

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Martin Karjus 164215IAPB

Veebipõhine platvorm lastele eesti keele lugemisoskuse treenimiseks ja hindamiseks

Bakalaureusetöö

Juhendaja: Einar Meister
PhD

Tallinn 2021

Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Autor: Martin Karjus

14.06.2021

Annotatsioon

Selle töö eesmärk on luua rakendus 6-10 aastastele lastele lugemisoskuse treenimiseks läbi karaoke stiilis lugemisharjutuste ja implementeerida admin-keskkond, kus saab lisada uusi harjutusi. Antud rakendus loob kindlustusfunktsionaalsuse kaudu ka võimalused eesti laste kõnekeele korpuse kogumiseks.

Töös on kirjeldatud lugemisoskuse tähtsust laste hariduses ning tutvustatud sarnaseid olemasolevaid lahendusi. Töö põhiosas on lahti seletatud rakenduse arenduskäik, kasutatud tehnilised vahendid ning demonstreeritud valminud funktsionaalsused. Töö tulemusena valmis veebirakenduse esialgne prototüüp, mida saab kasutada aadressil <https://readingkaraoke.xyz>. Rakenduse valideerimiseks viidi läbi pilootuuring.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 21 leheküljel, 8 peatükki, 12 joonist, 1 tabelit.

Abstract

Web-based Platform for Children for Reading Practice and Assessment in Estonian

The goal of this thesis is to create an application for 6–10 year-old children to train their reading skills through the use of karaoke style reading exercises and to implement an admin environment where new exercises can be added. Through its recording functionality it facilitates the compilation of children's speech in Estonian.

The thesis gives an overview of the importance of reading skills in child education and currently existing similar digital learning solutions. The main part gives an overview of the application development process and the technologies used, and demonstrates the created functionalities. This work resulted in a web application available at <https://readingkaraoke.xyz>. It was also tested in a small-scale pilot study.

The thesis is in Estonian and contains 21 pages of text, 8 chapters, 12 figures, 1 table.

Lühendite ja mõistete sõnastik

SPA	<i>Single-page application</i> , üheleherakendus
SQL	<i>Structured Query Language</i> , Struktuurpäringukeel
JPA	<i>Java Persistence API</i>
XML	<i>Extensible Markup Language</i> , laiendatav märgistuskeel
REST	<i>Representational state transfer</i>
ORM	<i>object-relational mapping</i> , object-relatsiooniline kaardistamine
API	<i>Application programming interface</i> , rakendusliides
CSS	<i>Cascading Style Sheets</i> , kaskaadlaadistik
HTML	<i>HyperText Markup Language</i> , hüperteksti märgistuskeel
WAV	<i>Waveform Audio File Format</i>
JSON	<i>JavaScript Object Notation</i>
GDPR	<i>General Data Protection Regulation</i> , Isikuandmete kaitse üldmäärus
TLS	<i>Transport Layer Security</i> , Transpordikihi turbeprotokoll
Droplet	DigitalOceani poolt pakutav serveriruumi rentimisteenus

Sisukord

1 Sissejuhatus.....	10
2 Taust ja eesmärk	12
2.1 Eesmärk	12
3 Olemasolevad lahendused ja probleemid	14
4 Rakenduse arendus	16
4.1 Arendusprotsess	16
4.2 Tagarakendus.....	16
4.2.1 Spring Boot.....	17
4.2.2 Maven	17
4.2.3 Lombok.....	17
4.2.4 Autosegmenteerija 2.0	17
4.2.5 EstNLTK.....	18
4.2.6 Andmebaas	18
4.2.7 Andmebaasi struktuur.....	18
4.3 Eesrakendus	19
4.3.1 Vue.....	19
4.3.2 Kujundus.....	20
4.4 Teenusepakkujad	20
5 Rakendus.....	21
5.1 Esileht	21
5.2 Sisselogimine ja registreerimine	21
5.3 Admin	23
5.3.1 Lause lisamine	23
5.3.2 Lausete haldamine	24
5.4 Harjutama	25
5.5 Tulemused.....	26
6 Valideerimine	27
6.1 Juhendajate tagasiside.....	27
6.2 Õppijate tagasiside.....	27

7 Edasiarendus võimalused.....	28
7.1 Juhendajate rollid ja õppijate grupid.....	28
7.1.1 Lausegrupid	28
7.2 Kõnetuvastuse ja muud rakendused.....	28
7.2.1 Korpuse kättesaadavaks tegemine lastekõne analüüsiks	29
7.2.2 Häälde analüüs	29
7.3 Rakenduse laiendamine teistele keeltele	30
8 Kokkuvõte.....	31
Kasutatud kirjandus.....	32
Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks.....	34
Lisa 2 – Tagasiside vorm	35

Jooniste loetelu

Joonis 1. Ekraanipilt: Kiirlugemise tarkvara lugemisharjutus lugemise soodustajaga (roheline kursor).	15
Joonis 2. Ekraanipilt Labl rakendusest.	15
Joonis 3. Andmebaasi skeem.	19
Joonis 4. Ekraanipilt esilehest.....	21
Joonis 5. Ekraanipilt registreerimise vaatest.	22
Joonis 6. Ekraanipilt sisselogimise vaatest.	22
Joonis 7. Ekraanipilt admin-alas lause lisamisest.....	24
Joonis 8. Ekraanipilt admin-alas lause halduse vaatest.	24
Joonis 9. Lugemise karaoke harjutus. Lugemisel olev sõna on punane.	25
Joonis 10. Lugemise karaoke harjutuse lõpp.	26
Joonis 11. Ekraanipilt tulemuste vaatest.....	26
Joonis 12. Ekraanipilt näidisest, kuidas võiks välja näha sõna valimine.	29

Tabelite loetelu

Tabel 1. Tagasiside vormi tulemused.	35
---	----

1 Sissejuhatus

Lugemine on üks olulisemaid oskusi, mida on vaja omandada noorelt, see on üheks peamiseks sammuks laste haridusredelil. Kui tähed selged ja lauseid veeritakse, on tähtis lugemist harjutada, et saavutada selles soravus. Kindlasti on neid, kellele tuleb see välja ilma suurema probleemita, kuid paljudel on vaja seda pikemalt harjutada.

On mitmeid rakendusi, mis on mõeldud nii lugema õppimiseks kui lugemise harjutamiseks, kuid eestikeelseid on vähe. Lisaks keskenduvad rakendused üldiselt neile, kes kas alles õpivad tähti, või neile, kes tahavad harjutada kiirlugemist.

Rakendust, mis keskenduks lugemisoskust omandavatele lastele, on paraku vähe.

Selle töö eesmärk on luua rakendus lugemisoskuse treenimiseks läbi karaoke stiilis ette antud lugemisharjutuste. Rakenduses on implementeeritud ka uute harjutuste lisamise funktsionaalsus. Admin-allas saavad administraatori õigustega kasutajad lisada lauseid koos ise sisse loetud lindistustega. Rakendus segmenteerib automaatselt sisse loetud laused ning loob sellest lugemisharjutused. Lisaks kogub rakendus kasutajate helinäiteid ning sellega on võimalik kaasa aidata eestikeelse laste kõnekorpuse suurendamisele.

Siin töös loodaval rakendusel on kaks kasutajatüüpi: õppijad ja juhendajad. “Õppijaks” nimetatakse siin töös kõiki kasutajaid, kes kasutavad rakendust lugemisoskuse treenimiseks. “Juhendajaks” nimetatakse siin töös kõiki kasutajaid, kelle roll on lugema õppijate toetamine ja juhendamine. Rakenduse põhiline sihtgrupp on lapsed vanuses 6-10, ehk need, kes omandavad lugemisoskust, kuid ei loe veel tingimata täiskasvanu tasemel. Lapsed, kelle lugemisoskus või arvutikasutamise oskus ei ole piisav, saavad rakendust kasutada juhendaja toel. Juhendajatel on ka võimalik hõlpsasti lisada uusi harjutusi.

Valminud rakendust testisime väikese pilootkatse korras mõlema sihtgrupi peal. Tagasiside oli positiivne ning avaldati lootust, et rakenduse arendamine jätkub. Siin töös loodud rakendus on küll pigem prototüüp, kuid see on hea alus laiema funktsionaalsusega

lugemise harjutamise platvormile, mida on võimalik tulevikus edasi arendada. Praegune versioon asub aadressil <https://readingkaraoke.xyz>.

2 Taust ja eesmärk

Varakult lugema õppimine toetab õppeedukust tulevikus. Uuringud on näidanud, et esimeses klassis hästi lugema õppinud (USA) lastel on keskkoolis tõenäolisemalt paremad lugemis- ja arusaamisoskused ning neil on suurem huvi kirjanduse vastu kui õpilastel, kellel algkoolis lugema õppimine võttis kauem aega [1]. Teises uurimuses, mis teostati Soomes, leiti, et edukas varajane tähtede omandamine esimeses klassis ennustas lugemisoskuse erinevusi algkooli lõpus [2].

Osad lapsed omandavad lugemise ja tekstide mõistmise oskusi kiiremini ja ka paremini kui eakaaslased. Üheks põhjuseks võivad olla erinevused kognitiivsetes võimetes [3], kuid mitte alati. Eesti algkoolilaste lugemisoskuse uuringus [4] ei leitud seost kirjandushuvi ja lugemisoskuse vahel, kuid tuvastati seos erinevate õpetamismeetodite ja laste tekstimõistmise oskuste vahel.

Lugemisoskuse hindamine on loomulikult keeruline [5][6]. Näiteks on leitud, et isegi standardiseeritud testid nii sarnastes keeltes nagu on eesti ja soome keel ei ole päris võrreldavad [7]. Seega, kui tahta kasutada laste lugemaõpetamisel digimeetodeid, tuleb lahendada vähemalt kaks probleemi: (1) õppevahend ise ja (2) meetod, kuidas hinnata õpilase progressi. Üks variant automaatselt lugemisoskust hinnata on kasutada masinõppelahendusi, kuid Eesti kontekstis puuduvad piisavalt suured laste kõnekorpused, mille peal treenida masinõppe lahendusi.

2.1 Eesmärk

Eesti keeles puudub tarkvara, mis aitaks 6–10 aastastel lastel lugema õppida ning lugemist harjutada. Selles töös luuakse lahendus lugemisoskuse treenimiseks lastele ning kasutatakse lugemise sujuvuse hindamiseks lihtsustatud mõõdikut, mis põhineb lugemiskiiruse mõõtmisel sõnade ja silpide arvu kaudu (vt [8]). See ei pruugi olla küll alati täpne, kuid aitab lugemisoskust kiirelt ja automaatselt hinnata. Selliseid digilahendusi on varem tehtud, kuid mitte eesti keele jaoks (vt peatükk 3).

Ka eesti kõnekeele korpustes on väikelapsed alaesindatud. Rahvusvahelise CHILDES korpuse eestikeelne osa¹ sisaldab küll laste kõne transkriptsioone, aga ainult kuni kolmeaastaste omi. Mahukaim eesti laste kõnekorpus [9], sisaldab 9–18 aastaste poiste ja tüdrukute (kokku 309) kõnesalvestusi. Siin töös välja töötatud rakendus loob lisaks eeldused koostada suurem eesti laste kõne korpus ühisloome (ingl. *crowdsourcing*) kaudu ning panustab seeläbi eesti kõnetehnoloogia arengusse. Korpust, mida oleks võimalik luua siin töös väljatöötatud veebirakenduse abil, saaks hiljem kasutada ka sellesama platvormi hilisemaks täiendamiseks ja muudeks kõnetehnoloogia rakendusteks (vt peatükk 7).

Käesolev töö võtab eeskujuks varasemaid lahendusi ja uurimistöid, kus on kasutatud karaoke stiilis visuaalseid lahendusi lugema õppimisel ning harjutamisel [10] [11] [12] [13]. Sellistes lahendustes näevad õppijad ekraanil lauseid või sõnu ning loevad või laulavad neid kaasa vastavalt visuaalsetele märguannetele. Siin töös ei kasutata küll laulmist ja muusikat, aga põhimõte on sama nagu karaokes – lapsed loevad lauseid kaasa, jälgides kasutajaliideses värviliselt märgitud sõnu.

Lisaks peab uute harjutuste lisamine olema lihtne, et see oleks kõigile erinevatele sihtgruppidele lihtsalt kasutatav.

Kokkuvõttes, siin töös loodud rakendus täidab kahte eesmärki – aitab lastel lugemist harjutada ning kogub keelenäiteid laste keelekorpuse täiendamiseks.

¹ CHILDES. Child Language Exchange System. <http://childes.psy.cmu.edu/data/>

3 Olemasolevad lahendused ja probleemid

Siin peatükis tuuakse välja olemasolevaid lugemise harjutamise lahendusi nii eesti kui inglise keeles.

Kiirlugemise tarkvara¹ on eestikeelne veebipõhine kiirlugemise harjutamise rakendus, mida kasutatakse ka Tartu Ülikoolis kiirlugemise kursusel [13]. Rakenduses on implementeeritud lugemisharjutused viiel erineval moel:

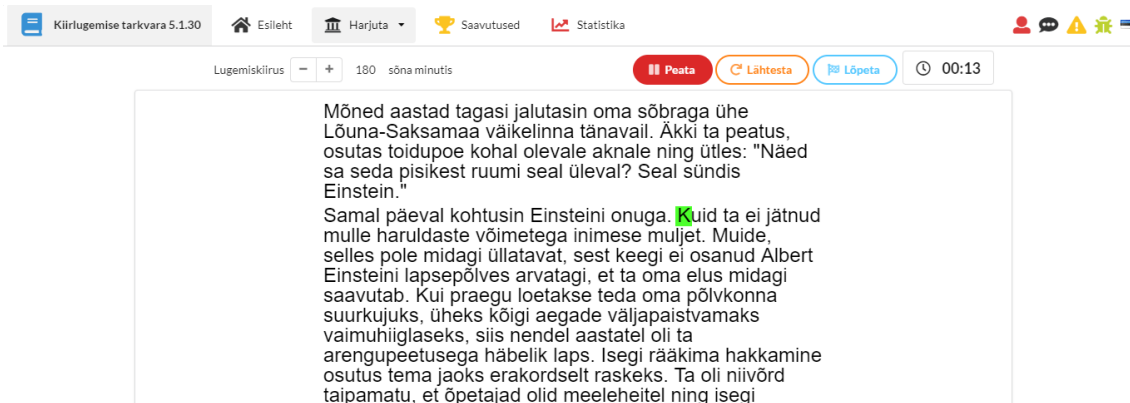
1. lugemistest – tavaline teksti lugemine stopperiga,
2. lugemise soodustaja (Joonis 1) – lugeja püüab teksti lugedes jälgida pilgujuhti,
3. keriv tekst – tekst kerib järjest ning lugeja peab teksti ära jõudma lugeda enne kui see vaateväljast kaob,
4. kustuv tekst – tekst kustub järjest lugemisväljalt ära,
5. sõnarühmad – lugeja püüab lugeda allajoonitud sõnagruppi.

Kõikidel varjantidel peale esimese saab kasutaja valida ise oma lugemiskiiruse, valides mitu sõna sekundis ta loeb ning see määrab visuaalse tagasiside kiiruse.

Peale teksti lugemist täidab kasutaja teksti omandamistaseme kontrolli, milleks on küsimustik. Pärast täitmist saab kasutaja tehtud harjutuse kohta tagasisidet – kui palju kulus aega ja mitu küsimust õigesti vastati. Rakendus on mõeldud eelkõige täiskasvanutele kiirlugemise õppimiseks, aga tekstide valikus on ka lastele suunatud tekste. Tekste saab ka ise lisada.

Teksti kiirus sõltub ainult tähtede arvust ja ei võta arvesse seda, kui kaua kulub tegelikkuses sõna lugemiseks aega.

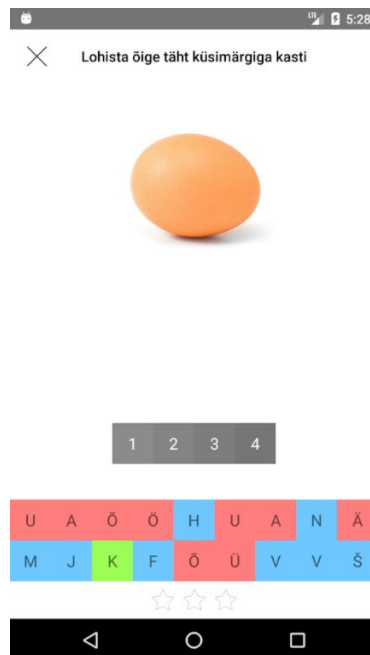
¹ kiirlugemine.keeleressursid.ee/



Joonis 1. Ekraanipilt: Kiirlugemise tarkvara lugemisharjutus lugemise soodustajaga (roheline kursor).

Homer¹ on inglisekeelne 2–8 aastastele suunatud rakendus IOS ja Android platvormidel. Selles saab õppida tundma tähti, sõnu ning harjutada lugemist. Õppimine toimub mängulises vormis interaktiivsete ülesannetega.

Labl² (Joonis 2) on eestikeelne 4+ aastastele lastele mõeldud tähtede ja sõnade õppimise rakendus Android ja IOS platvormidel, millega lapsed saavad õppida tähti, sõnu ja lauseid läbi mänguliste harjutuste.



Joonis 2. Ekraanipilt Labl rakendusest.

¹ apps.apple.com/us/app/learn-homer-reading-educational/id601437586

² play.google.com/store/apps/details?id=com.margusholland.labl

4 Rakenduse arendus

Selles peatükis on kirjeldatud rakenduse arendusprotsessi, peamisi arenduseks kasutatud teeke ja raamistikke.

4.1 Arendusprotsess

Rakenduse loomist alustati nõuete täpsustamisega ja kirjeldamisega ning vaadete visandamisega, nende alusel loodi git'i repositooriumisse esialgsed *issue*'d, et oleks lihtne jälgida töö kulgu. Sellele järgnes rakenduse arhitektuuri valik, lähtudes nõuetest ja autori isiklikust kogemusest. Nõudeid implementeeriti järk-järgult. Kui nõuded said implementeeritud, teostati nende manuaalne testimine ning valideerimine. Valideerimist teostati koos juhendajaga, kes andis tagasisidet, mille põhjal tehti implementeeritud funktsionaalsustele muudatusi.

Rakenduse kood on kirjutatud arendusvahendiga IntelliJ IDEA¹. Eesrakenduse Vue komponentide implementeerimisel on kasutatud Vue.js pistikprogrammi².

4.2 Tagarakendus

Tagarakendus on arendatud RESTful teenusena, mille kaudu saab veebirakendus suhelda tagarakendusega ja pärida läbi selle andmebaasi andmeid. Rakenduses toimub ka lisatud lausete automaatne töötlus ning kasutajatele tagasiside genereerimine.

¹ jetbrains.com/idea/

² plugins.jetbrains.com/plugin/9442-vue-js

4.2.1 Spring Boot

Tagarakenduse loomiseks kasutati Java Spring Boot¹ raamistiku. Spring Boot on Spring'i² peale loodud tarkvararakendus raamistik, mis lihtsustab tarkvaraarendust automaatsete Spring konfiguratsioonidega ja võimaldab ilma XML'i kasutamata teha konfiguratsiooni muudatusi.

4.2.2 Maven

Maven³ on avatud lähtekoodiga programmide halduse tööriist, mis aitab hallata ehitisi, tarkvara konfiguratsiooni ning teekide sõltuvusi. Teegid laeb Maven alla Maven repositooriumist projekti XML konfiguratsioonifaili järgi.

Projektis on kasutusel kolm konfiguratsioonifaili – tagarakenduse, eesrakenduse ning kogu projekti jaoks, et jooksutada vastava komponendi skripte.

4.2.3 Lombok

Selleks, et vähendada trafarettkoodi (ingl. *boilerplate code*) kirjutamist, on projektis kasutusel Lombok⁴. Lombok genereerib Java annotatsioonide põhjal protseduurid sisse baitkoodi kompileerimise hetkel. See hoiab lähtekoodi puhtamana ning loetavamana. Peamiselt on see kasutusel *getter*-, *setter*- ja tühjade konstruktormetodite genereerimiseks.

4.2.4 Autosegmenteerija 2.0

Autosegmenteerija 2.0 on veebirakendus, millega saab segmenteerida eestikeelset kõnet hääliku tasemel kasutades selleks kõnetuvastuse jaoks loodud akustilisi mudeleid. Teenus on kasutatav ka läbi API. Segmenteerija võtab sisendiks WAV formaadis audio faili ning audio failile vastava transkriptsiooni. Väljundiks on Praat⁵ TextGrid fail, milles on kaks kihti – sõnade ja häälikute hääldusintervallid, mõlemal kihil on märgitud ka pausid [14].

¹ spring.io/projects/spring-boot

² spring.io/

³ <https://maven.apache.org/>

⁴ projectlombok.org/

⁵ praat.org

4.2.5 EstNLTK

EstNLTK on avatud lähtekoodiga tööriistakomplekt eesti loomuliku keele töötlemiseks. Komplekt sisaldab mitmed funktsionaalsusi nagu nimeolemite tuvastust, morfoloogilist analüüsi, lõigu, lause ning sõna tokeniseerimist, automaatset silbitamist ja palju muud [15].

4.2.6 Andmebaas

Projektis kasutatakse PostgreSQL¹ andmebaasi, kuna see on avatud lähtekoodiga ja laialt levinud andmebaasi süsteem ning autoril on sellega eelnev kogemus nii tööl kui ülikoolis. Arenduse ajal oli kasutusel H2² andmebaas, mis võimaldab jooksutada kiiret lokaalset andmebaasi mälus, et arenduskeskkonda ei peaks eraldi paigaldama ega seadistama andmebaasi.

Andmebaasiga suhtlemiseks on kasutusel **Spring Data JPA**³(Java Persistence API) ja **Hibernate**⁴. Spring Data JPA on andmete juurdepääsu abstraktsioon, mis võimaldab lihtsamalt teha andmebaasipäringuid läbi JPA'd pakkuva teegi, milleks antud projektis on Hibernate. Spring Data JPA kasutab genereeritud funktsioone, et suhelda andmebaasiga ning seda kasutades ei ole vaja ise SQL päringuid kirjutada.

Hibernate on objekt-relatsioonilise kaardistamise (ingl. *object-relational mapping*, ORM) Java teek, mis võimaldab siduda domeeni mudeleid relatsioonandmebaasiga.

4.2.7 Andmebaasi struktuur

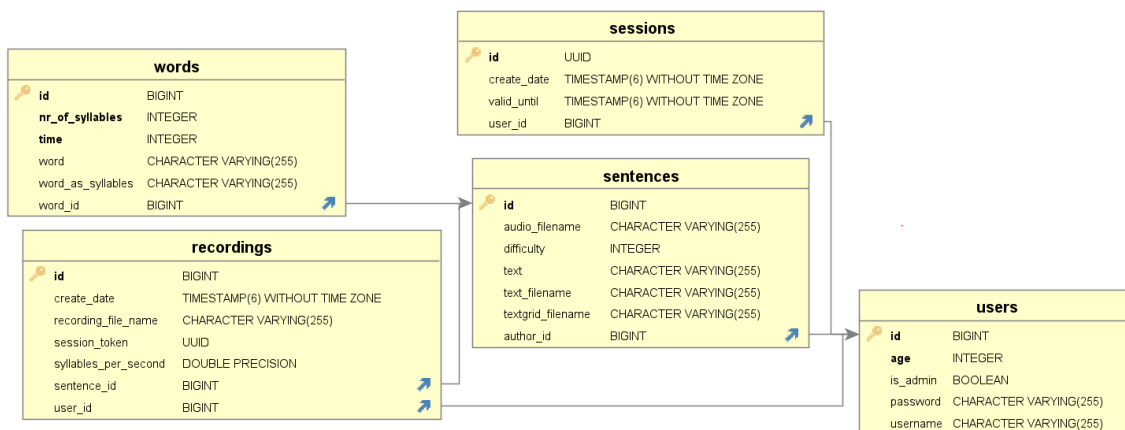
Andmebaasi skeem on genereeritud Spring Data JPA'ga annoteeritud domeeni objektide põhjal (vt. Joonis 3).

¹ postgresql.org

² h2database.com

³ spring.io/projects/spring-data-jpa

⁴ hibernate.org



Joonis 3. Andmebaasi skeem.

4.3 Eesrakendus

Eesrakenduse implementeerimiseks võeti kasutusele Javascript'i raamistik Vue.js¹ ning mitmeid teegid. Vue.js valiti kuna sellega saab luua komponendipõhist ja kergesti hallatavat ning arendatavat SPA (*single-page application*, üheleherakendus) kasutajaliidest; autoril on sellega varasem kogemus. Kasutajaliidese teekide haldamiseks kasutati NPM'i².

4.3.1 Vue

Vue.js on JavaScripti raamistik üheleheliste rakenduste loomiseks, mis võimaldab renderdada lehte reaktiivselt. Tänu sellele on võimalik kasutada veebilehel muutujaid, mis uuenevad koos muutuja väärtusega ning ka tingimusliku renderdamist, kasutades elementide märgistesse sisestatud tingimuslauseid.

¹ vuejs.org

² npmjs.com

Vue's on võimalik rakendada taaskasutatavaid komponente, mis koosnevad kolmest osast: mall (*template*), skript (*script*) ja stiil (*style*).

- Mallis kirjeldatakse komponendi visuaal kasutades HTML'i.
- Skriptis paikneb komponendi Javascript.
- Stiilis on defineeritud komponendis kasutusel olev CSS, mida on võimalik kapseldada ainult antud komponendi sisse.

4.3.2 Kujundus

CSS (ingl *Cascading Style Sheets*) kujunduse jaoks kasutatakse töös Bootstrap¹ ja Bootstrap-vue² ning Bootstrap'i teemat Minty³.

4.4 Teenusepakkujad

Rakendus on paigaldatud DigitalOceani⁴ pilvekeskkonda Droplet⁵ virtuaalmasinasse. Virtuaalmasina operatsioonisüsteemiks on Ubuntu 20.04 LTS (ingl *long-term support*) ning sellega on seotud staatiline IP aadress. Andmevahetus on krüpteeritud, TLS sertifikaadi pakkuja on Let's Encrypt⁶.

¹ getbootstrap.com

² bootstrap-vue.org

³ bootswatch.com/minty/

⁴ digitalocean.com

⁵ digitalocean.com/products/droplets/

⁶ letsencrypt.org

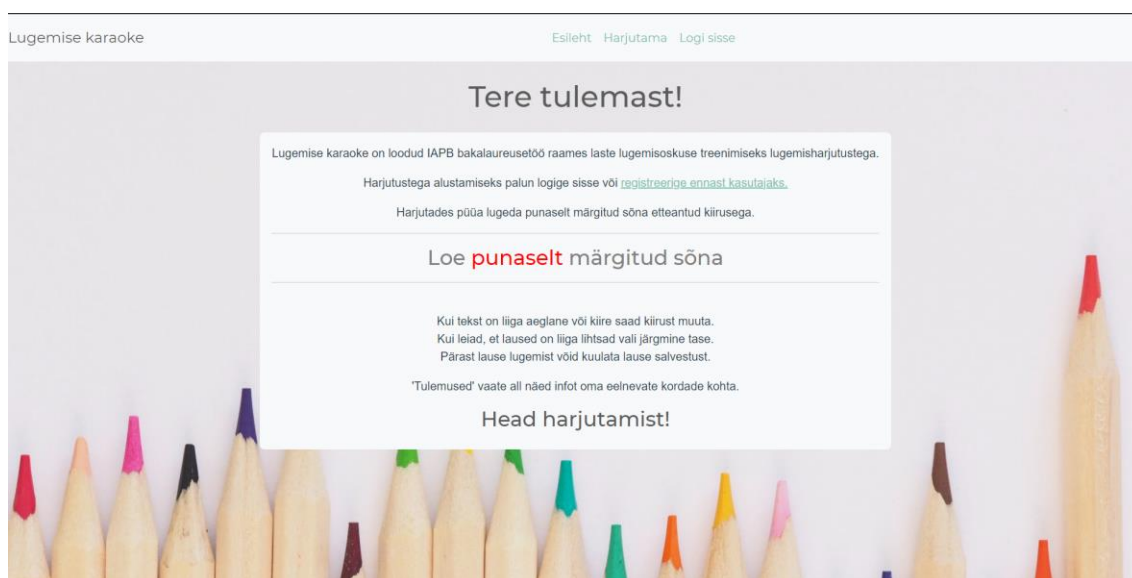
5 Rakendus

Siin peatükis on välja toodud implementeeritud esirakenduse vaated ja neid toetavad tagarakenduse funktsionaalsused.

Eesrakenduses on loodud vaated: Esileht, Harjutama, Admin, Tulemused, Sisselogimine ja Registreerimine.

5.1 Esileht

Esilehel paikneb lühike juhend, milles on lühidalt seletatud tavakasutajale, kuidas rakendust kasutada (Joonis 4). Admin kasutaja juhend asub admin alal.



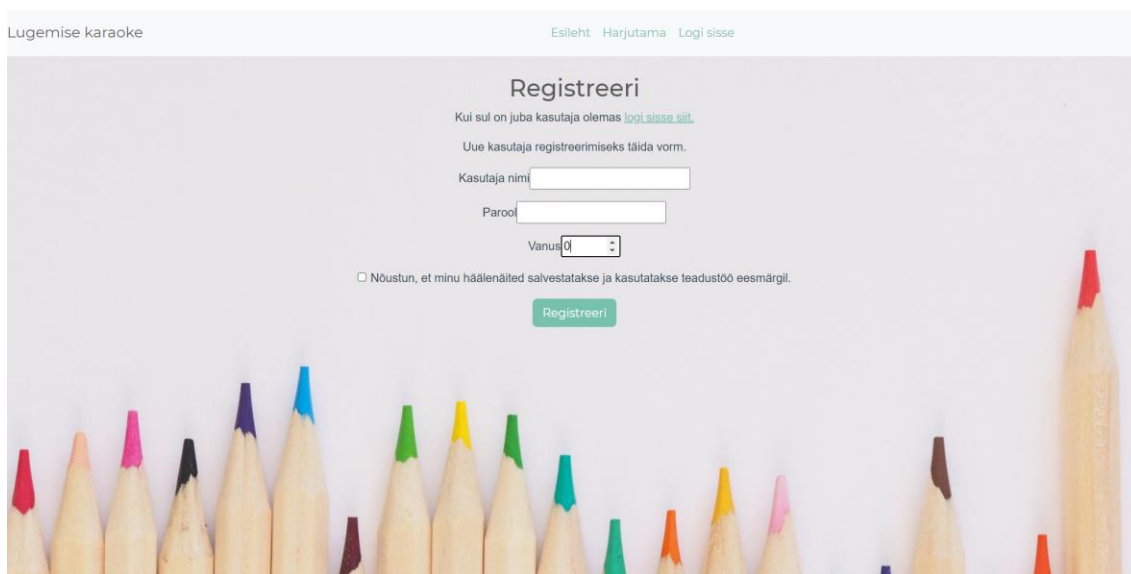
Joonis 4. Ekraanipilt esilehest.

5.2 Sisselogimine ja registreerimine

Kasutaja peab rakenduse kasutamiseks ennast registreerima (Joonis 5). Registreerimiseks on vaja sisestada kasutaja nimi, parool, vanus ning nõustuma, et kasutaja häälenäiteid salvestatakse ja kasutatakse teadustöö eesmärgil. Kasutustingimustega nõustumine on vajalik, kuna Euroopa Liidu Isikuandmete kaitse üldmäärus¹ (ingl *General Data Protection Regulation*) näeb ette, et kasutaja peab teadma, mis tema andmetega tehakse.

¹ <https://eur-lex.europa.eu/legal-content/ET/TXT/HTML/?uri=CELEX:32016R0679>

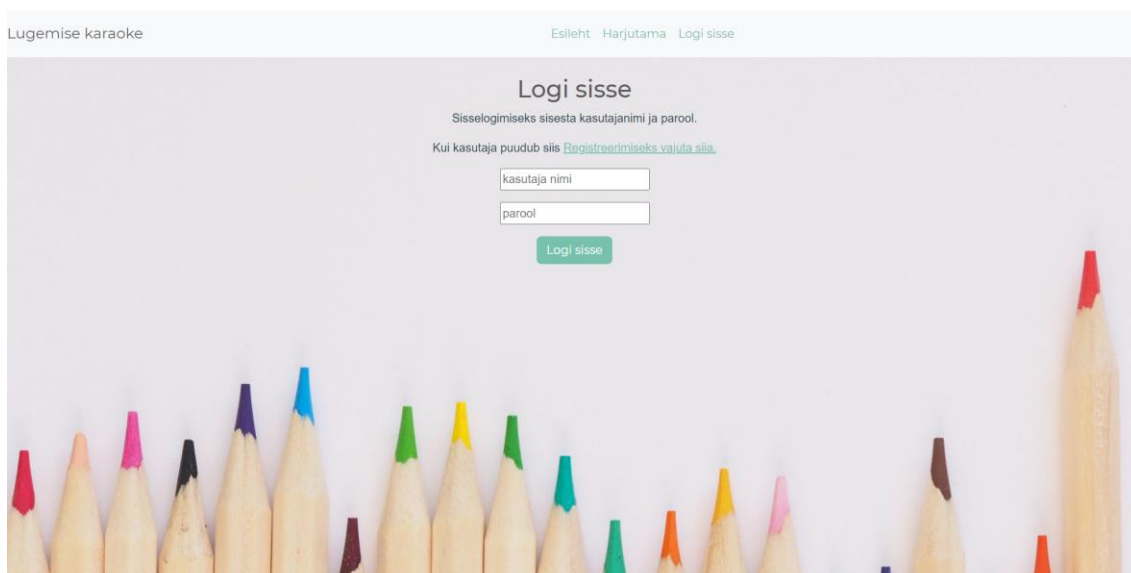
Rakendus ei küsi ega kogu kasutajalt isikuandmeid, millega oleks võimalik kasutaja isikut tuvastada, kaasa arvatud e-posti aadressi.



The screenshot shows the registration page for 'Lugemise karaoke'. At the top, there are navigation links: 'Esileht', 'Harjutama', and 'Logi sisse'. The main heading is 'Registreeri'. Below it, there is a message: 'Kui sul on juba kasutaja olemas [logi sisse siia](#).' followed by 'Uue kasutaja registreerimiseks täida vorm.' The form contains three input fields: 'Kasutaja nimi', 'Parool', and 'Vanus' (with a dropdown arrow). Below the fields is a checkbox labeled 'Nõustun, et minu häälennäited salvestatakse ja kasutatakse teadustöö eesmärgil.' and a green 'Registreeri' button. The background of the page features a row of colorful pencils.

Joonis 5. Ekraanipilt registreerimise vaatest.

Pärast registreerimist saab kasutaja sisselogida (Joonis 6) ning alustada lugemiskaraokega. Sisselogimisel alustatakse uut sessiooni ning see kestab kuni serveri poolt antud veebimärk (ingl. *webtoken*) aegub või kasutaja välja logib. Sessioone kasutatakse kasutajatele sessioonipõhiseks tagasiside andmiseks.



The screenshot shows the login page for 'Lugemise karaoke'. At the top, there are navigation links: 'Esileht', 'Harjutama', and 'Logi sisse'. The main heading is 'Logi sisse'. Below it, there is a message: 'Sisselogimiseks sisesta kasutajanimi ja parool.' followed by 'Kui kasutaja puudub siis [Registreerimiseks vajuta siia](#).' The form contains two input fields: 'kasutaja nimi' and 'parool'. Below the fields is a green 'Logi sisse' button. The background of the page features a row of colorful pencils.

Joonis 6. Ekraanipilt sisselogimise vaatest.

5.3 Admin

Admin-allas saavad ligi admin-õigustega kasutajad. Tavakasutajatele on admin-ala link peidetud. Admin-alaga seotud päringute tegemisel kontrollitakse serverrakenduses, kas kasutajal on toiminguteostamiseks õigus olemas.

Admin-allas on 2 alamlehte – lause lisamine ja haldamine.

5.3.1 Lause lisamine

Lause lisamisel (Joonis 7) tuleb sisestada lause, see häälega sisse lugeda ning valida keerukus. Keerukus on sisestaja subjektiivne hinnang lause keerukusele, keerukused on suunatud vanustele vastavalt:

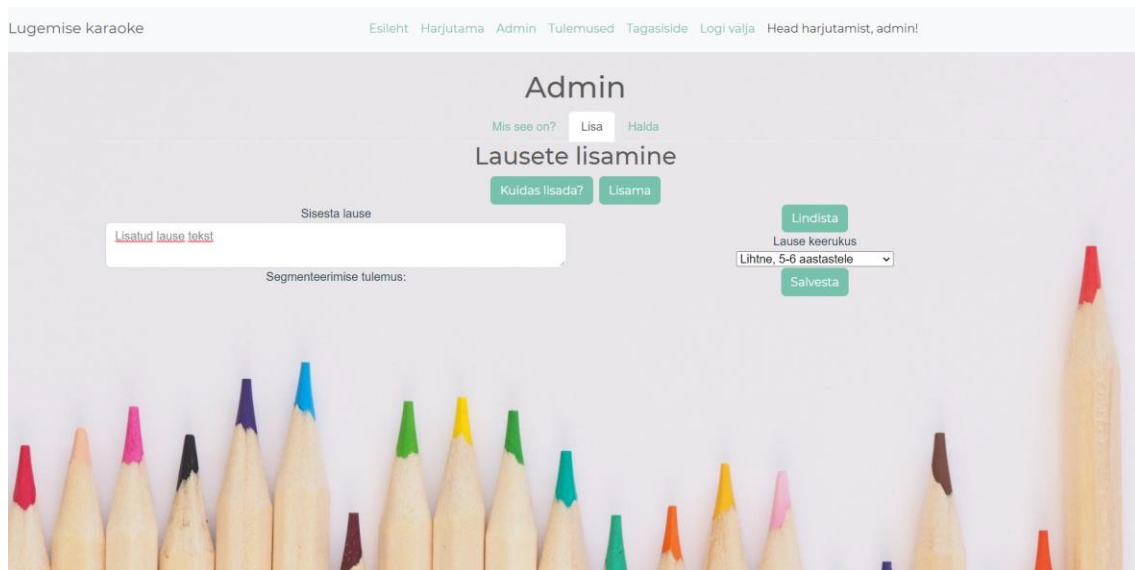
- kerge – suunatud 6–7 aastastele,
- keskmine – suunatud 8–10 aastastele,
- raske – suunatud üle 10-aastastele.

Keerukuse valimise juhend on välja toodud ka lisamise lehel.

Kui kasutaja salvestab lause, saadetakse see serverisse, kus kasutatakse Autosegmenteerija 2.0 teenust sisestatud audio ja teksti segmenteerimiseks TextGrid faili. Sellega leitakse lause sõnade kestused ja pausid lause lindistuses. Sõnade kestust kasutatakse lugemiskaraokes sõnade kuvamiseks õigel hetkel.

Seejärel silbitatakse sõnad kasutades EstNLTK silbitajat. Sõnad silbitatakse, et hinnata kasutaja lugemiskiirust.

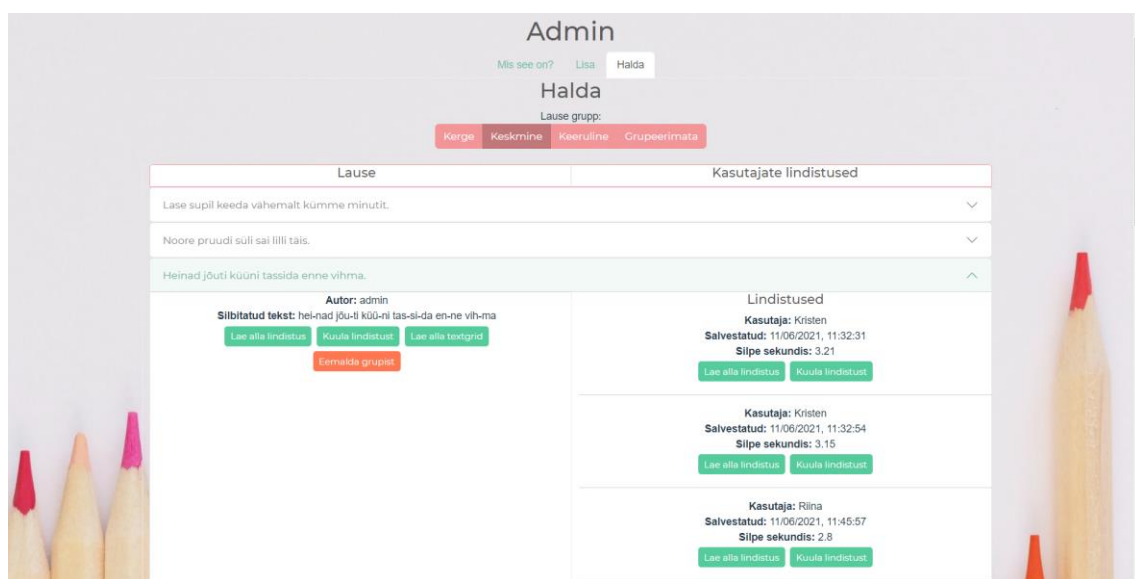
Kui lause on edukalt lisatud, saadetakse eesrakendusse tagasi TextGrid-faili allalaadimislink, juhaks kui kasutaja tahab tulemust üle kontrollida, ning kuvatakse lause silbitatud kujul. Kuna autosegmenteerija võtab pikemate lausete lisamisel mõnikümmend sekundit aega, ei pea kasutaja tulemust ära ootama, vaid saab hakata järgmisi lauseid lisama. Tulemused kuvatakse kasutajale tagantjärei samal ajal kui ta uusi lauseid lisab.



Joonis 7. Ekraanipilt admin-alas lause lisamisest.

5.3.2 Lausete haldamine

Lausete halduse lehel on näha kõik lisatud laused lausegruppide kaupa (Joonis 8). Praeguses rakenduses on implementeeritud ainult kolm gruppi raskusastmete kuvamiseks. Lausetel on näha silbitatud vorm ja autor. Lause lisaja lindistust ning genereeritud TextGrid faili saab alla laadida. Samuti on näha kasutajate harjutuste lindistusi, millal harjutamine toimus, mitu silpi sekundis kasutaja luges ning samuti saab lindistust alla laadida.



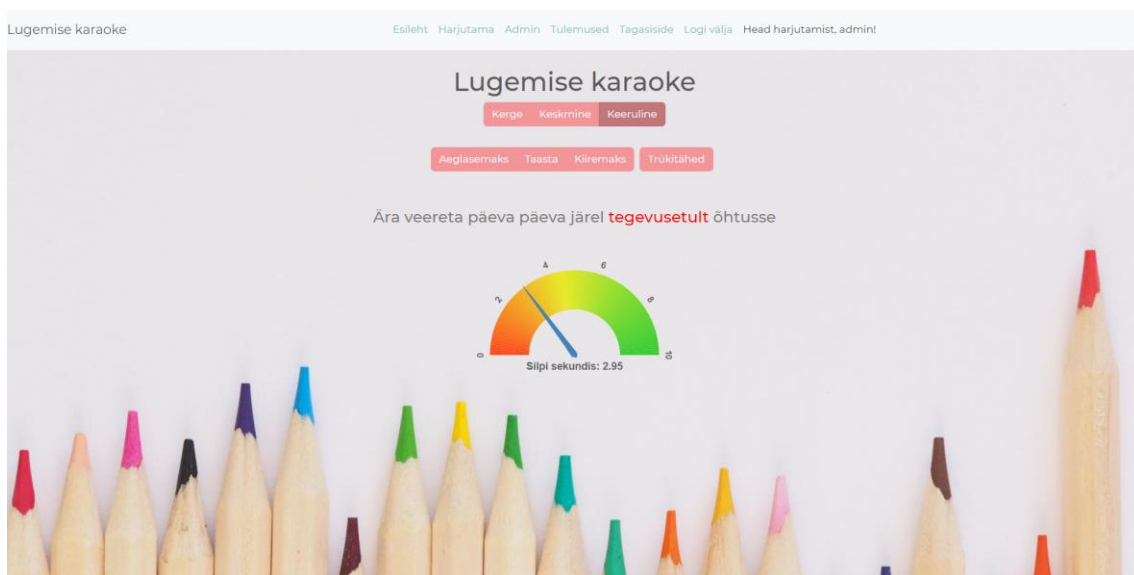
Joonis 8. Ekraanipilt admin-alas lause halduse vaatest.

5.4 Harjutama

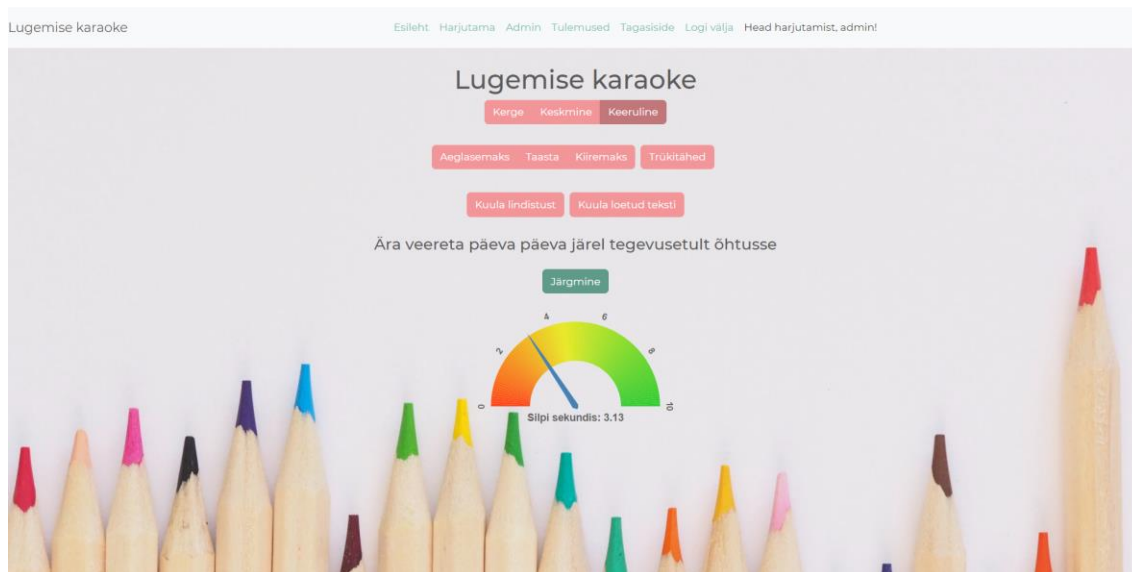
Sellel lehel paikneb lugemise karaoke, kus kasutaja saab harjutada lugemist. Kasutaja saab valida omale sobiva keerukuse, teksti kiiruse ning teksti kuvamise trüki- või kirjatähtedes. Esmane lause keerukus ja teksti kiirus määratakse lehele minnes vastavalt kasutaja vanusele.

Kui kasutaja vajutab nuppu “Alusta”, laetakse serverist talle valitud keerukusgrupist üks lause. Sõna, mida kasutaja peaks häälega lugema, märgitakse punaselt (Joonis 9). Sõna kestuse määrab lause autosegmenteerimise tulemusest saadud sõna kestus, mis korrutatakse valitud kiirusekordajaga. Lause all kujutatud spidomeetril näidatakse mitu silpi sekundis kasutaja loeb. Praegu ei ole rakenduses implementeeritud heli analüüs, vaid silbid sekundis näidatakse vastavalt valitud kiirusele ja silpide arvule. Peale lause lugemist saab kasutaja kuulata lause lisaja lindistust ning enda loetud lauset.

Loetud lause lindistus WAV formaadis ning silpide arv sekundis saadetakse salvestamiseks serverisse.



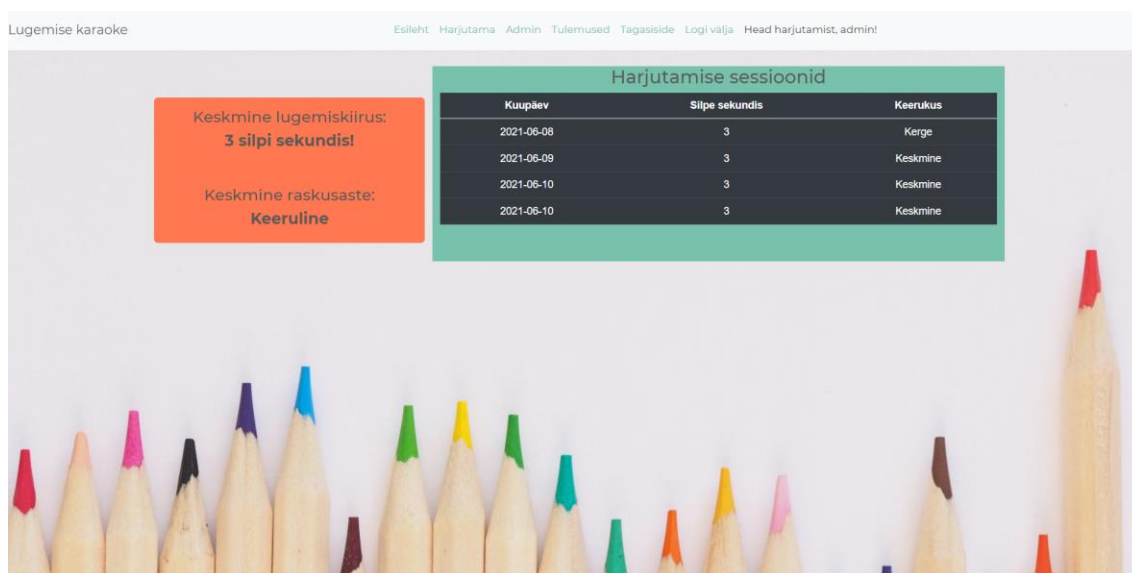
Joonis 9. Lugemise karaoke harjutus. Lugemisel olev sõna on punane.



Joonis 10. Lugemise karaoke harjutuse lõpp.

5.5 Tulemused

Tulemuste lehel saab kasutaja vaadata oma eelmiste harjutusessioonide keskmist loetud silpide arvu ning valitud keerukust ja ka kogu oma harjutuste ajaloo keskmist silpide arvu ja keerukust (Joonis 11).



Joonis 11. Ekraanipilt tulemuste vaatest.

6 Valideerimine

Rakendust valideeriti väiksemahulise pilootuuringus nii juhendajate kui õppijate sihtgruppidega – ühe logopeedi, ühe eesti keele õpetaja ja kolme lapsega vanustes 6, 10 ja 11. Kolm katses osalenut täitsid ka tagasiside vormi (vaata Lisa 2).

6.1 Juhendajate tagasiside

Kahe katses osalenud juhendaja tagasiside oli väga positiivne ja nad nägid rakenduses suurt potentsiaali õppevahendina. Näiteks katses osalenud logopeed kirjutas tagasisidevormis järgmist: „Logopeedina näen suurt kasutusvõimalust. Lapsed on erineva arengutasemega igas klassis ja see sobib kõigile.“ (vaata Lisa 2). Soovitati kaasata logopeedi või eripedagoogi, kes saaksid anda ideid, kuidas programmi täiendada. Eesti keele õpetaja soovitas ka noorematele ja kooliealistele, kes veel küll lugeda ei oska, aga õpivad, lisada funktsionaalsus üksikute sõnade ja piltide ühendamiseks.

6.2 Õppijate tagasiside

6-aastane ei saanud harjutustega üldse hakkama, sest ta tunneb ainult suuri tähti. Tekst oli ka liiga kiire isegi kõige madalamal lubatud kiirusel. Pärast tagasiside saamist viidi sisse täiendus, et kasutaja saaks soovi korral valida, kas teksti kuvatakse trükitähtedes või mitte ning lubati oluliselt madalam kiirus.

10-aastasel hea lugemisoskusega lapsele oli raske püsida ühtlases tempos, teisisõnu osad sõnad loeti aeglaselt ja teised kiiremine. See võib tulla sellest, et osad sõnad on lapsele tuttavamad kui teised. Üks võimalus seda aspekti täiendada oleks lisada rohkem näitelauseid, mis koosnevad ainult sõnadest, mis on tõenäolist lastele tuttavad. Seda tõenäosust oleks võimalik hinnata analüüsides sõnasagedusi lastekeele korpustes või laste kirjanduses (selline korpuseuuring aga eraldi uurimistööd).

11-aastane sai lihtsate ja keskmise keerukusega lausete lugemisega väga hästi hakkama ka siis, kui valis kiirema kiiruse. Keerulise raskusastmega lausete puhul oli vaja kõnetempot aeglasemaks teha. Testijale rakendus meeldis, kuid oleks tahtnud lugemiseks rohkem erinevaid lauseid. Tagasisides kirjutas ta järgmist: „Väga vajalik programm. See on huvitav programm.“ (vaata Lisa 2).

7 Edasiarendus võimalused

Lõputöö raames valmis töötav prototüüp, kuid kindlasti tasuks jätkata arendustega. Siin peatükis tuuakse välja mõned võimalikud edasiarenduse suunad.

7.1 Juhendajate rollid ja õppijate grupid

Praegu on rakenduses ainult tavakasutaja ja administraatori roll, kuid tulevikus saaks implementeerida lisaks veel juhendaja rolli logopeedide ja õpetajate jaoks ning luua tavakasutajatele grupid, kellele juhendaja rollist saaks harjutamiseks uusi harjutusi ette anda ja nende arengut jälgida. Nagu tagasisides välja toodud, peaks taolise funktsionaalsuse arendamist teostama vahetus koostöös juhendate sihtgrupi esindajatega ning kindlasti tuleks loodud lahendusi testida suurema valimi peal, kes esindavad erinevaid kasutajatüüpe.

7.1.1 Lausegrupid

Praegu on rakenduses ainult kolm lausegruppi keerukuse määramiseks, kuid tulevikus saaks lisada rohkem lausegruppe. Keerukuse võiks rakendus pakkuda edaspidi ise, et valik ei oleks enam nii subjektiivne. Loodud lausegruppe võiksid juhendajad saada jagada otse oma õpilastele.

7.2 Kõnetuvastuse ja muud rakendused

Antud töös rakendati lihtsamat meetodit lugemisoskuse hindamiseks, tulevikus tasuks kindlasti rakendada keerulisemaid ja täpsemaid meetodeid, et anda paremat tagasisidet kasutajatele [8] [16].

Lisaks oleks võimalik kasutada rakendusest saadud kõnekorpust valesti loetud või hääldatud teksti tuvastamiseks ning lugemiskvaliteedi paremaks hindamiseks, kasutades masinõppel põhinevaid meetodeid [17] [18] [19].

Rakenduses oleks võimalik implementeerida ka automaatne kiirusemuutmine vastavalt kasutaja eelnevalt loetud lausetele, et rihtida kiirus kasutajale sobivaks. Seda saaks teha mitte ainult lause, vaid ka sõna tasandil, sest nagu peatükis 6.2 välja toodi, siis ühtlases tempos püsimine erinevate lausete ja sõnadega võib tekitada probleeme.

7.2.1 Korpuse kättesaadavaks tegemine lastekõne analüüsiks

Rakendusest on võimalik juba praegu läbi API alla laadida helifaile, kuid edaspidi saaks API't täiendada lisaparaameetritega, et lindistuste või lindistuste gruppide leidmine oleks lihtsam.

7.2.2 Hääluse analüüs

Kui implementeerida hääluse analüüs saaks anda kasutajale paremat tagasisidet ning seda ka segmenteeritud helifailiga. Kuna segmenteeritud kuju on juba olemas saaks kasutajal lasta kuulata sõnu üks haaval ning välja tuua need, mis olid tal kehvasti häälstatud.

Kui häälstatud sõna tuvastamine piisavalt täpselt reaajas töötaks, saaks implementeerida mängulisi aspekte (ingl *gamification*) nagu näiteks interaktiivsed lood, kus kasutaja saab oma lugemisega valikuid teha ning kus vastavalt sellele, mis sõna kasutaja lugedes häälstatub, suunatakse ta järgmise lause juurde (Joonis 12). Taolised mängulised aspektid aitavad kasutajat motiveerida [10].



Joonis 12. Ekraanipilt näidistest, kuidas võiks välja näha sõna valimine.

7.3 Rakenduse laiendamine teistele keeltele

Praegu loodud kasutajaliides ja harjutused on eestikeelesed. Üks võimalik laiendus on liidese tõlkimine teistesse keeltesse, näiteks vene või inglise keel – siis oleks võimalik rakendust kasutada ka mitte-eestikeelsetel õppijatel eesti keele lugemise harjutamiseks. Rakendust on võimalik laiendada ka teistes keeltes lugema õppimiseks. See eeldab keeletehnoloogilise toe olemasolu vastavas keeles, sest harjutuste loomine ja lugeja oskuse hindamine sõltub keelespetsiifilistest lahendustest.

8 Kokkuvõte

Selle bakalaureusetöö eesmärgiks oli luua veebipõhine rakendus lastele lugemisoskuse treenimiseks „karaoke“ stiilis ja eesti laste kõnekorpuse jaoks helinäidete kogumiseks. Selleks uuriti, kuidas lapsed lugema õpivad ja sarnaseid olemasolevaid rakendusi.

Töös kirjeldati tehnilisi lahendusi, mida kasutati eesrakenduse ja tagarakenduse arendamisel. Demonstreeriti ka valminud funktsionaalsusi ning nende tööd.

Valminud veebirakendus on kasutajatele kättesaadav aadressil <https://readingkaraoke.xyz>. Rakenduses on administraatoril võimalik lisada lugemisharjutuste jaoks lauseid ja need häälega sisse lugeda. Selle põhjal luuakse lugemiskaraoke harjutus. Kasutajad saavad lugemiskaraoke lehel harjutada sisestatud lausete lugemist, valides selleks omale sobiva kiiruse ja lause raskusastme. Kasutajad saavad jooksvalt tagasisidet, mitu silpi sekundis nad loevad. Samuti on kasutajal võimalik näha kogu oma harjutuste seansside ajalugu. Administraatoril on võimalik kuulata harjutuste lindistusi.

Rakenduse jaoks koostati tagasiside küsitlus ning viidi läbi pilootuuring. Pilootuuringust selgus, et rakendusest on väga huvitatud nii juhendajad kui õppijad. Autor leiab, et siin töös valminud prototüübi põhjal võiks jätkata arendust ka tulevikus, kuna vajadus sellise teenuse järele on olemas.

Kasutatud kirjandus

- [1] A. E. Cunningham ja K. E. Stanovich, „Early reading acquisition and its relation to reading experience and ability 10 years later.“, *Dev. Psychol.*, kd 33, nr 6, lk 934–945, 1997, doi: 10.1037/0012-1649.33.6.934.
- [2] U. Leppänen, K. Aunola, P. Niemi, ja J.-E. Nurmi, „Letter knowledge predicts Grade 4 reading fluency and reading comprehension“, *Learn. Instr.*, kd 18, nr 6, lk 548–564, 2008, doi: <https://doi.org/10.1016/j.learninstruc.2007.11.004>.
- [3] R. Gillam ja J. Johnston, „Spoken and Written Language Relationships in Language/Learning-Impaired and Normally Achieving School-Age Children“, *J. Speech Hear. Res.*, kd 35, lk 1303–15, 1993, doi: 10.1044/jshr.3506.1303.
- [4] K. Uibu ja M. Männamaa, „Õpetamistegevused ja õpilaste tekstimõistmine üleminekul esimesest kooliastmest teise astmesse“, *Eesti Haridusteaduste Ajak. Est. J. Educ.*, kd 2, nr 1, lk 96–131, mai 2014, doi: 10.12697/eha.2014.2.1.05.
- [5] M. Kobayashi, „Method effects on reading comprehension test performance: text organization and response format“, *Lang. Test.*, kd 19, nr 2, lk 193–220, 2002, doi: 10.1191/0265532202lt227oa.
- [6] J. C. Alderson, *Assessing Reading*. Cambridge University Press, 2000. doi: 10.1017/CBO9780511732935.
- [7] P. Soodla, M. Vija, ja R. Pajusalu, „Eesti ja soome sõnalugemistestide võrdlus“, *Eesti Raken. Ühin. Aastaraam.*, kd 9, lk 279–296, 2013.
- [8] J. Duchateau, L. Cleuren, H. Van hamme, ja P. Ghesquière, „Automatic assessment of children’s reading level“, *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2007, kd 1, lk 1210–1213.
- [9] E. MEISTER ja L. Meister, „Eesti laste kõne I“, *Põhitooni Akust. Analüüs Keel Ja Kirjand.*, kd 60, nr 7, lk 518–533, 2017.
- [10] A. Gupta, „Karaoke: A tool for promoting reading“, *Read. Matrix Int. Online J.*, kd 6, nr 2, 2006.
- [11] R. Patel ja W. Furr, „ReadN’Karaoke: Visualizing Prosody in Children’s Books for Expressive Oral Reading“, *Proceedings of the SIGCHI Conference on Human*

- Factors in Computing Systems*, New York, NY, USA, 2011, lk 3203–3206. doi: 10.1145/1978942.1979417.
- [12] K. Reeder, J. Shapiro, J. Wakefield, ja R. D’Silva, „Speech Recognition Software Contributes to Reading Development for Young Learners of English“, *Int. J. Comput.-Assist. Lang. Learn. Teach.*, kd 5, lk 60–74, 2015, doi: 10.4018/IJCALLT.2015070104.
- [13] M. Siiber, „Tarkvara lugemiskiiruse arendamiseks“, Magistritöö, Tartu Ülikool, 2018.
- [14] T. Alumäe, O. Tilk, ja others, „Advanced Rich Transcription System for Estonian Speech“, *ArXiv Prepr. ArXiv190103601*, 2019.
- [15] S. Laur, S. Orasmaa, D. Särg, ja P. Tammo, „EstNLTK 1.6: Remastered Estonian NLP Pipeline“, *Proceedings of The 12th Language Resources and Evaluation Conference*, Marseille, France, mai 2020, lk 7154–7162. [Online]. Available at: <https://www.aclweb.org/anthology/2020.lrec-1.884>
- [16] J. Mostow ja M. Duong, „Automated Assessment of Oral Reading Prosody“, *Frontiers in Artificial Intelligence and Applications*, 2009, kd 200, lk 189–196. doi: 10.3233/978-1-60750-028-5-189.
- [17] J. Proença, C. Lopes, M. Tjalve, A. Stolcke, S. Candeias, ja F. Perdigão, „Automatic Evaluation of Reading Aloud Performance in Children“, *Speech Commun.*, kd 94, 2017, doi: 10.1016/j.specom.2017.08.006.
- [18] T. Baumann, „Learning to determine who is the better speaker“, 2018.
- [19] D. Bolaños, R. Cole, W. Ward, E. Borts, ja E. Svirsky, „FLORA: Fluent oral reading assessment of children’s speech“, *TSLP*, kd 7, lk 16, 2011, doi: 10.1145/1998384.1998390.

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Martin Karjus

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose "Veebipõhine platvorm lastele eesti keele lugemisoskuse treenimiseks ja hindamiseks", mille juhendaja on Einar Meister.
 - 1.1. reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.
3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

14.06.2021

¹ Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtes tamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud üks või enam isikut oma ühise loomingu tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtjaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. ja 1.2., siis lihtlitsents nimetatud tähtaja jooksul ei kehti.

Lisa 2 – Tagasiside vorm

Siin on esitatud tagasiside küsitluse tulemused (Tabel 1). Tagasiside vormi päises oli järgmine juhend: „Palun valige oma seisukoht väidetele skaalal 1 (ei nõustu üldse) kuni 5 (kindlasti nõustun)“.

Tabel 1. Tagasiside vormi tulemused.

Kasutaja	Logopeed	Eesti keele õpetaja	11a laps
Veebilehe disain oli hea	5	5	5
Lapsed olid liiga rasked	1	1	3
Juhend oli arusaadav	5	5	5
Harjutused olid sujuvad ja ilma probleemideta	5	5	5
Esialgsest valitud lugemiskiirus oli piisavalt kiire	5	5	3
Arvamused programmi kohta	Logopeedina näen suurt kasutusvõimalust. Lapsed on erineva arengutasemega igas klassis ja see sobib kõigile.	Suurepärane õppevara, ootan huviga selle kasutamise võimalust. Jõudu tööle ja kindlasti tasub edasi arendada.	Väga vajalik programm. See on huvitav programm.
Soovitused edasiarendusteks	Leidke omale logopeed või eripedagoog tiimi ja nad annavad veel ideid, kuidas programmi täiendada.	Nooremataele, kooliealistele, kes veel lugeda ei oska või kes veerivad, üksikute sõnade ja pildi ühendamine.	Rohkeim lauseid.