

# EESTI EMOTSIONAALSE KÕNE KORPUS: TEOREETILISED TOETUSPUNKTID\*

RENE ALTROV

Riikliku programmi „Eesti keele keeletehnoloogiline tugi” raames hakati Eesti Keele Instituudis 2006. aastal looma korpuspõhise kõnesünteesi jaoks eesti emotsionaalse kõne korpus.<sup>1</sup> Peale selle peaks korpus avama võimaluse kõnes avalduvaid emotsioone uurida. Et eesti süntesaator on tekst–kõne-süntesaator ning mõeldud peamiselt tekstide ettelugemiseks (vt Mihkla, Meister, Eek 2000: 309–320), siis kogutakse ka emotsionaalse kõne korpusesse loetud tekste. Eesmärk on saada korpusesse tajukatsetega kontrollitud viha, kurbuse ja rõõmu emotsioone väljendavaid lauseid ning lisaks neutraalset kõnet. Artiklis antakse ülevaade teoreetilistest lähtekohtadest, mida korpuse loomisel arvestatakse. Samuti kirjeldatakse eesti emotsionaalse kõne korpuse loomise etappe: emotsioonide, lugemismaterjali ja lugejate valikut ning emotsioonide tuvastamise tajukatseid.

## Varasemate uurimuste analüüs

**Emotsioonide avaldumine kõnes ja nende seos kultuuriga.** Kõnes väljenduvaid emotsioonide akustilisi tunnuseid on uuritud pikka aega. Psühhiaatria ja elektroakustika valdkonnas algasid empiirilised uuringud juba XX sajandi esimesel poolel (Skinner 1935: 81–137). 1960. aastatel oli huvi emotsioonide vastu eriti suur ja nüüdseks on põhiemotsioonide akustilised parameetrid suhteliselt hästi teada. Kuigi emotsioonid võivad avalduda kõnes semantilisel tasandil, kannab neid suuresti prosoodia. Paljude uurimuste järgi on emotsioonide tajumise tähtsaim parameeter helikõrgus. Teised emotsiooni tajumist mõjutavad tegurid on energia, kestus ja rääkimiskiirus (Murray, Arnott 1993: 1097–1108; Scherer 1989: 165–197; ten Bosch 2003: 213–225). Üldjoontes võib öelda, et emotsionaalsed efektid esinevad neljal kujul: energia, rääkimiskiiruse, helikõrguse ja hääle kvaliteedi muutumises (Toivanen, Väyrynen, Seppänen 2004: 383–412). Viie emotsiooni – viha, hirmu, kurbuse, murelikkuse ja rõõmu – akustiliste ja artikulaatorsete parameetrite seoste uurimine on näidanud, et paljudel juhtudel saab neid emotsioone eristada rääkimiskiiruse, lausungi kestuse ja artikulatsiooni täpsuse põhjal. Näiteks korreleerub kurbus aeglase kõnetempoga, hirmu korral on kõne keskmisest kiirem. Murelikkust väljendavates lausetes on eranditult keskmisest lühemad helitud klusiilid (Kienast, Paeschke, Sendlmeier 1999: 117–120).

\* Artikkel on saanud teoks tänu riiklikule programmile „Eesti keele keeletehnoloogiline tugi”.

<sup>1</sup> Projektis osalevad Rene Altrov, Hille Pajupuu ja Kairi Tamuri.

Emotsioonide väljendamine ja interpreteerimine on seotud kultuuriga. Kultuurierinevused ilmnevad isegi kõnetunnustes: erinevates kultuurides kajastuvad emotsioonid kõnes erinevalt (vt Kovescs 2000; Markus, Kitayama 2001: 119–137; Wierzbicka 1999). Selle tõttu on emotsioone pelgalt akustiliste tunnuste põhjal võimalik kindlamalt ära tunda vaid kultuurisiseselt. Keerukas on emotsiooni määratleda ka siis, kui seda ei edastata kuulaja emakeeles. Sellisel juhul sõltub tuvastamise täpsus konkreetsest emotsioonist: viha, hirmu, kurbust ja üllatust tuvastatakse paremini kui näiteks häbelikkust, domineerivust, rõõmu ja vastikust (Abelin, Allwood 2000: 110–113; Shigeno 1998: 281–284). Ka teise keele omandamist uurides on leitud emotsioonide ja kultuuri seoseid. Kui inimene omandab teise keele pärast teismeiga, siis erinevad need kaks keelt emotsionaalselt väljenduselt: esmalt omandatud keeles väljendatakse personaalset lähedust, teisena omandatud keeles distantsi ja eraldatust. Kakskeelsuse puhul on täheldatud emotsionaalse sõnavara keele- ja kultuurispetsiifilisust (vt Dewaele, Pavlenko 2002: 263–322; Pavlenko 2002: 45–78.).

Et kultuuri ja keele mõju emotsioonide tajumisele on oluline, siis ei saa tugineda teistes kultuurides teiste keelte põhjal tehtud emotsiooniuurimuste tulemustele: need saavad olla ainult suunda andvad.

**Emotsionaalse kõne andmekogud.** Usaldusväärsetel andmetel põhinevaid ja avalikult kasutatavaid emotsionaalse kõne andmebaase on veel väga vähe. Esimese suurema emotsionaalse kõne andmekogu *The Reading-Leeds* loomist alustati 1994. aastal (Greasley, Setter, Waterman, Sherrard, Roach, Arnfield, Horton 1995: 242–245).

2003. aastal avaldatud emotsioonide andmebaaside ülevaatest selgub, et enamasti on tegemist vaid väiksemahuliste andmestikega, mis on koondatud mingi üksikjuhu uurimiseks ja mis pole laialdaselt kasutatavad. Olemasolevate kogude probleemid on põhiliselt seotud andmete vähesuse, valiidsuse ja üldistatavusega (Douglas-Cowie, Campbell, Cowie, Roach 2003: 33–60).

Üldjuhul puutuvad uurijad kvaliteetse emotsioonide andmebaasi loomisel kokku nelja põhiprobleemiga: esimene neist puudutab andmebaasi ulatust ja käsitusala (kui palju võiks olla andmebaasis lugejaid/kõnelejaid ja emotsioonide tuvastajaid, tajukatsetes osalejaid, milliseid emotsioone andmebaas peaks hõlmama jms); teiseks tuleb otsustada, milline peab olema andmebaasi materjal (kas loomulik kõne, professionaalsete näitlejate või diktorite loetud tekstid või saadakse andmed mõnest muust juba olemasolevast allikast, näiteks telefonikõnedest); kolmas probleem on seotud esitatavate emotsioonide kontekstiga (milline on emotsiooniliste episoodide ajaline kestus ja emotsiooni edastamise viis, nt verbaalne või näoilmline vms); ja viimaks tuleb kindlaks määrata, millised deskriptorid lisatakse emotsionaalsele kõnele andmebaasis (Douglas-Cowie, Campbell, Cowie, Roach 2003: 33–60).

Kasutatavatest andmebaasidest sõltub otseselt kõnesünteesi kvaliteet. Louis ten Boschi sõnul saab tekst–kõne–sünteesi kvaliteeti (väärtust) hinnata lähtuvalt kahest tegurist: kõne arusaadavus ja kõne loomulikkus. Kui sünteesikõne arusaadavus on viimasel ajal oluliselt paranenud, siis vähem on olnud edu sünteesitava kõne loomulikkustamisel ning tundub, et selleni ei jõuta veel

nii pea (vt ten Bosch 2003: 213–225). Emotsionaalse kõne klassifitseerimise peamine probleem on realistlike andmete puudulikkus, näiteks tõelises situatsioonis vihaste inimeste korral (Huber, Batliner, Buckow, Noth, Warnke, Niemann 2000: 665–668).

Enamasti on emotsionaalse kõne sünteesi jaoks loodud andmebaasid üles ehitatud lausetele, mida loevad professionaalsed näitlejad ja teadustajad, kel palutakse ühtesid ja samu neutraalseid lauseid lugedes simuleerida erinevaid emotsioone (vt nt Abelin, Allwood 2000: 110–113; Zhao, Lu, Jiang, Wu 2000: 961–964). Lihtsaim viis koguda emotsionaalset kõnet ongi kasutada selleks näitlejaid. Head näitlejad suudavad esitada kõnet, mida kuulajad peavad usutavaks. On aga põhjust arvata, et näitlejate kõne ja loomuliku kõne vahel on süstemaatilisi erinevusi. Professionaalsete näitlejate või teadustajate tekitatud emotsioonid ei pruugi kajastada tegelikke emotsioone, pigem kipuvad näitlejad „üle mängima” ja liialdama, mistõttu emotsioonide väljenduslaad on ebaloomulik ja erineb ehtsast emotsioonist (Campbell 2000: 34–39; Douglas-Cowie, Campbell, Cowie, Roach 2003: 33–60; Scherer 2003: 227–256). Sünteeskõne loomulikustamiseks ei ole näitlejate emotsiooni-karakteristikute kasutamine ilmselt parim lahendus.

Uurimused näitavad, et usaldusväärseid emotsioonile iseloomulikke akustilisi tunnuseid võib leida ka loomulikus (s.o näitlemata) kõnes. Veelgi enam, ka loomulikus kõnes on emotsioonid kuulaja jaoks äratuntavad (Iida, Campbell, Higuchi, Yasumura 2003: 161–187). Loomulikus kõnes väljendatakse emotsioone mõõdukalt ja harva leiab kõnest täismahulisi puhtaid emotsioone. Näiteks hirmu, kurbuse ja viha kõrval on loomulikus kõnes olulised ka leebemad, n-ö emotsioonidega seotud seisundid, nagu sõbralikkus, huvitatus ja nauding, mida Klaus R. Scherer nimetab inimestevahelisteks hoiakuteks (Cowie, Cornelius 2003: 5–32; Scherer 2000). Loomuliku kõne andmebaasidest sisaldab täpselt määratlemata emotsioone näiteks Belfasti andmebaas (The Belfast Naturalistic Emotional Database; Douglas-Cowie, Cowie, Schroeder 2000: 39–44).

Siiski on valdav osa uurijaid keskendunud põhiemotsioonide väikesele hulgale. Need on emotsioonid, mille olemasolu on seotud ümbritsevas keskkonnas ellujäämisega. Teooria järgi on põhiemotsioonidel kindel sisu ja ülemaailmselt sarnane ja hästi äratuntav näoväljendus (vt Ekman 1972: 207–283; Scherer 1986: 143–165; Nölvak, Valk 2003: 169–191). Mõnikord on väidetud, et põhiemotsioonidest tulenevad ka kõik teised emotsioonid. Välja on pakutud väga erinev arv põhiemotsioonide kategooriaid: Paul Ekman kirjutab 15 emotsionaalsest seisundist, Richard Lazarus 17-st ja Ross Buck isegi 21-st. Üldiselt tunnistatakse, et põhiemotsioone on kuus: rõõm, kurbus, hirm, viha, vastikus ja üllatus. Hiljem on neile lisatud veel üks: põlgus (Ekman 1999: 301–320; Lazarus 1999; Buck 1999: 301–336).

Andmebaasiti on emotsioonide hulk väga erinev. Kõnesünteesiga seotud andmebaasides on enamasti põhiemotsioonid hirm, viha, kurbus ja rõõm, kuid sageli ka üllatus, vastikus ja „neutraalne” emotsioon.<sup>2</sup>

**Emotsioonide kõnes äratundmine.** Emotsioonid tuntakse valdavalt ära

---

<sup>2</sup> Emotsionaalse kõne süntesaatoritest ja neis kasutatavatest emotsioonidest leiab ülevaate veebilehelt <http://emosamples.syntheticspeech.de>, kus saab kuulata ka näiteid.

inimest kuulates ja nähes (Shigeno 1998: 281–284). Sünteeskõne puhul – kui sellega ei kaasne visualisatsiooni – tekib enamasti vajadus tunda emotsioon ära häälest. Senistes uurimustes on saadud emotsioonide äratundmise kohta küllaltki erinevaid tulemusi.

Klaus R. Scherer järeldas 1980. aastatel läbiviidud 30 uuringule tuginedes, et kuulaja tuvastab emotsiooni pelgalt hääle põhjal keskmiselt 60% juhtudest. Uurimuses hiljem neutraalse emotsiooni, viha, hirmu, rõõmu, kurbuse, vastikuse ja üllatuse tuvastamist Euroopa, Ameerika Ühendriikide, Aasia eri kultuurides, sai ta äratundmise määraks keskmiselt 66%, kusjuures Euroopas oli see madalam kui mujal: 62%. 2003. aastal esitas K. R. Scherer enda ja oma kolleegide varasemate uurimuste põhjal kokkuvõtte, mille järgi emotsioonide hääle järgi tuvastamine jääb 55–65% vahele (Scherer 1989: 165–197; Scherer, Banse, Walbott 2001: 76–92; Scherer 2003: 227–256).

Inglise, rootsi, hispaania, serbia ja araabia keele baasil emotsioonide tuvastamine (Abelin, Allwood 2000: 110–113; Al-Dakkak, Ghneim, Abou Zliekha, Al-Moubayed 2005; Dellaert, Polzin, Waibel 1996: 786–789; Montero, Gutierrez-Arriola, Palazuelos, Enriquez, Aguilera, Pardo 1998: 923–926; Slobodan, Zorka, Miodrag, Mirjana 2004: 77–81) on aga näidanud kõrgemaid protsente (vt tabelit 1). Paistab silma, et problemaatilisem on vastikuse äratundmine (väga madal tuvastusprotsent), kuid vastikuse kõrval või siis, kui vastikust pole lastudki tuvastada, on raske ära tunda ka rõõmu. Siinjuures on huvitav märkida, et serbia emotsioonide puhul aeti kõige sagedamini omavahel segi rõõmu ja viha ning neutraalset emotsiooni kurbusega. Samuti peeti küllaltki sageli hirmu ja neutraalset kõnet vihaks, kuid mitte kunagi vastupidi (Slobodan, Zorka, Miodrag, Mirjana 2004: 77–81).

Emotsioonide tuvastamise andmete nii suured erinevused võivad viidata emotsioonide kategoriseerimise, andmebaaside ülesehituse, andmete kogumise – näiteks simuleeritud (näideldud) kõne ja reaalsete andmete – ja nende interpreteerimise erinevustele, kuid uurimistulemustele võivad mõju avaldada ka emotsiooni edastaja isik, kultuur, sugu ja situatsioon (nt kas on tegemist monoloogi või dialoogiga) (Douglas-Cowie, Campbell, Cowie, Roach 2003: 33–60; Toivanen, Väyrynen, Seppänen 2004: 383–412).

Tabel 1.

### Emotsioonide tuvastamise protsent erinevates uurimustes

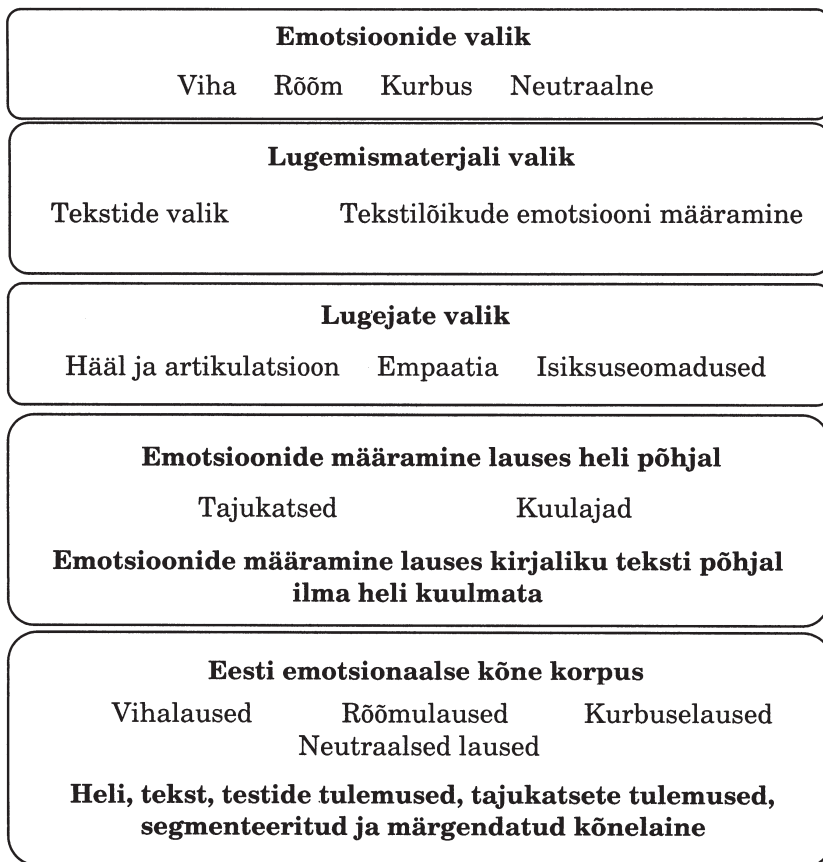
	Viha	Hirm	Rõõm	Kurbus	Vastikus	Üllatus	Neutraalne
Scherer jt 2001 (Euroopa, USA)	77	61	57	71	31	–	74
Dellaert jt 1996 (inglise)	96	64	88	80	–	–	–
Abelin, Allwood 2000 (rootsi)	83	66	92	69	12	83	–
Montero jt 1998 (hispaania)	89	–	75	90	–	–	89
Slobodan jt 2004 (serbia)	96	93	95	96	–	–	95
Al-Dakkak jt 2005 (araabia)	75	80	67	70	–	73	–

### Eesti emotsionaalse kõne korpuse teoreetiline kavand ja teostus

Viimase aja uurimustest on selgelt näha, et tekst-kõne-sünteesis on eelistus antud emotsionaalselt rikkale kõnele (Schröder 2001). Eesti tekst-kõne-sünteesis pole siiani emotsioone arvestanud. See pole ka võimalik olnud, sest

ei olnud emotsionaalse kõne korpust ega vastavaid akustilisi uuringuid. Eesti emotsionaalse kõne korpuse loomine annab võimaluse alustada eesti emotsionaalse kõne uurimist ja kasutada nii korpust kui ka uurimistulemusi kõnesünteesis.

Teoreetilistele teadmistele ja teiste emotsionaalse kõne



korpuste kogemustele tuginedes oleme eesti emotsionaalse kõne korpuse jaoks välja töötanud teoreetilise kavandi ning kontrollinud selle praktilise teostamise võimalikkust. Järgnevalt tutvustan nii kavandit kui ka teostust etappide kaupa (vt joonist 1).<sup>3</sup>

Joonis 1. Emotsionaalse kõne korpuse loomise etapid

**Emotsioonide valik.** Esimene samm korpuse loomisel on valida emotsioonide kategooriad, mida tahetakse hakata uurima ja kasutama. Eesti korpuse otsustasime koguda kurbust, viha ja rõõmu väljendavaid lauseid ja lisaks ka piisava hulga neutraalseid lauseid. Kogumist alustasime *viha*-lausetest, sest Ene Vainiku ja Heili Orava uurimuse järgi on viha eestlaste jaoks kõige tüüpilisem emotsioon (prototüüp), mis võib olla nii negatiivne

<sup>3</sup> Eesti emotsionaalse kõne korpuse kohta üksikasjalikumalt vt Altrov 2007.

(*viha on sotsiaalselt taunitav ja ebameeldiv tundeelamus*) kui ka positiivne (*viha annab jõudu*) (Vainik 2002: 537–553; Vainik, Orav 2005: 257–277). Võrdluseks võib tuua, et teistel rahvastel on emotsiooni prototüübiks näiteks rõõm (belglastel, prantslastel, itaallastel, šveitslastel), õnn (inglastel, kanadalastel) või hirm (hollandlastel). Üksnes soome mehed peavad eestlastega sarnaselt kõige tüüpilisemaks emotsiooniks viha (vt Frijda, Suprapti, Kaori, Reinout 1995: 121–143; van Goozen, Frijda 1993: 89–95; Tuovila 2005).

Neutraalseid lauseid võtsime korpusesse selleks, et tekiks võimalus kasutada korpust ka neutraalse kõne sünteesis (vt Matte 2006: 1–4).

Eesti korpus on kavandatud nii, et seda saaks täiendada teisi emotsioone väljendavate lausetega.

**Lugemismaterjali valik.** Et eesti süntesaator on mõeldud tekstide ettelugemiseks, siis mõtlesime ka korpuse materjali valides eeskätt tekstidele, mida süntesaatoril kõige tõenäolisemalt lugeda tuleb. Nõnda otsustasime koguda korpusesse materjali Eesti ajakirjandusest. Emotsiooni väljendavate üksiklausete asemel otsustasime valida tekstilõike, sest lõigu sõnum aitab teksti lugejal paremini ülal hoida konkreetset emotsionaalset seisundit (vt Iida, Campbell, Higuchi, Yasumura 2003: 161–187). Tekstid oleme valinud põhimõttel, et nad kajastaksid võimalikult laia teemaderingi ega oleks kõnekeelsed (silmas pidades taas eesti süntesaatori põhieesmärki lugeda ette kirjutatud tekste). Et tekstilõigu emotsioon paremini välja tuleks, on neid mõnevõrra kohandatud: lisatud (või tekstis asendatud) konkreetse emotsionaalse seisundi rõhutamiseks sõnu, lauseid ja väljendeid, emotsionaalseid laiendeid ning kirja-vahemärke.

Järgmine samm oli teada saada, kas vajalik emotsioon on tekstis olemas. Lasksime väljavalitud tekstilõike katsealustel (viiel mehel ja viiel naisel) omaette lugeda ja teksti põhjal emotsiooni olemasolu hinnata. Seejärel lindistasime need lõigud, milles katsealused valdavalt teatava kindla emotsiooni olemasolu kinnitasid.

Ideaalset korpuse suurust ei ole niisuguste korpuste loojad ja kasutajad veel määratlenud, kuid seni on osutunud täiesti piisavaks ühe kõneleja ühe emotsiooni kohta umbes 500-sõnalised korpused, mis enamasti on organiseeritud lausetena (vt Abe, Sagisaka, Umeda, Kuwabara 1990; Iida, Campbell, Higuchi, Yasumura 2003: 161–187). Sellest teadmisest lähtudes otsustasime ka Eesti emotsionaalse kõne korpusesse esmalt koguda ühe lugeja iga emotsiooni kohta teksti 500 eri sõna mahus, jättes võimaluse edaspidi emotsiooni väljendavate lausete ja kõnelejate hulka suurendada. Samuti seadsime eesmärgiks koguda korpusesse vähemalt 500 sõna mahus neutraalseid lauseid.

**Lugejate valik.** Emotsioonide väljendamiseks on üldjuhul järgmised viisid: spontaanne, näideldud või tahtlikult esilekutsutud väljendus. Olemasolevatesse emotsioonide andmebaasidesse on tavaliselt valitud kogenud rääkijate kõne (enamasti on kõnelejaks professionaalne näitleja või teadustaja, kes esitab neutraalseid sõnu, lauseid või tekstilõike kurvalt, rõõmsalt, vihaselt jne). On alust arvata, et näitlejate loetud tekstid on liiga intensiivsed ega kajasta emotsiooni autentselt. Et ka loomulikus (näitlemata)

kõnes on emotsioonid kuulaja jaoks piisavalt hästi äratuntavad ja et eesmärgiks on võimalikult loomutruu sünteeskõne, siis pidasime tähtsaks korpusesse saada näitlemata, loomulikke emotsioone (vt ka Campbell 2000: 34–39; Iida, Campbell, Higuchi, Yasumura 2003: 161–187).

Seetõttu otsustasime eesti emotsionaalse kõne korpusesse salvestada mitteprofessionaalsete lugejate tekste, seejuures silmas pidades, et nende artikulatsioon oleks korrektne ning hääli (kõrgus ja tämber) kuulajate jaoks meeldiv.<sup>4</sup> Potentsiaalsete lugejate hääle vastuvõetavust ja meeldivust laskisime proovisalvestuste põhjal kuulajatel hinnata (minimaalselt 10 kuulajat, võrdselt mehi ja naisi), artikulatsiooni korrektsuse üle otsustasid foneetikud.

Testisime ka lugejate empaatiavõimet. Empaatiat on omadus, mille abil suudetakse end teise inimesega samastada: näha ette tema kavatsusi ja käitumist ning kogeda samu emotsioone. Emotsioone kutsuvad esile ka kirjutatud tekstid. Seetõttu tuleks eelistada empaatiavõimega lugejaid, kes suudavad suurema tõenäosusega väljendada loetava tekstiga sama emotsiooni. Meie testisime lugejate empaatiavõimet eesti keelde tõlgitud Baron-Cohen–Wheelwrighti empaatiatestiga (2004: 163–175).

Et emotsioonid on seotud isiksuseomadustega, tuleks uurida ka lugejate isiksuseomaduste ja emotsioonide väljendamise seost. Viiefaktorilise mudeli järgi on kõik isiksuseomadused ehk seadumused mahutatavad viie isiksusedimensiooni raamistikku, kus kaks – neurootilisus (seadumus kogeda negatiivseid emotsioone hirmu, kurbust, süüd, viha jne) ja ekstravertsus (seadumus kogeda positiivseid emotsioone) – võiksid olla ühtlasi ka emotsioonide väljendamise eelduseks. Lugejate vajalike isiksuseomaduste määramiseks on võimalik kasutada isiksuseküsimustiku NEO-PI-R (Costa Jr, McCrae 1995: 21–50; vt ka Allik, Realo, Konstabel 2003: 43–44, 171–191; Möttus, Pullmann, Allik 2006: 49–157) eestikeelset versiooni EPIP-NEO. Korpuse teoreetilise kavandi katsetusest jätsime aga lugejate isiksuseomaduste määramise esi- algu välja, sest see nõuab pikaajalisemat uurimistööd.

Oluline on lugeja juhendamine: lugemissituatsioon peaks sarnanema tavaelus võimalikuga. Näiteks võib lugejat juhendada nii: „Palun loe tekst mulle ette, et ma saaksin seda edasi jutustada” (lugejale ei anta vihjet, et talt oodatakse emotsioonide väljendamist) või „Palun loe nii, et teksti meeoleu tuleks välja” (lugejale antakse vihje, mis võib soodustada emotsionaalsemat lugemist). Kasutusele peaks võtma sellise juhise, mis viib emotsionaalsema, ent mitte näideldud lugemiseni. Tööjuhisekatsetused tuleb teha enne konkreetsete väljavalitud lugejatega salvestuste alustamist. Korralikult läbimõeldud juhise on väga oluline: sellega antakse teada, mis lugejat ees ootab, milline on tema ülesanne, kas teda salvestatakse jms. Arusaamatu tööjuhise võib tekitada ärevust ja mõjutada ülesande täitmist (Bachman 1990: 123–124). Katsetasime erinevaid tööjuhiseid ja nende mõju emotsioonide väljendamisele. Pidama jäime sellisele juhisele, milles palusime lugejal endal otsustada, kuidas teksti lugeda, et selle meeoleu esile tuleks. Vihjeid, millist emotsiooni lugejalt ootame, me ei andnud.

Salvestatud tekstilõigud segmenteerisime lauseteks.

<sup>4</sup> Lugejate hääle meeldivus on oluline seetõttu, et eesti tekst–kõne–süntesaator hakkab kasutama selle lugeja loetud tekstide akustilisi andmeid.

**Tajukatsed ja kuulajad, lugemistest.** Eesti emotsionaalse kõne korpuse lausete emotsioon määratakse kindlaks tajukatsetega, milles kuulatakse kontekstist eraldatud lauseid. Kuulajal tuleb neis hinnata emotsiooni olemasolu (valikud: neutraalne, kurbus, viha, rõõm). Tajukatsete tulemused lisatakse korpuse lausetele ning need näitavad, mitu protsenti kuulajaist määras lause neutraalseks, kurbust, viha või rõõmu väljendavaks. Korpuse kasutaja otsustab ise, millise protsendiga tuvastatud emotsioonilauseid ta oma töös vajab. Sünteeskõne akustiliseks baasiks soovitatavate lausete emotsiooni tuvastusprotsent võiks olla 80 ja rohkem.

Tajukatsete kuulajate valikul lähtusime teadmised, et emotsiooni hääle järgi tuvastamine ilma rääkijat nägemata on kultuurispetsiifiline ja omandatav pikema aja jooksul. Emotsioone suudab vastavas kultuurikontekstis õigesti dekodeerida ainult niisugune kuulaja, kellel on selles kultuuris pikaajaline kogemus (Toivanen, Väyrynen, Seppänen 2004: 383–412). Seega pole otstarbekas valida kuulajateks liiga noori (nt tudengeid). Lisaks peavad kuulajad olema eesti rahvusest ning rääkima emakeelena eesti keelt, et nad oleksid võimelised eesti kultuuri emotsioone kuulamise teel tuvastama. Eesti emotsionaalse kõne korpuse projektis kasutamegi kuulajaid, kes on 30-aastased ja vanemad ning kelle emakeel ja hariduskeel on eesti keel.

Ka kuulajatele võiks teha isiksusetesti EPIP-NEO, et tekiks võimalus uurida isiksuseomaduste ja emotsioonide äratundmise seoseid. Varasemate uurimuste põhjal võib eeldada, et teiste inimeste emotsiooniväljendusi suudavad täpsemalt mõista kõrge avatusega inimesed (vt Valk 1999).

Et meie kõnekorpuse üks eesmärke on olla emotsionaalse kõnesünteesi akustiliseks baasiks, siis peame teadma, milliste lausete emotsioon tuntakse ära ainult helis sisalduvate emotsioonitunnuste põhjal ja milliste lausete emotsioonitajumist võib mõjutada lause tekst (sisu, süntaks, sõnavara jm). Seetõttu tekkis vajadus lasta lisaks tajukatsetele määrata samade lausete emotsioone ka pelgalt teksti põhjal, ilma heli kuulmata. Selles katses ei osalenud tajukatsetes kaasa teinud inimesed.

Kahe testi tulemuste võrdlemisel selgus, et kuulamise põhjal ja ainult lugemise põhjal määratud emotsioonid alati kokku ei lange. Sellest tulenevalt tekkis kaks liigendust:

1. Emotsiooni kannab ainult hää:
  - lugemise põhjal ei osata lause emotsiooni määrata, kuid kuulamise põhjal osatakse;
  - lugemise põhjal määratakse üks emotsioon (nt viha), kuulamise põhjal mõni teine (nt kurbus).
2. Emotsiooni määramist võis tajukatses mõjutada kuulatav tekst:
  - lugemise põhjal määratakse üks emotsioon (nt viha) ja kuulamise põhjal seesama emotsioon (viha).

Esimese liigenduse põhjal saadud heli moodustab sünteeskõne akustilise baasi. Teise liigenduse põhjal saadud lausete emotsioonide tajumisel mängib teatavat rolli tekst ning see avab võimaluse uurida, mis on tekstis sellist, mille põhjal me ühe või teise emotsiooni ära tunneme.

Eesti emotsionaalse kõne korpuses hoitakse heli, teksti, testide tulemusi, segmenteeritud ja märgendatud kõnelainet.



## Kokkuvõte: Eesti emotsionaalse kõne korpuse hetkeseis

Kirjeldatud meetodil, mille aluseks on varasematele uurimistulemustele ja teiste korpuste kogemustele toetuv teoreetiline kavand, on meie korpusesse praeguseks kogutud ja testitud kahe naisehääle 173 lauset: neist 79 on tunnistatud vihalauseks, 3 rõõmlauseks, 20 kurbuselauseks ja 25 neutraalseks lauseks. Kuigi korpus on alles katsetamise ja arendamise järgus, oleme ta teinud avalikult kättesaadavaks aadressil <http://urve.eki.ee:5000/>. Korpusest on võimalik praegu otsida viha, rõõmu, kurbust ja neutraalsust väljendavaid lauseid. Iga lause juures on näidatud emotsiooni (või neutraalsuse) tajuprotsent. Otsides võib ise ette anda protsendi, millest alates selle emotsiooniga tajutud lauseid soovitakse (vaikimisi on selleks protsendiks 51). Päringutes saab teha valiku: 1) otsida ainult neid lauseid, kus emotsiooni kannab vaid heli; 2) otsida ainult neid lauseid, mille emotsiooni tajumist on võinud mõjutada tekst. Laused kuvatakse tekstina ja neid saab kuulata lausele klikates. Ühe naishääle laused on segmenteeritud häälikuteks ja sõnadeks ning esialgselt märgendatud. Uurijatel on võimalik lausete heli, segmenteeringut ja märgendust (Praat TextGrid) endale salvestada [vt ka Altrov, H. Pajupuu (ilmumas)].

## Kirjandus

- Abe, M., Sagisaka, Y., Umeda, T., Kuwabara, H. 1990. ATR Technical Report TR-I-0116 Speech Database User's Manual. Kyoto: ATR Interpreting Telephony Research Laboratories.
- Abelin, A., Allwood, J. 2000. Cross Linguistic Interpretation of Emotional Prosody. – Proceedings ISCA w/s Speech and Emotion. Newcastle, Northern Ireland, lk 110–113.
- Al-Dakkak, O., Ghneim, N., Abou Zliekha, M., Al-Moubayed, S. 2005. Emotion Inclusion in an Arabic Text-to-Speech. – Paper presented at the 13<sup>th</sup> European Signal Processing Conference, 4.–8. 9. 2005. Antalya. <http://www.arehna.di.uoa.gr/Eusipco2005/defevent/papers/cr1886.pdf>.
- Allik, J., Realo, A., Konstabel, K. 2003. Isiksusepsühholoogia. Tartu: Tartu Ülikooli Kirjastus.
- Altrov, R. 2007. Emotsionaalse kõne korpuse loomine eesti keele tekst-kõne sünteesi jaoks. Tekstimaterjali evalvatsioon *viha* näitel. Tartu (magistritöö Tartu Ülikooli Filosoofiateaduskonna eesti ja soome-ugri keeleteaduse osakonnas). [http://dspace.utlib.ee/dspace/bitstream/10062/2739/1/altrov\\_rene.pdf](http://dspace.utlib.ee/dspace/bitstream/10062/2739/1/altrov_rene.pdf).
- Altrov, R., Pajupuu, H. (ilmumas). The Estonian Emotional Speech Corpus. Release 1. – The Third Baltic Conference on Human Language Technologies. Vilnius: Lietuvių kalbos institutas.
- Bachman, L. 1990. Fundamental Considerations in Language Testing. Oxford: Oxford University Press.
- Baron-Cohen, S., Wheelwright, S. 2004. The Empathy Quotient. An Investigation of Adults with Asperger Syndrome or High Functioning Autism

- and Normal Sex Differences. – *Journal of Autism and Developmental Disorders*, Vol. 34 (2), lk 163–175.
- Buck, R. 1999. Biological Affects. A typology. – *Psychological Review*, Vol. 106 (2), lk 301–336.
- Campbell, N. 2000. Databases of Emotional Speech. – *Proceedings ISCA w/s Speech and Emotion*. Newcastle, Northern Ireland, lk 34–39.
- Costa Jr, P. T., McCrae, R. R. 1995. Domains and Facets. Hierarchical Personality Assessment Using the Revised NEO Personality Inventory. – *Journal of Personality Assessment*, Vol. 64 (1), lk 21–50.
- Cowie, R., Cornelius, R. 2003. Describing the Emotional States That Are Expressed in Speech. – *Speech Communication*, Vol. 40 (1–2), lk 5–32.
- Dellaert, F., Polzin, T., Waibel, A. 1996. Recognizing Emotion in Speech. – *Proceedings of the International Congress on Spoken Language Processing*. Philadelphia, lk 786–789.
- Dewaele, J., Pavlenko, A. 2002. Emotion Vocabulary in Interlanguage. – *Language Learning*, Vol. 52 (2), lk 263–322.
- Douglas-Cowie, E., Campbell, N., Cowie, R., Roach, P. 2003. Emotional Speech. Towards a New Generation of Databases. – *Speech Communication*, Vol. 40, lk 33–60.
- Douglas-Cowie, E., Cowie, R., Schroeder, M. 2000. A New Emotion Database. Consideration, Sources and Scope. – *Proceedings ISCA w/s Speech and Emotion*. Newcastle, Northern Ireland, lk 39–44.
- Ekman, P. 1972. Universals and Cultural Differences in Facial Expressions of Emotion. – J. Cole (ed.), *Nebraska Symposium on Motivation 1971*. Lincoln, NE: University of Nebraska Press, lk 207–283.
- Ekman, P. 1999. Basic Emotions. – T. Dalgleish, M. Power (eds.), *Handbook of Cognition and Emotion*. New York: John Wiley, lk 301–320.
- Frijda, N. H., Suprapti, M., Kaori, S., Reinout, W. 1995. Emotions and Emotion Words. – J. A. Russell, J.-M. Fernandez-Dols, A. S. R. Manstead, J. C. Wellenkamp (eds.), *Everyday Concepts of Emotion. An Introduction to the Psychology, Anthropology and Linguistics of Emotion*. Dordrecht: Kluwer Academia, lk 121–143.
- Greasley, P., Setter, J., Waterman, M., Sherrard, C., Roach, P., Arnfield, S., Horton, D. 1995. Representation of Prosodic and Emotional Features in a Spoken Language Database. – *Proceedings of the XIIIth ICPHS*, Vol. 1. Stockholm: Stockholm University, lk 242–245.
- Huber, R., Batliner, A., Buckow, J., Noth, E., Warnke, V., Niemann, H. 2000. Recognition of Emotion in a Realistic Dialogue Scenario. – *ICSLP 2000. Proceedings of the Sixth International Conference on Spoken Language Processing*. Beijing, China, October 16–20. Vol. 1. Beijing, lk 665–668.
- Iida, A., Campbell, N., Higuchi, F., Yasumura, M. 2003. A Corpus-Based Speech Synthesis System with Emotion. – *Speech Communication*, Vol. 40, lk 161–187.
- Kienast, M., Paeschke, A., Sendlmeier, W. 1999. Articulatory Reduction in Emotional Speech. – *Proceedings of Eurospeech*. Budapest, lk 117–120.
- Kovescs, Z. 2000. *Metaphor and Emotion. Language, Culture, and Body in Human Feeling*. Cambridge, MA: Cambridge University Press.
- Lazarus, R. S. 1999. *Stress and Emotion. A New Synthesis*. New York: Springer.

- Markus, H. R., Kitayama, S. 2001. The Cultural Construction of Self and Emotion. Implications for Social Behavior. – W. G. Parrot (ed.), *Emotions in Social Psychology. Essential Readings*. Taylor and Francis: Psychology Press, lk 119–137.
- Matte, A. C. F. 2006. Neutral Speech Corpora – a Test for Neutrality. – *Speech Prosody 2006*, Dresden. *Speech Prosody 3rd International Conference Proceedings*. Vol. 1. Dresden: TUD Press, lk 1–4.
- Mihkla, M., Meister, E., Eek, A. 2000. Eesti keele tekst-kõne süntees: grafeem-foneem teisendus ja prosoodia modelleerimine. – T. Hennoste (toim), *Arvutuslingvistikalt inimesele*. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 1. Tartu: Tartu Ülikool, lk 309–320.
- Montero, J. M., Gutierrez-Arriola, J., Palazuelos, S., Enriquez, E., Aguilera, S., Pardo, J. M. 1998. Emotional Speech Synthesis. From Speech Database to TTS. – *Proceedings of the 5th International Conference on Spoken Language Processing*. Sydney, lk 923–926.
- Möttus, R., Pullmann, H., Allik, J. 2006. Toward More Readable Big Five Personality Inventories. – *European Journal of Psychological Assessment*, Vol. 22 (3), lk 49–157.
- Murray, I. R., Arnott, J. L. 1993. Toward the Simulation of Emotion in Synthetic Speech. A Review of the Literature on Human Vocal Emotion. – *Journal of Acoustic Society of America*, Vol. 93 (2), lk 1097–1108.
- Nõlvak, A., Valk, R. 2003. Isiksus ja emotsioonid. – J. Allik, A. Realo, K. Konstabel (toim), *Isiksusepsühholoogia*. Tartu: Tartu Ülikooli Kirjastus, lk 169–191.
- Pavlenko, A. 2002. Bilingualism and Emotions. – *Multilingua*, Vol. 21 (1), lk 45–78.
- Scherer, K. R. 1986. Vocal Affect Expression. A Review and a Model for Future Research. – *Psychological Bulletin*, Vol. 99, lk 143–165.
- Scherer, K. R. 1989. Vocal Correlates of Emotion. – H. Wagner, A. Masted (eds.), *Handbook of Psychophysiology. Emotion and Social Behavior*. London: Wiley, lk 165–197.
- Scherer, K. R. 2000. Emotion Effects on Voice and Speech. Paradigms and Approaches to Evaluation. Paper presented at the ISCA Workshop on Speech and Emotion, 5.–7. 9. 2000. Newcastle, Northern Ireland.
- Scherer, K. R. 2003. Vocal Communication of Emotion. A review of Research Paradigms. – *Speech Communication*, vol.40, lk 227–256.
- Scherer, K. R., Banse, R., Walcott, H. G. 2001. Emotion Inferences from Vocal Expression Correlate across Languages and Cultures. – *Journal of Cross-Cultural Psychology*, Vol. 32 (1), lk 76–92.
- Schröder, M. 2001. Emotional Speech Synthesis. A Review. – *Proceedings of Eurospeech*. Aalborg, lk 561–564.
- Shigeno, S. 1998. Cultural Similarities and Differences in the Recognition of Audio-Visual Speech Stimuli. – R. Mannell, J. Robert-Ribes (eds.), *Proceedings of the International Conference on Spoken Language Processing ICSLP'98*, Sydney, Australia, Vol. 1. Sydney, lk 281–284.
- Skinner, E. 1935. A Calibrated Recording and Analysis of the Pitch, Force and Quality of Vocal Tones Expressing Happiness and Sadness. – *Speech Monographs*, Vol. 2, lk 81–137.

- Slobodan, T. J., Zorka, K., Miodrag, D., Mirjana, R. 2004. Serbian Emotional Speech Database. Design, Processing and Evaluation. – SPECOM 2004. 9th Conference Speech and Computer, St. Petersburg, Russia, 20.–22. 9. 2004. St. Petersburg, lk 77–81.
- Zhao, L., Lu, W., Jiang, Y., Wu, Z. 2000. A Study on Emotional Feature Recognition in Speech. – ICSLP 2000. Proceedings of the Sixth International Conference on Spoken Language Processing. Beijing, China, October 16–20. Beijing, lk 961–964.
- ten Bosch, L. 2003. Emotion, Speech and the ASR Framework. – Speech Communication, Vol. 40 (1/2), lk 213–225.
- The Belfast Naturalistic Emotional Database. <http://www.idiap.ch/mmm/corpora/emotion-corporus>.
- Toivanen, J., Väyrynen, E., Seppänen, T. 2004. Automatic Discrimination of Emotion from Spoken Finnish. – Language and Speech, Vol. 47 (4), lk 383–412.
- Tuovila, S. 2005. Kun on tunteet. Suomen kielen tunnesanojen semantiikkaa. Acta Universitatis Ouluensis B 65. Oulu: Oulun yliopisto. <http://herkules oulu.fi/isbn9514278070/isbn9514278070.pdf>.
- Vainik, E. 2002. Millest on tehtud eestlaste emotsioonisõnavara? – Keel ja Kirjandus, nr 8, lk 537–553.
- Vainik, E., Orav, H. 2005. Tee tööd ja näe vaeva, ... aga ikka oled vihane. – Keel ja Kirjandus, nr 4, lk 257–277.
- Valk, R. 1999. Biases in Recognition of Emotions. Tartu (magistritöö Tartu Ülikooli psühholoogia osakonnas).
- van Goozen, S., Frijda, N. H. 1993. Emotion Words Used in Six European Countries. – European Journal of Social Psychology, Vol. 23, lk 89–95.
- Wierzbicka, A. 1999. Emotions Across Cultures. Oxford: Oxford University Press.

## **Theoretical Underpinning of The Corpus of Estonian Emotional Speech**

Keywords: Estonian, emotions, speech corpus, speech synthesis, emotion selection.

The Corpus of Estonian Emotional Speech is a project launched in 2006 under the state programme Language Technological Support for the Estonian Language. The article introduces the theoretical basis of the Corpus in progress and describes its four stages of development: emotion selection, selection of reading material, reader selection, and perception tests for emotion recognition. The Corpus material will include sentences carrying anger, sadness and joy (as recognized in perception tests) as well as neutral sentences.

*Rene Altrov (b. 1973), MA, Institute of the Estonian Language, extraordinary researcher, [rene.altrov@eki.ee](mailto:rene.altrov@eki.ee)*