

EMOTSIOONID – KÕNETEHNOLOOGIA OLEVIK JA TULEVIK

HILLE PAJUPUU

EmotSIONAalne kommunikatsioon polnud tõsiselt võetav uurimisvaldkond, kuni 1990. aastate keskpaigas lasi Massachusettsi Tehnoloogiainstituudi professor Rosalind Picard lendu mõiste *affective computing*, pidades selle all silmas vajadust anda arvutitele võime emotsioone ära tunda, mõista ja väljendada, et muuta nii inimese ja masina suhtlus loomulikumaks ja intelligentsemaks (Picard 1997).

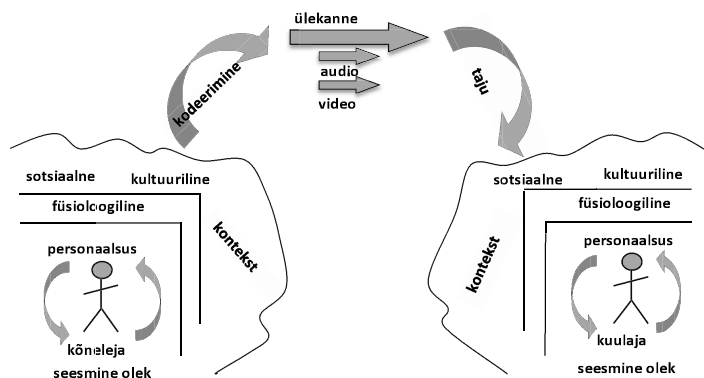
Ükskõik kui lihtsaks tehnoloogiliste vahendite kasutamine tehakse, tunnevad inimesed nendega kokku puutudes ikkagi frustratsiooni. Arvutikasutajad on tunnistanud, et nad vannuvad ja võivad muutuda isegi oma arvuti vastu vägivaldseks, kui selle töös tekib tõrkeid (MORI küsitlus 2002).¹ Inimese ja masina suhtluse uurimustes tuleks panustada sellele, et arvutid oskaksid sellist frustratsiooni ennetada või jooksvalt selle tekkimist märgata ning adekvaatselt reageerida. Ilma emotsioonitöötluseta pole see võimalik (Picard 2003, 2007).

Inimeste omavahelises suhtluses pole emotsioonide märkamine mingi probleem: teatakse, kuidas ja millal emotsioonile vastata. Ärrituse leevendamiseks piisab vahel vaid mõnest rahustavast lausest või vabandusest. Inimest, kes teise emotsioone ei suuda tõlgendada, peetakse emotsionaalseks invaliidiks. Inimese ja masina suhtluses on probleemiks nii arvuti mittereageerimine emotsioonidele kui ka ülereageerimine: kasutajat ärritavad ühtviisi mõlemad. Mõned meist ehk mäletavad Windows Office'i Clippitit (vt joonist 1), kirjaklambrikujulist animeeritud assistenti, kes passis peale, kui me arvutis kirjutasime, ning hüppas välja soovitusi jagama just siis, kui me sekkumist kõige vähem vajasime. Paljud kurtsid siis, et Clippit pole eriti intelligentne. Tegelikult oli Clippit vägagi intelligentne, mis puudutas Microsoft Office'i asju, ka sai ta väga hästi aru arvutikasutaja hetketegevusest, ent ta ei suutnud ette aimata kasutaja kavatsusi – Clippitil puudus emotsionaalne intelligentsus: ta ei märganud, et kasutaja on segamise pärast häiritud (ei tuvas-tanud kasutaja emotsiooni), ta ignoreeris kasutaja kasvavat ärritust (ei reageerinud sobiva emotsiooniga), ta lehvitas ja esitas enne lahkumist väikse



Joonis 1. Microsoft Office'i assistent Clippit.

¹ <http://www.ipsos-mori.com/researchpublications/researcharchive/970/Computer-Users-Suffer-PC-Rage.aspx>



Joonis 2. Emotsionaalse kommunikatsiooni mudel (McIntyre, Göcke 2008).

tantsu (oli emotsioonide väljendamises juhm). Lehitamine, naeratamine ja rõõmus tants võib olla lõbustav õiges kontekstis, stressirohketes tingimustes on see aga enam kui häiriv. Kuigi arvuteid on kerge panna välja nägema nii, nagu neil oleksid emotsioonid, on keeruline anda arvutile oskus aru saada, millal, kus ja kuidas sobivaid emotsioone näidata – ja millal neid MITTE näidata. Selgeks sai see, et nõuab palju õppimist, kuidas panna arvuti tasakaalukalt, respektieritaval ja intelligentsel viisil inimest kohtlema (Picard 2007, 2011).

Emotsionaalse kommunikatsiooni mudel on kujutatud joonisel 2. Emotsioonid on multimodaalsed, nad väljenduvad mitme kanali kaudu (näoilmed, hääl, žestid, sõnad). Kontekst (nt ümbrus, vestlus rühmas või telefoniga) määrab, milline neist kõige enam esile tuleb. Füsioloogiline mõju emotsioonide väljendamisele on näiteks vanusel, sool, tervise seisundil, kasvul. Oluline roll on ka kultuuril ja sotsiaalsetel teguritel: ühes kultuuris võib olla kombeks väljendada rõõmu valjuhäälselt, teises aga vaoshoitult, tuttavatega ollakse emotsionaalsem kui võrastega, mõjuda võib ka haridus. Emotsioonide väljendamine ja tõlgendamine sõltub veel ka personaalsetest omadustest – kas tegu on ekstraverdi või introverdigaga, lapse või täiskasvanuga – ning ka emotsionaalsest hetkeseisundist ja kogemustest (äsjased sündmused, nagu loteriivõit või töökaotus panevad emotsioone väljendama ja tõlgendama erinevalt) (Cowie jt 2005; McIntyre, Göcke 2006, 2008).

Inimese ja masina suhtluses on üheks partneriks arvuti. Õpetada arvuteid emotsioone ära tundma ja neile vastama on interdistsiplinaarne ülesanne, hõlmates informaatikat, psühholoogiat, filosoofiat, antropoloogiat, lingvistikat jne. Kõnetehnoloogia osa selles keerulises protsessis on panna arvutid kõnes emotsioone ära tundma (*emotion recognition from speech*) ja sobiva tundedetooniga rääkima (*emotional speech synthesis*).

Ühine eesmärk aktiveerib ja koondab kõnevaldkonna uurijad

Emotsioone tõlgendava ja väljendava arvuti idee on viimase 15 aastaga muutunud emotsioonid kõnevaldkonna eksootilisest uurimisteemast peavooluteemaks. Praktiline eesmärk – loomulikustada inimese ja masina suhtlust – on koondanud uurijaid diskuteerima emotsioonitöötamise probleemide üle nii teadusartiklites kui ka konverentsidel ja ärgitanud looma emotsiooniuurijate ühendusi.



Emotsioonide tõusmisele kõneuurimise keskmesse aitas kaasa 2003. aastal ilmunud Speech Communicationi erinumber, mille tänapäevalgi viidatavates artiklites leidsid käsitlust emotsioonide kirjeldamise ja defineerimisega seotud probleemid (Cowie, Cornelius 2003), kõnetehnoloogilisteks rakendusteks mõeldud emotsioonikorpuuste põhimõtted (Douglas-Cowie jt 2003), emotsioonituvastus kõnest (ten Bosch 2003), emotsionaalne kõnesüntees (Iida jt 2003), emotsionaalse kommunikatsiooni mudelid (Scherer 2003) ja teised emotsioonide akustika ja taju küsimused (Auberge, Cathiard 2003; Schröder 2003 jt).

Innustust lisas 2005. aastal toimunud esimene rahvusvaheline emotsioonitöötluste konverents (ACII2005) Pekingis, kus osales 134 ettekandega 140 inimest 32 riigist ja mille ettekanded leiab Springeri väljaandes LNCS.² Need regulaarsed konverentsid toimuvad nüüd emotsioonide uurimisele ja töötlemisele orienteeritud ülemaailmse assotsiatsiooni HUMAINE³ egiidi all (Tao, Tan 2009: v–vi). Sama assotsiatsioon toetab ka üleaastase keeleressursside konverentsi LREC emotsioonikorpuuste ja nende arendusega seotud tööpajajad. LREC-i 2008. aasta konverentsil oli üheks eriti esiletõstetud teemaks emotsioonide töötlemiseks loodavad ressursid.

Emotsiooniuurimise tähtsuse tõusu peegeldab ilmekalt näiteks kõneemotsioonituvastuses toimuv, milles võib selgelt näha kolme perioodi: mõned üksikud artiklid 1990. aastate teisel poolel (vähem kui 10 artiklit aastas), huvi kasv 2004. aastani (u 30 artiklit aastas) ja seejärel järsk tõus kuni siiani (üle 100 artikli aastas) (Schuller jt 2011).

Eesti hüppas selle ree peale 2006. aastal, alustades emotsionaalse kõne korpuuste loomist emotsioonide uurimiseks ja nende rakendamiseks kõnetehnoloogias; eeskätt peeti toona silmas emotsionaalset kõnesünteesi.⁴

Emotsiooni mõiste kõnetehnoloogias

Emotsiooni mõiste üle kõnetehnoloogiaalastes uurimustes eriti ei diskuteerita. Emotsiooni käsitatakse siin pigem laias tähenduses: emotsioon on see, „mis on enamasti kogu aeg olemas, ent puudub, kui inimesed on emotsioonitud” (Cowie jt 2011; Schuller jt 2011). Seda kontseptsiooni toetab Roddy Cowie ja tema kolleegide (2011) kirjeldatud katse, kus neljal inimesel lasti vaadata viietunnist videosalvestust aias toimetavatest ja suhtlevatest inimestest ning lõik-lõigult hinnata, milline kolmest valikuvariandist vastab kõige paremini muljele, mille salvestatud inimene selles lõigus jätab: kas emotsioon avaldub täiel määral, isik on keskmiselt emotsionaalne (= lõigus on emotsiooni elemente) või emotsioonitu. Kogu salvestusest tajuti 79 % keskmiselt emotsionaalsena, 14 % tugevalt emotsionaalsena ja vaid 7 % emotsioonivabana.

Emotsiooni määratlust võidakse aga sõltuvalt keeletehnoloogilisest eesmärgist kitsendada ja täpsustada, kasutades seejuures enamasti kas kategooria- või dimensioonipõhist lähenemist. Vt ka joonist 3.

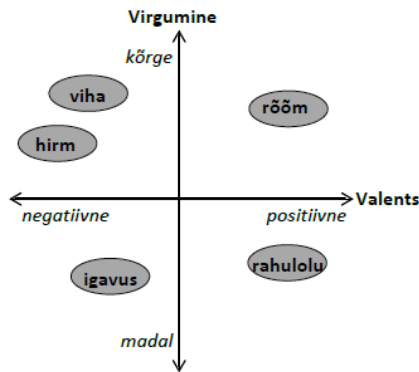
Esimesel juhul käsitatakse emotsioone kui diskreetseid kategooriaid (nimetatakse konkreetseid emotsioone). Vahet võidakse teha tugeva fooku-

² Lecture Notes in Computer Science.

³ <http://emotion-research.net/> (21. V 2012).

⁴ Riikliku programmi „Eesti keele keeletehnoloogiline tugi 2006–2010” projekt „Eesti emotsionaalse kõne korpus”.





Joonis 3. Kahedimensiooniline emotsiooniruum valentsi ja virgumise telgedega. Emotsioonikategooriad selles ruumis (Vogt jt 2008).

sega nn täismahulistel emotsioonidel (*full-blown, basic emotions*), nt viha, hirm, kurbus, rõõm jne, emotsiooniladsetel seisunditel (*emotion-related states*), nt kõhklus, autoritaarsus, huvi, väsimus jt, ning segaemotsioonidel (*mixed emotions, mixed states*), nt kui ühes episoodis täheldatakse üleminekut ühelt emotsioonilt teisele või ei tajuta ühte emotsiooni domineerivana. Täismahulisi emotsioone tuleb igapäevasuhtluses ette harva, võib-olla vaid kord päevas (Cowie jt 2011; Scherer jt 2010).

Diskreetsete emotsioonikategooriate arv on lahtine, kategooriate piirid ähmased ning igale neist ei pruugi leida kõnest füüsilisi korrelaate (nad ei pruugi kuuldelselt eristuda). Kategooriapõhiseid uurimistulemusi on tihti peale keeruline võrrelda, sest uuritakse erinevaid kategooriaid ning kategooriate sisu võib sõltuvalt keelest ka erineda (Cowie, Cornelius 2003, McIntyre, Göcke 2006).

Dimensioonipõhises lähenemises määratletakse emotsioon mitmemõttelises ruumis. Kaks kasutatavat dimensiooni on valents (*valence, pleasure*) ja virgumine (*arousal, activity/activation*). Valents kirjeldab emotsiooni positiivsuse või negatiivsuse astet ja virgumine aktiivsuse astet. Vahel kasutatakse ka kolmandat, kontrollidimensiooni (*control, potency, power* või *dominance*), mis näitab, kui võrd kontrolli all emotsioon on. Lisaks võidakse kasutada ka muid dimensioone, nt sõbralikkus (sõbralik vs. ebasõbralik), intensiivsus (madal vs. kõrge) jt (Cowie, Cornelius 2003; Cowie jt 2011).

Uurimistöodes kohtab ka dimensioonide ja kategooriate kombinatsioone, nt võidakse keskenduda negatiivse valentsiga emotsioonikategooriatele (vt nt Jaywant, Pell 2011).

Et tegemist on noore valdkonnaga, siis on mõistekasutus uurijati ja uurijarühmiti üsnagi erinev ja mõnevõrra segadust tekitav. Praktilise kasutuse ja teaduslikult põhjendatu vahel tasakaalu leidmiseks on alustatud emotsioonimärgenduse standardi loomist.⁵ Standardset märgenduskeelt saab kasutada nii emotsioonandmete käsitsi märgendamisel kui ka emotsioonide tuvastamisel ja genereerimisel.

⁵ Emotsioonimärgendusstandard EmotionML 1.0. <http://www.w3.org/TR/emotionml/> (21. V 2012).

Emotsioonikorpused ja nende sisu

Emotsioonitöötlus poleks mõeldav ilma emotsionaalse kõne korpuseta.

Emotsionaalseks kõnesünteesiks valitakse korpusest vajaliku emotsiooniga märgendatud kõneüksusi, mille jadast emotsionaalne kõne kokku pannakse (nn kõneüksuste valikul põhinev süntees, *unit selection synthesis*), või õpetatakse korpuse abil välja iga emotsiooni akustilistele parameetritele (nt intensiivsusele, kestusele, põhitoonile) vastav statistiline mudel ning kõnesignaali genereeritakse emotsioonikohaste mudelite parameetrite põhjal (nn statistiline süntees, *statistical parametric synthesis*) (Schröder 2009).

Emotsioonituvastuse süsteemides toimub kõnesignaalist andmete ekstraheerimine ja emotsiooniks klassifitseerimine. Emotsioonikorpuse kasutatakse siin tavaliselt selleks, et treenida iseõppivaid klassifitseerijaid, statistilisi mudeleid (El Ayadi jt 2011; Schuller jt 2011).

Emotsionaalse kõnesünteesi ja emotsioonituvastuse edukus sõltub korpuse sisust: mis seal on, kuidas märgendatud, millise kvaliteediga salvestatud jne.

Emotsioonikorpused võib jagada kolmeks suuremaks rühmaks kogutava emotsionaalse kõne laadi järgi (vt ka Schuller jt 2011; Vogt jt 2008).

1. Emotsioone lastakse näidelda (*prompted, simulated, acted emotions*). Kõnelejale, tavaliselt professionaalsele näitlejale, öeldakse, millist emotsiooni väljendada. Tihtipeale on sellised emotsioonid kontekstivälised, näitleja produtseerib üht ja sama semantiliselt identset lauset eri emotsioonidega. Enamasti on nii kogutud täismahulisi emotsioone, nagu viha, rõõmu, kurbust, vastikust jne. Sääraste korpuste miinus on see, et näideldud emotsioonid erinevad loomulikest, nad on ülepaisutatud, stereotüüpsed.

Kuulsamad siiani aktiivselt kasutatavad näideldud emotsionaalse kõne andmebaasid on pärit 1990. aastate teisest poolest: need on Berliini emotsionaalse kõne andmebaas BES (Burkhardt jt 2005) ja Taani emotsionaalse kõne andmebaas DES (Engberg jt 1997). Esimene neist sisaldab 10 inimese kõnet, kellel on lastud öelda emotsionaalselt neutraalse sisuga lauseid 7 täismahulise emotsiooniga, teises on need arvud vastavalt 4 ja 5.

2. Emotsioonid kutsutakse esile (*elicited, induced, stimulated emotions*). Selleks kavandatakse emotsioone tekitav olukord, nt lastakse osalejatel lahendada mingi emotsioone esilekutsuv ülesanne. Osalejatele ei öelda, et huvitatakse nende emtsioonidest. Need situatsioonid võidakse luua kahe inimese või inimese ja masina suhtluse vormis. Masina puhul võib olla tegu reaalse masinaga (robotiga) või täidab tema osa inimene (võlur Ozi meetod). Enamasti saadakse niimoodi kätte emtsioonilaadseid seisundeid (nt pettumus, frustratsioon, huvitatus, ükskõiksus), mitte täismahulisi emtsioone. Seega sarnanevad esilekutsutud emtsioonid rohkem reaalsete emtsioonidega kui näideldud emtsioonid. Emtsiooni võib esile kutsuda ka kirjaliku teksti lugemine (vt Velteni tehnika, nt Scherer ilmumas).

Esilekutsutud emtsioone sisaldavad näiteks SmartKom korpus (Schiel jt 2002) ja saksa AIBO emtsioonikorpus (Batliner jt 2004b), kus inimesi on salvestatud laboritingimustes täitmas ülesandeid, mis on mõeldud esile kutsuda rõõmu, empaatiat ja ärritust.

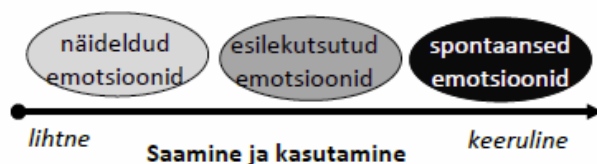
Kirjaliku tekstiga on emtsioonid esile kutsutud jaapani kõnesünteesi jaoks loodud emtsioonikorpuses. Eeldatud on, et tekstilõigu emtsioon teki-

tab ka lugejas vastava emotsiooni, mis kajastub tema hääles (Iida jt 2003). Samal põhimõttel on emotsioonid kogutud ka eesti emotsionaalse kõne korpusesse (Altrov, Pajupuu 2012).

3. Emotsioone kogutakse reaalses situatsioonides (*spontaneous, real, natural emotions*). Selleks salvestatakse igapäevaseid vestlusi ja TV vestlussaateid. Viimase aja trend on multimodaalsed korpused, kus märgendatakse nii kõneemotsioonid kui ka näiteks näoilmed. Üks selliseid avalikke ja seetõttu väga intensiivselt kasutatavaid korpuseid on saksakeelne Vera am Mittag (VAM), milles on salvestused samanimelisest TV jutusaatest. Korpus sisaldab 104 inimese 1421 videoklippi (emotsioonid pole märgendatud), 19 inimese 499 helilõiku (u 26 igalt kõnelejalt, emotsioonid on märgendatud) ja 20 inimese 1867 näopilti (u 94 igalt kõnelejalt, emotsioonid on märgendatud) (Grimm jt 2008). Reaalseid emotsioone on andmebaasidesse kogutud ka kõnekeskustega peetud telefonivestlustest (Ang jt 2002; Batliner jt 2004a; Devillers jt 2010) ja suhtlusest virtuaalagendiga (Schröder jt 2008).

Kui kirjutatud keele korpuste puhul räägitakse mitmeid miljoneid sõnu sisaldavatest kogudest, siis üle keskmise suurusega emotsionaalse kõne korpused sisaldavad „vaid” tuhat kuni paar tuhat kõnejuppi (lauset, väljendit, kõnevooru vms).

Emotsioonide uurimiseks ja töötlemiseks kasutatakse kõiki kolme tüüpi korpuseid. On selge seos korpusesse andmete kogumise keerukuse ja emotsioonianalüüsi ja -sünteesi keerukuse vahel: andmete loomulikkuse kasvuga muutuvad need ülesanded komplitseeritumaks (vt joonist 4).



Joonis 4. Emotsioonikorpuste tüübid, emotsioonide kogumise keerukus ja andmekasutuse keerukus kõnetehnoloogias (vt ka Vogt jt 2008).

Ülevaate 102-st emotsionaalset kõnet sisaldavast korpusest leiab Johannes Pittermanni ja tema kolleegide raamatu „Handling emotion in human-computer dialogues” lisast (2010: 237–272). Esindatud on neis 21 keelt. Kõik nimekirjas olevad korpused pole aga korralikult dokumenteeritud, paljudes neist on suures mahus (80 % ja rohkem) neutraalset kõnet. Autorid mainivad ka, et enamik korpuseid pole avalikult kättesaadavad või tuleb nende kasutamise eest võrdlemisi palju maksta. Tasuta kättesaadavate korpuste vähesus on emotsioonitöötlemises tõsiseks probleemiks, mistõttu on uurijad tänulikud iga lisanduva korpuse eest: korralikult dokumenteeritud ja märgendatud emotsioonikorpused võimaldavad statistilisi meetodeid katsetada mis tahes keele peal.

Eesti koht emotsionaalse kõne uurimise ja emotsioonitöötlaste maastikul

Süsteemaatiline kõneemotsioonide uurimine algas Eestis 2006. aastal, mil riikliku keeletehnoloogia programmi raames alustati Eesti Keele Instituudis emotsionaalse kõne korpuse loomist. Et tolleks ajaks oli maailmas juba emotsioonikorpuse olemas, neid oli kirjeldatud ning analüüsitud, siis sai teiste kogemusi arvesse võtta.

Eesti korpus loodi esmalt teoreetilise mudelina (vt Altrov 2008), milles vastati järgmistele küsimustele:

- mis on korpuse eesmärk;
- mis emotsioone peaks korpus sisaldama;
- kas korpus sisaldab näideldud, esilekutsutud või spontaanseid emotsioone;
- kas korpuse tekstid on kõneldud või loetud;
- milline on nende tekstide pikkus, sisu ja kontekst;
- kui suur peaks korpus olema;
- kes ja kuidas määrab korpuse kõneüksuse emotsiooni.

Leiti, et Eestis on vajadus sellise korpuse järele, mis võimaldaks kõnes ja kirjas olevaid emotsioone uurida ja oleks rakendatav ka kõnetehnoloogiliste ülesannete täitmisel. Et Eesti Keele Instituudis tegeldi parajasti kõneüksuste valikul põhineva tekst–kõne-süntheesiga (Mihkla jt 2008), siis nähti ka korpuse esimest rakendust just süntheesis, mis osaliselt määras nii korpuse tekstide valiku kui ka kõneüksuste segmenteerimise ja märgendamise tasandi.

Tolle aja korpused sisaldasid peamiselt näideldud emotsioone, kuigi näideldud emotsioonide reaalsele vastavuses juba kaheldi (Douglas-Cowie jt 2003). Uuemates korpustes oli hakatud näitlejate asemel kasutama ka n-ö tavalisi inimesi, kellele enam ei öeldud ette, mis emotsiooni etendada, vaid üritati emotsioon neis erinevate meetoditega esile kutsuda. Akemi Iidal kolleegidega (2003) oli õnnestunud mõõdukalt väljendunud jaapani emotsioonid kätte saada tavalise inimese loetud emotsionaalse sisuga tekstilõikudest. Et eesti süntesaator, nagu jaapani omagi, oli mõeldud kirjutatud tekstide ettelugemiseks, siis tundus mõistlik neil eeskuju võtta. Nii koguti ajakirjandusest hulk emotsionaalse värvinguga tekstilõike, mis lasti tavalisel meeldiva häälega naisterahval ette lugeda, ilma et talle oleks öeldud, et huvitatakse tema emotsioonidest. Need salvestised on korpuses lauseteks segmenteeritud ja läbinud kuulamistesti, milles iga lause emotsionaalsuse või neutraalsuse üle on otsustanud kuulajad. Lauset kuulates tuli valida üks vastusevariantidest: rõõm, kurbus, viha, neutraalne. Kuulajatele selgitati, et neil ei maksa oodata täismahulisi emotsioone: rõõmu all tuleks mõista ka tänulikkust, õnnelikkust, meeldivust, vaimustust; viha all pahameelt, irooniat, vastumeelsust, põlgust, kahjurõõmu, raevu; kurbuse all üksindust, trööstitust, murelikkust, lootusetust jne.

Korpuses iseloomustab igat lauset kuulamistesti tulemus: mida suurem protsent kuulajaist on lauses tajunud sama emotsiooni, seda selgemalt see emotsioon selles lauses avaldub.

Kuulamistestidega seoses vajas lahendamist küsimus, kui palju mõjutab kuulatava lause emotsiooni tajumist lause sisu. Et kindlalt eristada korpuses laused, mille emotsiooni kannab ainult hää, viidi läbi lugemistestid. Neis tuli lause emotsiooni üle otsustada teksti põhjal, ilma häält kuulmata. Ka lugemistesti tulemusi hoitakse korpuses iga lause juures.

Tabel 1.

**Emotsioonide klassifitseerimine korpuses
kuulamis- ja lugemistestide tulemuste põhjal***

Testid	Rõõm	Viha	Kurbus	Neutraalne	Ei oska öelda	Lause tüüp korpuses
<i>1. Ehkki Ott minu olemasolust midagi ei teadnud.</i>						
Kuulates	87,5	0,0	0,0	12,5	–	Rõõm, sisu ei mõjuta
Lugedes	4,0	0,0	32,0	32,0	32,0	
<i>2. Ükskõik, mida ma teen, ikka pole ta rahul!</i>						
Kuulates	0,0	14,3	80,0	5,7	–	Kurbus, sisu ei mõjuta
Lugedes	0,0	64,3	35,7	0,0	0,0	
<i>3. Täiesti mõistetamatu!</i>						
Kuulates	0,0	100,0	0,0	0,0	–	Viha, sisu mõjutab
Lugedes	0,0	83,0	0,0	11,0	5,6	

* Testitulemused on protsentides (Altrov, Pajupuu 2012).

Et kuulamis- ja lugemistesti tulemused ei lange alati kokku, on korpuse laused klassifitseeritud kaheks (vt tabelit 1):

1) laused, mille emotsiooni identifitseerimist kuulamistestis sisu ei ole mõjutanud (kuulamistesti tulemused erinevad lugemistesti omadest), tabelis 1 on sellisteks näideteks esimene ja teine lause;

2) laused, mille emotsiooni identifitseerimist kuulamistestis võib olla mõjutanud lause sisu (kuulamistesti ja lugemistesti tulemused langevad kokku), tabelis 1 on selliseks näiteks kolmas lause.

Eesti emotsionaalse kõne korpus on unikaalne, sest kõik korpuse laused on läbinud nii kuulamis- kui ka lugemistesti. Kuigi niisugune kahekordne testimine on aeganõudev, töötab ta korpuse valideerijana: korpuse kasutajad saavad olla kindlad, et selles olevate lausete emotsioon on äratuntav (Altrov, Pajupuu 2012).

See korpus on ka üks väheseid esilekutsutud emotsioone ja emotsiooni-laadseid seisundeid hõlmavaid kõnekorpuseid, mis on dokumenteeritud, avalikult ja tasuta kättesaadav ning piisavalt suure mahuga, et kasutada seda kõnetehnoloogias statistiliste mudelite õpetamisel, vt korpuse kodu <http://peeter.eki.ee:5000/>.

Korpus sisaldab hetkel ühe naishääle 1234 lauset, mis on klassifitseeritud rõõmuks, vihaks, kurbuseks või neutraalseks. Lausete heli (wav-formaat:

44.1 KHz, 16Bit, mono) ja teksti saab alla laadida ning salvestada. Segmenteeritud ja märgendatud on kolm tasandit: laused, sõnad, häälikud (*textgrid*-formaad). Päringuid saab teha nii testide kui ka lausete kohta. Alla laadida saab ka kogu korpuse tervikuna koos veebipõhise testimiskeskonnaga.

Korpust saab laiendada sõltuvalt eesmärgist ja keeletehnoloogilisest ülesandest: temasse saab lisada videoklippe, pilte, uusi kõnelejaid, jagada emotsioone enamatesse kategooriatesse ja määratleda neid dimensioonidel.

Korpuse materjali on juba kasutatud emotsioonide taju ja akustika uurimiseks. Silmapiiril on ka esimesed keeletehnoloogilised rakendused.

Emotsioonide taju

Emotsioonide taju kohta on eesti emotsionaalse kõne korpuse materjali põhjal tehtud mitmeid uurimusi. Kinnitust on leidnud, et emakeelsed kuulajad suudavad hääle järgi, ilma kõnelejat nägemata, hästi ära tunda ka mõõdukalt väljendunud emotsioone. 1234 lausest määras üle 50 % kuulajatest 700 lauset rõõmu, viha või kurbusega seotud emotsiooni kandvaks ja 208 neutraalseks (vt tabelit 2) (Altrov, Pajupuu ilmumas).

Tabel 2.

Kuulamistestis emotsionaalseks ja neutraalseks identifitseeritud laused

Kuulamistesti tulemused	Rõõm	Viha	Kurbus	Neutraalne
Emotsionaalseks ja neutraalseks identifitseeritud lausete arv	232	277	191	208
Keskmine identifitseerimisprotsent ja standardhälve	75,4 (14,5)	73,3 (14,6)	72,1 (14,7)	68,3 (11,9)

Üheks emotsioonidest arusaamise mõjutajaks on peetud vanust. Ühelt poolt on uurimused näidanud, et vanuse kasvades emotsioonidest arusaamine väheneb (Mill jt 2009; Paulmann jt 2008), kuid teisalt on emotsioonide väljendamine ja tõlgendamine kultuurikohane ning aja jooksul õpitav (vt Dewaele 2010; Nazir jt 2009). Rene Altrov ja Hille Pajupuu (2010) on uurinud nii vanuse kui ka kultuuri mõju emotsioonide tajumisele, püstitades hüpoteesi, et vanemad inimesed tunnevad emotsioone noortest paremini ära, sest on vastavas kultuuris pikemalt elanud. Katses lasti kolmel rühmal – noortel eestlastel, täiskasvanud eestlastel ja eesti keelt mitteoskavatel lätlastel – kuulata korpuse 35 eestikeelset lauset ning määrata nende emotsioon. Kõik kolm rühma erinesid emotsioonide määramises oluliselt. Noorte eestlaste ja täiskasvanud eestlaste arvamus läks kõige enam kokku kurbuse ja neutraalsuse määramisel. Noorte puhul võis täheldada tendentsi määrata täiskasvanutest rohkem lauseid neutraalseteks. Lätlased eesti emotsioone aga hääle järgi ära ei tundnud. Kuigi kurbuse määramisel saavutasid lätlased isekeskis üksmeele, määrasid nad kurbadeks teised laused kui eestlased.

Altrov ja Pajupuu (ilmumas) kordasid sama katset ka eesti keelt mitteoskavate itaallastega ning võrdlesid tulemusi lätlaste ja täiskasvanud eest-

laste omadega. Kui eestlased ja lätlased erinesid üksteisest oluliselt kõigi eesti emotsioonide määramises, siis itaallastel läks veidi paremini: nad tundsid ära eesti kurbuse. Itaallased ja lätlased sarnanesid omavahel eesti lausete viha ja neutraalsuse määramises, kuid need määrangud erinesid eestlaste endi omast.

Tulemustest saab järeldada, et mõõdukalt väljendunud emotsioonide tuvastamiseks ainult hääle järgi (nt raadiosaates) on oluline omandada kultuurispetsiifiline emotsioonide tõlgendamise oskus.

Emotsioonide akustika

Emotsioonide akustika uurimine Eestis on seni liikunud emotsioonide kirjeldamise ja modelleerimise suunas. Uurimuste eesmärk on rakendada teadmist peamiselt kõnesünteesis üksteisest eristuvate ja tajutavate emotsioonide genereerimiseks (Mihkla jt 2012). Statistiliste meetoditega on uuritud ja võrreldud korpuse rõõmu, viha, kurbuse ja neutraalsete lausete tempot, intensiivsust (valjust), vokaalide a , i ja u kvaliteeti (F1 ja F2) ning emotsionaalsete tekstilõikude pause (Tamuri 2010, 2011, 2012; Tamuri, Mihkla 2012).

Emotsionaalse kõne kiiruse iseloomustamiseks on kasutatud häälikute arvu sekundis. Rõõmu-, viha- ja kurbuselaudsed eristuvad üksteisest kõnekiiruse poolest. Kõige kiiremad on vihalaused ja kõige aeglasemad kurbuselaudsed: viha (17,5 h/s) > rõõm (17,1 h/s) > neutraalne (16,9 h/s) > kurbus (16,6 h/s). Sama tendentsi on täheldatud ka teiste keelte puhul (Tamuri, Mihkla 2012).

Ka lauseintensiivsus eristab emotsioone. Kõige intensiivsemaks on osutunud neutraalsed laused ja kõige vähem intensiivseteks kurbuselaudsed: neutraalne > viha > rõõm > kurbus. Võrreldes Schereri CPM mudeliga, mis ennustab emotsiooni mõju häälelisele väljendusele (Banse, Scherer 1996; Scherer 1986, 2009), võib eesti rõõmu liigitada nn vaikseks rõõmuks (vastandina lustlikule rõõmule, mis eeldab intensiivsuse tõusu) (Tamuri 2011). Eestlaste poolt vihaks määratud laused ei liigitu Schereri mõistes ei kuumaks ega külmaks vihaks – mõlemal juhul pidanuks intensiivsus tõusma. Madal intensiivsus eesti viha puhul on arvatavasti kultuuri- (või ka isiku)spetsiifiline ja viitab vajadusele selle emotsiooni väljendust edasi uurida ja tõlgendada.

Emotsionaalsete lausete vokaalide uurimine on näidanud, et teistest eristub ainult kurbuse puhul a : ta on kaotanud oma kvaliteedi ja muutunud ebamäärase kõlaga vokaaliks (Tamuri 2012).

Kairi Tamuri (2010) on uurinud ka emotsionaalsete tekstilõikude pauside vastavust kirjaliku teksti erinevatele kirjavahemärkidele ning leidnud, et kõik emotsioonid eristuvad üksteisest vaid lauselõpu punktidele vastavate pauside keskmistelt kestustelt. Lühimad on vihalausete ja rõõmulausete vahelised pausid ja kõige pikemad kurbuselause vahelised.

Seega teame eesti emotsiooniväljenduse kohta praeguseks seda, et kurvana räägitakse vaiksemalt, aeglasemalt, pikemate pausidega ja lödvemalt artikuleerides, nii nagu väga paljudes teisteski keeltes.

Rõõmu ja viha puhul räägitakse kiiremalt, kuid neutraalsest kõnest vaiksemalt. Neutraalse kõne kiirus on mõõdukas, ent väljendutakse valjemalt. Ka see näib olevat kultuurispetsiifiline: kui eestlane räägib kõva häälega, ei pruugi ta sugugi vihane olla!

Näideldud kõne puhul on oluliseks emotsioonide eristajaks peetud põhitooni (hääle kõrgust), kuid tema eristusvõime ei pruugi olla samavõrd oluline esilekutsutud ja spontaansete emotsioonide puhul (vt Vogt, Andre 2005). Eesti emotsioonide põhitooni kohta andmeid kirjandusest veel ei leia.

Emotsionaalne kõnesüntees ja emotsioonituvastus

Maailmas on emotsionaalne kõnesüntees ja emotsioonituvastus jõudnud juba arvestatavale tasemele.

Emotsionaalses kõnesünteesis on suhteliselt häid tulemusi saadud üksuste valikul põhineva meetodiga, kuid see meetod eeldab küllaltki suuri emotsioonikorpusi, kus iga sünteesitava emotsiooni jaoks on piisav hulk märgendatud kõnet, mille üksustest „kokku laduda” vajaliku emotsiooniga uus kõne. Kuigi selline „taasesitatud” kõne on loomulikuga sarnane, saab sellega väljendada vaid korpuses esindatud emotsioone. Nendest erinevate emotsioonide loomine kõneparameetreid muutes on praktiliselt võimatu.

Viimasel ajal on suuremat tähelepanu pälvinud statistiline süntees. Selle eelis on, et mudelite õpetamiseks vajaminevad emotsioonikorpused ei pea olema kuigi suured. Tulemuseks on sujuv loomuliku meloodiaga kõne, kuid selle arusaadavus pole sageli nii hea kui üksuste valikul saadud sünteeskõnel.

Hetkel ollakse olukorras, kus selget võitjat pole: nii statistiline kui ka üksustepõhine süntees on arenemisvõimelised ja nende kombinatsioonidelgi tundub olevat tulevikku (Barra-Chicote jt 2010; Schröder 2009).

Emotsioonituvastuses kasutatakse põhiliselt statistilisi meetodeid. Treeningkorpuse sisu määrab meetodi kasutuspiirkonna: kui tuvastajat on õpetatud näideldud kõnet sisaldava korpuse peal, siis emotsioonituvastus spontaansest kõnest ei pruugi õnnestuda, ja vastupidi (Fernandez, Picard 2011).

Eestis alustati kõneemotsioonidega tegelemist märgatavalt hiljem kui mujal. Hilise stardi miinus on see, et teistele järelejõudmiseks tuleb tohutult pingutada, tempokaks liikumiseks on vähe inimesi ja ressursse. Kuni ollakse maas, ollakse teaduslikust diskussioonist väljas. Kuid on ka plusspool: õppida saab teiste kogemustest ja vigadest, kasutada ning kohandada juba loodud vabavara. See on päris suur ajavõit.

Emotsionaalse kõne korpusega jõuti teistele järele üpris hiljuti, mil rahvusvahelist areeni jagati ala tipptegijatega.⁶ Eesti emotsionaalse kõne korpust tutvustati kui paindlikku tehnoloogiliste arengusuundadega kohanduvat korpust.

Emotsionaalse kõnesünteesiga Eestis alles alustatakse, seda peetakse suureks väljakutseks. Kõnesünteesimeetoditest on 40 aasta jooksul Eesti Keele Instituudis läbi proovitud praktiliselt kõik: alustatud formantsünteesist ja jõutud välja statistilise sünteesini. Mis meetodit kasutada emotsionaalseks sünteesiks, seda parajasti kaalutakse. Olemasolev emotsioonikorpus võimaldab proovida mitut meetodit (Mihkla jt 2012).

Emotsioonituvastus Eestis sai alguse värske riikliku keeletehnoloogia programmi projektiga „Kõne ja teksti emotsionaalsuse statistilised mudelid” ning peaks andma esimesi tulemusi 2014. aasta lõpuks. Kasutusse on võetud Euroopa ühisprojekti SEMAINE sensitiiivse tehiskuulaja⁷ platvormi kõne-emotsioonituvastuse osa, mille keelespetsiifilisi mooduleid kohandatakse eesti andmetele. Suurem tulevikuplaan on sulatada kokku emotsioonisüntees ja emotsioonituvastus ning luua empaatiline tehissuhtleja.

Kuigi Eestis on kombeks oodata mis tahes vallas ülikiireid tulemusi, näitavad teiste maade kogemused, et tegemist on keeruliste ülesannetega, mis vajavad aega nii teoreetiliseks läbitöötamiseks kui ka teostuseks, katsetamiseks ja arendamiseks.

⁶ ES3, vt <http://emotion-research.net/sigs/speech-sig/es12>

⁷ Sensitive Artificial Listener (SAL), vt <http://semaine.opendfki.de/wiki/SEMAINE-3.1>

Rosalind Picard, kes aastaid on jälginud ja suunanud emotsioonitöötlustes toimuvat, on nüüd, 50-aastasena kirjutanud, et tüüpilise emotsiooniuurija teekond ei ole mitte niivõrd ronimine Everesti tippu kui pigem ronimine mööda väga pikka redelit tipu suunas. See kestab kümneid aastaid ja seetõttu on tähtis olla kindel, et redel saaks püsti pandud väärt paika: kui lõpuks jõuad viimasele pulgale, näeksid vaadet, mis väärrib seda tohutut jõupingutust (Picard 2010).

Artikkel on valminud tänu projektidele EKT1 „Kõne ja teksti emotsionaalsuse statistilised mudelid” ja SF0050023s09 „Eesti keele alusuuringud keeletehnoloogiliste rakenduste teenistuses”.

Kirjandus

- Altrov, Rene 2008. Eesti emotsionaalse kõne korpus: teoreetilised toetuspunktid. – Keel ja Kirjandus, nr 4, lk 261–271.
- Altrov, Rene, Pajupuu, Hille 2010. Estonian Emotional Speech Corpus: Culture and age in selecting corpus testers. – I. Skadiņa, A. Vasiļjevs (toim), Human Language Technologies – The Baltic Perspective. – Proceedings of the Fourth International Conference Baltic HLT 2010, Amsterdam: IOS Press, lk 25–32.
- Altrov, Rene, Pajupuu, Hille 2012. Estonian Emotional Speech Corpus: Theoretical base and implementation. – L. Devillers, B. Schuller, A. Batliner, P. Rosso, E. Douglas-Cowie, R. Cowie, C. Pelachaud (toim), 4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals (ES3). Istanbul, lk 50–53.
- Altrov, Rene, Pajupuu, Hille (ilmumas). Recognition of emotions in intercultural communication. – Proceedings of the 4th International Adriatic-Ionian Conference „Across Languages and Cultures”.
- Ang, Jeremy, Dhillon, Rajdip, Shriberg, Elizabeth, Stolcke, Andreas 2002. Prosody-based automatic detection of annoyance and frustration in human–computer dialog. – Proceedings of Interspeech. Denver, CO, lk 2037–2040.
- Auberge, Veronique, Cathiard, Marie 2003. Can we hear the prosody of smile? – Speech Communication, kd 40, nr 1–2, lk 87–97.
- Banse, Rainer, Scherer, Klaus R. 1996. Acoustic profiles in vocal emotion expression. – Journal of Personality and Social Psychology, kd 70, nr 3, lk 614–636.
- Barra-Chicote, Roberto, Yamagishi, Junichi, King, Simon, Montero, Juan Manuel, Macias-Guarasa, Javier 2010. Analysis of statistical parametric and unit selection speech synthesis systems applied to emotional speech. – Speech Communication, kd 52, nr 5, lk 394–404.
- Batliner, Anton, Hacker, Christian, Steidl, Stefan, Nöth, Elmar, Haas, Jürgen 2004a. From emotion to interaction: lessons from real human–machine dialogues. – E. André, L. Dybkjær, W. Minker, P. Heisterkamp (toim), Affective Dialogue Systems. Berlin: Springer, lk 1–12.
- Batliner, Anton, Hacker, Christian, Steidl, Stefan, Nöth, Elmar, D’Arcy, Shona, Russell, Martin, Wong, Michael 2004b. „You stupid tin box” – children interacting with the AIBO robot: A cross-linguistic emotional speech corpus. – Proceedings of the 4th International Conference of Language Resources and Evaluation LREC 2004. Lisbon, lk 171–174.

- Burkhardt, Felix, Paeschke, Astrid, Rolfes, Miriam, Sendlmeier, Walter, Weiss, Benjamin 2005. A database of German emotional speech. – Proceedings of Interspeech, Lisbon, lk 1517–1520.
- Cowie, Roddy, Cornelius, Randolph R. 2003. Describing the emotional states that are expressed in speech. – Speech Communication, kd 40, nr 1–2, lk 5–32.
- Cowie, Roddy, Douglas-Cowie, Ellen, Cox, Cate 2005. Beyond emotion archetypes: Databases for emotion modelling using neural networks. – Neural Networks, kd 18, nr 4, lk 371–388.
- Cowie, Roddy, Sussman, Noami, Ben-Ze'ev, Aaron 2011. Emotions: concepts and definitions. – R. Cowie, C. Pelachaud, P. Petta (toim), Emotion-Oriented Systems: The HUMAINE Handbook. Berlin–Heidelberg: Springer, lk 9–31.
- Devillers, Laurence, Vaudable, Christophe, Chastagnol, Clément 2010. Real-life emotion-related states detection in call centers: a cross-corpora study Department of Human-Machine Interaction. – Proceedings of Interspeech. Chiba, lk 2350–2353.
- Dewaele, Jean-Marc 2010. Emotions in Multiple Languages. New York: Palgrave Macmillian.
- Douglas-Cowie, Ellen, Campbell, Nick, Cowie, Roddy, Roach, Peter 2003. Emotional speech: Towards a new generation of databases. – Speech Communication, kd 40, nr 1–2, lk 33–60.
- El Ayadi, Moataz, Kamel, Mohamed S., Karray, Fakhri 2011. Survey on speech emotion recognition: Features, classification schemes, and databases. – Pattern Recognition, kd 44, nr 3, lk 572–587.
- Engberg, Inger S., Hansen, Anya Varnich, Andersen, Ove, Dalsgaard, Paul 1997. Design, recording and verification of a Danish emotional speech database. – EURO-SPEECH-1997, lk 1695–1698.
- Fernandez, Raul, Picard, Rosalind 2011. Recognizing affect from speech prosody using hierarchical graphical models. – Speech Communication, kd 53, nr 9–10, lk 1088–1103.
- Grimm, Michael, Kroschel, Kristian, Narayanan, Shrikanth 2008. The Vera am Mittag German audio–visual emotional speech database. – Proceedings of IEEE International Conference on Multimedia and Expo (ICME). Hannover, lk 865–868.
- Iida, Akemi, Campbell, Nick, Higuchi, Fumito, Yasumura, Michiaki 2003. A corpus-based speech synthesis system with emotion. – Speech Communication, kd 40, nr 1–2, lk 161–168.
- Jaywant, Abhishek, Pell, Marc D. 2011. Categorical processing of negative emotions from speech prosody. – Speech Communication, kd 54, nr 1, lk 1–10.
- McIntyre, Gordon, Göcke, Roland 2006. Researching emotions in speech. – P. Warren, C. I. Watson (toim), Proceedings of the 11th Australian International Conference on Speech Sciences & Technology. New Zealand: University of Auckland, lk 264–269.
- McIntyre, Gordon, Göcke, Roland 2008. The composite sensing of affect. – C. Peter, R. Beale (toim), Affect and Emotion in Human-Computer Interaction. Berlin–Heidelberg: Springer, lk 104–115.
- Mihkla, Meelis, Piits, Liisi, Nurk, Tõnis, Kiissel, Indrek 2008. Development of a unit selection TTS system for Estonian. – F. Čermak, R. Marcinkevičienė, E. Rimkutė, J. Zabarskaitė (toim), Proceedings of the Third Baltic Con-

- ference on Human Language Technologies: The Third Baltic Conference on Human Language Technologies, Kaunas Lithuania, October 4–5, 2007. Vilnius: Vytauto Didžiojo Universitetas, Lietuvių kalbos institutas, lk 181–187.
- Mihkla, Meelis, Hein, Indrek, Kalvik, Mari-Liis, Kiissel, Indrek, Sirts, Risto, Tamuri, Kairi 2012. Estonian speech synthesis: Applications and challenges. – Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference „Dialogue”. Moscow: RGGU, lk 443–453.
- Mill, Aire, Allik, Jüri, Realo, Anu, Valk, Raivo 2009. Age related differences in emotion recognition ability: a cross-sectional study. – *Emotion*, kd 9, nr 5, lk 619–630.
- Nazir, Asad, Enz, Sybille, Lim, Mei Yui, Aylett, Ruth, Cawsey, Alison 2009. Culture–personality based affective model. – *AI & Society*, kd 24, lk 281–293.
- Paulmann, Silke, Pell, Marc D., Kotz, Sonja A. 2008. How aging affects the recognition of emotional speech. – *Brain and Language*, kd 104, nr 3, lk 262–269.
- Picard, Rosalind W. 1997. *Affective Computing*. Cambridge–Massachusetts: The MIT Press.
- Picard, Rosalind W. 2003. Affective computing: challenges. – *International Journal of Human-Computer Studies*, kd 59, nr 1–2, lk 55–64.
- Picard, Rosalind W. 2007. Toward machines with emotional intelligence. – G. Matthews, M. Zeidner, R. D. Roberts (toim), *The Science of Emotional Intelligence: Knowns and Unknowns*. Oxford: Oxford University Press, lk 419–474.
- Picard, Rosalind W. 2010. Emotion Research by the people, for the people. – *Emotion Review*, kd 2, nr 3, lk 250–254.
- Picard, Rosalind W. 2011. Robots, autism, and god. – *The Veritas Forum* (<http://www.youtube.com/watch?v=xsFH2YUsTls&feature=relmfu>).
- Pittermann, Johannes, Pittermann, Angela, Minker, Wolfgang 2010. *Handling Emotions in Human-Computer Dialogues*. Dordrecht–Heidelberg–London–New York: Springer.
- Scherer, Klaus R. 1986. Vocal affect expression: a review and a model for future research. – *Psychological Bulletin*, kd 99, nr 2, lk 143–165.
- Scherer, Klaus R. 2003. Vocal communication of emotion: A review of research paradigms. – *Speech Communication*, kd 40, nr 1–2, lk 227–256.
- Scherer, Klaus R. 2009. The dynamic architecture of emotion: Evidence for the component process model. – *Cognition & Emotion*, kd 23, nr 7, lk 1307–1351.
- Scherer, Klaus R. (ilmumas). Vocal markers of emotion: Comparing induction and acting elicitation. – *Computer Speech & Language*.
- Scherer, Stefan, Siegert, Ingo, Bigalke, Lutz, Meudt, Sascha 2010. Developing an expressive speech labeling tool incorporating the temporal characteristics of emotion. – *Proceedings of the 7th International Conference on Language Resources and Evaluation*. La Valetta, lk 1172–1175.
- Schiel, Florian, Steininger, Silke, Türk, Ulrich 2002. The SmartKom multimodal corpus at BAS. – *Proceedings of the 3rd Conference on Language Resources and Evaluation*, Gran Canaria, lk 200–206.
- Schröder, Marc 2003. Experimental study of affect bursts. – *Speech Communication*, kd 40, nr 1–2, lk 99–116.
- Schröder, Marc 2009. Expressive speech synthesis: Past, present, and possible future. – J. Tao, T. Tan (toim), *Affective Information Processing*. London: Springer, lk 111–126.

- Schröder, Marc, Cowie, Roddy, Heylen, Dirk, Pantic, Maja, Pelachaud, Catherine, Schuller, Björn 2008. Towards responsive sensitive artificial listeners. – Proceedings of the 4th Internat. Workshop on Human–Computer Conversation, Bellagio.
- Schuller, Björn, Batliner, Anton, Steidl, Stefan, Seppi, Dino 2011. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. – Speech Communication, kd 53, nr 9–10, lk 1062–1087.
- Tamuri, Kairi 2010. Kas pausid kannavad emotsiooni? – Eesti Rakenduslingvistika Ühingu Aastaraamat, nr 6. Tallinn: Eesti Keele Sihtasutus, lk 297–306.
- Tamuri, Kairi 2011. Emotsionaalse kõne valjus. – Eesti Keele Instituudi talvseminar 21.–22. dets. 2011. Tallinn.
- Tamuri, Kairi 2012. Kas emotsioonid peegelduvad formantides? – Eesti Rakenduslingvistika Ühingu Aastaraamat, nr 8. Tallinn: Eesti Keele Sihtasutus, lk 231–243.
- Tamuri, Kairi, Mihkla, Meelis 2012. Emotions and speech temporal structure. – Linguistica Uralica, kd XLVIII, lk 209–217.
- Tao, Jianhua, Tan, Tieniu (toim) 2009. Affective Information Processing. London: Springer.
- ten Bosch, Louis 2003. Emotions, speech and the ASR framework. – Speech Communication, kd 40, nr 1–2, lk 213–226.
- Vogt, Thurid, Andre, Elisabeth 2005. Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. – IEEE International Conference on Multimedia & Expo (ICME2005), lk 474–477.
- Vogt, Thurid, Andre, Elisabeth, Wagner, Johannes 2008. Automatic recognition of emotions from speech: A review of the literature and recommendations for practical realisation. – C. Peter, R. Beale (toim), Affect and Emotion in Human-Computer Interaction, LNCS 4868. Heidelberg: Springer, lk 75–91.

Emotions – the Present and Future of Speech Technology

Keywords: affective computing, classification of emotions, emotional speech, emotional speech corpora

One of the challenges of human-computer interaction is how to make a computer recognize emotions from human speech and react adequately. The article provides a survey of the role of speech technology in affective computing, explains the concept of emotion in that field, introduces different approaches to emotion classification (categorical and dimensional, combined), and describes different emotion corpora (acted, elicited, spontaneous) as well as the methods of collecting emotional speech for each kind of corpus.

The paper highlights recent and ongoing work at the Institute of the Estonian Language in affective computing, in particular the creation and content of the Estonian Emotional Speech Corpus. A brief survey is given of the available research on the acoustics and perception of Estonian emotions as well as of the current trends in emotional speech synthesis and in emotion recognition from speech.

Hille Pajupuu (b. 1956), PhD, Institute of the Estonian Language, senior researcher, hille@eki.ee