

KAS FORMANDID PEEGELDAVAD EMOTSIOONE?

Kairi Tamuri

Ülevaade. Formante peetakse üheks emotsiooni peegeldavaks akustiliseks parameetriks – kõneleja emotsionaalne seisund võib põhjustada muutusi häälikute formantstruktuuris. Uurimuse eesmärgiks on teada saada, kas ja kuidas mõjutab lausungi emotsioon eestikeelse etteloetud emotsionaalse kõne lühikesi vokaale ning ettelugeja artikulatsiooni täpsust. Selleks mõõdeti etteloetud emotsionaalse ning neutraalse kõne lühikeste rõhuliste vokaalide *a*, *i*, *u* esimest ja teist formanti. Uurimistulemustest selgus, et emotsioonipaariti (k.a neutraalne kõne) on statistiliselt olulised erinevused olemas vokaalide *a* ja *i* esimeste formantide (F1) keskmistes. Teine formant (F2) emotsioonide eristamisel olulist rolli ei mängi. Vokaalide redutseerumist ja artikulatsiooni täpsust uuriti eukleidilise kauguse arvutamise abil. Tulemused näitasid, et viha- ja rõõmu-lausungites jääb kõneleja artikulatsiooni täpsus võrreldes neutraalse kõnega enam-vähem samaks, kurbuse-lausungites see aga langeb.*

Võtmesõnad: emotsionaalne kõne, akustika, formantanalüüs, artikulatsiooni täpsus, eesti keel

1. Sissejuhatus

Paljud emotsioonisünteesi ja -tuvastusega seotud uurimused on näidanud, et igal emotsioonil on just talle omased akustilised tunnused. Kuulajad suudavad emotsioonid kõnes ära tunda ainuüksi heli põhjal, ilma kõnelejat nägemata (nt telefoni-vestlustes) (Bachorowski 1999). Kõnelõigu emotsiooni ollakse võimelised määrama isegi siis, kui kõik selle lõigu sõnad on tähenduseta (Goudbeek jt 2009). See, et kuulajad suudavad ainult hääle järgi emotsioone usaldusväärsetl ära tunda, toobki kaasa oletuse, et hääle kannab kogu informatsiooni kõneleja emotsionaalse seisundi kohta¹ ja et emotsioonide häälelised väljendused on erineva akustilise mustriga (nt Paeschke jt 1999, Banse, Scherer 1996).

* Artikkel on valminud tänu projektidele EKT1 ja SF0050023s09.

¹ Seda vaid siis, kui kuulaja ei näe kõnelejat, sest vastasel juhul muutub akustiliste tunnuste domineerimine emotsiooni äratundmisel küsitavaks.

1.1. Formandid

Üheks akustiliseks parameetriks, mida emotsionaalse kõne puhul uuritakse, on formandid. Formantide abil saab teha järeldusi selle kohta, kuidas emotsioonid üksteisest akustiliselt erinevad ning kuidas mõjutavad emotsioonid häälikute kvaliteeti ja kõneleja artikulatsiooni täpsust.

Formandid on akustilise energia kontsentratsiooni piirkonnad kõnespektris. Kõikidel helilistel häälikutel on oma formantstruktuur ning nende foneetilist kvaliteeti tajutakse harilikult esimesest kolmest formandist (F1, F2 ja F3).

Formantide asukohad ei ole fikseeritud, vaid sõltuvad kõnetrakti suurusest ja kujust. Kõnetrakt muutub, kui varieerub häälduselundite (nt keel, lõug ja huuled) asend. Samuti muudavad kõnetrakti suurust ja kuju emotsioonid. Kui kõneleja emotsionaalne seisund muutub, siis muutuvad tema näoilme ning häälduselundite asend ja lihaspinge (Ververidis, Kotropoulos 2006, Johnstone, Scherer 2000). Naeratades öelduna kõlab lausung akustiliselt üsnagi erinevalt sellest, kui öelda sama asja kulmu kortsutades (Tartter 1980). See omakorda tekitab kõnesignaalis akustilisi muutusi, mis peegelduvad formantmuustrites.

Uurimistulemused on näidanud, et emotsioonid mõjutavad kõige rohkem F1 ja F2 väärtusi (nt Tolkmitt, Scherer 1986).

1.2. Artikulatsiooni täpsus ja redutseerunud häälikud

Põhjus, miks kuulatakse foneetilisi erinevusi kahe vokaali vahel, peitub selles, et vokaalidel on erinev asukoht kahemõõtmelises vokaaliruumis, mida iseloomustavad vokaali kõrgus ehk vertikaalne dimensioon ja vokaali eespoolsus-tagapoolsus ehk horisontaalne dimensioon. Need kaks mõõdet on seotud keele asukohaga suus vokaali moodustamisel. Jonathan Harrington (2010: 46) kirjeldab vokaale kahemõõtmelise jaotusega $F_1 \times F_2$. Vokaali kõrguse kasv seostub F_1 väärtuse vähenemisega, vokaali tagapoolsuse suurenemine seostub aga F_2 kahanemisega. Seega: vokaali *i* $F_1 <$ vokaali *e* F_1 -st ja vokaali *a* $F_2 <$ vokaali *ä* F_2 -st.

Kui vokaaliruum kitseneb, siis kõneleja artikulatsiooni täpsus väheneb. Vokaalid, mida hääldatakse vähem selgelt, kipuvad redutseeruma, s.t nad kaotavad oma kvaliteeti. Kui ilmneb kõrvalekalle vokaali *n*-ö normaalsest positsioonist akustilises ruumis, siis nimetatakse seda vokaali tsentraliseerumiseks (ingl *centralization*). Tsentraliseerumise puhul moodustub vokaal kõneleja vokaaliruumi keskpunkti lähedal ja sarnaneb švaa-häälikuga, millel pole selget kvaliteeti. (Harrington 2010: 190–193)

Uurimistulemustest (Tolkmitt, Scherer 1986) on selgunud, et näiteks stressis või depressioonis inimene ei artikuleeri helilisi häälikuid sama jõupingutusega nagu neutraalse kõne puhul. Sellisel juhul häälikud redutseeruvad ning artikulatsiooni täpsus langeb. Harrington (2010: 190–193) kirjutab, et selge kõne puhul asetsevad vokaalid teineteisest kaugemal ning kuulajal on erinevate häälikute vahel kergem vahet teha. Vokaaliruum muutub suuremaks ning artikulatsiooni täpsus tõuseb.

2. Uurimisküsimus ja -materjal

Formantanalüüsi eesmärk on leida vastused kahele küsimusele: 1) kas eestikeelse etteloetud emotsionaalse kõne lühikesi vokaale mõjutab lausungi emotsioon ning 2) kas emotsioonid mõjutavad ettelugeja artikulatsiooni täpsust.

Uurimistöö akustiline baas on Eesti Keele Instituudi eesti emotsionaalse kõne korpus ehk EEKK². Korpus toetub seisukohale, et emotsioonid on hääle põhjal piisavalt hästi ära tuntavad ka loomulikus, mittenäideldud kõnes, ja et mittenäideldud kõne on loomuliku kõnesünteesi eeldus (Iida jt 2003). Korpuse loomisel on lähtutud printsibist, et lõigu sisu kutsub lugejas esile mingi emotsiooni. Seetõttu pole lugejale ette öeldud, mis emotsiooniga teksti lugeda. Korpus sisaldab ühe naishääle etteloetud ajakirjanduslikke tekstilõike, mis on segmenteeritud lauseteks, sõnadeks ja häälikuteks. Korpuses sisalduvate lausete emotsioonid (viha, rõõm ja kurbus) või neutraalsus on määratud tajukatsetega (vt Altrov 2008).

Uurimismaterjalis on EEKK laused, mille emotsiooni või neutraalsust on tajunud üle 50% kuulajaist. Formantanalüüsiks valiti neist lausetest välja kolm lühikest vokaali: *a*, *i* ja *u* (vokaale on kokku 2395, täpsemat jaotust vt tabelist 1). Need on eesti keeles sagedasti esinevad vokaalid, mis moodustavad kolmnurkse vokaaliruumi (*a* on madal tagavokaal, *i* kõrge eesvokaal, *u* kõrge tagavokaal), mille abil saab hästi kirjeldada kõneleja artikulatsiooni täpsust. Vokaalide juures oli oluline ka nende rõhuline positsioon. Selle tingis soov leida vokaale, mis oleksid võimalikult vähe redutseerunud, sest on teada, et mitterõhuliste silpide häälikud kipuvad oma kvaliteeti kaotama (vt nt Moon, Lindblom 1994). Kuna eesti keeles asub rõhk sõnades tavaliselt esimesel silbil, siis pärinevad väljavalitud vokaalid sõnaalgulistest positsioonidest, mille struktuur on VC, CVC või CCVC³.

Tabel 1. Formantanalüüsi materjal

Vokaal	Emotsioon	Vokaalide arv
<i>a</i>	viha	395
	rõõm	279
	kurbus	244
	neutraalne	319
	KOKKU	1237
<i>i</i>	viha	220
	rõõm	146
	kurbus	154
	neutraalne	145
	KOKKU	665
<i>u</i>	viha	138
	rõõm	133
	kurbus	115
	neutraalne	107
	KOKKU	493
<i>a, i, u</i>	KOKKU	2395

² Vt <http://peeter.eki.ee:5000/> (28.09.2011).

³ V – vokaal; C – konsonant.

3. Meetod

Uurimisküsimustele vastuse leidmiseks mõõdeti esmalt etteloetud emotsionaalse ning neutraalse kõne lühikeste rõhuliste vokaalide *a*, *i* ja *u* esimest ja teist formanti. Harrington (2010: 172–183) kirjutab, et otsus vokaali kvaliteedi kohta tehakse akustiliste andmete põhjal mingil kindlal ajahetkel ja harilikult on selleks koht, mis ühtib vokaali kestuse keskpunktiga (ingl *vowel target*). Suuresti valitakse vokaali keskpunkt formantide mõõtmiseks seepärast, et naaberhäälikute kontekstiline mõju on siis mõõdetavale vokaalile kõige väiksem ning selles punktis on vokaal kõige stabiilsem, s.t formantväärtused muutuvad kõige vähem.

Vokaalide redutseerumist ja artikulatsiooni täpsust mõõdeti eukleidilise kaugusega⁴, et näha, kas vokaalid on võrreldes neutraalse kõnega tsentraliseerunud või mitte ning kas vokaaliruum on võrreldes neutraalse kõnega emotsiooniti suurenenud või vähenenud. (Vt Harrington 2010: 191–195) Et eukleidilist kaugust mõõta, määrati esmalt kõneleja vokaaliruumi keskpunkti (ingl *centroid*). Selleks arvutati välja neutraalne häälik *x*, võttes arvesse kõigi korpuses leiduvate ja uurimismaterjali tingimustele vastavate sõnaalguliste struktuuride VC, CVC või CCVC lühikeste vokaalide keskmised.

Akustiliste parameetrite mõõtmisel kasutati programmi Praat⁵ ning kõne andmebaaside süsteemi EMU⁶ abi. Formantväärtuste automaatse arvutamise vead korrigeeriti käsitsi ning pärast seda analüüsiti mõõtmistulemusi statistikakeskkonnas R⁷. R-i abil viidi läbi ANOVA-analüüs⁸.

Klaus R. Scherer (1986) on välja töötanud emotsiooniakustika valdkonnas mudeli, mis ennustab emotsiooni mõju häälelisele väljendusele. Schereri *component process model* e CPM võtab arvesse psühholoogilisi ja füsioloogilisi mõjusid, mis kaasnevad emotsiooni väljendamisega, ning näitab, et on olemas emotsioonispetsiifilised akustilised muustrid. Scherer kirjeldab emotsiooni kui kohanemismuutuste (ingl *adaptive changes*) sarja, mis on vastastikusel seoses. Emotsiooni tekkimisest impulsi saanud närvisüsteem mõjutab nii kõneleja hingamist kui ka tema kõneorganite lihasinget ning see toob kaasa erinevusi kõnesignaali akustikas. Nt midagi väga ebameeldivat avaldub tihti neelu ja kõri pitsituses, mille tagajärjel läheb kõnetrakt pingesse ning väljatulev hääli on kõrgema sagedusega (Thompson, Balkwill 2006).

CPM-mudel ennustab muu hulgas, kuidas muutuvad emotsioonide mõjul vokaalide F1 ja F2 väärtused võrreldes neutraalse kõnega. Võrdlus antud töö uurimistulemuste ja CPM-mudeli ning varasemate uurimuste vahel esitatakse 6. peatükis.

4. Emotsioonidega kaasnevad muutused formantväärtustes

Senised uurimused emotsioonide formantväärtuste kohta pole andnud päris ühesid tulemusi. Põhjus võib peituda erinevas uurimismaterjalis (näideldud *vs.* esile kutsutud emotsiooniga *vs.* spontaanne kõne; heli *vs.* heli + pilt), erinevas uuritavas

⁴ Eukleidilise kauguse valem: $d(x, V_{em}) = \sqrt{(F1x - F1V_{em})^2 + (F2x - F2V_{em})^2}$, kus *x* on neutraalne vokaal ja *V_{em}* vaadeldava emotsiooni vokaali keskmine.

⁵ <http://www.praat.org/> (28.09.2011).

⁶ <http://emu.sourceforge.net/> (28.09.2011).

⁷ <http://www.r-project.org/> (28.09.2011).

⁸ Tabeli genereerimiseks kasutati R'i skripti (vt http://peeter.eki.ee:5000/anova_au.R), lähteandmeid vt http://peeter.eki.ee:5000/vokaalid_au.csv (28.09.2011).

keeles ja kultuuris ning erinevas emotsioonidele antavas sisus (nt viha all võib mõista nii raevu kui ka lihtsalt ärritust, rõõmu all rahulolu või vaimustust, kurbuse all pettumust või muret). (Burkhardt jt 2006, McIntyre, Roland 2006, Wilting jt 2006, Douglas-Cowie jt 2003, Scherer 2003) Samuti erineb viis, kuidas emotsioone kirjeldatakse. Domineerivad kaks lähenemist – kategoriaalne (ingl *category approach*) ja dimensiooniline (ingl *dimensional approach*). Esimene lähenemine määratleb emotsioonid kategooriatena, nt kurbus, viha ja rõõm. Teine lähenemine kasutab emotsioonide kirjeldamiseks dimensioone. Kaks enam kasutatavat dimensiooni on valents (positiivne vs. negatiivne) (ingl *valence*) ja virgumine (madal vs. kõrge) (ingl *arousal*). (McIntyre, Roland 2006, Cowie, Cornelius 2003) Kui panna emotsioonid kahemõõtmelisele teljestikule, siis on rõõm kõrge virgumise ja positiivse valentsiga emotsioon, viha kõrge virgumise ja negatiivse valentsiga emotsioon ning kurbus madala virgumise ja negatiivse valentsiga emotsioon.

Varasemad uurimused prantsus- (Goudbeek jt 2009), inglise- (Murray, Arnott 2008, Tartter 1980), saksa- (Kienast, Sendlmeier 2000, Paeschke jt 1999) ja jaapanikeelse (Iida jt 2003) emotsionaalse kõne kohta on näidanud, et esinevad mingid üldised tendentsid teatud emotsioonide ja nende akustiliste korrelaatide vahel.

4.1. Viha ja selle võrdlus teiste emotsioonidega

Võrreldes teiste põhiemotsioonide ja neutraalse kõnega on viha puhul täheldatud vokaalide kõrgemat F1 sagedust (Goudbeek jt 2009, Murray, Arnott 2008, Juslin, Laukka 2003) ning vokaalil *a* oluliselt madalamat F2 sagedust (Goudbeek jt 2009). Samuti on viha-emotsiooniga vokaalide juures märgitud nende kõrgemat artikulatsiooni täpsust (Murray, Arnott 2008, Juslin, Laukka 2003, Kienast, Sendlmeier 2000, Paeschke jt 1999).

Viha kohta on saadud ka teistsuguseid tulemusi. Iida jt (2003) uurimusest selgus, et kõrgema helikõrguse ja kiirema tempoga emotsioonidel (nagu viha) on vokaaliruum kitsenenud ning vokaalid rohkem redutseerunud.

4.2. Rõõm ja selle võrdlus teiste emotsioonidega

Rõõmu-lausungite vokaalidel on võrreldes teiste põhiemotsioonide ja neutraalse kõnega täheldatud F1 ja F2 väärtuste tõusu, kusjuures F1 on enamikul vokaalidel rohkem mõjutatud kui F2 (Juslin, Laukka 2003, Kienast, Sendlmeier 2000, Tartter 1980).

Martijn Goudbeek kolleegidega (2009) märkis vokaalide *a*, *i* ja *u* F1 keskmise tõusu kõrge virgumisega emotsioonide (nt rõõm) puhul, kuid võrreldes madala virgumisega (nt kurbus) emotsioonidega oli neil vokaalil *a* madalam F2 keskmine. Positiivse emotsiooniga (nt rõõm) vokaalidel oli F2 keskmine kõrgem kui negatiivse emotsiooniga (nt viha) vokaalidel.

Rõõmu-lausungite vokaalide artikulatsiooni täpsuse kohta on saadud kaheksuguseid andmeid – artikulatsiooni täpsus kas tõuseb (nt Murray, Arnott 2008, Juslin, Laukka 2003, Johnstone, Scherer 2000) või on sarnane neutraalse kõne artikulatsiooni täpsusega (Kienast, Sendlmeier 2000).

4.3. Kurbus ja selle võrdlus teiste emotsioonidega

Kurbuse-lausungite vokaalidele on võrreldes teiste põhiemotsioonide ja neutraalse kõne vokaalidega omane F1 keskmise langus (Juslin, Laukka 2003). Märgitud on ka seda, et negatiivse valentsiga emotsioonide (nt kurbus) vokaalidel on F2 keskmine madalam kui positiivsete emotsioonide (nt rõõm) vokaalidel. (Goudbeek jt 2009)

Kurbuse-lausungite vokaalidele on omane ka madal artikulatsiooni täpsus (Murray, Arnott 2008, Juslin, Laukka 2003, Kienast, Sendlmeier 2000).

Emotsioonide tekitatud muutusi heliliste häälikute formantväärtustes on Juslin ja Laukka (2003) selgitanud sellega, et viha ja rõõm on kõrge virgumisega emotsioonid, mis panevad kõneleja rääkimisel rohkem pingutama (ta püüab selgemalt artikuleerida). Lihaspinge tekitab kõris aga pitsituse. Pingutamine ja pitsitus viivad kõnetrakti lühenemiseni ning toovad kaasa kõrgema F1 ja täpsema artikulatsiooni. Rõõmu puhul muudab kõnetrakti lühemaks ka naeratamine (Tartter 1980). Madala emotsionaalse virgumisega kurbuse puhul pole aga kõris sellist lihaspinget ja kõneleja räägib lõdvestunumalt. See tingib madalama F1 ja ebatäpsema artikulatsiooni.

5. Uurimistulemused: vokaalide *a*, *i*, *u* esimese ja teise formandi väärtused emotsiooniti

Et saada teada, kas eestikeelse etteloetud emotsionaalse kõne lühikeste vokaalide F1 ja F2 väärtused on mõjutatud oluliselt lausungi emotsioonist võrreldes neutraalse kõnega ning kas emotsionaalse kõne vokaalide formantväärtused erinevad üksteisest oluliselt, mõõdeti vokaalide *a*, *i*, *u* F1 ja F2 väärtused ning arvutati emotsiooniti nende keskmised (vt tabel 2). Siin ja edaspidi on väärtused esitatud nii hertsides kui ka barkides (psühhoakustiline skaala).

Tabel 2. Vokaalide *a*, *i*, *u* esimese ja teise formandi keskmised emotsiooniti

Vokaal	Emotsioon	F1 (Hz)	F1 (Bark)	F2 (Hz)	F2 (Bark)
<i>a</i>	viha	615	5,9	1412	10,7
	rõõm	644	6,1	1440	10,8
	kurbus	558	5,4	1460	10,9
	neutraalne	635	6	1448	10,9
<i>i</i>	viha	386	3,9	2652	14,9
	rõõm	382	3,8	2658	14,9
	kurbus	359	3,6	2642	14,9
	neutraalne	403	4	2660	14,9
<i>u</i>	viha	427	4,3	1298	10,1
	rõõm	418	4,2	1272	10
	kurbus	410	4,1	1334	10,3
	neutraalne	430	4,3	1274	10

Tabelist 2 selgub, et erinevused emotsionaalse kõne ja neutraalse kõne vokaalide F1 keskmiste vahel on väikesed. Kõige enam eristub neutraalsest kõnest kurbuse-emotsioon (vokaal *a* 635 Hz / 6 Bark vs. 558 Hz / 5,4 Bark ja vokaal *i* 403 Hz / 4 Bark vs. 359 Hz / 3,6 Bark). Emotsiooniti erinevad F1 keskmised kõigil kolmel vokaalil, kuid ka siin ei ole erinevused alati suured. Kõige rohkem erineb teistest emotsioonidest kurbus vokaalide *a* ja *i* puhul. Vokaali *a* juures viha vs. kurbus (615 Hz / 5,9 Bark vs. 558 Hz / 5,4 Bark) ja rõõm vs. kurbus (615 Hz / 5,9 Bark vs. 558 Hz / 5,4 Bark) ning vokaal *i* juures viha vs. kurbus (386 Hz / 3,9 Bark vs. 359 Hz / 3,6 Bark).

Erinevused emotsionaalse kõne ja neutraalse kõne vokaalide F2 keskmistes on samuti väikesed. Kõige enam eristub neutraalsest kõnest kurbuse-emotsioon vokaali *u* juures (1274 Hz / 10 Bark vs. 1334 Hz / 10,3 Bark). Emotsiooniti erinevad F2 keskmised vokaalide *a* ja *u* puhul. Vokaali *a* juures erineb kõige rohkem viha kurbusest (1412 Hz / 10,7 Bark vs. 1460 Hz / 10,9 Bark) ning vokaali *u* juures rõõm kurbusest (1272 Hz / 10 Bark vs. 1334 Hz / 10,3 Bark).

Et saada teada, kas emotsioonirühmade erinevused on statistiliselt olulised, tehti formantväärtustele ANOVA-analüüs. Tulemused on esitatud tabelites 3 ja 4.

Tabel 3. ANOVA-analüüsi tulemused emotsionaalsete lausete lühikeste vokaalide *a*, *i*, *u* F1 kohta. (Statistiliselt oluline erinevus on märgitud halli taustaga)

Vokaal	Emotsioonipaar	DF	F value	Pr(>F) Hertzi skaala	Pr(>F) Barki skaala
<i>a</i>	viha vs. rõõm	672	6,43	0,0115	0,03
	viha vs. neutraalne	712	4,01	0,0457	0,04
	viha vs. kurbus	637	20,17	0,0001	0,01
	rõõm vs. neutraalne	596	0,65	0,4211	0,69
	rõõm vs. kurbus	521	33,04	0,0001	0,01
	neutraalne vs. kurbus	561	35,42	0,0001	0,01
<i>i</i>	viha vs. rõõm	364	0,45	0,5010	0,44
	viha vs. neutraalne	363	5,70	0,0174	0,02
	viha vs. kurbus	372	14,77	0,0001	0,01
	rõõm vs. neutraalne	289	6,48	0,0115	0,01
	rõõm vs. kurbus	298	6,81	0,0095	0,01
	neutraalne vs. kurbus	297	27,41	0,0001	0,01
<i>u</i>	viha vs. rõõm	269	0,45	0,5013	0,39
	viha vs. neutraalne	243	0,11	0,7380	0,78
	viha vs. kurbus	251	2,11	0,1478	0,14
	rõõm vs. neutraalne	238	0,78	0,3778	0,31
	rõõm vs. kurbus	246	0,38	0,5384	0,60
	neutraalne vs. kurbus	220	2,59	0,1089	0,11

Tabelist 3 nähtub, et emotsioonipaariti on statistiliselt olulised erinevused ($p < 0,05$) vokaalide F1 keskmistes nii *a* kui ka *i* puhul (v.a vokaali *i* viha vs. rõõm). Vokaali *u* puhul pole erinevused F1 keskmistes emotsiooniti olulised.

Kui võrrelda emotsioone neutraalse kõnega, erinevad sellest oluliselt vokaali *a* puhul viha- ja kurbuse-emotsioon ning vokaali *i* juures kõik kolm emotsiooni.

Tabel 4. ANOVA-analüüsi tulemused emotsionaalsete lausete lühikeste vokaalide *a*, *i*, *u* F2 kohta. (Statistiliselt oluline erinevus on märgitud halli taustaga)

Vokaal	Emotsioonipaar	DF	F value	Pr(>F) Hertzi skaala	Pr(>F) Barki skaala
<i>a</i>	viha vs. rõõm	672	1,77	0,1839	0,28
	viha vs. neutraalne	712	3,69	0,0553	0,05
	viha vs. kurbus	637	4,42	0,0358	0,07
	rõõm vs. neutraalne	596	0,13	0,7195	0,49
	rõõm vs. kurbus	521	0,55	0,4600	0,50
	neutraalne vs. kurbus	561	0,25	0,6176	0,93
<i>i</i>	viha vs. rõõm	364	0,07	0,7926	0,84
	viha vs. neutraalne	363	0,19	0,6667	0,63
	viha vs. kurbus	372	0,19	0,6610	0,40
	rõõm vs. neutraalne	289	0,02	0,8948	0,81
	rõõm vs. kurbus	298	0,31	0,5753	0,40
	neutraalne vs. kurbus	297	0,48	0,4896	0,30
<i>u</i>	viha vs. rõõm	269	0,71	0,4015	0,38
	viha vs. neutraalne	243	0,51	0,4758	0,40
	viha vs. kurbus	251	1,14	0,2859	0,32
	rõõm vs. neutraalne	238	0,00	0,9623	0,96
	rõõm vs. kurbus	246	3,27	0,0718	0,07
	neutraalne vs. kurbus	220	2,55	0,1118	0,09

Tabelist 4 selgub, et emotsiooniti on statistiliselt oluline erinevus ($p < 0,05$) F2 keskmistes vaid vokaali *a* hertsid puhul, kus F2 eristab viha kurbusest ($p = 0,0358$). Barki skaalal statistiliselt olulist erinevust siin ei ole. Ülejäänud F2 keskmised ei erine oluliselt ühegi uuritud vokaali puhul ei emotsiooniti ega ka võrdluses neutraalse kõnega.

6. Formantanalüüsi tulemuste võrdlus CPM-mudeli ja varasemate uurimustega

CPM-mudel ennustab, et võrreldes neutraalse kõnega tõuseb viha-lausungites heliliste häälikute F1 ja langeb F2. Kui kõrvutada see ennustus antud uurimuse tulemustega, siis selgub, et neutraalse kõne vokaalidest erinesid viha puhul *a* ja *i*, mille F1 keskmised vastupidiselt CPM-i ennustusele langesid. Vokaalide *a*, *i* ja *u* F2 keskmised neutraalse kõne F2-st ei erinenud. Varasemates uurimustes on viha puhul täheldatud nii vokaalide kõrgemat F1 sagedust (Goudbeek jt 2009, Murray,

Arnott 2008, Juslin, Laukka 2003) kui ka vokaali *a* oluliselt madalamat F2 sagedust (Goudbeek jt 2009). Neid tulemusi käesolev uurimus ei kinnitanud.

Rõõmu puhul ennustab CPM-mudel F1 kahanemist. Antud uurimuse tulemused näitavad, et rõõmu-lausungites erineb neutraalse kõne vokaalide F1-st statistiliselt oluliselt vaid *i* F1, mille väärtus tõesti langeb. Teised uurimused on näidanud, et rõõmu vokaalide F1 ja F2 väärtused tõusevad (Juslin, Laukka 2003, Kienast, Sendlmeier 2000, Tartter 1980) ning neil on madala virgumisega emotsioonidest (nt kurbus) vokaali *a* F2 keskmine madalam. Samuti on rõõmu-lausungite vokaalidel F2 keskmine kõrgem kui negatiivse emotsiooniga (viha ja kurbus) lausungite vokaalidel. (Goudbeek jt 2009) Käesoleva uurimuse tulemustes polnud rõõmu puhul F2 väärtustes emotsiooniti ega võrdluses neutraalse kõnega statistiliselt olulisi erinevusi. Seega ka rõõmu puhul ei ühtinud antud uurimuse tulemused varasematega.

Kurbuse kohta ennustab CPM-mudel, et võrreldes neutraalse kõnega heliliste häälikute F1 keskmine tõuseb ning F2 keskmine langeb. Antud uurimuse tulemused näitavad, et kurbuse-emotsiooni puhul erinevad vokaalide F1 keskmised statistiliselt oluliselt neutraalse kõne vokaalidest *a* ja *i*, mille F1 väärtused langevad. Varasemates uurimustes on kurbuse-lausungite vokaale iseloomustatud neutraalsest kõnест madalama F1 keskmisega (Juslin, Laukka 2003). Negatiivse valentsiga kurbuse puhul on märgitud ka vokaalide madalamat F2 keskmist võrreldes positiivsete emotsioonidega (nt rõõm) (Goudbeek jt 2009). Käesolevast uurimusest selgub, et ühegi vokaali F2 väärtused ei erine statistiliselt oluliselt emotsiooniti (v.a vokaalil *a* viha vs. kurbus Hertzi skaalal, vt tabel 4) ega võrreldes neutraalse kõnega.

Tulemustest võib järeldada, et rõõmu ja viha väljendatakse eesti keeles teisiti kui keeltes, mille kohta on seni uurimusi tehtud. Osaline sarnasus on aga kurbuse väljendamises. Samas peab silmas pidama seda, et erinevuste põhjus võib peituda ka erinevas uurimismaterjalis.

7. Artikulatsiooni täpsus emotsiooniti

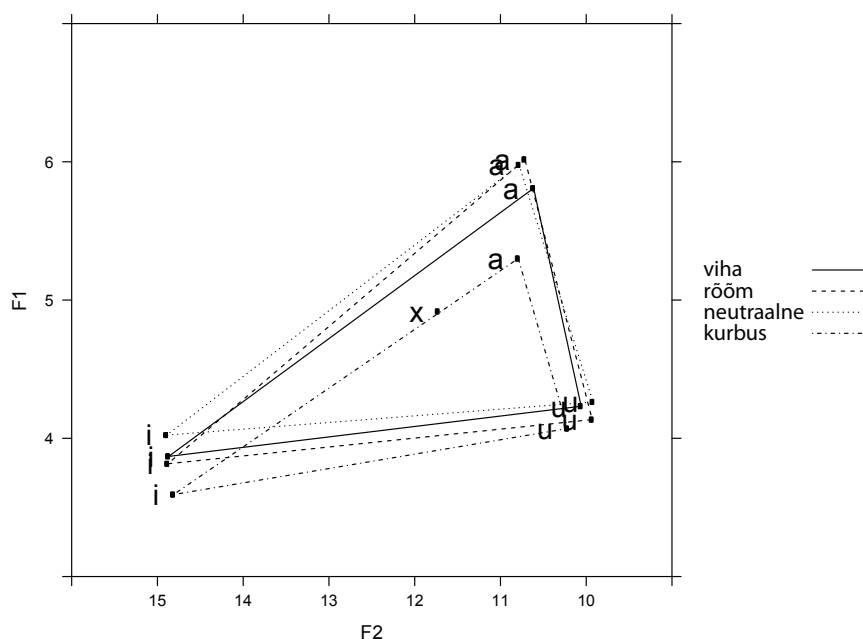
Selleks, et saada vastus küsimusele, kas emotsioonid mõjutavad ettelugeja artikulatsiooni täpsust, arvutati esmalt välja lugeja neutraalne häälik *x* ($F1 = 511 \text{ Hz} / 4,9 \text{ Bark}$ ja $F2 = 1745 \text{ Hz} / 11,7 \text{ Bark}$) ning seejärel eukleidiline kaugus emotsionaalse ja neutraalse kõne vokaalide *a*, *i* ja *u* keskmiste ning neutraalse hääliku *x* vahel kasutades valemit $d(x, V_{em}) = \sqrt{(F1x - F1V_{em})^2 + (F2x - F2V_{em})^2}$, kus *x* on neutraalne vokaal ja V_{em} vaadeldava emotsiooni vokaali keskmine.

Eukleidilise kauguse mõõtmistulemused on esitatud tabelis 5.

Tabel 5. Eukleidiline kaugus vokaalide ja neutraalse hääliku *x* vahel (Hertzi skaala / Barki skaala)

Vokaal	<i>x</i> – viha	<i>x</i> – rõõm	<i>x</i> – kurbus	<i>x</i> – neutraalne
<i>a</i>	348,9/1,4	332,7/1,5	288,8/0,9	321,8/1,4
<i>i</i>	915,6/3,3	922,1/3,4	909,8/3,4	921,3/3,3
<i>u</i>	454,8/1,7	482,1/1,8	423,2/1,6	477,9/1,8

Tabelist 5 selgub, et emotsionaalse kõne vokaalide *a*, *i* ja *u* artikulatsiooni täpsus emotsiooniti ja võrreldes neutraalse kõnega eriti ei erine. Erandiks on kurbuse-emotsiooni *a*, mis on võrreldes neutraalse kõne ja teiste emotsioonidega tsentraalsele häälikule *x* märkimisväärselt lähemal. Kurbuse-emotsiooni *a* artikulatsiooni täpsus on seega muutunud ning vokaal on oma kvaliteedi kaotanud. Vt joonis 1.



Joonis 1. Emotsioonid vokaaliruumis Barki skaalal

Kui võrrelda antud uurimuse artikulatsiooni täpsuse tulemusi varasemate uurimustega, saab öelda, et kurbuse puhul need osaliselt kattuvad – paljud uurijad on kurbuse-lausungites märkinud artikulatsiooni täpsuse langemist (Murray, Arnott 2008, Juslin, Laukka 2003, Kienast, Sendlmeier 2000). Viha kohta on saadud erinevaid tulemusi. Osa uurijaid väidab, et viha puhul artikulatsiooni täpsus tõuseb (Murray, Arnott 2008, Juslin, Laukka 2003, Kienast, Sendlmeier 2000, Paeschke jt 1999, Scherer 1986), teised aga märgivad artikulatsiooni täpsuse langemist (Iida jt 2003). Antud uurimuse tulemuste põhjal võib öelda, et viha puhul artikulatsiooni täpsus võrreldes neutraalse kõnega oluliselt ei muutu.

Rõõmu-lausungite vokaalide artikulatsiooni täpsuse kohta on saadud samuti kahesuguseid andmeid – artikulatsiooni täpsus on kas tõusnud (nt Murray, Arnott 2008, Juslin, Laukka 2003, Johnstone, Scherer 2000) või on olnud sarnane neutraalse kõne artikulatsiooni täpsusega (Kienast, Sendlmeier 2000). Käesoleva uurimuse andmed olulist artikulatsiooni täpsuse tõusu rõõmu-lausungites võrreldes neutraalse kõnega ei näita.

Kurbuse artikulatsiooni täpsuse vähenemise tendents kinnitab seda, et kurbust väljendatakse keeleti sarnaselt. Teiste emotsioonide kohta saadud tulemuste vasturääkivusi võib seletada nii uurimismaterjali erinevuse kui ka nende emotsioonide kultuurispetsiifilisema väljenduslaadiga.

8. Kokkuvõte

Uurimuse eesmärgiks oli teada saada, kas ja kuidas mõjutab lausungi emotsioon eestikeelse etteloetud emotsionaalse kõne lühikesi vokaale ning ettelugeja artikulatsiooni täpsust.

Formantanalüüs näitas, et eesti emotsioonid viha, rõõm ja kurbus mõjutasid vaadeldud lühikestest vokaalidest *a*, *i* ja *u* vokaale *a* ja *i*. Emotsioonide mõju neile vokaalidele väljendus nende F1 väärtuste muutumises. Kõige selgemini eristus teistest emotsioonidest kurbus. Vokaalide *a*, *i* ja *u* F2 väärtusi emotsioonid statistiliselt oluliselt ei mõjutanud.

Artikulatsiooni täpsuse mõõtmised näitasid, et emotsioonidel on sellele väikene mõju olemas. Kuigi viha- ja rõõmu-lausungites jäi artikulatsiooni täpsus võrreldes neutraalse kõnega enam-vähem samaks, siis kurbuse-lausungites see langes.

Viidatud kirjandus

- Altrov, Rene 2008. Eesti emotsionaalse kõne korpus: teoreetilised toetuspunktid. – Keel ja Kirjandus, 4, 261–271.
- Bachorowski, Jo-Anne 1999. Vocal expression and perception of emotion. – Current Directions in Psychological Science, 8 (2), 53–57. <http://dx.doi.org/10.1111/1467-8721.00013>
- Banse, Rainer; Scherer, Klaus R. 1996. Acoustic profiles in vocal emotion expression. – Journal of Personality and Social Psychology, 70 (3), 614–636. <http://dx.doi.org/10.1037/0022-3514.70.3.614>
- Burkhardt, Felix; Audibert, Nicolas; Malatesta, Lori; Türk, Oytun; Arslan, Levent M.; Auberge, Veronique 2006. Emotional prosody – does culture make a difference? – Proceedings of Speech Prosody. Dresden, Germany. May 2–5, 2006.
- Douglas-Cowie, Ellen; Campbell, Nick; Cowie, Roddy; Roach, Peter 2003. Emotional speech: Towards a new generation of databases. – Speech Communication, 40 (1-2), 33–60. [http://dx.doi.org/10.1016/S0167-6393\(02\)00070-5](http://dx.doi.org/10.1016/S0167-6393(02)00070-5)
- Goudbeek, Martijn; Goldman, Jean Philippe; Scherer, Klaus R. 2009. Emotion, dimensions and formant position. – INTERSPEECH-2009, 1575–1578.
- Harrington, Jonathan 2010. – Phonetic Analysis of Speech Corpora. Chichester: Wiley-Blackwell.
- Iida, Akemi; Campbell, Nick; Higuchi, Fumito; Yasumura, Michiaki 2003. A corpus-based speech synthesis system with emotion. – Speech Communication, 40 (1-2), 161–187. [http://dx.doi.org/10.1016/S0167-6393\(02\)00081-X](http://dx.doi.org/10.1016/S0167-6393(02)00081-X)
- Johnstone, Tom; Scherer, Klaus R. 2000. Vocal communication of emotion. – M. Lewis, J. Haviland (Eds.). Handbook of Emotion, 2nd ed. New York: Guilford, 220–235.
- Juslin, Patrik N.; Laukka, Petri 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? – Psychological Bulletin, 129 (5), 770–814. <http://dx.doi.org/10.1037/0033-2909.129.5.770>
- Kienast, Miriam; Sendlmeier, Walter F. 2000. Acoustical analysis of spectral and temporal changes in emotional speech. – R. Cowie, E. Douglas-Cowie, M. Schroeder (Eds.). Speech and Emotion: Proceedings of the ISCA workshop. Newcastle, Co. Down, 92–97.
- McIntyre, Gordon; Roland, Göcke 2006. Researching emotions in speech. – Proceedings of the 11th Australian International Conference on Speech Sciences & Technology, 264–269.
- Moon, Seung-Jae; Lindblom, Björn 1994. Interaction between duration, context, and speaking style in English stressed vowels. – Journal of the Acoustical Society of America, 96 (1), 40–55. <http://dx.doi.org/10.1121/1.410492>

- Murray, Iain R.; Arnott, John L. 2008. Applying an analysis of acted vocal emotions to improve the simulation of synthetic speech. – *Computer Speech and Language*, 22 (2), 107–129. <http://dx.doi.org/10.1016/j.csl.2007.06.001>
- Paeschke, Astrid; Kienast, Miriam; Sendlmeier, Walter F. 1999. FO-contours in emotional speech. – *Proc. ICPhS 99, San Francisco, Vol. 2*, 929–933.
- Scherer, Klaus R. 1986. Vocal affect expression: A review and a model for future research. – *Psychological Bulletin*, 99 (2), 143–165. <http://dx.doi.org/10.1037/0033-2909.99.2.143>
- Scherer, Klaus R. 2003. Vocal communication of emotion: A review of research paradigms. – *Speech Communication*, 40 (1-2), 227–256. [http://dx.doi.org/10.1016/S0167-6393\(02\)00084-5](http://dx.doi.org/10.1016/S0167-6393(02)00084-5)
- Tartter, V. C. 1980. Happy talk: Perceptual and acoustic effects of smiling on speech. – *Perception and Psychophysics*, 27 (1), 24–27. <http://dx.doi.org/10.3758/BF03199901>
- Thompson, William F.; Balkwill, L.-L. 2006. Decoding speech prosody in five languages. – *Semiotica*, 158 (1/4), 407–424. <http://dx.doi.org/10.1515/SEM.2006.017>
- Tolkmitt, Frank J.; Scherer, Klaus R. 1986. Effect of experimentally induced stress on vocal parameters. – *Journal of Experimental Psychology: Human Perception and Performance*, 12 (3), 302–313. <http://dx.doi.org/10.1037/0096-1523.12.3.302>
- Ververidis, Dimitrios; Kotropoulos, Constantine 2006. Emotional speech recognition: Resources, features, and methods. – *Speech Communication*, 48 (9), 1162–1181. <http://dx.doi.org/10.1016/j.specom.2006.04.003>
- Wilting, Janneke; Krahmer, Emiel; Swerts, Marc 2006. Real vs. acted emotional speech. – *Proceedings of Interspeech 2006 ICSLP, Pittsburgh, PA, USA*, 805–808.

Võrgumaterjalid

- ANOVA-analüüsi lähteandmed. http://peeter.eki.ee:5000/vokaalid_aiu.csv (28.09.2011).
- ANOVA-analüüsi tabeli genereerimiseks kasutatud statistikakeskkonna R skript. http://peeter.eki.ee:5000/anova_aiu.R (28.09.2011).
- Boersma, Paul; Weenink, David 2011. Praat: doing phonetics by computer, Version 5.2.35 [Computer program]. <http://www.praat.org/> (28.09.2011).
- Eesti emotsionaalse kõne korpus. <http://peeter.eki.ee:5000/> (28.09.2011).
- Kõne andmebaaside süsteem EMU. <http://emu.sourceforge.net/> (28.09.2011).
- Statistikakeskkond R. <http://www.r-project.org/> (28.09.2011).

Kairi Tamuri (Eesti Keele Instituut) uurimisvaldkond on emotsionaalse kõne akustika.
kairi.tamuri@eki.ee

DO FORMANTS SPEAK OF EMOTIONS?

Kairi Tamuri

Institute of the Estonian Language

Formants belong to the acoustic parameters addressed by the research of emotional speech. The aim of the present study was to find out how, if at all, the utterance emotion of Estonian read emotional speech may affect short vowels and articulatory precision.

The acoustic basis of the study was the Estonian Emotional Speech Corpus (EESC) of the Institute of the Estonian Language. The Corpus contains non-acted speech (journalistic passages) read by a female voice, which have been segmented into sentences, words and speech sounds. The emotional colouring (anger, joy, sadness) or neutrality of the Corpus sentences has been ascertained by perception tests. The research material consisted of such Corpus sentences whose emotion or neutrality had been confirmed by more than half of the listeners. Three short vowels, *a*, *i* and *u*, were picked from those sentences and subjected to formant analysis. Those three are frequent Estonian vowels defining a triangular vowel space (*a* being a low back vowel, *i* a high front vowel and *u* a high back vowel) enabling a fine description of a speaker's articulatory precision.

To solve the research question, measurements were conducted on the first and second formants (F1 and F2) of the short stressed vowels *a*, *i*, *u* as segmented from read Estonian emotional and neutral speech. The results showed that when measured pairwise (partners including neutral sentences) significant differences are revealed between the F1 means of *a* and *i*. F2, however, plays no significant role in emotion differentiation. Vowel reduction and articulatory precision were investigated by using Euclidian distance. According to the results, in the case of angry and joyous utterances articulatory precision was more or less the same as in neutral ones, while a fall signalled sad utterances.

Keywords: emotional speech, acoustics, formant analysis, articulatory precision, Estonian