

Emootiot äänessä. Tunneilmaisun akustiset ominaisuudet ja vastaanotto

Teija Waaramaa-Mäki-Kulmala

Lectio praecursoria puheopin väitöskirjaksi tarkoitetun tutkimuksen *Emotions in voice. Acoustic and perceptual analysis of voice quality in the vocal expression of emotions* tarkastustilaisuudessa Tampereen yliopistossa 8.5.2009. Vastaväittäjänä toimi professori Olli Aaltonen (Helsingin yliopisto) ja kustoksena professori Anna-Maija Korpijaakko-Huuhka.

Puheentuoton perusedellytykset

Ihmisen ääni on hämmästyttävän joustava väline viestien ja tunteiden ilmaisemisessa sekä tiedonvaihdossa ihmisten välillä. Kyky vastaanottaa ja tuottaa tietoisesti puheääntä on ihmiselle lajityypillinen ominaisuus, jota on edeltänyt pitkä kehityshistoria. Ensiksikin puheen synty on edellyttänyt kykyä tuottaa erilaisia, toisistaan erottuvia ääniteitä. Tämä puolestaan on edellyttänyt ääniväylän *anatomista* kehittymistä juuri tähän tarkoitukseen. (Fitch 2000.) Ihmisen ääniväylä on suhteellisen pitkä, ja kurkunpää sijaitsee melko matalalla ääniväylässä verrattuna muihin kädellisiin. (Kuva ääniväylästä löytyy esimerkiksi osoitteesta <http://www.abdn.ac.uk/langling/resources/midsagsectionbw.jpg>.) Tällaisen rakenteen ansiosta kielellä on tilaa liikkua vertikaalisesti ylös ja alas sekä horisontaalisesti sivusuunnassa suhteellisen laajasti (Lieberman 1969; 1972). Nämä liikkeet ovat mahdollistaneet niin ikään laajan formanttirakenteiden käytön. Formantit syntyvät kurkunpään yläpuolisessa osassa ja ovat siten ääniväylän resonansseja. Ne ovat puheentuoton tärkein ominaisuus. Esimerkiksi kuiskaus on tunnistettavissa juuri sen oman akustisen formanttirakenteeseensa ansiosta, vaikka kuiskauksessa ihminen ei puhukaan niin sanotusti ”ääneen”. (Fitch 2000.)

Toisena puheen edellytyksenä on ollut äänen tuottoon liittyvän *neuraaliverkoston* kehittyminen. Hermoyhteyksiä on tarvittu riittävä määrä paitsi puheen tuottamiseen myös sen *kontrollointiin* ja edelleen puheen *imitointiin*. Ihmisen erityistä kykyä imitoida kuulemiaan ääniä on pidetty perusedellytyksenä monipuolisen *sanaston* syntymiselle, mikä on tyypillistä kaikille ihmisten puhumille kielille. Kolmanneksi ihmisellä on täytynyt olla kyky *oppia* puheen ääniteitä. Lopuksi vielä on tarvittu edellytyksiä puheen *vastaanottamiseen*. (Fitch 2000; MacNeilage 1998.)

Puheen artikulaation perusliikkeet saattavat pohjautua alun perin syömisen liikkeisiin, suun jatkuvaan avaamiseen ja sulkemiseen, pureskeluun, imemiseen ja nuolemiseen. Vähitellen nämä liikkeet ja niihin liittyvät äänet ovat alkanet saada viestinnällisiä merkityksiä. Näihin saattaa pohjautua myös puheen tavujen rytmii. (MacNeilage 1998.)

Nykyisen käsityksen mukaan on mahdollista, että Neandertahlin ihmisellä on ollut suuremmat aivot kuin nykyihmisellä, mutta se ei kuitenkaan auttanut Neandertahlin ihmistä menestymään. Neandertahlin ihmisen menehtymiseen on jopa esitetty syyksi hänen lyhyttä ääniväyläänsä, joka oli yhtä pitkä kuin nykyihmisen lapsen ääniväylä. Näin ollen Neandertahlin ihminen ei ole pystynyt tuottamaan yhtä suurta äänteiden määrää kuin nykyihminen ja siksi hänen kommunikaatiokykynsä äänen avulla on siis ollut selvästi rajoittuneempi. Lyhyt ääniväylä on myös todennäköisesti tuottanut taajuudeltaan suhteellisen korkeita ääniä aikuiseen nykyihmiseen verrattuna. Nämä ominaisuudet ovat merkinneet, että Neandertahlin ihmisen puhe ei ole voinut kehittyä samoin kuin nykyihmisen. (Lieberman mm. 1969; 1972.)

Äänisymbolismi

John Ohalan (1997) *frekvenssikooditeorian* mukaan yhteys puheäänien ja niiden merkitysten välillä ei ole satunnainen. *Äänisymbolismia* tavataan ihmisen ohella myös muilla eläinlajeilla. Äänisymbolismin keskeinen ajatus on se, että pieni lähde (ihminen tai eläin) tuottaa korkeataajuisen äänen ja suuri lähde vastaavasti matalataajuisen äänen. Korkeataajuinen ääni tulkitaan yleisesti vähemmän uhkaavaksi ja alistumista viestiväksi. Matalataajuinen ääni koetaan enemmän uhkaavaksi ja ylemmyyttä osoittavaksi.

Tutkimuksissa on havaittu, että kuuntelijoiden arviointi puhujan ruumiin koosta ja hänen emotionaalisesta tilastaan liittyvät toisiinsa. Arvioinnit näyttävät perustuvan sekä ääniväylän pituuteen että äänen perustaajuuteen. Kun tutkijat tuottivat eräässä kokeessa synteettisiä vokaaleja pidennetyllä ääniväylällä ja madalletulla äänenkorkeudella, niin kuuntelijat arvioivat vokaalien olevan peräisin suurikokoisesta ihmisestä. Kun ääniväylää lyhennettiin ja äänenkorkeutta nostettiin, kuuntelijat kokivat äänilähteen pieneksi. Nämä erot vaikuttivat myös siihen, kuultiinko näytteiden sisältävän vihaa vai iloa. (Chuenwattanapanithi & al. 2008; Xu & Chuenwattanapanithi 2007.)

Pelineuronit

Aivojen Brocan aluetta on yleisesti pidetty puheen motoriikkaa kontrolloivana alueena. Brocan alueen on kuitenkin todettu aktivoituvan myös *käden* liikkeistä sekä pelkästään jo liikkeen *mielikuva*sta. Todennäköisesti Brocan alueen ensisijainen tehtävä onkin ollut toisten ihmisten tekemien liikkeiden *havainnointi*

ja *ennakointi* jo ennen itse puheen kehittymistä. Tämä liikkeitä peilaava järjestelmä näyttäisi olevan se silta, joka yhdistää ihmisen aivoissa toisen ihmisen motorisen toiminnan ennakoimisen ja havainnoinnin sekä oman motorisen toiminnan tuottamisen. Näitä *peilineuroneja* on tarvittu myös puheen artikulaatio-liikkeiden tunnistamiseen, tuottamiseen ja lopulta varsinaisen puheen synty-miseen. (Gallese & al. 1996; Rizzolatti & Arbib 1998; Rizzolatti & al. 1996.) Käsillä ja kasvoilla, etenkin suulla ja kielellä, onkin ihmisen aivoissa huomattavan suuri edustuksellinen osuus suhteessa muihin ruumiinosiin. Kuva niin kutsutusta homunculuksesta osoittaa, miten suuri tämä edustus on (ks. kuva <http://www.juergenhaenggi.ch/Bilder/Homunculus.png>).

On esitetty myös, että ihmisen imitointikyvyn taustalla saattavat olla erityiset audiovisuaaliset peilineuronit, jotka eivät ole erikoistuneet havainnoimaan tilaa ja tilassa tapahtuvia liikkeitä, vaan ne ovat erikoistuneet kuulohavaintoon perustuvan toiminnan käsittelyyn. Tämän teorian mukaan ihminen kykenee tunnistamaan motoriikkaan perustuvan liikkeen pelkästään kuulohavaintonsa avulla, näkemättä puhujaa. Toiminnan havaitseminen tai sen ennakoiminen on peilineuroneille siis samaa kuin toiminnan simuloiminen ”sisäisesti”. Audiovisuaaliset peilineuronit tunnistavat näin ollen abstraktin toiminnan merkityksiä ja tarkoitusta. Ne ovat ikään kuin auditiivinen portti ihmisen käyttämään kieleen ja sen merkityksiin. Audiovisuaalisten peilineuronien ”löytyminen” osoittaa, että peilineuronien toiminta ei rajoitu pelkästään liikkeen havainnointiin, vaan peilineuronit toimivat useilla muillakin aivojen alueilla, kuten tunteiden välittymisessä. (Gallese 2004; Kohler & al. 2002.)

Simuloitumisella on keskeinen osuus ihmisen episteemisessä maailman-käsityksessä. Se ei ole tahdonalainen kognitiivinen tapahtuma, jonka avulla ihmiset tulkitsisivat piilomerkityksiä toistensa käyttäytymisestä, vaan se on osa ihmisaivojen perustoimintaa. Juuri tämän toimintamekanismin ansiosta ihminen on me-keskeinen yksilö, jolle interpersonaaliset toiminnot ovat perusinhimillisiä, universaaleja ilmiöitä. (Gallese 2004.)

Taustatunne

Ihmisen puhe sisältää aina enemmän tai vähemmän emotionaalista informaatiota. Silloin kun puheesta ei havaita positiivista tai negatiivista sävyä eli valenssia, voidaan sanoa, että kyseessä on neutraali valenssi. Kuuntelukokeet ovat osoittaneet, että jos puhe on tarkoitettu neutraaliksi mutta jos se ei kuuntelijoiden mielestä kuitenkaan sisällä vähäisintäkään positiivista vivahdetta, niin vastaanottaja kokee sen helposti pelkästään negatiiviseksi, tyykyksi. (Waaramaa-Mäki-Kulmala 2009.)

Täysin neutraalin tunnetilan mahdollisuus onkin kyseenalaistettu voimakkaasti. Tunnetun neurologin Antonio Damasion (2007) mukaan ihminen on aina jonkinlaisessa emotionaalisisessa tilassa. Tätä olotilaa hän kutsuu *taustatunteeksi*.

Damasion (2007) mukaan ihminen voi olla neutraalissa tunnetilassa vain silloin, kun hän on tiedoton. Izard (2007) ajattelee samoin: ihminen on jatkuvassa emotionaaliossa tilassa ja tämä tila ohjaa hänen tietoisuuttaan. Taustatunne vaikuttaa ihmisen mieleen, käyttäytymiseen ja sosiaaliseen kyvykkyyteen. Se vaikuttaa myös emotionaaliossa informaation havainnointiin ja vastaanottoon. Jos ihminen on esimerkiksi taipuvainen näkemään asiat vihamielisinä, hän löytää vastaavia sävyjä ja vihjeitä sieltäkin, missä muut eivät niitä havaitse.

Ihmisen päämäärätavoitteisessa toiminnassa on aina mukana tunnekomponentteja. Taustatunne kuvaa yksilön yleistä emotionaaliossa asennoitumista, jossa emootioiden havaitseminen on kaksisuuntainen prosessi. Toisaalta emootiot ohjaavat ja valikoivat informaatiota. Tällöin prosessi on proaktiivinen. Kuitenkin samalla tietoisuuden ja käyttäytymisen suhde prosessiin on reaktiivinen. (Izard 2007.)

Emootion taustalla on siis paitsi ihmisen evoluutioon liittyvä pitkä kehitys, myös monimutkainen kognitiivinen prosessointi aivoissa. Näin ollen tunne ei voi olla järjen vastakohta, vaan se on pikemminkin järjen syy ja seuraus. Izard (2007) onkin sanonut, että normaaliossa ihmismielessä kognitiota ei ole ilman emootiota.

Emosfääri

Puhujasta syntyviin käsityksiin vaikuttavat sekä kulloinenkin tilanne että myös aiemmat käsitykset ja kokemukset. Pienilläkin puheen vivahde-eroilla voi olla keskeinen merkitys mielleyhtymien, konnotaatioiden ja henkilökohtaisten assosiaatioiden syntymiselle. Näitä vivahteita on luonnollisesti sitä helpompi havaita, mitä paremmin puhujat tuntevat toisensa (Birkett & al. 2007).

Yksilön intra- ja interpersonaaliossa aspektien sekä ympäröivän kulttuurin konventioiden lisäksi myös universaalit, ihmisen evoluutiossa kehittyneet tavat muokkaavat tunteiden ilmaisua ja vastaanottoa. Tätä yksilöllisen emotionaaliossa kehityksen koordinaatistoa ja sen päälle rakentuvaa yksilöllisten kokemusmaailman hermeneuttista kehää voidaan kutsua *emosfääriksi* (vrt. Juri Lotmanin *semiosfääri* [1984], 2005).

Universaali intrapersonaaliossa A	Universaali interpersonaaliossa B
Kulttuuriossa intrapersonaaliossa C	Kulttuuriossa interpersonaaliossa D

Kuvio 1. Emosfäärin nelikenttä (Waaramaa-Mäki-Kulmala 2009)

Emosfääriin liittyy sekä yksilöllinen että yleinen, ulkoinen ja sisäinen, jotka ovat jatkuvassa vuorovaikutuksessa keskenään rajojensa välillä ja vieläpä rajojensa ylitse. Emosfääri on jatkuvassa aktiivisessa toiminnassa, mikä heijastelee ihmismielessä kulloinkin meneillään olevia useita simultaaneja emotionaalisia prosesseja. Toiset näistä prosesseista ovat pitempi- ja toiset lyhyempikestoisia, toiset voimakkaampia ja toiset taas heikompia vaikutukseltaan. Emosfääriä voi kuitenkin dominoida jokin yksittäinen tunnetila, kuten hyvä tai paha mieliala, suru tai ilo. Emosfääri merkitsee siis jatkuvaa tulkitsemisen, valitsemisen, unohtamisen ja muistamisen prosessointia. Tällaisten tapahtumien tuloksena emosfääriin ydinrakenne voi muuttua esimerkiksi (teoreettisesti) niin, että odottamaton negatiivinen kokemus saattaa vaikuttaa yksilön mentaaliseen terveydentilaan horjuttamalla sitä tai jopa johtamalla hänen persoonallisuutensa muutokseen (tämä olisi intrapersonaalinen tapahtuma) ja edelleen hänen sosiaalisten suhteittensa muuttumiseen (tämä olisi puolestaan interpersonaalinen muutos emosfäärissä). Samanlainen tapahtumaketju voi olla mahdollinen vastaavasti positiivisen kokemuksen kohdalla. Tällä tavoin emosfääriin dialogi, vuoropuhelu, on merkityksiä ja emotionaalista tietoa välittävää vuorovaikutusta.

Emosfääriin ajatuksen mukaan tunteet eivät voi syntyä, eivätkä ne voi olla olemassa eristettyinä, vaan ne tarvitsevat välttämättömän kontekstinsa. Itse tunteetkaan eivät voi olla olemassa muutoin kuin suhteessa toisiinsa: ilo tunnustetaan iloksi, koska on olemassa myös suru, ja viha tunnustetaan vihaksi, koska tiedämme, mitä on hellyys. Vain määrittelemällä muut tunteet erilaisiksi, tunne saa omat ominaisuutensa.

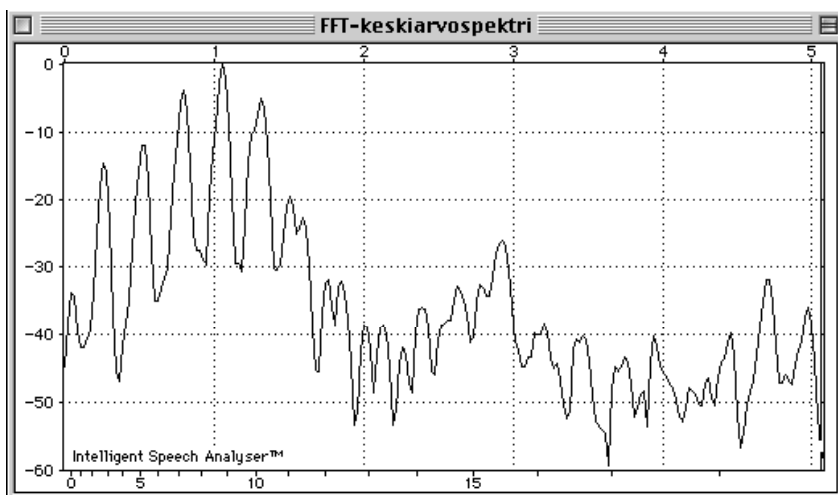
Äänenlaatu emootioilmaisuiissa

Tähänastinen puheilmaisun emootiotutkimus on keskittynyt lähinnä äänen prosodiikkaan, kuten perustaajuuteen, äänen voimakkuuteen, keston, kesto-suhteisiin lauseen sisäisellä sekä puhe- ja artikulaatiotempoon. Näillä muuttujilla on todettu olevan vaikutusta emootioiden välittämisessä.

Viime aikoina on tutkimuksessa alettu kiinnittää enemmän huomiota *äänenlaadun* akustisiin ominaisuuksiin ja niiden merkitykseen emotionaalisessa informaatioissa. Äänenlaatu tarkoittaa yhtäältä yksilön taikka sosiaalisen tai kulttuurisen ryhmän äänenkäyttötapoja. Toisaalta se tarkoittaa yksilöllistä auditiivista äänenväriä tai sen jotakin ominaispiirrettä. Ominaispiirteet erottavat puhujat toisistaan, mutta ne voivat myös viestiä esimerkiksi puhujan emotionaalisesta tilasta. Akustisesti äänenlaatu ilmenee ennen kaikkea puhesignaalin energian jakautumisena taajuusasteikolle, jota voidaan kuvata spektrillä. Esimerkiksi alla olevista kuvioista 2 ja 3 näkyy, miten surun ilmaisussa käytetään tyypillisesti vähemmän äänienergiaa kuin vihan ilmaisussa. (Laver 1980; Fant 1970.)



Kuvio 2. Vokaali [a:] ilmaistuna surussa¹



Kuvio 3. Vokaali [a:] ilmaistuna vihassa

Verbaalisesti neutraalin ilmaisun on todettu herättävän laajalti erilaisia mielleyhtymiä pelkästään äänenlaadun muutosten vuoksi. Äänenlaatu näyttää olevan se muuttuja, joka välittää hienojakoisempaa informaatiota kuin vaikkapa helposti havaittavat suuret äänenkorkeuden muutokset, jotka ovat tyypillisiä korkean

¹Horisontaalinen akseli kuvaa kuvioissa 2 ja 3 taajuutta (Hz) ja vertikaalinen akseli amplitudia (dB). Spektrit on tehty ISA (Intelligent Speech Analyser) -ohjelmalla, jonka on kehittänyt DI Raimo Toivonen.

intensiteetin tunteissa. Äänenlaadun on lisäksi todettu olevan tärkein sosiaalisia eli kulttuurisia emootioita erotteleva muuttuja. Onkin esitetty, että äänenlaatu pitäisi nostaa tärkeydessä äänen prosodisten piirteiden rinnalle. (Campbell & Mokhtari 2003; Laukkanen & al. 1997.)

Nyt tarkastettavalla väitöstutkimuksella on pyritty selvittämään, voiko äänenlaadun perusteella havaita vokaalin mittaisesta signaalista emootiota tai sen valenssia ja voiko yhdeltä sävelkorkeudelta tuotetusta vokaalista tunnistaa eri tunnetiloja. Lyhyiden ääninäytteiden tutkimisen ideana on ollut se, että niissä eivät pääse vaikuttamaan nuo kaikkein ilmeisimmät emootioiden välittäjät eli perustajuus, voimakkuus ja kesto eikä muu nonverbaalinen viestintä tai sanojen merkitys. Tavoitteena on siis ollut päästä avaamaan arvoitusta äänenlaadun merkityksestä emootioilmmaisussa.

Kirjallisuus

- Birkett P. B., Hunter M. D., Parks R. W., Farrow, T. F., Lowe H., Wilkinson, L. D. & Woodruff, P. W. 2007. Voice familiarity engages auditory cortex. *Neuroreport* 18 (13), 1375–1378.
- Campbell, N. & Mokhtari, P. 2003. Voice quality: the 4th prosodic dimension. *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, 3.–9.8.2003, 2417–420.
- Chuenwattanapranithi, S., Xu, Y., Thipakorn, B. & Maneewongvatana, S. 2008. Encoding emotions in speech with the size code. A perceptual investigation. <<http://www.phon.ucl.ac.uk/home/yi/publications.html>> Viitattu 27.10.2008.
- Damasio, A. 2007. Brain and mind: from medicine to society. Conference lecture (video) in ”Brain and mind: from medicine to society”, Barcelona, Spain. 24th of May 2007. <<http://www.youtube.com/watch?v=KbacW1HVZVk&NR=1>> Viitattu 9.5.2008.
- Fant, G. 1970. Acoustic theory of speech production. With calculations based on X-ray studies of Russian articulations. (2nd ed.). The Hague: Mouton.
- Fitch, W. T. 2000. The evolution of speech: a comparative review. *Trends in Cognitive Sciences* 4 (7), 258–267.
- Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. 1996. Action recognition in the premotor cortex. *Brain* 119, 593–609.
- Izard, C. E. 2007. Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science* 2 (3), 260–280.
- Kohler, E., Keyers, C., Umiltá, M. A., Fogassi, L., Gallese, V. & Rizzolatti, G. 2002. Hearing sounds, understanding actions: Action representation in mirror neurons. *Science* 297, 846–848.
- Laukkanen, A.-M., Vilkmán, E., Alku, P. & Oksanen, H. 1997. On the perception of emotions in speech: the role of voice quality. *Scandinavian Journal of Logopedics, Phoniatrics, Vocology* 22 (4), 157–168.

- Laver, J. 1980. The phonetic description of voice quality. Cambridge: Cambridge University Press.
- Lieberman, P. 1972. The speech of primates. The Hague: Mouton.
- Lieberman, P., Klatt, D. H. & Wilson, W. A. 1969. Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science* 164, 1185–1187.
- Lotman, Y. 2005. On the semiosphere. *Sign System Studies* 33 (1), 207–229.
- MacNeilage, P. F. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21, 499–546.
- Ohala, J. J. 1997 Sound symbolism. Proceedings of the 4th Seoul International Conference on Linguistics (SICOL). 11. –15.8.1997, 98–103.
- Rizzolatti, G. & Arbib, M. A. 1998. Language within our grasp. *Trends in Neurosciences* 21 (5), 188–194.
- Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. 1996. Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* 3, 131–141.
- Waaramaa-Mäki-Kulmala, T. 2009. Emotions in voice; Acoustic and perceptual analysis of voice quality in the vocal expression of emotions. Academic dissertation. *Acta Universitatis Tamperensis* 1399. Tampere University Press.
- Xu, Y. & Chuenwattanapranithi, S. 2007. Perceiving anger and joy in speech through the size code. 16th International Congress of Phonetic Sciences (ICPHS). Saarbrücken, 6. –10.8.2007.