustista signaalia ei kuitenkaan tarkastella objektiivisesti — sitä itse asiassa modifioidaan sen mukaan, minkälainen sen odotusten mukaan pitäisi olla — vaan on kehäpäätelmän vaara, ja pahimmassa tapauksessa tutkijan ennakkokäsitykset saattavat aiheuttaa tuloksiin systemaattisia virheitä. Vokaalien sointivärin havaitsemisen selittämiseksi on välttämätöntä pyrkiä etenemään

täysin algoritmisesti akustisesta signaalista kohti perseptuaalisfoneettista avaruutta.

Vokaalien psykoakustisessa kuvauksessa pyritään esitystapaan, jonka parametrit vastaavat niitä ulotteita, joiden perusteella ihmisen on periaatteessa mahdollista tehdä havaintoja ja päätelmiä vokaalien sointiväristä eli laadusta. Edetessään ulkokorvasta kohti keskushermostoa ääniärsyke kokee matkalla joukon siirtojärjestelmän rakenteellisista ja toiminnallisista ominaisuuksista aiheutuvia muutoksia. Siksi kuulijan havaitsema sisäinen vaste eroaa ärsytyksen aiheuttaneesta ulkoisesta, akustisesta signaalista. Vokaalien psykoakustinen kuvaus eroaa siis akustisesta kuvauksesta siten, että siinä otetaan huomioon ne tunnetut rajoitukset ja muutokset, jotka kuuloradan ns. siirtofunktio aiheuttaa signaalin ominaisuuksien erotettavuudessa. Pidän selviönä, että kielitieteellisestikin orientoituneen vokaalien kuvauksen tulee mahdollisuuksien mukaan perustua tämän mukaiseen vokaalien luonnehdintaan. Ei liene hedelmällistä vedota vokaalien kuvauksessa ominaisuuksiin, jotka eivät ole psykoakustisesti realistisia — jotka esim. pahoin vääristyvät tai kokonaan häviävät matkalla pitkin kuuloratаa.

Vokaalien psykoakustinenkaan kuvaus ei välttämättä ilmaise kaikkea sitä, mikä lopulta vaikuttaa vokaalien havaitsemisessa. Kun sanotaan psykoakustisen kuvauksen pyrkivän ottamaan huomioon ne ulotteet, jotka ovat periaatteessa vastaanottajan käytettävissä, tarkoitetaan sitä, että yritetään ottaa huomioon kuulemista yleensä koskevat rajoitukset. Saattaa olla, että kielellinen kuuleminen eroaa muunlaisesta kuulemisesta, ja edelleen, että kuulijan puhuma kieli muovaa havaintoa. On ts. mahdollista, että vokaalien sointivärin havaitseminen poikkeaa ei-kielellisten ääniärsykkeiden havaitsemisesta ja että se on jopa kielikohtaista. Jos nämä toistaiseksi melko spekulatiiviset ajatuskulut osoittautuvat todeksi, silloin vokaalien täydellinenkään psykoakustinen kuvaus ei ole identtinen ehkä lopulta hahmottuvan havaintofoneettisen kuvauksen kanssa, jossa jokainen parametri saa juuri kyseisen kielen mukaisen painotuksen. Iivonen (1982: 73) mainitsee yhtenä psykoakustiselta kuvaustavalta vaadittavana ominaisuutena sen, että kielten välisen vertailun mahdollisuus säilyy, ja tässä suhteessa kielittäinen kuvaustapa menisikin liian pitkälle. Sikäli kuin kielellinen kuuleminen eroaa ei-kielellisestä, tulisi tavoitteeksi asettaa kielelliseen kuulemiseen perustuva mutta yksityisistä kielistä riippumaton psykoakustiikka. Näin saataisiin kielellisen kuulemisen erityisluonteen huomioon ottava objektiivinen perusta mm. kieltenväliselle vertailulle. Joka tapauksessa nykyisiinkin psykoakustiikan tuloksiin perustuvaa vokaalien kuvausta on pidettävä selvänä edistyksenä aiempiin puhtaasti akustisiin kuvausmenetelmiin nähden. Näin on katsottava, vaikka olemassa oleva tieto kuuloradan siirtofunktiosta perustuu suurelta osin ei-kielellisten ärsykkeiden havaitsemiseen. Psykoakustinen kuvaus ottaa

huomioon — toteutuksensa mukaan suuremman tai pienemmän — osan kuuloradan signaaliin aiheuttamista muutoksista, se on yksityisistä kielistä riippumaton, ja sen objektiivisuutta voidaan lisätä tekemällä analyysi mahdollisimman algoritmiseksi.

Kielentutkijaa ei välttämättä kiinnosta se, missä kohden kuuloratata ja mistä syystä siirtofunktion eri muunnokset tapahtuvat, vaan pikemmin niiden kokonaisvaikutus. Juuri tätä on pyritty mallintamaan seuraavassa selostettavassa analyysimenetelmässä. Vokaalien psykoakustisen laadun määrittämistä ja siirtofunktiota ovat Suomessa tarkemmin käsitelleet ainakin Iivonen (1982) ja Karjalainen (1982b).

2. Algoritmisen menetelmän kuvaus

Menetelmään on pyritty sisällyttämään kirjallisuudesta saamani käsityksen mukaan tärkeimmät niistä akustiseen signaaliin sovellettavista psykoakustisesti motivoituneista muunnoksista, jotka simuloivat ääniärsykkeelle kuuloradassa tapahtuvia transformaatioita. Samat muunnokset on yleensä otettu huomioon muissakin vokaalin sointivärin havaitsemista mallintavissa menetelmissä, koska niiden on todettu johdonmukaisesti parantavan menetelmien suorituskykyä (jolloin vertailukriteerinä on tietenkin käytetty ihmisen käyttäytymistä koeoloissa). Sen sijaan käsillä olevassa mallissa ei ole otettu mukaan esim. suhteellisen helposti toteutettavaa fooni—sooni-muunnosta eikä taajuustason peittoilmiötä, koska näiden lisävaiheiden on todettu vain vähän parantavan psykoakustisia malleja tai jopa heikentävän niitä (Bladon ja Lindblom 1981, Blomberg ym. 1982). Joka tapauksessa äännöksistä rajattujen vokaalin osien aallonmuodot, analyysin välivaiheet ja lopulliset spektrit ovat tallessa digitaalisessa muodossa, joten analyysi voidaan tehdä täsmälleen samasta aineistosta paremmaksi osoittautuvalla tavalla. Psykoakustisilta implikaatioiltaan toisiaan vastaavat mallit voivat erota matemaattiselta toteutukseltaan; tässä tutkimuksessa on päädytty kriittisiä kaistoja vastaavien kiinteiden suodattimien käyttöön lähinnä sen vuoksi, että tuloksena on intuitiivisesti selväpiirteinen vokaalin psykoakustisen spektrin esitys. Parametrien määrä on siinä eksplisiittisesti ilmaistu, ja niitä on helppo käsitellä tilastollisesti (esim. ortogonaalisten, toisistaan riippumattomien muuttujien määrän selville saamiseksi).

Psykoakustisten spektrien likiarvoon päästään seuraavien muunnosten kautta:

1. Vokaaliäännöksen analysoitavan kohdan digitaalisesti tallennettu aallonmuoto (amplitudi—aika-kuvaus) muutetaan spektrimuotoon (amplitudi—frekvenssi-kuvaukseksi) Fast Fourier Transform (FFT) -algoritmia käyt-

TAULUKKO 1. Käytettyjen digitaalisten suodattimien (kaistojen) keskitaajuudet (f_m), kaistanleveydet (f_b) ja rajataajuudet (f_c).

Kaista n:o	f_m Hz	f_b Hz	f_c Hz
			200
1	250	100	300
2	350	100	400
3	450	110	510
4	570	120	630
5	700	140	770
6	840	150	920
7	1000	160	1080
8	1170	190	1270
9	1370	210	1480
10	1600	240	1720
11	1850	280	2000
12	2150	320	2320
13	2500	380	2700
14	2900	450	3150
15	3400	550	3700
16	4000	700	4400

tävän suodatinpakan avulla.[2] Aikaikkunana on 12,8 millisekunnin Hammingin ikkuna. Suodattimet kattavat taajuusalueen 200 Hz:stä 4,4 kHz:iin, ja niiden keskitaajuudet, kaistanleveydet ja rajataajuudet (taulukko 1) vastaavat psykoakustisessa kirjallisuudessa esitettyjä ns. kriittisten kaistojen arvoja siten, että käytetyt 16 kaistaa ovat identtiset Zwickerin ja Feldtkellerin (1967: 74) kaistojen 3—18 kanssa. Saman kriittisen kaistan sisällä esiintyvä akustinen energia summautuu ja vaikuttaa esim. vokaalien sointivärin havaitsemisessa yhtenä kokonaisuutena; kriittisen kaistan käsitteellä on myös selvät anatomiset ja fysiologiset vastineensa sisäkorvan rakenteessa (Zwicker ja Feldtkeller 1967, Schaft 1970). Tämän muunnoksen avulla otetaan huomioon kuulon fysiologinen taajuusasteikko: koska kukin kaistoista on yhden kriittisen kaistan eli Barkin levyinen, suodatinpakka muuntaa lineaarisen taajuusasteikon Barkin asteikoksi. Samalla tulee otetuksi huomioon kriittisen kaistan mukainen taajuusresoluutio. Muunnoksessa saadaan jokaista vokaaliäännöstä kohti 16 tunnuslukua, jotka ilmaisevat kunkin suodattimen kohdalla esiintyneen akustisen energian ns. RMS-amplitudin.

[2] Suodatinpakka toteutettiin ohjelmalla FIR Windowed Filter Design Program — Window (Rabiner, McGonegal & Paul) ja suodatus ohjelmalla Fastfilt — An FFT Based Filtering Program (Allen), kumpikin julkaistu teoksessa Programs for digital signal processing, IEEE Press, New York 1979. Äänitykset tehtiin kaiuttomassa studiossa käyttäen Brüel & Kjaerin Impulse Precision Sound Level Meter Type 2209 -mikrofonia, Otari MTR-10 -nauhuria ja Agfa PEM 468 -ääninauhaa (nopeudella 38 cm/s). Digitointi, suodatus ja spektrin muunnokset tehtiin LSI 11/23 -tietokoneella ja tilastollinen käsittely Turun yliopiston laskentakeskuksen DEC-20 -tietokoneella.

2. Äänen subjektiivinen voimakkuus eli kuuluvuus riippuu ärsykkeen äänenpainetasosta, ja tämän mukaisesti edellisessä vaiheessa saadut RMS-arvot muunnetaan kuuluvuuden havaintoa paremmin vastaaviksi desibeliarvoiksi. Vokaalien kokonaisäänenpainetasojen erojen eliminoimiseksi lasketaan kaikkien 16 kaistan db-arvojen keskiarvo, määritellään se nollaksi ja ilmaistaan kunkin kaistan db-arvo poikkeamana 0 db:n tasosta. Menettely säilyttää kaistojen väliset voimakkuussuhteet, ja vastaava normaalistus kokonaisvoimakkuuden suhteen toteutetaan tavalla tai toisella kaikissa vokaalien analysointimenetelmissä.

3. Äänenpainetason ja kuuluvuuden vastaavuus on erilainen eri taajuuksilla, ja tämä otetaan huomioon tekemällä db-arvoihin ns. vakioäänekkyyskäyrästön mukaiset korjaukset. Yksinkertaisuuden vuoksi korjaukset tehdään kaikissa tapauksissa 50 foonin isofonikäyrän mukaisesti kunkin kaistan keskitaajuuden mukaiseen määrään. Tämä katsottiin riittävän tarkaksi likiarvoksi normaalien puheäänekkyystasojen ja suodatinpakan kattaman taajuusalueen kannalta. Analyysin lopputuloksena on vokaalin 16-parametrinen spektriesitys, jossa kukin luku ilmaisee yhden kriittisen kaistan äänekkyystason.

Valitut suodatinpakan ala- ja ylärajataajuudet määräytyvät puheena olleista julkaistuista kriittisten kaistojen arvoista ja teknisistä rajoituksista. Jos analyysi olisi ulotettu vielä alempiin kriittisiin kaistoihin, ei ehkä enää olisikaan analysoitu vokaalien vaan käytetyn äänentallennus- ja -toistojärjestelmän ominaisuuksia. Toisaalta digitoinnissa käytetty näytteenottotaajuus asettaa rajan taajuusalueen yläpäähän.

3. Aineiston keruu

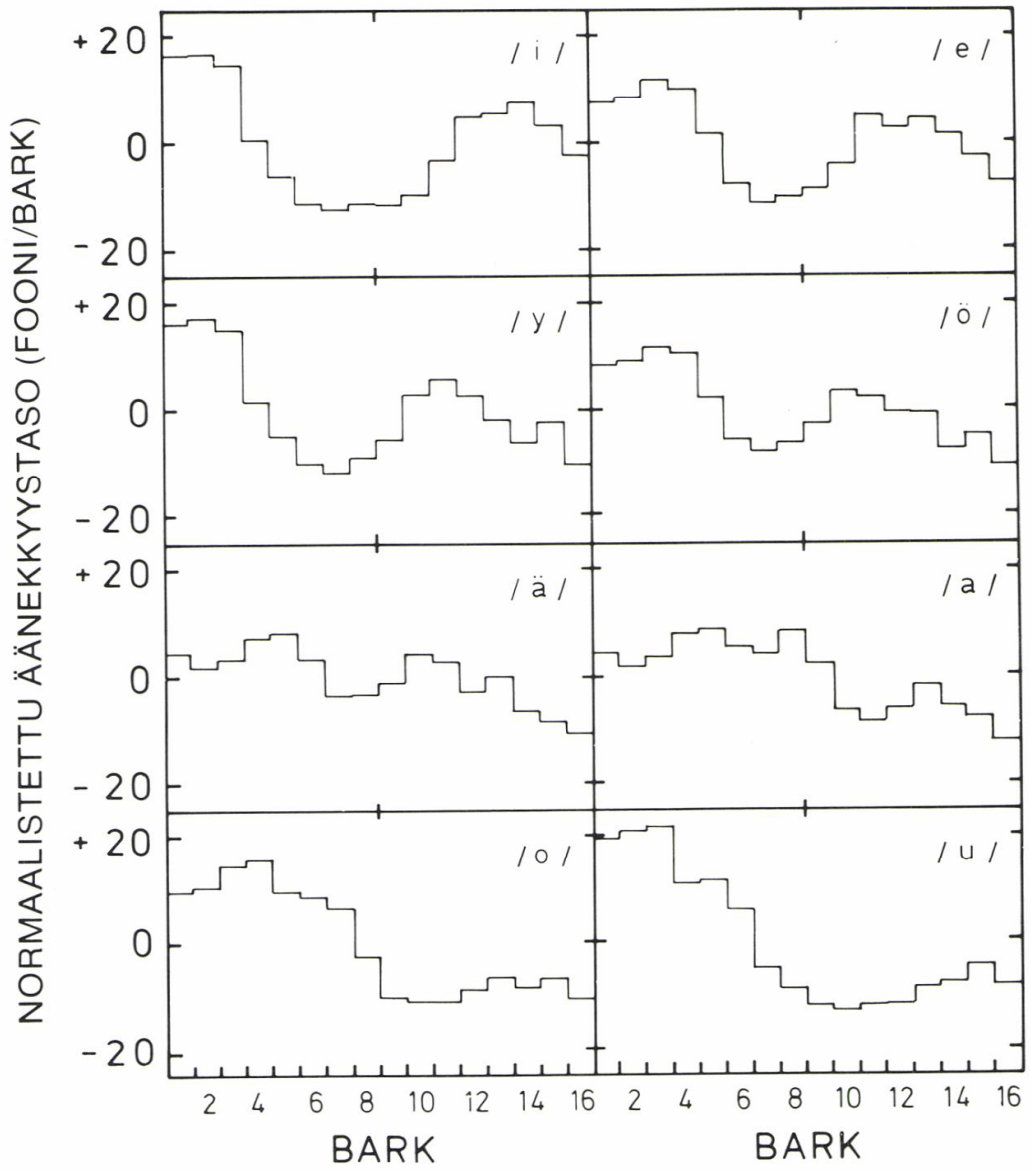
Tutkimuksen aineistona ovat kehyslauseessa »Lue sana h_tti uudelleen» esiintyvät suomen lyhyet monoftongit. Puhunnokset äänitettiin Turun yliopiston fonetiikan laboratoriossa studioluokan välineitä käyttäen. Koehenkilöinä oli 8 miestä ja 8 naista iältään n. 20—45 vuotta, ja kriteerinä koehenkilöksi valinnassa oli saatavillaolo. Mukana oli sekä ei-akateemisen että akateemisen koulutuksen saaneita; osa jälkimmäisistä oli saanut foneettista koulutusta. Koehenkilöt lukivat lauseet äänen korteista; heille annettiin ohjeeksi toistaa koko lause, jos he huomaisivat tehneensä virheen, ja muuten puhua siten kuin heistä tuntui luontevimmalta. Muita ohjeita ei annettu, ja puhujat erosivat toisistaan käsittääkseni suuresti sekä äänen voimakkuuden että puhenopeuden ja ääntämisen huolellisuuden suhteen; näitä eroja ei ole pyritty kvantifioimaan. Korteissa kukin vokaali esiintyi kaikkiaan kymmenen kertaa; kortit oli saatettu satunnaiseen järjestykseen paitsi kunkin vokaalin ensimmäistä ja viimeistä toistoa; nämä oli järjestetty korttipakan al-

kuun ja loppuun, eikä niitä analysoitu. Tutkittuja vokaaleja oli siis kaikkiaan 2 (puhujaryhmää) × 8 (puhujaa) × 8 (vokaalifoneemia) × 8 (toistoa) = 1 024 kpl. Koska kaikki vokaalit esiintyivät identtisessä ympäristössä ja koska /h/:n laatu pikemmin määräytyy samassa tavussa seuraavan vokaalin mukaan kuin päinvastoin, on aihetta olettaa, että tutkitut vokaalijaksot edustavat suomen lyhyiden monoftongien melko puhdasta, kontekstin vaikutuksista riippumatonta laatua. Vokaalia seuraavan /t/:n vaikutus ei puolestaan voi olla kovinkaan suuri vokaalin alkupuolelle sijoittuvassa otoskohdassa, jonka määrittelyä selvitän aivan kohta.

Sekunnin pituinen jakso kehyslauseen keskeltä otettiin tarkasteltavaksi käsivaraisesti kuuntelun perusteella, alipäästösuodatettiin 4,5 kHz:n taajuudella, tallennettiin digitaalisesti 10 kHz:n näytteenottotaajuudella käyttäen 12 bitin A/D-muunninta ja siirrettiin näyttöpäätteen kuvaputkelle visuaalista tarkastelua varten. Aallonmuodosta pyrittiin löytämään kohta, jossa /h/:lle tyypillinen henkäyssointi oli jo muuttunut vokaalin vuodottomaksi fonaatioksi. Pääasiallisena vihjeenä tässä toimituksessa käytettiin ääniaallon värähtelyn amplitudia, ja otoskohta määritettiin kursorin avulla alkamaan siitä, missä edellä tapahtunut amplitudin jyrkkä nousu alkoi laantua. Tätä aikapistettä seuraavat 512 näytepistettä tallennettiin levyyn analyysiä varten. Koska näytteenottotaajuus oli 10 kHz, vastaa tallennettu jakso 51,2:ta millisekuntia. Varsinaisessa suodatuksessa analysoitiin tallennetun jakson alusta 12,8 millisekunnin pituinen jakso, mikä vastaa miehillä keskimäärin n. puoltatoista ja naisilla 2—3:a glottispulssia.

Näytteenottokohdan määrittäminen oli kieltämättä jossain määrin mielivaltaista ääniaalloissa esiintyneen vaihtelun vuoksi. Se onkin tämän tutkimuksen ainoa vaihe, joka analyysivaiheessa vaati tutkijan omaa harkintaa: kaikilta muilta osin koko analyysi tapahtui koneellisesti ja täysin algoritmisesti etukäteen ohjelmoitujen periaatteiden mukaan. Koko ja vain alun perin tallennettu materiaali on mukana tuloksissa, ja kustakin vokaaliäännöksestä otettiin vain yksi näyte tietämättä siinä vaiheessa mitään sen spektriominaisuuksista. Formanttimittauksissa ei liene tavatonta, että osa alkuperäisestä aineistosta joudutaan karsimaan pois, kun äännöksissä esiintyy teorian kannalta odotuksenvastaisia epäsäännöllisyyksiä; tämä materiaalinhukka tulee sen lisäksi, että pelkkien formanttiparametrien mittaaminen aiheuttaa joka tapauksessa suurta informaatiokatoa. Tähän viittaa esim. seuraava Iivosen (1979: 67) selostus: »All the individual formant values (F1 and F2) *which could be measured* were drawn on a formant chart – – » (korostus KS:n). Iivonen toimii tässä tietysti täysin korrektisti ja asiallisesti viitatessaan ongelman olemassaoloon; samoin hän tekee käsitellessään juuri mittaamisen ongelmia seikkaperäisesti useissa kirjoituksissaan, mutta usein asia sivuutetaan vaitiololla niin, että tutkimusselostuksia lukeva voi vain arvailla, kuin-

KUVA 1. Suomen lyhyiden monoftongien psykoakustiset keskiarvospektrit, miespuhujat. Kukin spektri perustuu 64 tuotokseen.



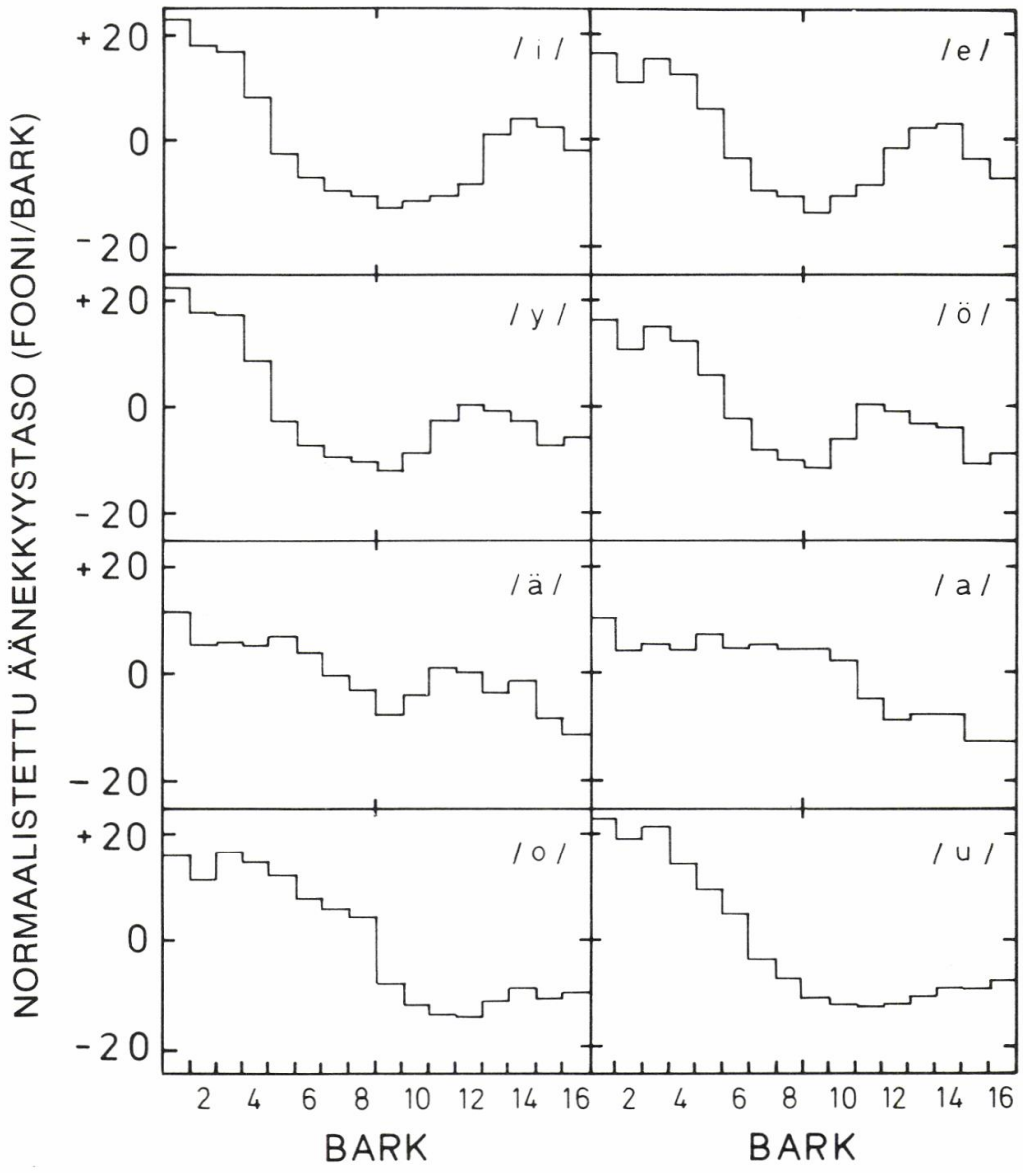
ka suppeaan ja tarkasti valikoituun aineistoon julkaistut tulokset lopulta perustuvat — vai perustuvatko ollenkaan.

4. Tuloksia

4.1. Yleistä

Miesten tuotoksista lasketut vokaalien keskiarvospektrit on esitetty kuvassa 1 ja naisten kuvassa 2. Miesten ja naisten keskiarvospektrit ovat silmämääräisesti pitkälti toisensa kaltaiset, ja on mahdollista tehdä kummankin ryhmän vokaaleja koskevia yleisluonteisia huomioita. Vaikka perinnäisessä vo-

KUVA 2. Suomen lyhyiden monoftongien psykoakustiset keskiarvospektrit, naispuhujat. Kukin spektri perustuu 64 tuotokseen.



kaalien akustisessa analyysissä keskeisellä sijalla olevat formantit eivät useasti näy spektreissä selvinä energiahuippuina, ei ole vaikea todeta selviä vastaavuuksia esitystapojen välillä. Samalla on kiintoisaa ja rohkaisevaa todeta, että monet formanttianalyysin pulmat saavat ratkaisun psykoakustisesti realistisemmassa analyysi- ja kuvausmenetelmässä. Seuraavassa esitetyt psykoakustisia spektrejä koskevat väitteet pitävät paikkansa — paitsi keskiarvospektreihin — suureen osaan yksityisten tuotosten spektreistä, joten kyseessä eivät ole esim. keskiarvojen laskennan yhteydessä syntyneet matemaattiset harhat.

Ensiksikin takavokaalien, eritoten /u/:n (tyyppisten vokaalien) kolmatta

formanttia on usein vaikeaa löytää, vaikka spektrografi — toisin kuin ihmisen kuulorata — korostaa signaalin ylempiä taajuuksia (n. 6 db/oktaavi). Kuvien 1 ja 2 spektrien perusteella on ilmeistä, että psykoakustisesti /u/:ssa ei olekaan mitään systemaattista energiahuippua perinnäisesti määritellyn F3:n kohdalla. Esim. miesten /u/:n kaistan 15 kohdalla esiintyvä paikallinen maksimi on taajuudeltaan liian korkea käydäkseen kolmannesta formantista, ja jos se olisikin tulkittava neljänneksi formantiksi, missä silloin sijaitsee F3? Naisten /u/:n osalta tilanne on formanttiteorian kannalta yhtä vaikea.

Toiseksi: ei-väljien etuvokaalien F2:n ja F3:n löytäminen ja erottaminen toisistaan — ja ylemmistä formanteista — on ongelmallista jo puhtaasti analyysiteknisesti; pulmallista se on myös siksi, että kertynyt tieto vokaalien havaitsemisesta osoittaa etuvokaalien ylempien formanttien vaikuttavan havainnossa usein yhtenä kokonaisuutena. Tämä on tausta Fantin kehittämälle ns. efektiivisen kakkosformantin (F2′) käsitteelle, jossa tavoitellaan tavanomaisessa formanttianalyysissä pysytellen perseptuaalisesti merkityksellistä ylempien formanttien painotettua keskiarvoa (ks. esim. Carlson ym. 1975). Nyt käsillä olevan tyyppisissä menetelmissä ei ole tarpeen painiskella kyseisten formanttien määrityksen eikä niiden keskinäisen painotuksen kanssa; vokaaleissa tulee usein esiin vain yksi leveä, selvää huippukohtaa vailla oleva energiakasauma ylempien kaistojen kohdalla, ja näin kuvauksen suhde perseptiosta saatuihin tuloksiin on huomattavasti suorempi.

Kolmanneksi: pyöreiden takavokaalien kahta alinta formanttia on usein vaikeaa paikantaa ja erottaa toisistaan vetoamatta aiempiin mittaustuloksiin tai teoreettisiin tietoihin. Taas voidaan todeta, että psykoakustisessa kuvauksessa noille vokaaleille onkin tyypillistä leveähkö, selviä maksimikohtia sisältämätön energiakasauma taajuusalueen alapäässä. Kaikissa edellä käsitellyissä tapauksissa spektrin muoto kuitenkin erottelee vokaaleja toisistaan, ja ongelmiksi käsitetyt seikat osoittautuvat vanhemman tutkimusmenetelmän tutkimusvälineestä johtuviksi näennäisprobleemeiksi.

Seuraavissa jaksoissa käsittelen saamiani alustavia tuloksia lähinnä osoittaakseni, mihin eri tarkoituksiin menetelmää voidaan käyttää. Palaan teemoihin tarkemmin toisaalla.

4.2. Vokaalien automaattisesta luokittelusta

Tutkittavien vokaalien aposteriorisella automaattisella luokittelulla on oma itseisarvonsa sikäli, että luokittelun onnistumisen perusteella voidaan arvioida analyysin systemaattisuutta ja vokaalien ominaispiirteiden kuvaajiksi saatujen keskiarvospektrien edustavuutta objektiivisella tavalla (nim. laskemalla esim. oikeiden luokitusten osuudet). Samalla luokittelusta saadaan alustavaa

tietoa mm. vokaalien ja yleisemmin puheen automaattista koneellista tunnistusta varten. Tällä puolestaan on merkitystä myös puheen havaitsemista koskevien mallien testaamisessa eksplisiitillä tavalla.

Vokaalispektrien luokittelu perustuu Plompin (1970) formuloimaan kompleksisten äänten spektraalisen etäisyyden mittaan, jossa kahden stationaarisen äänen s_i ja s_j etäisyys D_{ij} lasketaan kaavasta

$$D_{ij} = \sqrt[p]{\sum_{n=1}^{m} | L_{in} - L_{jn} |^p} \quad \text{, jossa}$$

L_{in} = äänen i kaistan n äänenpainotaso desibeleinä,
m = kaistojen lukumäärä ja
p = muuttuja, joka voi saada erilaisia arvoja.

Käsillä olevassa sovelluksessa L_{in} tarkoittaa spektrin i kaistan n foonilukemaa, yhden Barkin levyisten kaistojen määrä on 16 ja muuttujalla p on arvo 2, jolloin kyseessä on ns. euklidinen etäisyysmitta. Tällöin etäisyys D_{ij} voidaan käsittää kahden pisteen väliseksi etäisyydeksi 16-ulotteisessa avaruudessa, jonka koordinaatteina ovat kaistojen fooniarvot. Edellä on jo mainittu se, että subjektiiviset arviot vokaalien välisistä etäisyyksistä korreloivat paremmin koko spektrin muodon kuin formanttien taajuuksiin perustuvien laskennallisten etäisyyksien kanssa, ja koko spektriin perustuvista etäisyysmitoista juuri euklidisen etäisyysmitan (p = 2) on todettu tuottavan parhaat tulokset (ks. esim. Karjalainen 1982b: 106). Tätä mittaa on käytetty kaikissa tuonnempana selostettavissa luokitteluissa ja etäisyyslaskelmissa.

Spektrien luokitteluissa kutakin yksityistä tapausta verrataan luokittelukategorioina toimiviin referenssispektreihin ja tapaus luokitetaan siihen kategoriaan, johon sen laskettu etäisyys on pienin; haluttaessa voidaan esim. tulostaa kunkin tapauksen etäisyydet kaikkiin luokittelukategorioihin ja täten mm. kvantifioida luokittelun virhemarginaali ja arvioida mahdollisten normaalistusalgoritmien tehokkuutta tarkemmin kuin jos käytettävissä ovat vain esim. oikeiden luokitusten määrät.

Vokaalien automaattinen luokittelu tehtiin ensi vaiheessa erikseen miesten ja naisten aineistoissa siten, että kummassakin ryhmässä luokittelukategorioina käytettiin ryhmän tuotoksista laskettuja vokaalien keskiarvospektrejä. Luokittelujen tulokset näkyvät taulukoista 2 (miehet) ja 3 (naiset). Miesten aineistossa oikeita luokituksia saatiin kaikkiaan 499/512 eli 97,5 %; vokaalikohtainen minimimäärä oli 59/64 eli 92,2 % ja maksimimäärä 64/64 eli 100 %. Vastaavat prosenttiluvut naisilla ovat 96,3, 87,5 ja 100. Myös luokit-

TAULUKKO 2. Vokaalispektrien luokittelu vokaalien keskiarvospektrien perusteella. Miesten aineisto.

		luokiteltu							
		/i/	/e/	/y/	/ö/	/ä/	/a/	/o/	/u/
	/i/	64							
	/e/		62		2				
puhuttu	/y/			63	1				
	/ö/		4		60				
	/ä/					64			
	/a/					5	59		
	/o/					1		63	
	/u/								64
	yht.	64	66	63	63	70	59	63	64

TAULUKKO 3. Vokaalispektrien luokittelu vokaalien keskiarvospektrien perusteella. Naisten aineisto.

		luokiteltu							
		/i/	/e/	/y/	/ö/	/ä/	/a/	/o/	/u/
	/i/	64							
	/e/	4	60						
puhuttu	/y/			64					
	/ö/		4	1	59				
	/ä/		3	1	4	56			
	/a/						64		
	/o/							63	1
	/u/				1				63
	yht.	68	67	66	64	56	64	63	64

telun virheet ovat enimmäkseen odotuksenmukaisia sikäli, että vokaalit on luokiteltu virheellisesti jonkin foneettisen ominaisuuden suhteen naapurivokaaliksi (vrt. myös kohdassa 4.5 esitettyihin psykoakustisiin etäsyyksiin). Selvästi odotuksenvastaisia ovat vain yhden miesten /o/-tuotoksen luokittelu /ä/:ksi, yhden naisten /ä/-tuotoksen luokittelu /y/:ksi ja yhden naisten /u/-tuotoksen luokittelu /ö/:ksi — näissä saattaa olle kyse esim. vokaaliin sekoittuneesta hälystä.

Oikeiden luokitusten määrät ja virheluokitustenkin johdonmukaisuus ovat nähdäkseni osoitus siitä, että algoritmisesti lasketut keskiarvospektrit edustavat käytetyn psykoakustisen mallin rajoissa hyvin suomen monoftongien ominaispiirteitä. Virheluokitukset kasaantuivat tietyille puhujille siten, että miehistä yhden osalle tuli kolmasosa kaikista virheistä (ja nämä koskivat kaikki samaa vokaalia ja samaa virheluokitusta) ja puolelle puhujista tuli yli 90 % miesten virheistä; naisilla taas yhden puhujan osalle tuli yli puolet virheluokituksista. Virheet — paitsi noita odotuksenvastaisia tapauksia — selittynevät siis pääosin puhujien välisten erojen ja keskiarvon matemaattisen

luonteen yhteisvaikutuksesta. Esimerkiksi erotteluanalyysin luokittelufunktiot ottavat aritmeettista keskiarvoa paremmin huomioon luokiteltavien ryhmien yksittäiset, epäsäännölliset tapaukset, ja alustavat erotteluanalyysit parantavatkin kauttaaltaan tässä esitettyjä luokittelutuloksia. Parannukset eivät kuitenkaan ole kovin suuria ainakaan prosenttiyksikköinä ilmaistuina, mutta tämä ehkä johtuu oikeiden luokittelujen jo alun perin suurista osuuksista.

4.3. Puhujien automaattisesta luokittelusta

Eri puhujien saman vokaalifoneemin tuotokset näyttivät spektrien alustavan visuaalisen tarkastelun perusteella usein sisältävän puhujakohtaisia ominaisuuksia vokaalin identiteetistä kertovan informaation lisäksi. Yksityisten spektrien tarkastelussa on kuitenkin vaikeata erottaa toisistaan vokaali-informaatiota, puhujan ääniväylästä johtuvia konstantteja ominaisuuksia ja puhujan sisäistä satunnaista vaihtelua. (Äänneympäristö, yksi vokaalin toteutuksessa systemaattista vaihtelua aiheuttava tekijä, pysyi tässä tutkimuksessa koko ajan muodollisesti vakiona.) Laskemalla jokaisen puhujan kunkin vokaalifoneemin tuotosten keskiarvo ja vähentämällä siitä saman vokaalin koko aineistosta laskettu keskiarvo saadaan todennäköisesti erotukseksi se, mikä on tyypillistä kunkin puhujan kyseisten vokaalien tuotoksille erotuksena muista puhujista. Tällä tavoin lasketut henkilöiden vokaalikohtaiset profiilit poikkeavatkin toisistaan hyvin mutkikkaalta näyttävällä tavalla sekä puhujittain että samalla puhujalla vokaaleittain. Erityisesti on syytä korostaa sitä, että saman puhujan eri vokaaleista lasketut profiilit eroavat toisistaan huomattavasti. Esim. kuulijan todennäköisesti suorittamassa automaattisessa puhujan normaalistuksessa ei siis ole kyse yksinkertaisesta, eri vokaalien suhteen vakiona pysyvästä korjauksesta, vaan korjauksessa on otettava huomioon ainakin vokaalin summittainen laatu. Seuraavassa selostetaan alustavia luokitteluja, joiden tarkoituksena oli saada käsitys siitä, missä määrin yksityiset spektrit sisältävät puhujakohtaista tietoa.

Ensiksi kokeiltiin puhujien luokittelua kunkin puhujan koko aineistosta saatujen keskiarvospektrien perusteella, taas miehillä ja naisilla erikseen. Tämä tehtiin puolittain pilanpäiten, mutta saadut tulokset ovat mielestäni hyvinkin yllättäviä. Miesten osalta saatiin näet oikeiden puhujaluokitusten määräksi 168/512 (32,8 %) ja naisten 254/512 (49,6 %). Kuitenkin koko aineistosta lasketut kunkin puhujan keskiarvospektrit — joissa siis kaikki vokaalit ovat mukana vaikuttamassa — ovat varsin etäällä kaikista yksityisten vokaaliäännösten spektreistä, eivätkä ne siis edusta mitään konkreettisesti esiintyvää (täysin satunnaisessa luokittelussa todennäköinen oikeiden luokitusten määrä olisi kummassakin ryhmässä 64/512 eli 12,5 %). Puhujien

luokiteltavuudessa on tässä kieltämättä vaikeatulkintaisessa luokittelussa kuitenkin suuria eroja: parhaassa tapauksessa puhujan tuotokset luokiteltiin puhujan suhteen oikein 44 kertaa 64:stä (68,8 %) ja huonoimmassa kerran (1,6 %)!

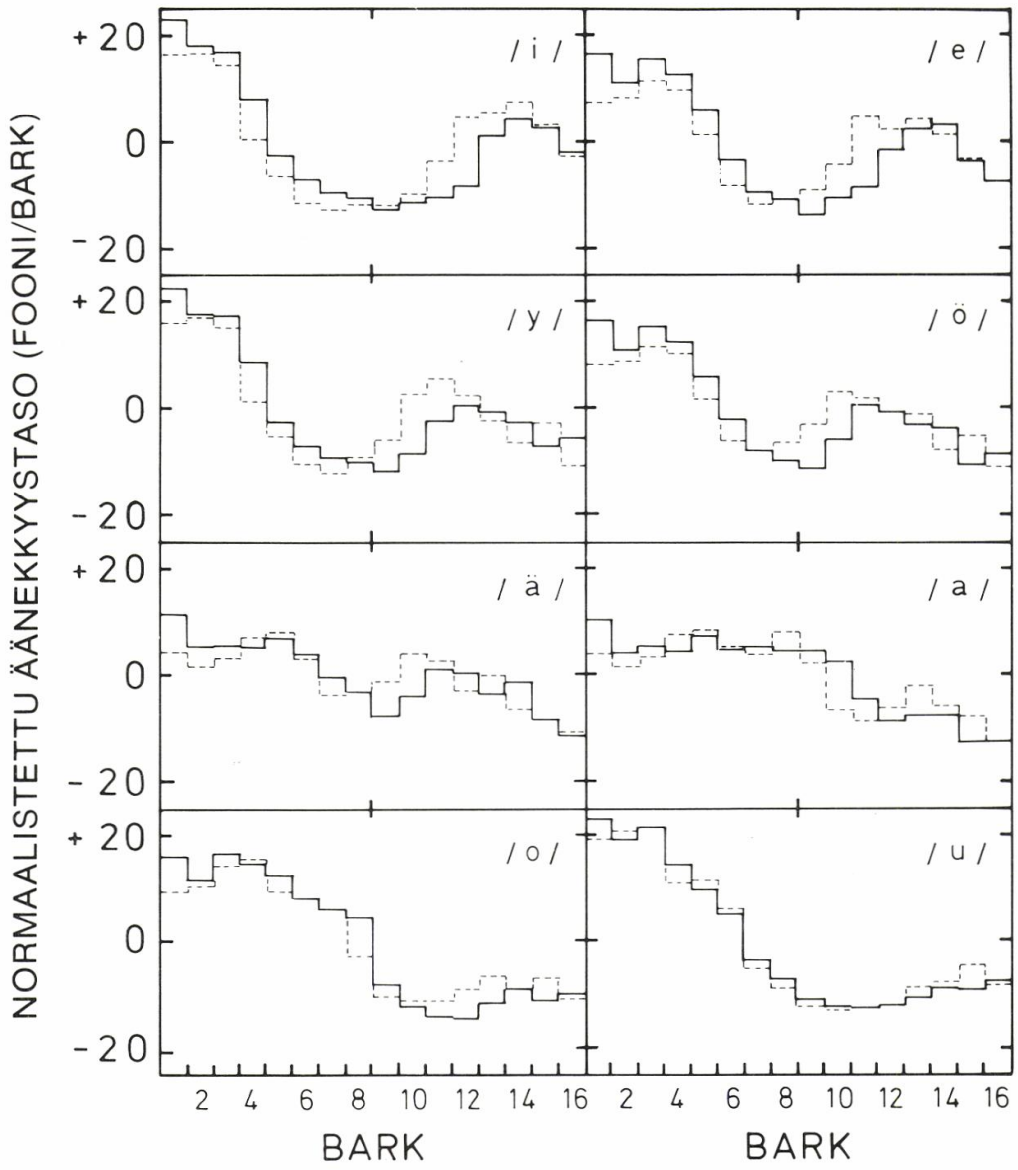
Jos vokaalimuuttujaa pidettiin vakiona, ts. luokitukset tehtiin kunkin vokaalin osalta erikseen — luokittelukategorioina kussakin vokaalissa puhujien kahdeksan toiston keskiarvot —, saatiin miehillä puhujan oikean luokittelun keskiarvoksi 462/512 (90,2 %) ja naisilla 470/512 (91,8 %). Heikoimmin luokiteltu puhuja luokiteltiin oikein kummassakin ryhmässä 55 kertaa 64:stä (85,9 %). Kiintoisaa on, että kummassakin ryhmässä puhujien oikea luokittelu oli maksimaalista (100 %) /ö/-vokaaleissa ja minimaalista /u/-vokaalissa (miehet 70,3 % ja naiset 67,2 %). On siis ilmeistä, että samalla kun /u/-tuotokset ovat luotettavimmin luokiteltavissa vokaali-informaation suhteen, ne sisältävät vähiten tietoa puhujan ominaisuuksista; /ö/, joka kaiken kaikkiaan oli heikoimmin oikein luokiteltu vokaali, sisältää taas luotettavinta tietoa puhujasta. Väliin jäävissä vokaaleissa ei selvää tendenssiä ole havaittavissa.

4.4. Miesten ja naisten vokaalien eroista

Kuvassa 3 kummankin ryhmän vokaalien keskiarvospektrit on esitetty päällekkäin vertailun helpottamiseksi. Suoraviivaisin sukupuolten välinen ero on ehkä se, että kaistan 1 fooniarvot ovat naisilla miesten vastaavia arvoa suuremmat. Tämä aiheutuu mitä ilmeisimmin siitä, että perussävel osuu naisilla juuri tämän kaistan sisälle (ks. taulukkoa 1). Tätä eroa lukuun ottamatta voidaan etuvokaalien osalta karkeasti arvioida, että naisten spektrien yhden Barkin lineaarinen siirto taajuusasteikossa alaspäin toisi ne melko lähelle miesten spektrejä (vrt. Klatt 1982: 186, 191). Takavokaaleissa erot ovat mutkikkaammat ja varsinkin /u/:ssa kaikkiaan melko vähäiset. Tässäkin voidaan havaita /u/:n sisältämän ei-kielellisen informaation suhteellinen niukkuus: /u/:n puhujan henkilön ja sukupuolen mukaisen normaalistuksen määrä on ilmeisesti pienempi kuin missään muussa vokaalissa.

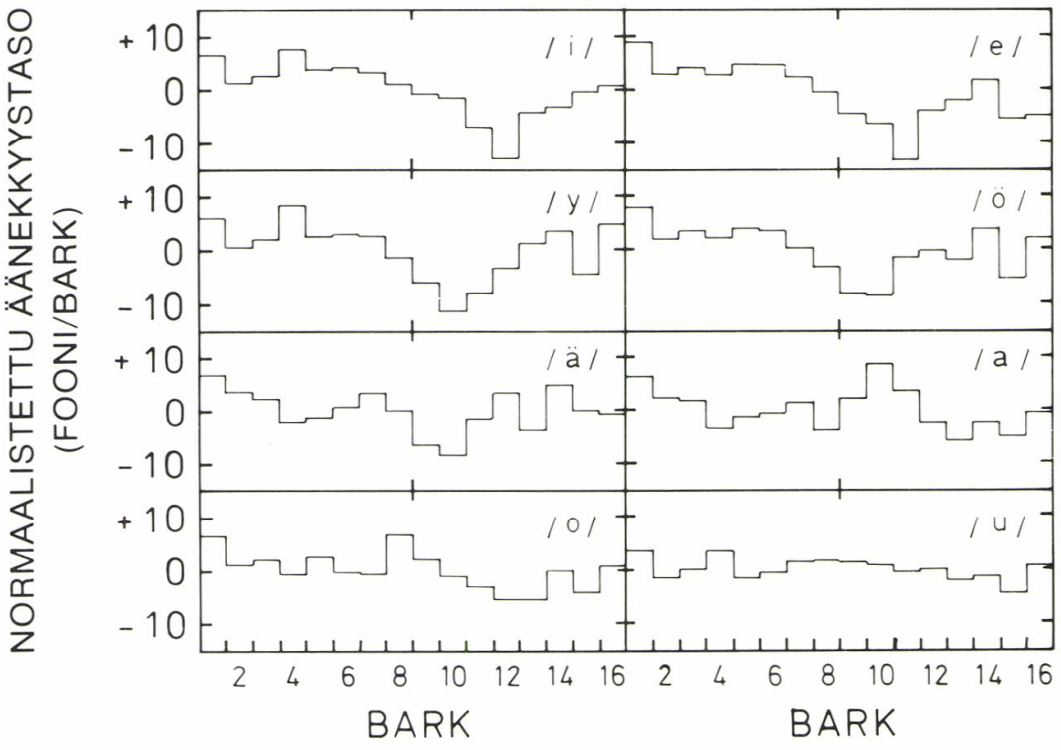
Ryhmien väliset erot ilmenevät hiukan toisella tavalla kuvassa 4, jossa näkyvät vokaaleittain miesten ja naisten keskiarvospektrien erot kullakin kaistalla. Positiivinen arvo tarkoittaa, että naisten spektrissä fooniluku on suurempi kuin vastaavan kaistan arvo miehillä. Erotusspektrien voidaan todeta sisältävän melko suuren määrän systematiikkaa. Ensiksikin käyrät pyrkivät olemaan positiivisia alempien kaistojen kohdalla ja negatiivisia ylempien kohdalla, ts. naisten spektreissä ylempien taajuuksien vaimennus on runsaampaa kuin miehillä. Toiseksi: ei-väljien etuvokaalien kesken erotusspektrit ovat melko samanlaiset kaistojen 1 ja 8 välillä. Kolmanneksi: saman

KUVA 3. Miesten (----) ja naisten (—) tuotoksista lasketut keskiarvospektrit.



väljyysasteen etuvokaaleissa erotusspektrit ovat samalla alueella käytännöllisesti katsoen identiset, ts. sukupuolen mukainen normaalistus on sama /i/:ssä ja /y/:ssä aina kaistaan 7 asti, ja vastaavasti /e/:ssä ja /ö/:ssä. Neljänneksi: myös /a/:n ja /ä/:n erotusspektrit ovat hyvin lähellä toisiaan kaistaan 7 saakka. Mainituissa saman väljyysasteen pareissa erotusspektrien erot kaistojen 1 ja 7 välillä ovat keskimäärin vain 0,75 foonia, ja suurimmillaankin (parien /e/—/ö/ ja /a/—/ä/ kaistalla 7) alle 2 foonia. On hyvin mahdollista, että näitä systemaattisia vastaavuuksia käytetään puheen vastaanotossa puhujan sukupuolesta johtuvien erojen kompensoimiseen: kuulija tekee niiden mukaiset korjaukset vastaanottamiensa vokaalien spektreihin (etuvokaaleissa lähinnä spektrin ylempien kaistojen kohdalla). Toistaiseksi olen vasta pyrkinyt selvittämään

KUVA 4. Miesten aineiston ja naisten aineiston keskiarvospektrien erotukset.



tämän normaalistuksen tarvetta eri vokaaleissa siten, että olen tarkastellut luokittelun onnistumisen riippuvuutta luokittelukategorioina toimivista keskiarvospektreistä. Aineiston kummankin puhujaryhmän vokaalispektrit on luokiteltu käyttäen sekä saman sukupuolen että koko aineiston tuotoksista laskettuja keskiarvospektrejä luokittelukategorioina. Oikeiden luokitusten prosenttiset osuudet kussakin kolmessa tapauksessa on esitetty vokaaleittain taulukossa 4, jossa kukin luku ilmaisee sukupuolten oikeiden tunnistusten keskiarvoa kyseisessä luokittelutilanteessa ja viimeinen rivi parhaimman ja huonoimman luokittelun eroa prosenttiyksikköinä. Luokittelujen virheitä en tässä käsittele: ne ovat ennustettavissa miesten ja naisten keskiarvospektrien erojen perusteella.

TAULUKKO 4. Oikeiden luokitusten määrä prosentteina eri luokittelutapauksissa. SS = luokittelukategorioina käytetty saman sukupuolen tuotoksista laskettuja keskiarvospektrejä, KA = luokittelukategorioina koko aineistosta lasketut keskiarvospektrit, VS = luokittelukategorioina vastakkaisen sukupuolen tuotoksista lasketut keskiarvospektrit, ja MAX-MIN = parhaimman ja huonoimman luokitusprosentin erotus.

	/i/	/e/	/y/	/ö/	/ä/	/a/	/o/	/u/	x̄
SS	100,0	95,3	99,2	93,0	93,8	96,1	98,4	99,2	96,9
KA	95,3	81,2	87,5	86,7	91,4	96,9	98,4	99,2	92,1
VS	68,8	42,2	68,0	60,9	68,8	86,7	93,8	96,1	73,2
MAX-MIN	31,2	53,1	31,2	32,1	25,0	10,2	4,6	3,1	23,8

Taulukosta 4 voidaan havaita, että takavokaaleissa luokittelun onnistuminen on paljon riippumattomampaa käytetyistä luokittelukategorioista kuin etuvokaaleissa. Normaalistuksen tarve on takavokaaleissa selvästi pienempi; erityisesti voidaan taas panna merkille /u/:n sisältämän vokaali-informaation pysyvyys erilaisissa oloissa. Vokaalien automaattinen, puhujan sukupuolen huomioon ottava tunnistus saattaisi siksi tapahtua niin, että vokaali ensin alustavasti luokitetaan, jolloin takavokaaleissa päästäisiin ilmeisesti heti melko hyviin tuloksiin. Tämän jälkeen varsinkin etuvokaalien luokitusta tarkennettaisiin alustavan luokituksen ja edellä käsiteltyjen systemaattisten vastaavuuksien perusteella.

Aineiston spektrit on myös luokiteltu sukupuolen mukaan kunkin vokaalifoneemin tuotosten osalta erikseen, luokittelukategorioina kussakin tapauksessa kyseisen vokaalin miesten ja naisten aineistoista erikseen lasketut keskiarvospektrit. Miehillä oikeita luokituksia saatiin kaikkiaan 484/512 (94,5 %), naisilla 485/512 (94,7 %). Kuten jo sopii odottaa, valtaosa virheluokituksista koski /u/-tuotoksia.

Vokaalien, puhujien ja puhujan sukupuolen luokittelusta saadut tulokset osoittavat psykoakustisten spektrien sisältävän verrattomasti enemmän tietoa kuin perinnäisesti käytetyt muutaman alimman formantin arvot. Tätä voidaan pitää yhtenä lisäargumenttina koko spektrin muodon huomioon ottavan kuvauksen puolesta. Lisäksi on muistettava, että spektrografi vain vaivoin soveltuu naisten ja etenkin lasten vokaalien analysointiin. Nyt käytetyssä menetelmässä naisten tuotokset on analysoitu täsmälleen samaa algoritmia käyttäen kuin miestenkin tuotokset, eikä tulosten luotettavuudessa — kun sitä arvioidaan jälkikäteen tehtyjen luokittelujen valossa — näytä olevan eroa. Tulevaisuudessa aineistoa on tarkoitus täydentää myös lasten tuotoksilla. Silloin käytettävissä oleva aineisto tarjonnee hyvän pohjan koko normaalistuksen tarkemmalle tutkimukselle. Tavoitteena voisi aluksi pitää sitä, että vain 8:aa referenssispektriä (yksi vokaalifoneemia kohti) käyttäen saavutettaisiin vähintään yhtä hyvät vokaalien ja puhujan sukupuolen oikeat luokitteluprosentit kuin edellä selostetuissa erillisissä luokituksissa. Varsinaisesti koetukselle kaavailtu tunnistusalgoritmi joutuu tietenkin vasta sitten, kun sitä sovelletaan alkuperäisen puhujajoukon ulkopuolelle, saati kun se itse paikantaa analysoitavan kohdan. Vasta tällöin voidaan puhua ensi askelista varsinaisen puheen automaattisen tunnistuksen saralla.

4.5. Vokaalien psykoakustisista etäisyyksistä

Sikäli kuin käytetty suodatinpakka on kattamansa taajuusalueen puolesta vokaalien tunnistuksen ja erottelun kannalta riittävä, omaksuttu analyysimenetelmä psykoakustisesti todenmukainen ja otos tilastollisesti edustava,

lasketut vokaalien keskiarvot kuvastavat vokaalien relevantteja ominaisuuksia perifeerisen kuulon tasolla. Yhä kiistelty kielellisen ja ei-kielellisen kuulon erilaisuus ja edellisen mahdollinen kielikohtaisuus saattavat merkitä joitakin muutoksia, mutta toistaiseksi paras käytettävissä oleva arvio vokaalien havainnon kannalta tärkeistä ominaisuuksista perustuu välttämättä psykoakustiikan yleistä kuulemista koskeviin tutkimustuloksiin. Tässä mielessä on taulukoiden 2 ja 3 psykoakustisia keskiarvospektrejä pidettävä parhaina tässä tutkimuksessa käytettävissä olevina arvioina suomen vokaalien perseptuaalisfoneettisista ominaisuuksista.

Euklidisen etäisyysmitan avulla laskettiin vokaalien keskiarvospektrien väliset etäisyydet miesten ja naisten osalta erikseen (taulukot 5 ja 6). Taulukkoja keskenään vertaillessa voi panna merkille, että miesten aineistossa etäisyydet ovat kauttaaltaan suuremmat kuin naisten aineistossa; eron suuruus vaihtelee vertailtavan vokaaliparin mukaan. Tämä saattaa johtua siitä, että suodatinpakka ei naisten vokaalien erottelun kannalta ulotu yhtä riittävän ylös kuin miesten (vrt. kohtaan 4.4 edellä). Muuten etäisyydet kummassakin ryhmässä ovat hyvin samansuuntaiset. Niinpä kummassakin ryhmässä /i/:n ja /a/:n välinen etäisyys on suurin ja /e/:n ja /ö/:n etäisyys pienin; jälkimmäisestä tosin varsinkin naisilla on hyvin pieni matka seuraavaksi pienimpään etäisyyteen.

Vokaalien psykoakustinen kuvaus on määritelmän mukaan yksityisistä kielistä riippumaton, ja tässä suhteessa se on oivallinen perusta esimerkiksi kontrastiivisille vertailuille ja äänteellisten universaalien tutkimukselle (vrt. kontrastiivisessa kielentutkimuksessa esitettyyn vaatimukseen, että vertailuilla täytyy olla objektiivinen perusta, ns. *tertium comparationis*). Varsinkin vokaalijärjestelmien universaalien tutkimuksessa on vedottu siihen, että vokaalien väliset etäisyydet määräävät systeemien rakennetta (Liljencrants ja Lindblom 1972, Crothers 1978). Yhtenä ongelmana vokaalijärjestelmien rakenteen ennustamisessa on ollut mm. se, että käytetyt teoriat ennustavat (generoivat) typologisen tiedon valossa liiallisen määrän suppeita vokaaleja [i]:n ja [u]:n väliin: tämä ulote tulee liian täyteen verrattuna vokaalien korkeusulotteeseen. Epäsuhta empiirisen typologisen tiedon ja teoreettisten mallien ennustusten välillä johtuu siitä, että formanttimittausten perusteella — vaikka ne olisi muutettu mel-asteikollekin (1 mel = 1/100 Bark) — [i]:n ja [u]:n välinen etäisyys näyttää paljon suuremmalta kuin etäisyydet korkeusulotteen ääripäiden välillä, ja tällöin suppeiden äärivokaalien väliin jää näennäisesti paljon tilaa. Taulukoista 5 ja 6 nähdään, että koko spektrin huomioon ottavassa psykoakustisessa kuvauksessa ainakin suomen /i/:n ja /u/:n välinen etäisyys on selvästi pienempi kuin /i/:n ja /ä/:n tai /u/:n ja /a/:n väliset etäisyydet. Tämä on selvästi paremmassa sopusoinnussa typologisen tiedon kanssa; väljyysasteita on maailman kielissä yleensä enemmän

TAULUKKO 5. Vokaalien keskiarvospektrien väliset psykoakustiset etäisyydet. Miesten aineisto.

	/i/	/e/	/y/	/ö/	/ä/	/a/	/o/
/e/	22,8						
/y/	25,1	21,4					
/ö/	32,6	16,1	18,8				
/ä/	43,8	26,3	33,2	17,9			
/a/	50,0	38,2	44,7	30,8	21,5		
/o/	46,9	38,1	43,6	32,5	31,8	25,3	
/u/	41,2	39,5	40,0	36,5	42,3	41,1	22,3

TAULUKKO 6. Vokaalien keskiarvospektrien väliset psykoakustiset etäisyydet. Naisten aineisto.

	/i/	/e/	/y/	/ö/	/ä/	/a/	/o/
/e/	17,9						
/y/	17,6	17,1					
/ö/	26,7	15,5	15,6				
/ä/	36,4	26,0	29,2	19,3			
/a/	47,4	40,7	42,0	33,4	22,0		
/o/	41,1	35,7	39,6	33,0	32,6	29,4	
/u/	29,8	27,3	28,1	25,8	35,1	39,0	19,9

kuin takaisuusasteita pyöreys—laveus-vastakohta mukaan lukien (Crothers 1978).

Taulukoista 5 ja 6 voidaan todeta, että vokaalien /y/, /ö/ ja /ä/ psykoakustiset etäisyydet toisiinsa, /y/:n ja /ö/:n etäisyydet muihin — varsinkin saman väljyysasteen — etuvokaaleihin ja /ä/:n etäisyys /a/:han ovat lyhimpien etäisyyksien joukossa. Lisäksi on mainituille vokaaleille ominaista se, että ne ovat lähellä useaa vokaalia. Onkin selvää, että jos nämä vokaalit poistettaisiin suomen vokaalijärjestelmästä, sekä minimaaliset että keskimääräiset vokaalien väliset etäisyydet kasvaisivat. Tässä mielessä /y/:tä, /ö/:tä ja /ä/:tä voidaan pitää psykoakustisesti heikkoina vokaaleina: ne aiheuttavat psykoakustisessa avaruudessa tungoksen. Olen esittämässäni ns. palataalisen vokaaliharmonian selityksessä ottanut lähtökohdaksi noiden vokaalien perseptuaalisen heikkouden (Suomi 1983a, 1983b), ja tältä osin ovat lasketut etäisyydet sopusoinnussa aiempien väitteideni kanssa. Käsitykseni kyseisten vokaalien perseptuaalisesta heikkoudesta perustui kuitenkin paremman tiedon puutteessa niiden sijaintiin F2:n ulotteella, ja tässä suhteessa olen joutunut täsmentämään teoriaani uuden tiedon valossa (Suomi 1984).

LÄHTEET

Bladon, Anthony 1982: Arguments against formants in the auditory representation of speech. — The representation of speech in the peripheral auditory system (ed. by R. Carlson & B. Granström), Elsevier Biomedical Press, Amsterdam, s. 95—102.

——— & Lindblom, Björn 1981: Modeling the judgment of vowel quality differences. — Journal of the Acoustical Society of America 69 s. 1414—1422.

Blomberg, Mats, Rolf Carlson, Kjell Elenius & Björn Granström 1982: Experiments with auditory models in speech recognition. — The representation of speech in the peripheral auditory system (ed. by R. Carlson & B. Granström), Elsevier Biomedical Press, Amsterdam, s. 197—201.

Carlson, Rolf, Gunnar Fant & Björn Granström 1975: Two formant models, pitch and vowel perception. — Auditory analysis and perception of speech (ed. by G. Fant & M. Tatham), Academic Press, London, s. 55—82.

——— & Grandström 1982: Towards an auditory spectrograph. — The representation of speech in the peripheral auditory system (ed. by R. Carlson & B. Granström), Elsevier Biomedical Press, Amsterdam, s. 109—114.

Crothers, John 1978: Typology and universals of vowel systems. — Universals of human language, Vol. 2, Phonology (ed. by J. Greenberg), Stanford University Press, Stanford, California, s. 93—152.

Fant, Gunnar 1970: Acoustic theory of speech production. Second printing. Mouton, The Hague.

Iivonen, Antti 1979: On the problems of vowel study utilizing acoustic methods. — Fonetiikan päivät — Jyväskylä 1978, Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 18, s. 57—81.

——— 1982: Vokaalien psykoakustisesta laadusta. — X Fonetiikan päivät Tampereella 20.—21. 3. 1981, Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitoksen julkaisuja 7 s. 73—115.

Karjalainen, Matti 1982a: Formanttiparametrien mittauksesta ja analyysistä. — X Fonetiikan päivät Tampereella 20.—21. 3. 1981, Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitoksen julkaisuja 7 s. 123—137.

——— 1982b: Puheen perifeerisen kuulemisen laskennallisista malleista. — XI Fonetiikan päivät — Helsinki 1982, Helsingin yliopiston fonetiikan laitoksen julkaisuja 35 s. 89—118.

Klatt, Dennis 1982: Speech processing strategies based on auditory models. — The representation of speech in the peripheral auditory system (ed. by R. Carlson & B. Granström), Elsevier Biomedical Press, Amsterdam, s. 181—196.

Ladefoged, Peter 1967: Three areas of experimental phonetics. Oxford University Press, London.

Liljencrants, Johan & Lindblom, Björn 1972: Numerical simulation of vowel quality systems: the role of perceptual contrast. — Language 48 s. 839—862.

Papçun, George 1980: How do different people say the same vowels? Discriminant analyses of four imitation dialects. — UCLA Working Papers in Phonetics 48, University of California, Los Angeles.

Plomp, Reinier 1970: Timbre as a multidimensional attribute of complex tones. — Frequency analysis and periodicity detection in hearing (ed. by R. Plomp & G. Smoorenburg), Sijthoff, Leiden, s. 397—411.

Pols, Louis, H. Tromp & Reinier Plomp 1973: Frequency analysis of Dutch vowels from 50 male speakers. — Journal of the Acoustical Society of America 53 s. 1093—1101.

Scharf, Bertram 1970: Critical bands. — Foundations of modern auditory theory (ed. by J. Tobias), Academic Press, New York, s. 157—202.

Suomi, Kari 1983a: Palatal wovel harmony: a perceptually motivated phenomenon? — Nordic Journal of Linguistics 6 s. 1—35.

——— 1983b: Itämerensuomen vokaaliharmoniasta, neutraaleista vokaaleista ja keskivokaaleista. — Virittäjä 87 s. 508—517.

SUOMI, KARI 1984: A revised explanation of the causes of palatal vowel harmony based on psychoacoustic spectra. — Nordic Journal of Linguistics 7 (painossa).

ZWICKER, EBERHARDT & FELTKELLER, RICHARD 1967: Das Ohr als Nachrichtenempfänger. Hirzell, Stuttgart.

On determining the psychoacoustic quality of vowels

An algorithmic method and results for Finnish monophthongs

KARI SUOMI

Reasons are discussed for the growing dissatisfaction with the conventional method of vowel analysis in terms of the frequencies of the lowest formants, and an alternative method is advocated in which the shape of the whole spectrum is taken into account. The particular method used in these experiments attempts to simulate the major transformations that, according to the findings of general psychoacoustics, take place in the auditory channel; the output of the analysis is an approximation of the peripheral representation of vowel quality in the inner ear in terms of the loudness levels of 16 adjacent critical bands in phons/Bark. Except for the initial visual determination of the sampling point in the audio wave, the analysis is fully algorithmic.

The data, consisting of eight repetitions of the eight Finnish monophthongs in a constant frame sentence spoken by eight male and eight female adults, were analysed and the resultant psychoacoustic spectra were classified by machine on the basis of minimum computed Euclidean distances between each token spectrum and various average spectra acting as classificatory categories.

General differences between the psychoacoustic spectra and corresponding spectrographic displays are commented on, with the observation that several of the analytical problems notoriously connected with formant frequency measurements disappear in the present method. The results obtained in the machine classifications of the spectra with respect to vowel phoneme identity, speaker identity and speaker sex indicate that the psychoacoustically transformed spectra contain much more information than the few traditionally extracted formant frequency parameters, and this can be taken as a further argument in favour of the whole spectrum approach to vowel quality. Hence, the latter approach could be profitably adopted also in more clearly linguistically oriented applications. Finally, the computed Euclidean distances among the average vowel spectra are discussed with a view to their implications — to be further elaborated elsewhere — for the explanation of typological universals of vowel systems and of the so-called palatal vowel harmony.