

# **Eesti emotsionaalse kõne korpus (EEKK)**

**<http://urve.eki.ee:5000>**

Hille Pajupuu

Rene Altrov

Eesti Keele Instituut 2007

# Täitjad ja kaasaaitajad

## Töölepinguga

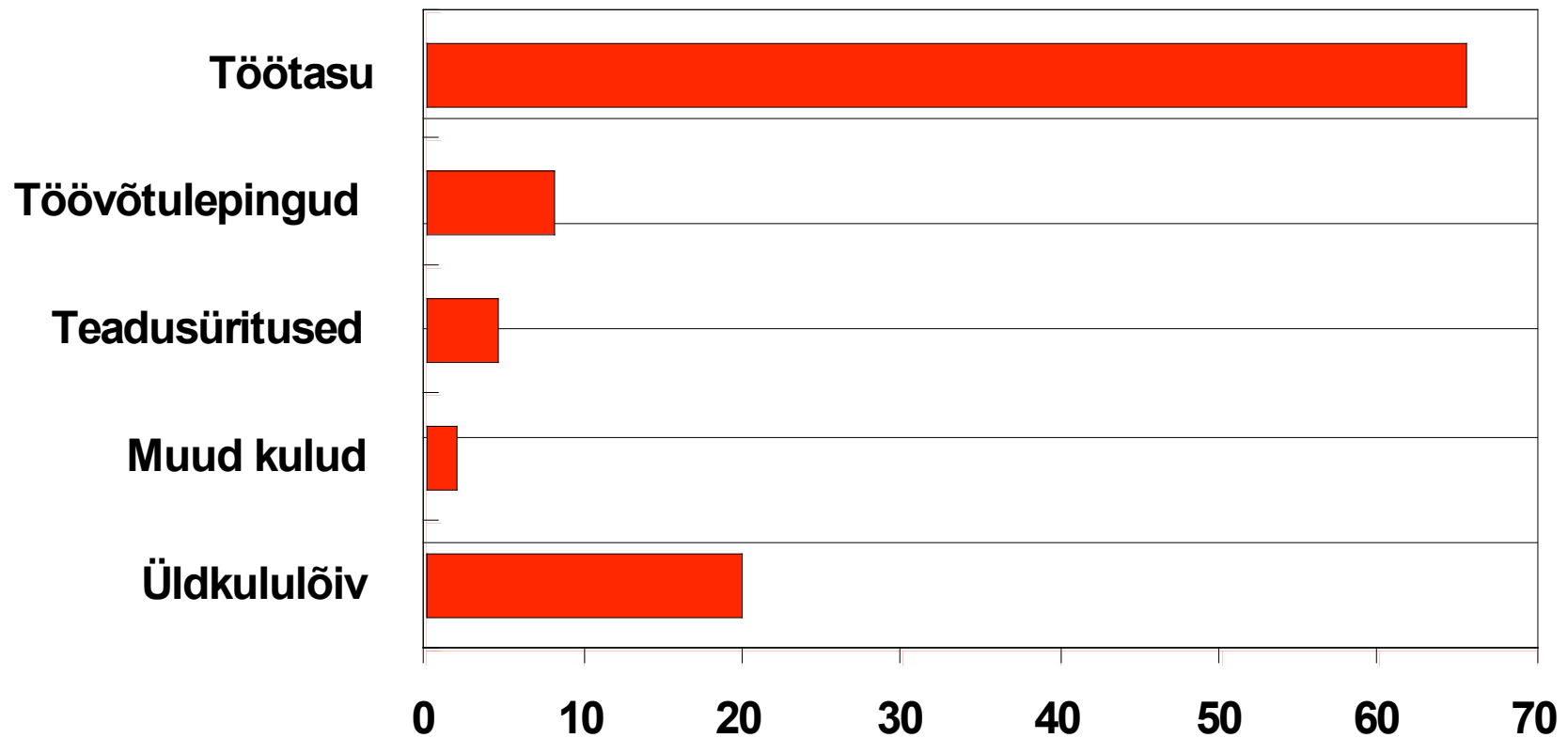
- Hille Pajupuu, PhD, projektijuht (0,1)
- Rene Altrov, MA, erakorraline teadur (1,0)
- Kairi Tamuri, MA, analüütik (1,0)

---

## Töövõtulepinguga

- Programmeerija (2006)
- Tõlkijad, lugeja (2007)

# Projekti rahajaotus 2007 (protsentides)



## Korpuse eesmärk:

- 1) olla korpuspõhise emotsionaalse tekst-kõne sünteesi akustiline baas;
- 2) olla usaldusväärne andmekogu kõnes avalduvate emotsioonide uurimiseks.

# Põhiotsused

- Lähtume seisukohast, et emotsioonid on hääle põhjal piisavalt hästi ära tuntavad loomulikus, mittenäideldud kõnes, ja et mittenäideldud kõne on loomuliku kõnesünteesi eeldus.
- Korpus peab sisaldama lauseid, mille emotsioonid on tajukatsetega kontrollitud.
- Korpus peab olema igati laiendatav: lugejatega, lausetega, emotsioonidega (s.h emotsioonilaadsete seisunditega), spontaanse kõnega jne.

# Korpuse loomise etapid

## Emotsioonide valik

viha rõõm kurbus neutraalne

## Lugemismaterjali valik

emotsioone kandvad tekstilõigud

## Lugejate valik

meeldiv hääl, hea artikulatsioon, empaatia

**Lugemine, salvestamine, lauseteks segmenteerimine**

## Kuulajate valik

täiskasvanud eestlased, min 30 in

**Kuulamistest** – kontekstita lause emotsiooni määramine heli põhjal

viha? rõõm? kurbus? neutraalne?

**Lugemistest** – lause emotsiooni määramine heli kuulmata

viha? rõõm? kurbus? neutraalne?

## Korpuse sisu

Viha-, rõõmu- kurbuselaused, neutraalsed laused

**Segmenteeritud ja märgendatud kõnelaine, tekst,  
kuulamis- ja lugemistesti tulemused**

# Kolm põhiemotsiooni – mida nende all mõelda

**viha** pahameel, iroonia, vastumeelsus, põlgus, kahjurõõm, raev

**rõõm** tänulikkus, õnnelikkus, meeldivus, vaimustus

**kurbus** üksindus, trööstitus, murelikkus, lootusetus

**neutraalne** = tavaline kõne, ilma eriliste emotsioonideta

# Kuulamistesti tulemuste aruandlus korpuses

***No kuidas ma saan seda kommenteerida?*** 

N	Neutr	Rõõm	Kurbus	Viha
32	9,4%	0,0%	6,2%	<b>84,4%</b>

***Lõin talle lahtise käega vastu õlga.*** 

N	Neutr	Rõõm	Kurbus	Viha
32	<b>78,1%</b>	3,1%	3,1%	15,6%

***Tuleb välja, et ma olen tööandja oma.*** 

N	Neutr	Rõõm	Kurbus	Viha
32	9,7%	25,8%	25,8%	38,7%

# Kuulamis- ja lugemistesti andmed koos

Kuulamise ja lugemise põhjal määratud emotsioonid ei lange alati kokku.

## KAKS RÜHMA:

### I. Emotsiooni kannab ainult hääl:

a) Teksti põhjal ei suudeta lause emotsiooni määrata, aga kuulates suudetakse.

**Viha 87,5%:** *Peaminister peab tegema oma suu lahti ja midagi ütleva.* 

b) Kirjutatud teksti põhjal määratakse üks emotsioon, kuulamise põhjal teine.

Nt pidas järgnevat lauset teksti põhjal **81,8%** vihaks, kuulates tajuti lauses aga kurbust:

**Kurbus 63,3%:** *Katkev internet, tarduv ja ruuduline telepilt, kättesaamatu tehnilise toe telefon, mittetöötav koduleht...* 



# Kuulamis- ja lugemistesti andmed koos

## II. Emotsiooni tajumist mõjutab tekst:

Lugemistestis ja kuulamistestis määratakse lausele üks ja sama emotsioon:

**Viha 90,6%:** *Igasugustel peksta ma ennast ka ei lase.* 📢

Heli kuulmata määrati lause emotsioon vihaks **92%**

**Kurbus 80,6%:** *Ma olen sügavalt nõrdinud.* 📢

Heli kuulmata määrati lause emotsioon kurbuseks **58,6%**

# Lausete kaheks rühmaks eraldamise olulisus

- **Laused, kus emotsiooni kannab ainult hääl**

Just seda osa korpusest, kus emotsioon peitub akustikas, mitte tekstis, pakume sünteesi akustiliseks baasiks.

- **Laused, mille emotsiooni tajumist mõjutab tekst**

See osa korpusest annab võimaluse uurida, kuidas kirjaliku teksti põhjal emotsiooni ära tuntakse, milline on nende lausete süntaks, sõnavara jne.

Uurimistulemuste rakendus sünteesis – võimaldab teksti põhjal otsustada, milliseid lauseid pidada vihaks, rõõmuks, kurbuseks jne.

# Päringud

**Korpusest saab otsida viha, rõõmu või kurbuse emotsiooni kandvaid lauseid ja neutraalseid lauseid.**

- Laused kuvatakse tekstina ja need on klikates kuulatavad
- Lause ees kuvatakse emotsiooni tajuprotsent.
- Tajuprotsendi saab ette anda, alates millest lauseid otsitakse.

**Päringutes saab teha piirangu:**

- a) kuvada ainult need laused, kus tekst ei mõjuta lause emotsiooni tajumist;
  - b) kuvada ainult need laused, kus tekst mõjutab lause emotsiooni tajumist.
- Lausete heli, segmenteeringu ja märgenduse (Praat TextGrid) saab endale salvestada.

# Korpuse tehniline teostus

Tarkvara on loodud platvormina korpuse sisu (kõne, segmenteerimise ja märgenduse) hoidmiseks ja päringute tegemiseks, emotsionaalse kõne tajukatsete koostamiseks, läbiviimiseks, tulemuste analüüsiks.

Kasutajaliides on eesti-, inglise- ja soomekeelne.

Korpus on teostatud veebipõhise rakendusena kasutades vabavara:

**Linux, PostgreSQL, Python, Praat.**

# Korpuse praegune seis

- Korpus on katsetamise ja arendamise faasis.
- Korpus sisaldab hetkel kahte naishäält ning nende loetud ja tajukatsetega kontrollitud lauseid: **79** vihalauset, **3** rõõmlauset, **20** kurbuselauset ja **25** neutraalset lauset.
- Ühe naishääle laused on segmenteeritud häälikuteks (Praat). Alustatud on märgendamist.
- Võimalus lausete heli, segmenteeringut ja märgendust (Praat TextGrid) endale salvestada.

# Korpuse tulevik

- **Korpuse laiendamine lausetega**
- **Korpuse emotsioonide täpsem märgendamine**
  - (1) põhiemotsiooni sees;
  - (2) lubada kuulajal ise välja pakkuda, mis emotsiooniga tema meelest tegu on

## **Neutraalse kõne osakaalu oluline tõstmine**

Sünteesis, ka emotsionaalse kõne sünteesis ei saa läbi neutraalse kõneta.

Kuna difoonidest loomulikku sünteeskõnet ja veel vähem emotsionaalset kõnet kokku ei pane, siis on vaja suuremaid kõneüksusi.

SALVESTADA eesti-X sõnastiku 40 000 märksõna koos põhivormidega, segmenteerida, märgendada.

- **Sünteesi päringuliidese lisamine**

# **Eesti emotsionaalse kõne korpuse põhimõte**

Olla ka oma arenguetappides avalikult kasutatav.