

TARTU ÜLIKOOL

Arvutiteaduse instituut

Informaatika õppekava

**Jürgen Leppsalu**

# **Eestikeelse häälassistendi raamistiku loomine**

**Bakalaureusetöö (9 EAP)**

Juhendaja: Sven Aller

Tartu 2021

## **Pealkiri**

### **Lühikokkuvõte:**

Bakalaureusetöö raames loodi eestikeelse häälassistendi raamistik nutiseadme rakendusena. Kui häälassistendi terviklik süsteem suudab käskluse põhjal tuvastada tegevuse ja selle teostada, siis raamistik sooritab vaid esimese osa sellest protsessist: käskluse seostamise tegevusega. Tarkvara testiti vabatahtlikega, kes esitasid süsteemile häälkäsklusi ning selle käigus hinnati toimingute tuvastamise täpsust ning uuriti süsteemi kitsaskohti. Kergemad vead, mis esinesid testide käigus, parandati lõputöö raames. Süsteemi edasiseks parendamiseks pakuti välja edasiarendusvõimalusi, mis laiendaksid loodud töö funktsionaalsust.

**Võtmesõnad:** Virtuaalassistent, häälassistent, kõnesüntees, kõnetuvastus, loomuliku keele töötlus, Androidi rakendus

**CERCS:** P170 Arvutiteadus, arvutusmeetodid, süsteemid, juhtimine (automaatjuhtimisteooria)

## **Title in English**

### **Abstract:**

The goal of the bachelor's thesis was to create an Estonian virtual assistant framework as a smart device application. A virtual assistant can determine an action based on the user's speech and execute it, whereas the framework is only capable of deciding what action the user requested. The software was tested with volunteers, who gave voice commands to the application. This offered some insight into the accuracy of the predictions and the weaknesses of the system. Simple problems, that emerged whilst testing, were removed from the software in the scope of this thesis. To further improve the system additions were proposed, which would improve usability of the application.

**Keywords:** Virtual assistant, voice assistant, speech synthesis, speech recognition, natural language processing, Android app.

**CERCS:** P170 - Computer science, numerical analysis, systems, control

# Sisukord

<b>1. Sissejuhatus</b>	<b>4</b>
<b>2. Häälassistendid</b>	<b>5</b>
<b>3. Häälassistendi raamistiku tagasüsteem</b>	<b>6</b>
3.1 Kõne transkribeerija	6
3.1.1 Transkribeerija kasutamine Dictate.js tarkvara abil	7
3.1.2 Transkribeerija kasutamine Kõnele-service tarkvara abil	9
3.2 Teksti analüsaator	9
3.3 Kõnesüntesaator	11
<b>4. Häälassistendi raamistiku eessüsteem</b>	<b>13</b>
4.1 Nutiseadme rakenduse platvorm	13
4.2 Häälassistendi rakenduse töövoog	14
<b>5. Häälassistendi raamistiku süsteemianalüüs</b>	<b>16</b>
5.1 Metoodika	16
5.2 Tulemused	16
5.2.1 Vead ja kitsaskohad	18
5.2.2 Lihtsamate parandustega süsteemi tulemused	21
5.2.3 Edasiarendused	22
<b>6. Kokkuvõte</b>	<b>23</b>
<b>Viidatud kirjandus</b>	<b>24</b>
<b>Lisad</b>	<b>26</b>
<b>Litsents</b>	<b>36</b>

## 1. Sissejuhatus

Eesti riigil on soov käia kaasaegsete keeletehnoloogiliste arengutega kaasas, sest see soodustab eesti keele pikaajalist püsijäämist [1]. Häälassistendid<sup>1</sup> on üks selline keeletehnoloogia valdkond, mis viimastel aastatel on funktsionaalsuse poolest jõudsalt arenenud ning mille kasutajaskond kasvab pidevalt [3]. Nimetatud süsteemide ingliskeelsed variandid on saadaval mitme erineva tootja poolt nii rakenduste kujul nutiseadmes kui ka nutikõlaritesse sisse ehitatud funktsionaalsusena [3], kuid eestikeelseid häälassistente pole. Leidub rakendusi nagu Dikteeri [4] ja Kõnele [5], kuid nende peamine ülesanne on transkribeerida kõne tekstiks ning sisestada see kasutaja soovitud tekstivälja või otsingumootorisse.

Bakalaureusetöö peamine eesmärk on luua nutiseadme rakendusena eestikeelse häälassistendi raamistik. Kui virtuaalassistendi terviklik süsteem suudab käskluse põhjal tuvastada tegevuse ja selle teostada, siis raamistik sooritab vaid esimese osa sellest protsessist: käskluse seostamise tegevusega. Loodud tarkvara testitakse vabatahtlikega, kes esitavad süsteemile häälkäsklusi, ning selle käigus hinnatakse toimingute tuvastamise täpsust ja uuritakse süsteemi kitsaskohti. Sekundaarne eesmärk on dokumenteerida, kuidas saab kasutada eesti keele keeletehnoloogilisi tööriistu.

Bakalaureusetöö teises peatükis kirjeldatakse ingliskeelsete häälassistentide olemust ja kasutamist. Kolmandas peatükis tutvustatakse töö käigus loodud ja olemasolevaid keeletehnoloogilisi tarkvaralahendusi, mis moodustavad häälassistendi raamistiku tagasüsteemi (ingl *backend*). Neljandas peatükis antakse ülevaade eessüsteemist (ingl *frontend*), põhjendades nutiseadme äpi jaoks valitud operatsioonisüsteemi ning kirjeldades rakenduse töövoogu. Viiendas peatükis tutvustatakse katseid, mille käigus vabatahtlikud testisid rakenduse tööd, ja analüüsitakse saadud tulemusi.

---

<sup>1</sup> Häälassistent - rakendus, mis suudab häälkäskluse sisu põhjal sooritada soovitud tegevuse [2]

## 2. Häälassistendid

Virtuaalassistent on tarkvara, mis võtab kasutajalt sisendi loomulikus keeles, olgu see siis sõnastatud küsimusena või käsklusena, ning selle põhjal sooritab soovitud tegevuse või tagastab kasutajale vastuse. Sisend võib olla nii suuline kui ka kirjalik, kuid oluline on, et suhtlus seadmega oleks teise inimesega suhtluse sarnane. Kuigi definitsioonist lähtuvalt ei ole sisendi tüübil piiranguid, siis kasutajate tagasisidest on näha selgeid eeliseid häälassistenditel tekstipõhiste virtuaalassistentide ees. Olulisteks plussideks nimetavad kasutajad, et häälkäskluste andmine on kiireim meetod seadmega suhtlemiseks ning lisaks sellele on hääl kasutatav ka siis kui käed või nägemine on muude tegevustega hõivatud. [6]

Tänapäeval on häälassistentidega turul paljud suured tehnoloogia ettevõtted nagu Apple, Google, Microsoft ja Amazon. Enamik inimesi on teadlikud selliste süsteemide olemasolust ning paljud kasutavad neid, mida näitab muuhulgas Hispaania telekommunikatsiooni firma Telefonica uurimus [7]. Uurimismeeskonna veebipõhise küsitluse käigus selgus, et ainult kümnendik vastanutest pole kunagi kuulnud häälassistenditest. Enamik vastanutest ehk 55,9% teadsid, mida sellised süsteemid teevad, ning 27,9% kasutavad sellist tarkvara. Uuringu andmed süsteemide kasutamise osas kinnitavad, et leidub arvestatava suurusega kasutajaskond, kes saab kasu sellest tehnilisest lahendusest iga päev või vähemalt mitu korda nädalas. Küsitluses osales 3749 inimest erinevatest Euroopa ja Lõuna-Ameerika riikidest ning Ameerika Ühendriikidest.

Häälassistente kasutatakse mitmel erineval moel. Olulised kasutusjuhud tuuakse välja ülikooli University of Michigan, tarkvaraettevõtte Mozilla ja telekommunikatsiooni firma Verizoni uurimusest [8], mille raames analüüsiti virtuaalassistentide tootjate logisid, viidi läbi küsitlusi ja korraldati intervjuusid. Nendest selgus, et kasutajad kasutavad peamiselt seadmega kõnelemise võimalust, et otsida internetist küsimustele vastuseid, mängida muusikat ja et suhelda erinevate IoT seadmetega nagu nutilambid. Oluliselt leidis kasutust veel hääle abil seadme helitugevuse reguleerimine, taimeri seadmine, seadmelt nalja küsimine, alarmi seadmine ning seadme enda seadete muutmine.

Nendest uurimustest võib järeldada, et häälassistenditel on olulisi eeliseid tekstipõhiste virtuaalassistentide ees. Samuti selgusid põhilised kasutusjuhud, millega pöörduakse häälassistentide poole. Neid teadmisi arvestatakse bakalaureusetöös raamistiku loomisel, et rakendus oskaks seostada kasutajate kõne ja inimeste poolt enim soovitud tegevusi.

### 3. Häälassistendi raamistiku tagasüsteem

Virtuaalne häälassistendi raamistik peab töötamiseks edukalt lahendama mitu erinevat ülesannet: kõne salvestama, kõne transkribeerima, transkriptsioonist eraldama tegevuse ja andma kasutajale tagasisidet tuvastatud toimingu kohta. Seda kõike ei saa teha seadmes endas. Üks piirang on näiteks mälu, mille kasutust nutiseadmete operatsioonisüsteemid jälgivad [9,10]. Kui töötamise käigus ületab rakendus ette määratud mälumahu, siis rakenduse töö lõpetatakse operatsioonisüsteemi poolt. Selliste probleemide tõttu tuleb jagada raamistik eessüsteemiks ja tagasüsteemiks. Eessüsteemi ehk nutiseadme rakenduse eesmärk on kasutaja kõne salvestada ja anda tagasisidet nii süsteemi seisundi kui ka tuvastatud tegevuste kohta. Tagasüsteemi moodustavad serverid, mis hoolitsevad kõne transkribeerimise, tehiskõne loomise ja teksti analüüsi eest.

Tagasüsteemi moodustavad kolm serverrakendust:

- TTÜ Küberneetika Instituudi foneetika- ja kõnetehnoloogia labori kõne transkribeerija;
- Käesoleva bakalaureusetöö käigus loodud tekstianalüsaator;
- Eesti Keele Instituudi kõnesüntesaator.

Nende tööd ja kasutamist kirjeldatakse järgnevates peatükkides.

#### 3.1 Kõne transkribeerija

Kasutaja kõnest informatsiooni eraldamiseks, tuleb esmalt kõnelus teisendada tekstiks. Helifaili transkribeerimist võimaldab eestikeelse kõne transkribeerija server, mis on loodud TTÜ Küberneetika Instituudi foneetika- ja kõnetehnoloogia labori poolt ning mille looja on Tanel Alumäe. Serverrakendus kasutab vabavaralist Kaldi kõnetuvastuse tööriista ning selle pistikprogrammi GStreamer Server. Tarkvara ja pistikprogramm suudavad eestikeelse kõne mudeli põhjal transkribeerida eestikeelse kõne tekstiks. Akustiline mudel, mille põhjal Kaldi transkriptsioone koostab, on treenitud, kasutades eestikeelsete helifailide korpust, kuhu on kogutud ligikaudu 135 tunni jagu helifailide koos transkriptsioonidega. Transkribeerimise server töötab ülem-alluv-mudeli järgi, kus on üks põhiprogramm ja palju veebitööliseid, mis võtavad kasutaja poolt saadetud audio, edastavad selle põhiprogrammile ja tagastavad saadud vastuse kasutajale. Selline mudel tagab selle, et paljude samaaegsete päringute korral tullakse transkribeerimise ja kõikide klientide teenindamisega edukalt toime. [11]

Kui on põhjust lokaalselt server tööle panna, siis on kogu Tanel Alumäe tarkvara saadaval vabavaraliselt Kaldi GStreamer Server repositooriumis [12]. Repositooriumi dokumentatsiooni kohaselt on süsteemi paigaldamiseks kaks varianti. Esimene variant on laadida alla José Eduardo Silva loodud tarkvara, mis tuleb käivitada platvormis Docker<sup>2</sup>. See loob transkribeerimise süsteemi jaoks vajaliku keskkonna ning kasutaja saab selles keskkonnas käivitada serveri programmi ja selle veebitöölised. Kui serveri aadressil saata

---

<sup>2</sup> Docker - tarkvara, mis võimaldab rakendusi arendada ja avalikustada konteinerites, et tarkvara projektide sõltuvuste ja keskkonna ülesseadmine oleks võimalikult mugav [13]

helifail koos vajalike parameetritega tagastatakse helifaili transkriptsioon päringu tegijale. Juhised süsteemi ülesseadmiseks ning kasutamiseks [14] on head ja täpsed, mistõttu sai lõputöö autor süsteemi isiklikus arvutis tööle. Teine variant on laadida alla Tanel Alumäe enda repositoorium, paigaldada endale dokumentatsioonis kirjeldatud sõltuvused ja kasutada süsteemi samamoodi nagu esimese variandi puhul. Selle paigaldamine nii Windowsi kui Linuxi masinasse on vaevarikas ning lõputöö autor ei suutnud seda meetodit kasutades süsteemi tööle saada. Ebaedu põhjuseid oli mitmeid. Esiteks tuleb installeerida palju Python 2.7 sõltuvusi, millest mõned pole dokumentatsioonis korrektselt kirjeldatud. Näiteks ei ole pakette YAML ja JSON võimalik pip<sup>3</sup> tööriista abil leida. On analoogsed paketid PyYAML ja PyJSON, kuid kuna süsteemi tööle ei saadud, pole võimalik kinnitada, kas need on süsteemi jaoks sobivad lisad. Teiseks tuleb iseseisvalt kompileerida Kaldi raamistik ja Kaldi GStreamer laiendus. Kaldi installeerimine nõuab palju järjestikuseid samme ning autori kogemuse kohaselt võttis programmi kompileerimise protsess paar tundi. Kui need tegevused on edukalt sooritatud, siis tuleb veel alla laadida soovitud keele akustiline mudel ja keele mudel, enne kui saab testida, kas transkribeerimise süsteem töötab. Kuna Tanel Alumäe kirjeldab dokumentatsioonis protsessi samuti „keerulisena” [14], siis on soovitatav hoiduda sellest variandist.

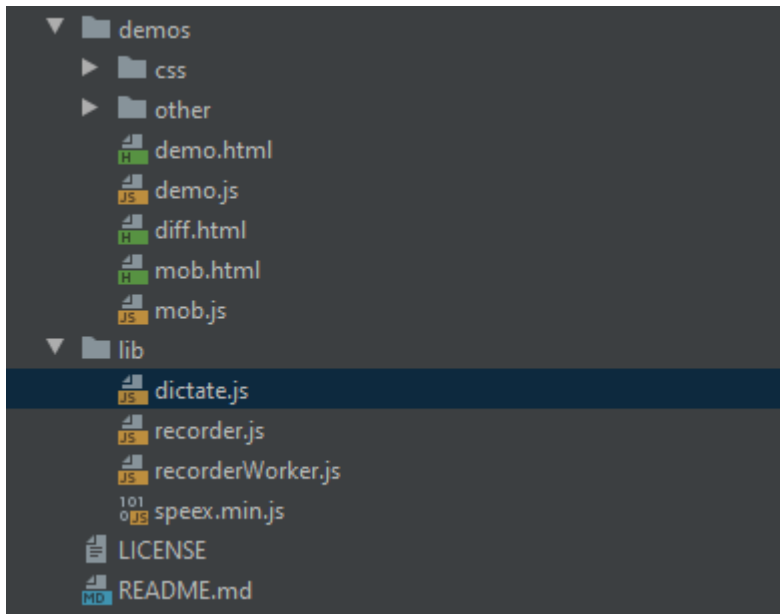
Sageli puudub vajadus serveri rakendust ise hallata. Selle asemel saab kasutada valmislahendusi, mis teostavad ise infovahetuse avaliku eestikeelse transkribeerimise serveri ja rakenduse vahel. Järgnevalt kirjeldataksegi kuidas selliseid võimalusi kasutada.

### **3.1.1 Transkribeerija kasutamine Dictate.js tarkvara abil**

Dictate.js [17] on Kaarel Kaljuranna loodud JavaScripti teek, mis võimaldab veebirakendustel suhelda transkribeerimise serveriga. Selle kasutamiseks on esmalt vaja alla laadida Dictate.js repositoorium [17]. Joonisel 1 on näha kõik failid, mida repositoorium sisaldab. Alamkaust „lib” sisaldab serveriga suhtlemiseks vajalikku tarkvara. Alamkaust „demos” sisaldab endas Kaarel Kaljuranna, Dictate.js autori, loodud demorakendusi, mis konverteerivad kõne tekstiks ja kirjutavad selle tekstikasti.

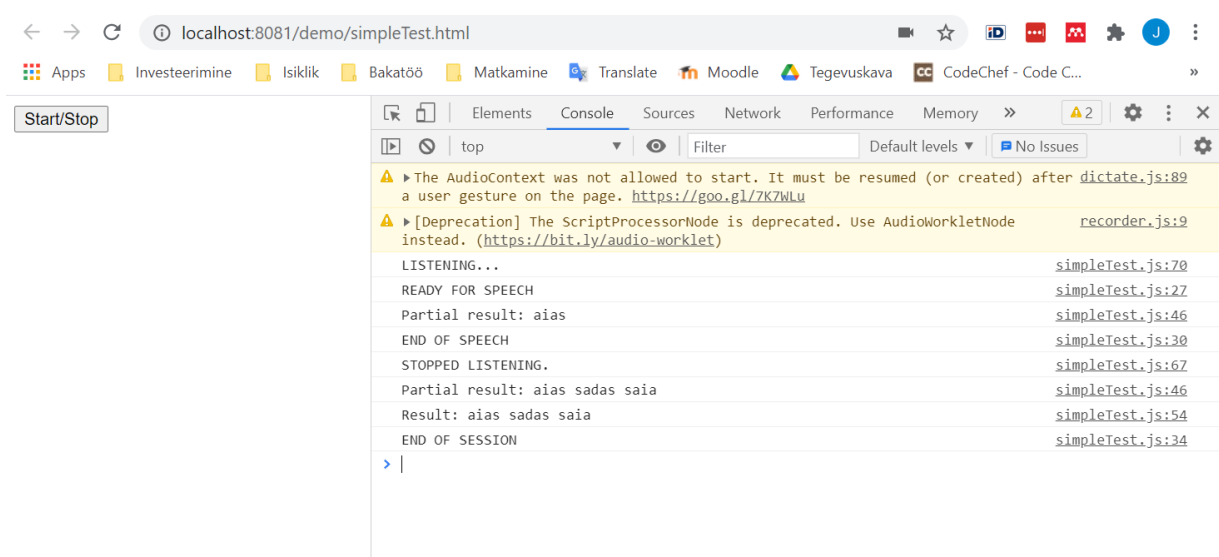
---

<sup>3</sup> pip - Pythoni paketi haldur [15]



Joonis 1. Dictate.js rakenduse failipuu.

Olemasolevad testrakendused sisaldavad endas palju koodi, mis on seotud veahalduse ja kasutajale tagasiside andmisega. Need pole transkribeerimise loogika mõistmise seisukohalt olulised. Parema ülevaate andmiseks valmistati lõputöö käigus lihtsakoeline JavaScripti programm simpleTest.js, mis nupu vajutamisel kuulab kasutaja kõne, transkribeerib selle ning talletab transkriptsiooni muutujas. Tulemus (vt Joonis 2) logitakse ka konsoolis, et oleks viis veenduda transkribeerimise edukuses. Näidiskirandus koos paigaldamise ja käivitamise juhendiga on avalik GitHubi repositooriumis <https://github.com/leppsalujyrgen/Kone-transkribeerimine-JavaScriptis>.

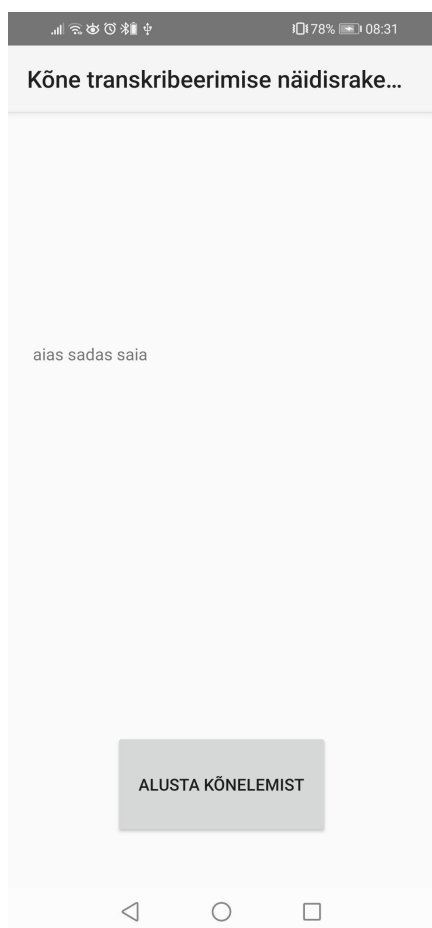


Joonis 2. JavaScripti näidiskirandus transkribeeris kõneletud fraasi „aias sadas saia”.

### 3.1.2 Transkribeerija kasutamine Kõnele-service tarkvara abil

Kõnele-service on Androidi rakendus, mis pakub kõne transkribeerimise teenust teistele rakendustele [18]. See on saadaval eraldiseisva äpina Google Play poes, kuid bakalaureusetöös lõimiti teenuse ja virtuaalassistendi rakenduse lähtekood, et virtuaalassistendi proovimiseks ei peaks paigaldama mitut erinevat rakendust.

Kõnele-service teenuse katsetamiseks loodi ühe tekstikasti ja ühe nupuga minimalistlik Androidi rakendus (vt Joonis 3). Kui nupule vajutatakse, hakkab rakendus salvestama ja transkribeerima kõne. Kui kasutaja lõpetab rääkimise, siis peatatakse mikrofoni salvestamine ning kuvatakse ekraanile kõne tekstina. Veahaldusega rakendus ei tegele, sest äpi eesmärk on pakkuda lihtsat kõne transkribeerimise koodinäidist Androidis. Rakendus on koos paigaldamise ja kasutusjuhendiga avalikus repositooriumis <https://github.com/leppsalujyrgen/Kone-transkribeerimine-Android-rakenduses>.



Joonis 3. Androidi näidisrakendus transkribeeris kõneletud fraasi „aias sadas saia”.

### 3.2 Teksti analüsaator

Teksti analüüsimise serverrakendus on loodud lõputöö käigus. Seda kasutatakse Keeleressursside Keskuse serveris ning ligipääs rakendusele käib veebiaadressi <http://prog.keeleressursid.ee/veebid/haalassistentvova> kaudu. Programmi eesmärk on võtta sisendiks käsklus teksti kujul ning seostada see tegevusega. Tegevusi, mida süsteem tunneb,

on kuus: meedia esitamine ja peatamine, helitugevuse tõstmine ja vähendamine, meeldetuletuse seadmine ja otsingu teostamine.

Teksti analüüsimiseks tuleb saata serveri URL-ile POST päring koos JSON lastiga, mis sisaldab välja nimega „speech”, mille väärtus on käsklus sõnena. Vastusena tagastatakse JSON last, mis sisaldab alati välja nimega „action”, mille väärtus sõltub käsu olemusest. Päringu näide on kuvatud Joonisel 4, kus on näha nii POST päring kui ka serverilt saadud vastus.

## Run Curl Commands Online

Execute Curl commands directly from your browser. Learn Curl with live Curl examples. Test APIs with ReqBin Online Curl Client.



The screenshot shows the ReqBin online curl client interface. At the top, there are navigation options: File, Generate Code, Tools, Share, and another Generate Code. Below this, there are tabs for Curl, Raw, US, and a Run button. The main area is divided into two sections: the left section shows the curl command used, and the right section shows the response.

```
curl -X POST -H "Content-Type: application/json" -d '{"speech": "ära mängi enam muusikat palun"}' http://prog.keeleressursid.ee/veebid/haalassistentvova
```

Status: 200 (OK) Time: 408 ms Size: 0.02 kb

Content (1) Headers (7) Raw (9) HTML Timings

```
{"action": "stop_media"}
```

Joonis 4. ReqBin veebirakendusega [21] tehtud POST päring tekstianalüüsi serverile.

Teksti analüüsimiseks kasutatakse keeletöötlus teeki EstNLTK. See tööriist võimaldab muuhulgas teisendada sõnu algvormi, leida sõnadele sünonüüme ja eraldada tekstist ajalisi viiteid [22]. Serveri programm, saades sisendiks käskluse teksti, kasutab EstNLTK teeki ja mustrisobitus meetodit tegevuse tuvastamiseks. Mustrisobitus meetod on infoeraldus viis, kus iga väljundi jaoks on süsteemis defineeritud sisendlause ülesehitus [23]. Teisisõnu kui programmile esitatakse sõne, mis vastab struktuuri poolest ühele väljundile, siis otsustab süsteem selle väljundi tagastada kasutajale. Tihti aga võivad käsklused sobida mitme mustriga, mistõttu peab rakendus otsustama ühe või teise korralduse tüübi kasuks. Selle probleemi lahendamiseks kasutatakse serveriprogrammis tegevuste hierarhiat, mis mitme tegevuse mustri esinemise puhul, otsustab kõige üldisema tegevuse kasuks. Võib tunduda intuiitsem, et mida konkreetsem on tegevus, seda rohkem peaks toimingut eelistama teistele, kuid tegelikkuses võivad universaalsemad käsklused nagu otsingu tegemise käsk sisaldada endas täpsema tegevuse näiteks meedia mängimise mustrit. Seetõttu tuleb liigitamisel pooldada universaalsemaid tegevusi.

Mustrid, mida serverrakendus tegevuste tuvastamiseks kasutab, on lihtsad. Igal tegevuse jaoks on loodud hulk sõnadest, mida otsitakse kasutaja käsklusest. Kui käsu transkriptsioonis leidub hulgas esinev sõna, siis esiteks kontrollitakse, kas rakendus pole teksti töötlemise käigus üldisemat tegevust tuvastanud ning kui ei ole, siis teiseks määratakse väljundiks hulgaga seotud tegevus. Helitugevuse ja meediapleieri mõjutamise käskude puhul peame tuvastama ka lauses eituseid, sest kasutajad võivad esitada häälassistendile soovi eitaval kujul näiteks „ära mängi enam muusikat” või „ära mängi nii valjult muusikat”. Lisamustriteta määraks süsteem need käsklused valesti, tuvastades meediapleieri käivitamise ja helitugevuse

tõstmise toimingut. Sellepärast muudetakse lauses eituse esinemise korral helitugevuse tõstmine, helitugevuse vähendamine, meedia mängimine ja meedia peatamine vastand tegevuseks. Meeldetuletuse määramisel otsitakse lisamustrina, EstNLTK tööriistu kasutades, lausest ajalisi viiteid, mille leidmisel seostatakse käsklus meeldetuletuse seadmisega, eeldusel, et rakendus pole eelnevalt tuvastanud universaalsemat tegevust.

Tegevusega seotud sõnade hulk peab sisaldama vaid tuvastamist vajava sõna algvormi. See lihtsustus on võimalik EstNLTK vahendi tõttu, mis suudab iga sõna, mis esineb tekstis, teisendada algvormi. Tabelis 1 on võimalik näha kõiki võtmesõnu, mis seostavad lauset tegevusega, ja teksti analüsaatori tagastatavat väärtust iga tegevuse tuvastamise korral. Kui tegevust ei tuvastata, siis serveri tagastatav väärtus on „null”. Tabeli 1 viimases veerus kujutatakse ka võtmesõnu, millega täiendati testijärgselt (vt ptk 5.2.1 ja ptk 5.2.2) programmi.

Tegevus	Analüsaatori tagastatav väärtus	Võtmesõnad	Pärast testimist lisatud võtmesõnad
Meedia esitamine	start_media	alustama, mängima	esitama
Meedia peatamine	stop_media	lõpetama, peatama	kinni, välja
Helitugevuse tõstmine	increase_volume	tõstma, suurendama, valjemalt, valjem, vali, kõvemalt, kõvem, kõva, tugevamalt, tugevam, tugev, kõrgem, kõrge	juurde
Helitugevuse vähendamine	decrease_volume	nõrgemalt, nõrgem, nõrk, madalamalt, madalam, madal, vaiksemalt, vaiksem, vaikne	vähendama, alandama tasemalt, tasem, tasa
Otsingu tegemine	search	otsima, leidma, googeldama, {kõik küsisõnad va. kui}	kui
Meeldetuletuse seadmine	reminder	teavitama, tuletama	-

Tabel 1. Tekstianalüsaatori tuvastatavad tegevused koos võtmesõnade ja tagastatavate väärtustega.

### 3.3 Kõnesüntesaator

Kõnesüntesaator on tarkvara, mis võtab sisendiks teksti ning loob selle põhjal helifaili, mis imiteerib inimkõnet [19]. Eestikeelse teksti sünteesimine kõneks on võimalik aastast 2002, kui väljastati kõnesüntesaatori esimene versioon Eesti Keele Instituudi, Küberneetika Instituudi ja ettevõtte Filosoft OÜ poolt [20]. Sellest ajast on tarkvara täiendatud ning viimane versioon, 01.05.2021 seisuga, on avalikustatud 2019. aasta märtsis.

Kõnesünteesi serverirakendust saab kasutada aadressil <https://teenus.eki.ee/konesyntees>, kust leiab muuhulgas ka kasutusjuhendi, kuidas veebiteenusele tuleb päringuid esitada. Kõik parameetrid edastatakse URL-is ehk vastuse saamiseks on vaja ainult teha veebiaadressile korrektses formaadis päring. Kui soovitakse kuulda lauset „Tere-tere, vana kere!”

tehiskõnena, siis tuleb teha GET päring või minna brauseris aadressile [https://teenus.eki.ee/konesyntees?tekst=""Tere-tere, vana kere!](https://teenus.eki.ee/konesyntees?tekst=). Vastusena saadakse JSON last kahe väljaga: esimese nimi on „mp3url” ja teise nimi „wavurl”. Mõlema väärtused on veebiaadressid tehiskõne helifailidele, mille rakendus lõi, ning URL-e külastades kuuleb masinat kõnelemas näitena toodud fraasi. Soovi korral saab lisaparameetriga muuta veel masinkõne diktsiooni kiirust. Kasutajajuhendi eeskujul prooviti ka sünteesitud hääle muutmist, kuid see parameeter väljundit ei muutnud.

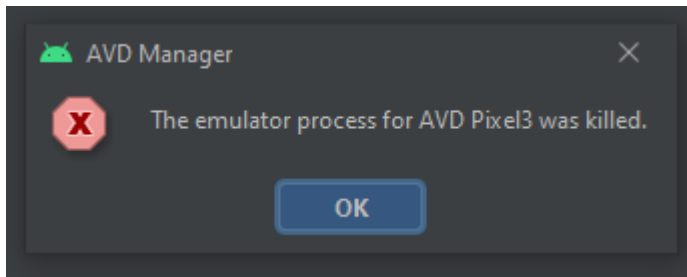
## 4. Häälassistendi raamistiku eessüsteem

### 4.1 Nutiseadme rakenduse platvorm

Töö käigus loodi rakendus, mida saab kasutada nutiseadmetes. Nendes enim esinevad operatsioonisüsteemid on Android ja iOS, olles kasutuses ligikaudu 99% seadmetest [24]. Android on 2020. aasta seisuga kõige populaarsem nutiseadmete operatsioonisüsteem nii mujal maailmas [24] kui ka Eestis [25]. Eesti Vabariigis hõivab Android nutitelefonide turust 67,78%, mis on märgatavalt suurem, kui lähim konkurent iOS turuhõivega 31,36%. Kuna Android on Eesti kasutajate seas populaarne, siis tehti eestikeelse häälassistendi rakendus sellele operatsioonisüsteemile. Leiduvad programmeerimise raamistikud nagu React [26], mis võimaldavad luua nutiseadme rakendust, mis töötab mõlemal operatsioonisüsteemil, kuid platvormist sõltumatu arenduse suurim vajakajäämine on kohati puudulik süsteemi funktsioonide kasutamise võimalus ning puuduv tugi vähem levinud kolmanda osapoole tarkvara osas [27]. Need on olulised takistused virtuaalassistendi puhul, sest terviklik virtuaalassistendi lahendus peab pääsema ligi paljudele seadme riistvaralistele ja tarkvaralistele elementidele ning käesoleva süsteemi puhul tuleb kasutada ühte kolmanda osapoole tarkvara, Kõnele-service, mis töötab Android, kuid mitte iOS platvormil. Seetõttu otsustati töö raames keskenduda ühele operatsioonisüsteemile.

Ettevõtte Google on mitmel olulisel viisil lihtsustanud rakenduste loomist Android platvormile. Üks neist on Android Studio IDE [28], mis on tarkvara arenduskeskkond Androidi rakenduste loomiseks. Selle olulised võimalused on jälgida rakenduse ressursikasutust, emuleerida erinevaid nutiseadmeid ja lubada arendajal muuta rakenduse kujundust mugava graafilise kasutajaliidese abil. Teiste IDE-dega sarnaselt suudab Android Studio IDE automaatselt lõpetada arendaja kirjutamisel olevat koodilõiku ning pakkuda integreeritud tööriistu versioonihalduse ja koodi silumise jaoks. Lisaks nendele võimalustele on loodud Android SDK [29], mis võimaldab arendajal teke kasutades luua hõlpsalt rakenduse jaoks vajalikke komponente, vaateid, teavitusi ja kasutada operatsioonisüsteemi teenuseid [29].

Android Studio puuduseks on nutiseadmete emulaatori halb töökindlus. Tõsine probleem on emulaatori võimetus salvestada mikrofoniga heli [30]. Lahendus on rakendust katsetada füüsilises seadmes näiteks arendaja enda nutitelefonis, kus saab heli salvestada mikrofoniga. Teine puudus emulaatoriga on see, et tõrgete esinemisel antakse kohati kasutajale ebamäära tagasisidet. Emuleerimise protsessi käivitamisel võib saada teate, et tegevus on lõpetatud (vt Joonis 5) kuvamata töö katkemise põhjust või lisainformatsiooni. Kasutajatel on vea põhjustanud graafikakiirendi vale seadistus, probleemid arvuti viirusetõrje tarkvaraga või aegunud Android SDK tööriistad [31], kuid veateade ei viita ühelegi konkreetsele probleemile. Lõputöö käigus esines seesama tõrge ning viga oli üldse kaustas, kuhu salvestati emulaatori failid. Nimelt kausta teekonnas esines täht 'ü', mida Androidi IDE ära ei tundnud, ja seega ei leitud ka vajalikku faili ülesse. Selliste probleemide puhul peaks kasutajale tagasiside olema selgem, kuid hetkel on veateated väga üldised. Probleemidest hoolimata on seadmete emuleerimine kasulik, et hinnata iga võimaliku ekraanitüübi korral kasutajaliidese korrektsust.



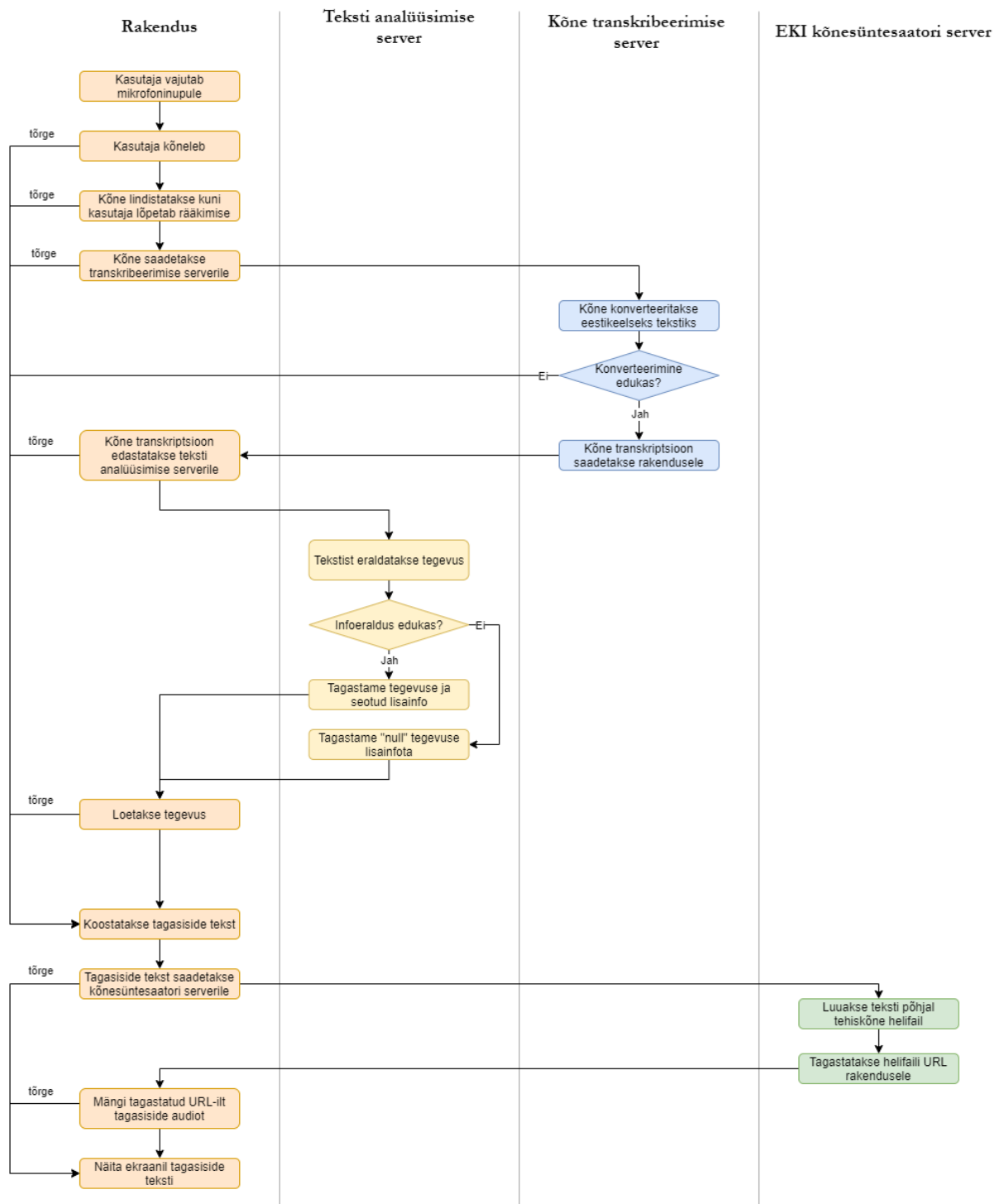
Joonis 5. Emulaatori ebamäärane veateade Android Studio arenduskeskkonnas.

## 4.2 Häälassistendi rakenduse töövoog

Eestikeelse häälassistendi raamistiku tööd juhib nutiseadme rakendus. See suhtleb kõikide serveritega ning otsustab vastuste põhjal, mida järgmisena tuleb teha. Raamistiku tavapärase töövoog algab, kui kasutaja avab rakenduse ning vajutab kasutajaliideses mikrofoni nupule (vt Lisa 1). Seejärel kuulatakse kasutaja kõne kuni rääkimine lõpeb ning ühtlasi peatatakse mikrofoni salvestamine. Kõnele-service tarkvara tegeleb kõne transkribeerimisega ning tagastab vastuse rakendusele. Kõne tekst edastatakse teksti analüüsimise serverile, mis eraldab käsklusest tegevuse ning tagastab selle äpile. Rakendus otsustab vastuse põhjal, millist tegevust on vaja täitma hakata, kuid kuna raamistik toimingut seadmes ei soorita, siis antakse kasutajale tagasiside justkui oleks soovitud tegevus sooritatud edukalt (vt Lisa 4). Täpsem töövoog kirjeldus on näha Jooniselt 6. Rakendus on avalikult saadaval GitHubis veebiaadressil <https://github.com/leppsalujyrgen/eestikeelse-haalassistendi-raamistik>.

Kuigi tavaolukorras peaks rakendus töötama korrektselt, võib rakenduse töö käigus tekkida tõrkeid. Seetõttu peab rakendus tegelema veahaldusega. Virtuaalassistendi äpp suudab kasutajale anda adekvaatset tagasisidet, kui mikrofon ei tööta korrektselt, seade ei pääse internetti ja serveritega ei saada ühendust, kusjuures on täpsustatud millise serveriga probleem esineb. Muude vigade puhul tagastatakse üldine veateade, mis juhendab kasutajat pöörduma tehnilise toe poole.

Tagasisidet antakse nii kasutajaliideses teksti kujul kui tehiskõne juhistena. Kuna seadme ekraanil on ruum piiratud, siis kuvatakse sinna kompaktne teade rakenduse tegevuste kohta. Kõne teel tagasiside andmisel on samuti pikkuse piirang, sest korduvaid teavitusi on tüütu pikalt kuulata, kuid sellegipoolest saab heliliselt kirjeldada äpi olekut detailsemalt kui ekraanil. Seetõttu on igal võimalikul tagasiside tüübil üks tekst, mida näidatakse ekraanil teatena, ja teine analoogne tekst, mida esitatakse masinkõnena. Teavitus jääb ekraanile kuniks vajutatakse mikrofoni nuppu, et uut häälkäsklust sisestada.



Joonis 6. Häälassistendi raamistiku töövoog.

## **5. Häälassistendi raamistiku süsteemianalüüs**

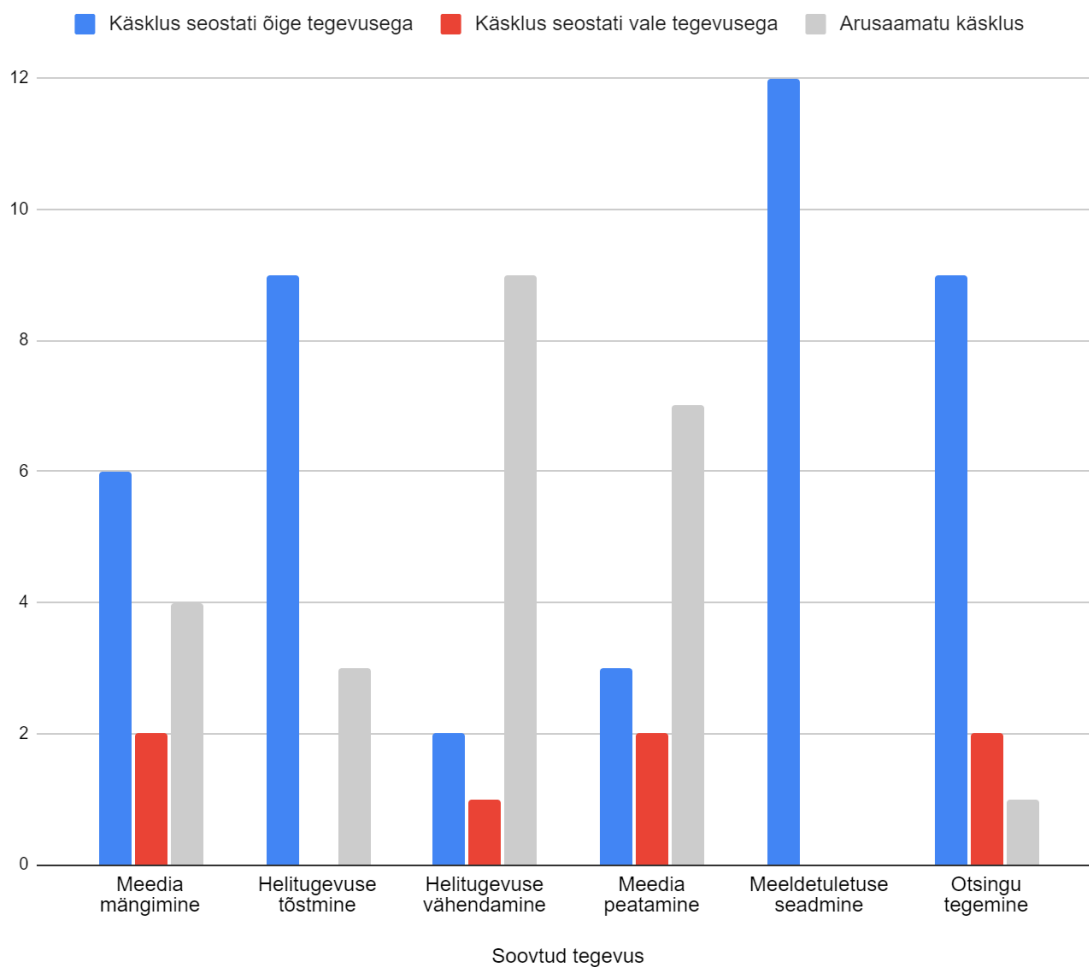
### **5.1 Metoodika**

Häälassistendi raamistiku tööd oli vaja hinnata. Selleks korraldati katsed nelja vabatahtlikuga, kes esitasid rakendusele käsklusi, ning vaadati, kas süsteem tuvastab õige tegevuse käskluse transkriptsiooni põhjal. Süsteem tuvastab kuut tegevust: helitugevuse suurendamine ja vähendamine, meedia mängimine ja peatamine, otsingu tegemine ning meeldetuletuse seadmine. Sellest tulenevalt paluti katseisikutel koostada iga toiminguga 3 sõnastuselt erinevat käsklust ja kirjutada need ankeeti (vt Lisa 5-10). Seejärel esitati korraldused häälassistendile ning jäädigi ootama rakendusepoolset tagasisidet tuvastatud tegevuse osas. Rakendusele esitatud käskluse transkriptsioon ja sellega seostatud toiming salvestati ka tekstianalüüsi serveri logifaili. Lisaks korrektsete käskluste koostamisele paluti vabatahtlikel mõelda veel 3 sellist käsku, mis ei vasta ühelegi eelnimetatud tegevusele, et hinnata süsteemi võimet tuvastada tundmatuid palveid. Viimasena küsiti osalejatelt, milliseid teisi tegevusi peaks eestikeelne häälassistent tundma, et uurida, mida potentsiaalsed kasutajad süsteemilt veel tahta võivad.

Selleks, et mitte suunata katsealuseid kasutama kindlaid sõnu, prooviti katsete käigus esitada küsimusi kindlas vormis. Kõigepealt kirjeldati üldiselt olukorda, mida häälassistent peaks oskama lahendada, ning seejärel küsiti kasutajalt, mida peaks ütleva häälassistendile, et jõuda soovitud tulemuseni. Näiteks kui sooviti, et katsealune mõtleks 3 käsklust, mis kirjeldaksid helitugevuse vähendamist, esitati talle järgmise situatsiooni kirjeldus ja küsimus: “Meediapleier mängib, aga loo vahetudes taipate, et nüüd on kõlaritest kostuv heli liiga vali Teie jaoks. Mida ütleksite häälassistendile?”.

### **5.2 Tulemused**

Katsete kokkuvõtte on Joonisel 7. Tulpdiagrammist on näha, et läbiviidud katsetes iga tegevuse kategooria puhul seostatakse kasutaja kõne pigem õige tegevusega kui muu suvalise tegevusega. Kõige kindlamalt suudab süsteem tuvastada, kui kasutaja räägib meeldetuletuse seadmisest, otsingu tegemisest või helitugevuse suurendamisest. Siiski on näha, et sageli ei seostata kõne transkriptsiooni ühegi tegevusega, mida kirjeldavad tulbad nimega “Arusaamatu käsklus”.

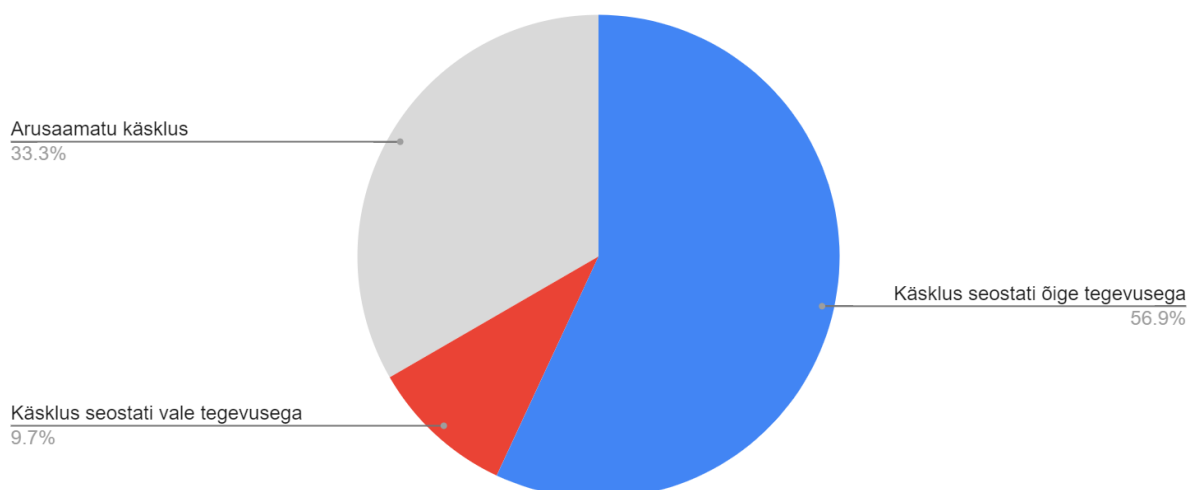


Joonis 7. Katsete käigus õigete ja valede tegevustega seostatud käskluste arv.

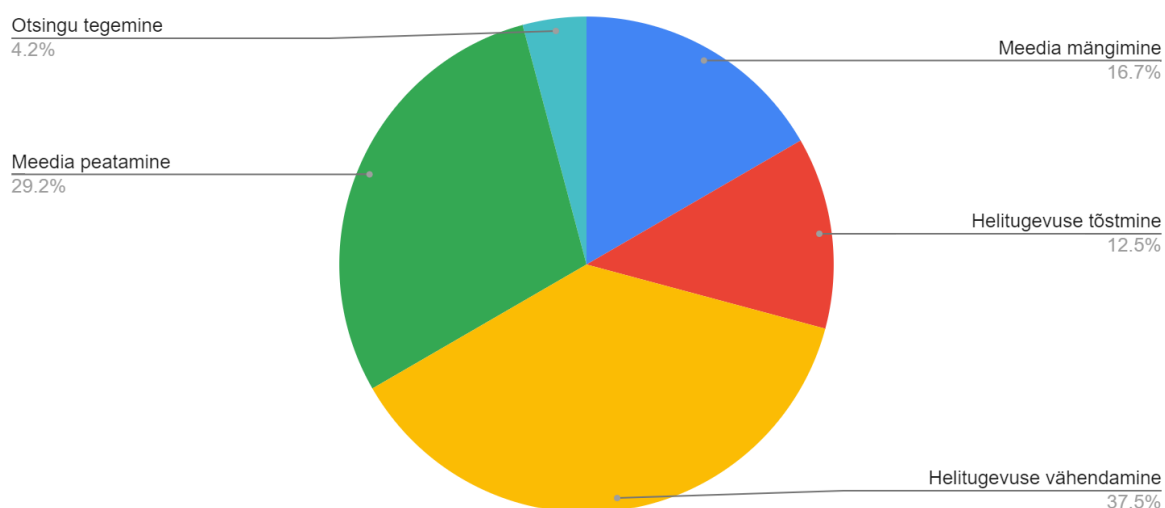
Joonis 8 näitab, et need ütlused, mida häälassistent ei suuda liigitada, moodustavad kolmandiku katsetes esitatud käskudest. Selle jagunemine erinevate tegevuste kategooriate vahel on kajastatud Joonisel 9. Selgelt on süsteemil raskusi meedia peatamise ja helitugevuse vähendamise käskude mõistmisega, sest üheskoos moodustavad need 66,7% kõikidest liigitamata käsklustest. Parandades nende käskluste tuvastamist suureneks oluliselt leitud õigete tegevuste arv.

Osalejate koostatud käsklused, mis ei tohtinud vastata ühelegi süsteemi poolt tuvastatavale tegevusele, jäeti süsteemi poolt liigitamata 11 korral 12-st. See on positiivne tulemus, mis näitab, et praeguste andmete korral ei hakka süsteem tundmatut käsku seostama suvalise tegevusega.

Kuna valim oli väike, siis ei saa tulemuste põhjal teha üldistusi, kuidas kasutajad suhtlevad häälassistendiga ja millised on häälassistendi raamistiku tugevused. Sellest hoolimata võib katsete tulemuste põhjal leida nõrkusi loodud süsteemis ning neid analüüsida. Lihtsamad probleemid saab lahendada käesoleva töö raames ning keerulisemate puhul leida potentsiaalseid lahendusi, mida edasiarendustena implementeerida.



Joonis 8. Katsete käigus õigete ja valede tegevustega seostatud käskluste osakaal.



Joonis 9. Liigitamata käskluste ülevaade, arvestades iga tegevuse kategooriat, kuhu arusaamatu käsk oleks pidanud kuuluma.

### 5.2.1 Vead ja kitsaskohad

Teksti analüüsimisel kasutatakse mustrisobitus meetodit, mis toetub programmis defineeritud sõnade nimekirjale, et seostada käsklus toiminguga. Seega kui häälassistendile öeldu ei sisalda ühtegi sõna, mis tekstianalüüsi programmi jaoks tuntud on, klassifitseeritakse kasutaja sisend tundmatu käsklusena. See probleem esines katsetes palju, eriti helitugevuse vähendamise korral. Mõned näited on Tabelis 2, kus sõnad, mida ei tuvastata, on „esita”, „vaiksemaks” ja „alanda”.

Esitatud käsklus	Käskluse transkriptsioon	Leitud tegevus	Soovitud tegevus
VOVA esita lugu "Good News".	nova esita lugu goodnews	null	start_media
Muuda helitugevust 10% vaiksemaks.	muuda helitugevust 10% vaiksemaks	null	decrease_volume
Alanda helitugevust	alanda helitugevust	null	decrease_volume

Tabel 2. Puuduliku sõnavara tõttu tuvastamata jäänud käsud.

Kõne vigane transkriptsioon on samuti üks põhjus, miks kasutaja soovitud operatsiooni ei tuvastata. Tabelis 3 on näha, et mõne käskluse puhul transkribeeritakse kõne niivõrd valesti, et õige tegevuse leidmine on võimatu. Kergemate eksimuste puhul kompenseerib mustrisobitus meetod vigase sisendteksti puudujäägi, sest süsteem vajab ainult teatud sõnade korrektset transkribeerimist, et korraldusest eraldada õige tegevus. Näitena toodud vigade korral pole aga sellestki abi.

Katsete käigus esines paaril korral olukord, kus häälkäskluse osa „ma ei” transkribeeriti kui „mai”, mis põhjustas vale tegevuse tuvastamise (vt Tabel 4). Seda konkreetset juhtu saaks käsitleda tulevastes versioonides, kus „mai” esinemise korral vaadatakse paralleelselt lauset, kus „mai” on asendatud sõnapaariga „ma ei” ning otsustatakse kumb on lause kontekstis loogilisem. Mõistuspärasema lause järgi tegevuse määramine parandab õige tegevuse leidmise tõenäosust.

Lisaks nendele probleemidele ei suuda eestikeelse kõne transkribeerija võõrkeelseid nimesid ning sõnasid korrektselt transkribeerida (vt Tabel 5). See on mõistetav, kuid täisväärtusliku häälassistendi jaoks on vajalik nendeski olukordades saada öeldust korrektne kirjalik esitus, sest näiteks paljud populaarsed muusikapalad ja taskuhäälingud on ingliskeelsete nimedega. Samuti kasutasid katseisikud otsingu tegemisel kõnes ingliskeelseid termineid ja nimetusi (vt Tabel 6).

Esitatud käsklus	Käskluse transkriptsioon	Leitud tegevus	Soovitud tegevus
Heli kinni.	hereetiline	null	stop_media
Volüüm maha.	ilmamaa	null	decrease_volume
Heli tasemaks.	heli tase max	null	decrease_volume

Tabel 3. Vigased transkriptsioonid, kust õige tegevuse leidmine on võimatu.

Esitatud käsklus	Käskluse transkriptsioon	Leitud tegevus	Soovitud tegevus
Ma ei taha enam muusikat kuulata.	mai tahan muusikat kuulata	reminder	stop_media
Ma ei taha enam muusikat kuulata.	mai taha enam muusikat kuulata	reminder	stop_media

Tabel 4. Vigased transkriptsioonid, kust õige tegevuse leidmine on abimeetodiga võimalik.

Esitatud käsklus	Käskluse transkriptsioon	Leitud tegevus	Soovitud tegevus
Mängi Drake „Pound cake“.	mängid raidkivi.ee	start_media	start_media
Pane Joe Rogani podcast käima.	reparatsioon on a boot cast käima	null	start_media
Mängi "Summer of 69".	mängis sambarose sixty nine	start_media	start_media
Leia seis Manchester Unitedi ja Tottenhami vahel.	leia s.a.s. manchesteri näidete ja tottenhami vahel	search	search
Kes on filmi Titanic režissöör?	kes on filmi titaanid režissöör	search	search

Tabel 5. Transkriptsioonid, mis sisaldavad inglisekeelseid sõnu või nimesid.

Meeldetuletuse seadmisega oli kaks erisugust probleemi ning mõlemad tulenesid teksti analüüsimise serverirakendusest. Esimese vea põhjustas teksti analüüsimise otsustusmehhanism, mis pidi valima tegevuste hierarhia järgi tegevuse. Nimelt otsustati ennatlikult tagastada nutiseadmele meeldetuletuse seadmise tegevus vaatamata, kas käsk sisaldab otsingut käivitavat tegusõna. Seega mehhanism eksis defineeritud hierarhia vastu ning tagastas vale toiming (vt Tabel 6). Muudel juhtudel selliseid vastuolusid ei tekkinud. Teise vea puhul palus kasutaja rakendusel mängida selle aasta pop-lugu, kuid meedia mängimise asemel otsustati meeldetuletuse seadmise tegevuse kasuks (vt Tabel 7). Selle olukorra lahendamise nõuab täpsemate mustrite loomist, mis näiteks muusikapala mängimise palve korral välistaks meeldetuletuse ja otsingu tegevuse.

Esitatud käsklus	Käskluse transkriptsioon	Leitud tegevus	Soovitud tegevus
VOVA otsi tänaseid NBA tulemusi.	loova otsi tänaseid tulemusi	reminder	search

Tabel 6. Käsklus, mille korral tagastati vale toiming ebakorrekse tegevuste hierarhia järgimise tõttu.

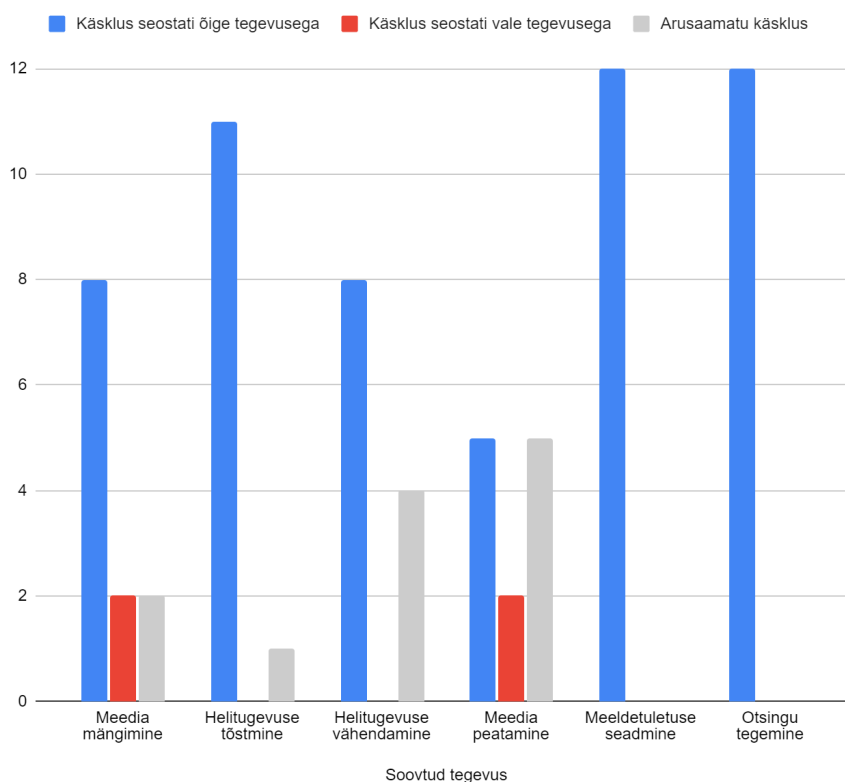
Esitatud käsklus	Käskluse transkriptsioon	Leitud tegevus	Soovitud tegevus
Mängi mõni selle aasta pop-lugu.	mängi mõni selle aasta poplugu	reminder	start_media

Tabel 7. Käsklus, mille korral tagastati vale toiming tegevuste hierarhiat järgides.

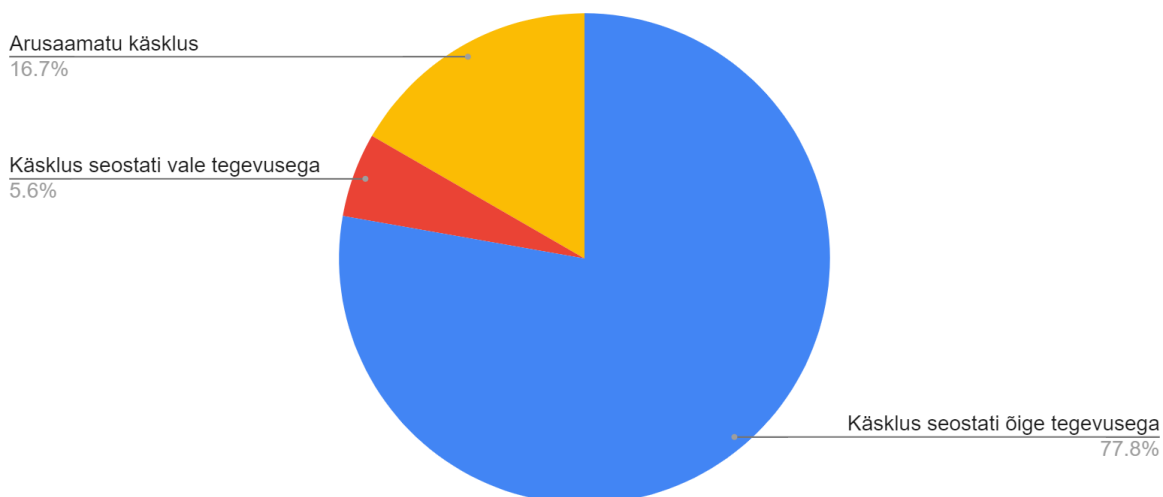
### 5.2.2 Lihtsamate parandustega süsteemi tulemused

Lihtsamad vead, mis ilmnesid katsete käigus, parandati tekstianalüüsi programmis. Nendeks olid puudujäägid süsteemile teadaolevate sõnade osas ning meeldetuletuse seadmise klassifitseerimisel eksimine iseenda tegevuste hierarhia vastu. Parandatud tekstianalüüsi programmi sisestati katsete käigus talletatud käskluse transkriptsioonid ning vaadati, milliseid tegevusi analüüsirakendus tagastab. Tulemused (vt Joonis 10 ja Joonis 11) näitavad, et tõsisid tegevuste õige liigitamine, samal ajal kui valesti liigitatud ja arusaamatute käskluse arv vähenes.

Vabatahtlike tagasisidest saadi veel teada, et praegune häälassistendi kasutajaliides, mis enne uue häälkäsu sisestamist sundis kuulama lõpuni tehiskõnelist tagasisidet, polnud mugav. Selleks täiendati Androidi rakendust nii, et iga kord kui häälkäskluse sisestamiseks vajutatakse mikrofoni nuppu, tühistatakse käimasolev transkribeerimise või tagasiside andmise protsess ja alustatakse uuesti kasutaja kuulamist.



Joonis 10. Parandatud teksti analüüsivõime mudeli õigete ja valede tegevustega seostatud käskluse arv.



Joonis 11. Parandatud teksti analüüsimise mudeli käigus õigete ja valede tegevustega seostatud käskluste osakaal.

### 5.2.3 Edasiarendused

Häälassistendi raamistiku süsteemil on mitmeid edasiarenduse võimalusi. Esimene võimalus, kuidas süsteemi parendada, on leida viis, kuidas kõnest eraldada võõrkeelseid termineid ja nimesid. Eriti oluline on see inglise keele puhul, sest katsetes esines olukordi, kus osalejad kasutasid otsingu tegemisel ja meediapleieri käivitamisel ingliskeelseid väljendeid ja nimetusi. Rakenduse kasutajad võivad soovida ka näiteks venekeelseid või soomekeelseid nimesid kasutada, seega võiks edasiarendus olla laiendatav teistele keeltele samasuguse mudeli põhjal.

Teine edasiarendusviis on lisainformatsiooni eraldamine käsklusest, mis täpsustab tegevuse olemust. Näiteks kui kasutaja soovib reguleerida helitugevust teatud protsendi võrra, siis oleks hea, kui süsteem suudaks selle lisaparametri tagastada nutiseadme rakendusele. Teised olukorrad, kus täiendavat teavet oleks tarvis muust sisendtekstist eraldada, on otsingu tegemisel päringu tuvastamine, meeldetuletuse puhul kirjelduse ja korrektse tähtaja leidmine ning meedia esitamise korral soovitud loo eraldamine.

Raamistiku funktsionaalsust saab laiendada, et tuvastatud tegevused realiseeritaks nutiseadmes. Tähelepanu tuleb siinkohal pöörata kindlasti meeldetuletuse seadmise toimingule, sest ebakorrekse kõne- või tekstitöötlemise korral võib rakendus seada kalendrisse soovimatuid sündmuseid ning põhjustada seeläbi kasutajale ebameeldivusi. Probleemi vältimiseks võib küsida kasutajalt kinnitust meeldetuletuse aja ja sündmuse kohta enne, kui teavitus kalendrisse kirjutatakse. Teine variant oleks talletada rakenduses tegevuste ajalugu ning lubada inimesel tühistada seadmes tehtud muudatused äpi kasutajaliidesest.

## 6. Kokkuvõte

Bakalaureusetöö käigus loodi häälassistendi Androidi rakendus ja teksti analüüsimise serverrakendus. Koos Tallinna Tehnikaülikooli kõne transkribeerimise serveri ja Eesti Keele Instituudi kõnesüntesaatori serveriga moodustavad need komponendid virtuaalassistendi raamistiku, mis suudab kasutaja kõne põhjal tuvastada kuut erinevat tegevust ja anda kasutajale asjakohast tagasisidet. Tundmatute käskluste puhul ja erinevate vigade esinemisel teavitatakse kasutajat ning pakutakse kasutajale võimaluse korral lahendusi probleemide likvideerimiseks. Kasutajaga suhtlemine käib rakenduses tekstilise ja tehiskõnena esitatud tagasiside kaudu. Autori loodud tarkvara ja eelnimetatud serverite kasutamist kirjeldatakse töö käigus vastavate koodinäidete ja selgitustega.

Raamistikku testiti 4 osalejaga, kes esitasid virtuaalassistendi rakendusele käsklusi. Uuriti, kas süsteem tuvastab kasutaja soovitud tegevuse ning enamikel kordadest suudeti õige tegevus tuvastada. Esialgse tekstianalüüsi programmi puhul seostati käsklused õige tegevusega 56,9 protsendil kordadest ning vale tegevus määrati 9,8 protsendil käsklustest. Katsetejärgsed parandused tõstsid käskude sidumist õige tegevusega 77,8 protsendile ning vähendasid vale tegevuse määramist 5,6 protsendini.

Bakalaureusetööl on mitmeid edasiarenduse võimalusi. Esimene neist on täiendada virtuaalassistendi raamistiku Androidi rakendust, et seade soovitud tegevuse tuvastamisel ka teostab vajaliku toimingut. See muudaks raamistiku täisväärtuslikuks virtuaalassistendiks. Teine võimalus on uurida, kuidas tuvastada kõnest võõrkeelseid väljendeid ja nimesid, sest nende tundmine on rakendusele vajalik näiteks laulu otsimisel. Viimane variant on täiendada teksti analüüsimise programmi, et käskluste transkriptsioonist saaks leida ka tegevusi täpsustavaid parameetreid.

## Viidatud kirjandus

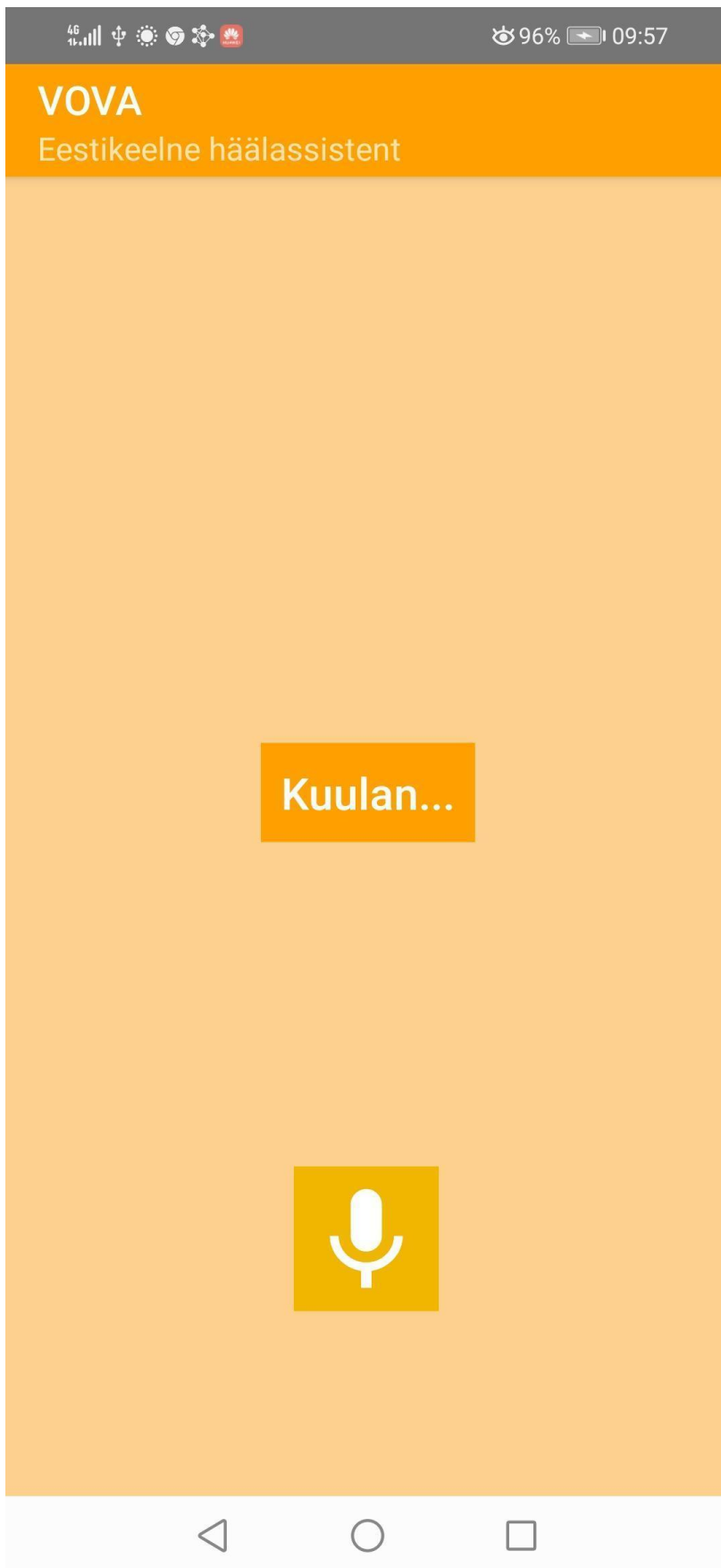
- [1] Eesti keele arengukava 2021–2035.  
[https://www.hm.ee/sites/default/files/eesti\\_keeles\\_arengukava\\_2021-2035\\_29.10.2020\\_riigikogusse.pdf](https://www.hm.ee/sites/default/files/eesti_keeles_arengukava_2021-2035_29.10.2020_riigikogusse.pdf) (07.05.2021)
- [2] McLaughlin M. What a Virtual Assistant Is and How It Works. *LifeWire*, 2020.  
<https://www.lifewire.com/virtual-assistants-4138533> (25.04.2021)
- [3] Liu R. The History And Growth of Voice AI and Search. *Medium*, 2019.  
<https://medium.com/cornertechandmarketing/the-history-and-growth-of-voice-ai-and-search-a0b7b4d37986> (25.04.2021)
- [4] Dikteeri. <http://bark.phon.ioc.ee/dikteeri/> (25.04.2021)
- [5] Kõnele kasutusjuhend. [http://kaljurand.github.io/K6nele/docs/et/user\\_guide.html](http://kaljurand.github.io/K6nele/docs/et/user_guide.html)  
(25.04.2021)
- [6] The rise of Virtual Personal Assistants.  
<https://ati.ec.europa.eu/reports/technology-watch/rise-virtual-personal-assistants> (07.05.2018)
- [7] Garcia M. P., Lopez S. S., Donis H. jt. 2018. Everybody is talking about Voice Activated Virtual Assistants, but how are people really adopting and using them? Lessons from a multi-country study. Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI). 1-3.
- [8] Ammari T., Kaye J., Tsai J., Bentley F. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Transactions on Computer-Human Interaction*. 10-14.
- [9] Overview of memory management.  
<https://developer.android.com/topic/performance/memory-overview> (25.04.2021)
- [10] Reducing Your App's Memory Use.  
[https://developer.apple.com/documentation/xcode/improving\\_your\\_app\\_s\\_performance/reducing\\_your\\_app\\_s\\_memory\\_use](https://developer.apple.com/documentation/xcode/improving_your_app_s_performance/reducing_your_app_s_memory_use) (25.04.2021)
- [11] Alumäe T. 2014. Full-duplex speech-to-text system for Estonian. *Human Language Technologies - the Baltic Perspective : Proceedings of the Sixth International Baltic Conference*. 3–10.
- [12] Kaldi GStreamer server. <https://github.com/alumae/kaldi-gstreamer-server> (04.05.2021)
- [13] Docker.  
<https://www.ibm.com/cloud/learn/docker#:~:text=Docker%20is%20an%20open%20source%20containerization%20platform,the%20code%20in%20any%20environment> (04.05.2021)
- [14] docker-kaldi-gstreamer-server. <https://github.com/jcsilva/docker-kaldi-gstreamer-server>  
(04.05.2021)

- [15] pip. <https://pypi.org/project/pip/> (04.05.2021)
- [17] dicat.js. <https://github.com/Kaljurand/dicate.js> (04.05.2021)
- [18] Kõnele service. <https://github.com/Kaljurand/K6nele-service> (04.05.2021)
- [19] Eesti keele tekst-kõne süntees.  
<http://www.eki.ee/keeletehnoloogia/projektid/syntees/tnks.html> (04.05.2021)
- [20] Sünteesimeetodite tutvustus ja allalaadimine.  
<https://www.eki.ee/heli/index.php/k%C3%B5nes%C3%BCntees> (04.05.2021)
- [21] Run Curl Commands Online. <https://reqbin.com/curl> (04.05.2021)
- [22] Working with text. <https://estnltk.github.io/estnltk/1.4.1/tutorials/text.html> (04.05.2021)
- [23] Kitani T., Eriguchi Y., Ilara M. 1994. PATTERN MATCHING IN THE TEXT EXTRACT INFORMATION EXTRACTION SYSTEM. Proceedings of the 15th conference on Computational linguistics. 1064-1070.
- [24] Mobile Operating System Market Share Worldwide.  
<https://gs.statcounter.com/os-market-share/mobile/worldwide> (17.03.2021)
- [25] Mobile Operating System Market Share Estonia.  
<https://gs.statcounter.com/os-market-share/mobile/estonia#monthly-200901-201710>  
(17.03.2021)
- [26] React Native. <https://reactnative.dev/> (17.03.2021)
- [27] Klubnikin A. Cross-platform vs Native Mobile App Development: Choosing the Right Development Tools for Your Project. *Medium*, 2017.  
<https://medium.com/all-technology-feeds/cross-platform-vs-native-mobile-app-development-choosing-the-right-dev-tools-for-your-app-project-47d0abafee81> (17.03.2021)
- [28] Meet Android Studio. <https://developer.android.com/studio/intro> (17.03.2021)
- [29] Sarkar A., Goyal A., Hicks D., Sarkar D., Hazra S. Android Application Development: A Brief Overview of Android Platforms and Evolution of Security Systems. 2019. Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). 73-79.
- [30] MediaRecorder overview.  
<https://developer.android.com/guide/topics/media/mediarecorder> (17.03.2021)
- [31] ERROR Android emulator gets killed.  
<https://stackoverflow.com/questions/36841461/error-android-emulator-gets-killed>  
(17.03.2021)

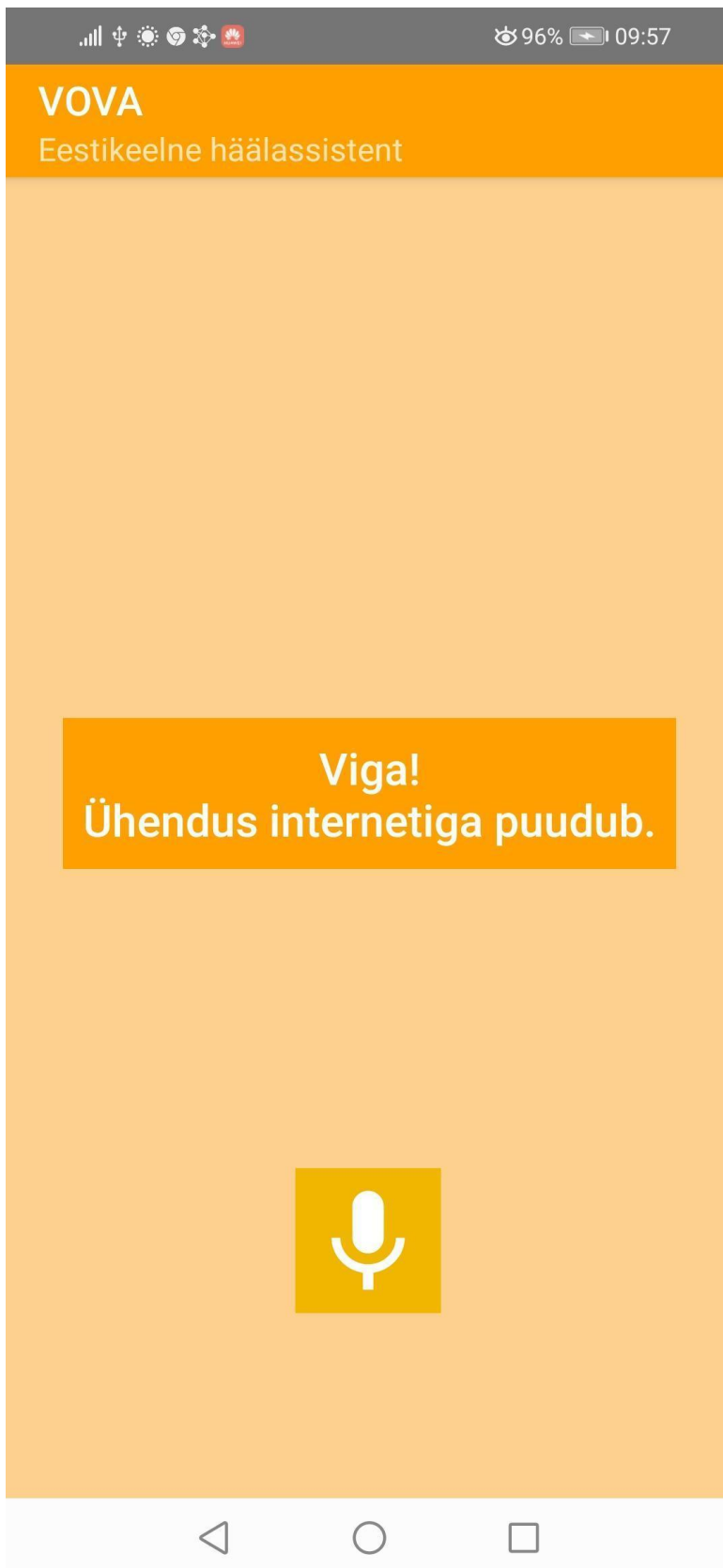
## Lisad



Lisa 1. Androidi rakenduse kasutajaliides avamisel.



Lisa 2. Androidi rakenduse kasutajaliides peale mikrofoninupu vajutamist.



Lisa 3. Androidi rakenduse kasutajaliides, kui kõne transkribeerimisel tuvastati internetiühenduse puudumine.



Lisa 4. Androidi rakenduse kasutajaliides kui tuvastati meediapleieri käivitamise tegevus.



## Häälassistendi töö hindamine

Küsimused Vastused 4

4 vastust

Aktsepteerib vastuseid 

Kokkuvõte

Küsimus

Individuaalne

## Täisnimi

4 vastust

Kaarel Leppsalu

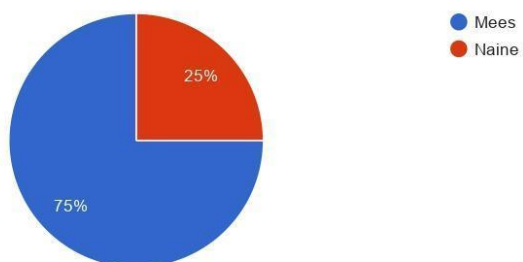
Martin Valdo Randpere

Stella-Mirell Polakene

Kaspar Leppsalu

## Sugu

4 vastust



Lisa 5. Katses osalejatele esitatud ankeedi esimene lehekülj koos tulemustega.

## Vanus

4 vastust

15

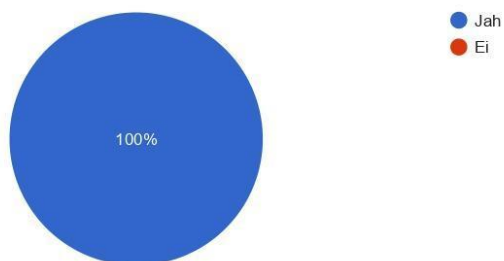
22

21

25

## Kas lubate kasutada oma antud vastuseid ja tagasisidet Jürgen Leppsalu bakalaureusetöös?

4 vastust



## Katsed häälassistendi rakendusega

Istute oma toas ja ümbruses on kuidagi vaikne. Soovite kuulata telefonist head lugu või taskuhäälingut. Mida ütleksite häälassistendile?

4 vastust

VOVA mängi "Good News".

Hei, VOVA mängi laulu "Good News".

VOVA esita lugu "Good News".

Mängi muusikat

Pane Joe Rogani podcast käima

Mängi Drake „Pound cake“

Mängi summer of 69, mängi midagi kurba, pane lugu

Ava muusikaplayer.

Mängi suvalist lugu minu playlist'ist.

Mängi mõni selle aasta pop-lugu.

Lisa 6. Katses osalejatele esitatud ankeedi teine lehekülj koos tulemustega.

Teie soov sai täidetud. Käivitati soovitud meediavorm ning see kostub kõlaritest. Kahjuks aga on see natuke liiga vaikne Teie jaoks. Mida ütleksite häälassistendile?

4 vastust

Tõsta heli 30% võrra.  
Tõsta volüümi 80% peale.  
Suurenda helitugevust 20%.

Heli kõvemaks  
Tõsta volüümi  
Liiga vaikne on, valjemaks heli palun

pane kõvemaks, volüümi juurde, tümm põhja

Suurenda helitugevust.  
Keera volüümi juurde.  
Mängi lugu kõrgema helitugevusega.

Meediapleier mängib, aga loo vahetudes taipate, et nüüd on kõlaritest kostuv heli liiga vali Teie jaoks. Mida ütleksite häälassistendile?

4 vastust

Langeta heli 15% võrra.  
Muuda helitugevust 10% vaiksemaks.  
Lase muusikat 20% vaiksemalt.

Liiga vali on heli  
Vähenda volüümi  
Kõlar karjub liialt

pane vaiksemaks, volüüm maha, heli tasemaks

Vähenda helitugevust  
Alanda helitugevust  
Mängi muusikat vaiksemalt.

Lisa 7. Katses osalejatele esitatud ankeedi kolmas lehekülj koos tulemustega.

Telefon on nüüd mänginud meediat juba pikalt ja Teie ei jaksa seda enam kuulata. Mida ütleksite häälassistendile?

4 vastust

VOVA pane muusika pausile.

Vaigista laul.

Lülita muusika välja.

Pane muusika kinni

Pane heli vaig

Ma ei taha enam muusikat kuulata

heli kinni, muusika kinni, peata muusika

Lõpeta muusika mängimine

Lõpeta lugu

Ma ei taha enam muusikat kuulata.

Teile on tulnud teade sündmusest või kohustusest mida tahate meeles pidada. Soovite, et telefon seaks meeldetuletuse. Mida ütleksite häälassistendile? Palun mõelge aeg ja soovi korral ka sündmus ise välja.

4 vastust

VOVA lisa meeldetuletus homseks kell 10.

Lisa alarm homseks kell 10.

Tuleta meelde mulle homme kell 10.

Pane kalendrisse, et kolmapäeval kell kaks päeval on trenn

Sea meeldetuletus, et reedel kell 6 hommikul on rehvivahetus

Pane nädalplaani kirja igapäevane õhtul kõndimine kell 22

pane meeldetuletus homme kella 15ks, tuleta meelde, et reedel on ema sünnipäev, sea meeldetuletus aasta aja pärast

Sea meeldetuletus: näitude võtmine 30. aprillil kell 10

Sea meeldetuletus: hambaarsti aeg 10. mail kell 12

Sea meeldetuletus: bussiaeg homme kell 7 õhtul

Lisa 8. Katses osalejatele esitatud ankeedi neljas lehekülj koos tulemustega.

Teil on tekkinud küsimus või soov leida infot internetist. Mida ütleksite häälassistendile? Palun mõelge konkreetne küsimus või infopäring ise välja.

4 vastust

VOVA otsi tänaseid NBA tulemusi.  
Leia seis Manchester Unitedi ja Tottenhami vahel.  
Otsi informatsiooni hariliku kuuse kohta

Kui pikk on Hiina müür  
Kes oli USA eelmine president  
Mis kuupäeval rünnati USA kaksiktorne

millal on suur reede? kes on eesti president? kui vana on uku suviste?

Kes on filmi Titanic režissöör?  
Millal on nähtav järgmine päikesevarjutus Tartus?  
Mis on Rootsi pealinn?

Katsed häälassistendi rakendusega

Mida ütleksite häälassistendile, kui soovime, et õeldu/käsk ei vastaks ühelegi tegevuse kategooriale?

4 vastust

Räägi mulle üks nali.  
Loe ette sõnumit kasutajalt Jürgen Leppsalu.  
VOVA jäta meelde järgnevad sõnad: tomat, piim, sealiha.

Vaigista sissetulevad kõned  
Helista emale  
Ava interneti browser

ava kaamera, lülita ennast välja, tee pilt

Ava äpp Postimees  
Kuva homne ilmateade  
Kustuta äpp Delfi

Lisa 9. Katses osalejatele esitatud ankeedi viies lehekülj koos tulemustega.

Milliseid teisi tegevusi peaks Teie arvates eestikeelne häälassistent veel tundma?

3 vastust

helistamine kellelegi

Timeri seadmine meeldetuletustest eraldi

Rakenduste avamine

Kui käskluse esitamine läks sassi, siis võiks olla viis peatada käskluse tuvastamine.

Taimer/ stopperi seadistamine.

Lisa 10. Katses osalejatele esitatud ankeedi kuues lehekülj koos tulemustega.

## **Litsents**

### **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Jürgen Leppsalu,

1. Annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose Eestikeelse häälassistendi raamistiku loomine, mille juhendaja on Sven Aller, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

*Jürgen Leppsalu*

**06.05.2021**