

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Hanna Raudsepp IAIB206845

**HÄÄLE AKUSTILISTE TUNNUSTE VISUALISEERIMISE
RAKENDUS**

Bakalaureusetöö

Juhendaja: Einar Meister
PhD

Tallinn 2025

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: Hanna Raudsepp

16.01.2025

Annotatsioon

Hääle akustiliste tunnuste visualiseerimise rakendus

Käesoleva bakalaureusetöö eesmärk oli luua töölauarakendus, mis lihtsustab GeMAPS akustiliste tunnuste kasutamist foneetilistes uuringutes. Rakendus võimaldab helifailidest GeMAPS-tunnuste eraldamist, nende visualiseerimist erinevatel diagrammitüüpidel (ajatelje-, histogrammi-, karp-, radari- ja vokaalikaardi diagrammid) ning sarnasuse analüüsi, et leida kõige sarnasemad salvestised. Helifailide töötlemisel (OpenSMILE) kasutatakse TextGridide andmeid, et võimaldada foneemide ja sõnade tasemel analüüsi. Tunnuste ajaraamilised väärtused salvestatakse MongoDB dokumendipõhises andmebaasis. Kasutajaliidese arendamiseks kasutatakse PyQt raamistikku ning interaktiivsete graafikute genereerimiseks rakendatakse Plotly teeki. Analüüsimeetoditena rakendatakse klasterdamist (KMeans) ja koosinussarnasuse arvutamist.

Lõputöö on kirjutatud eesti keeles ning sisaldab teksti 34 leheküljel, 6 peatükki, 5 joonist, 10 tabelit.

Abstract

Voice Acoustic Feature Visualization Tool

The aim of this bachelor's thesis was to create a desktop application for analyzing acoustic features of GeMAPS. The application allows extracting GeMAPS features from audio files, visualizing them on different chart types (timeline, histogram, box, radar and vocal map charts) and similarity analysis to find the most similar recordings. Audio File Processing (OpenSMILE) uses data from TextGrids to allow specific phoneme and word analysis. The extracted feature values are stored in a MongoDB document-based database. The user interface is created in the PyQt Python framework and the Plotly library is used to generate interactive graphs. Both clustering (KMeans) and cosine similarity calculation are used as analysis methods.

The thesis is written in Estonian and is 34 pages long, including 6 chapters, 5 figures and 10 tables.

Lühendite ja mõistete sõnastik

CSV	Comma-Separated values
DataFrame	Tabelilaadne andmestruktuur
eGeMAPS	Extended Geneva Minimalistic Acoustic Parameter Set
GeMAPS	Geneva Minimalistic Acoustic Parameter Set
HTML	HyperText Markup Language
JSON	JavaScript Object Notation
LLD	Low-Level Descriptors
MATLAB	Numbriline andmetöötluskeskkond
MongoDB	Dokumendipõhine andmebaas
PCA	Principal Component Analysis
Plotly	Pythoni teek andmete visualiseerimiseks
PyQt	Pythoni teek graafiliste kasutajaliideste loomiseks
TextGrid	Failiformaat kõnesalvestuste märgendamiseks
WAV	Pakkimata helifaili formaat

Sisukord

1	Sissejuhatus	10
2	Olemasolevad rakendused	11
3	Rakenduse nõuded	13
3.1	Funktsionaalsed nõuded	13
3.2	Funktsionaalsete nõuete kasutusjuhud	13
4	Rakenduse arendus	18
4.1	Tunnuste eraldamine ja TextGridide töötlus	18
4.1.1	OpenSMILE	18
4.1.2	TextGridide töötlemine	19
4.2	Andmete salvestamine MongoDB andmebaasis	19
4.3	Kasutajaliidese loomine PyQt raamistikuga	21
4.4	Visualisatsioonide loomine	22
4.5	Sarnasuse analüüs	23
4.5.1	Klasterdamine	23
4.5.2	Koosinussarnasuse arvutamine	23
4.5.3	Koosinussarnasuse arvutamine PCA-ga	24
4.6	Testimine	24
4.6.1	Testitav rakendus	24
5	Tulemused ja analüüs	25
5.1	Rakenduse funktsionaalsus ja vaated	25
5.1.1	Rakenduse kasutajaliidese tutvustus	25
5.1.2	Kasutusjuhtude analüüs	26
5.2	Tulemuse võrdlus olemasoleva lahendusega	29
5.3	Tagasiside	31
5.4	Võimalused edasiarenduseks	32
6	Kokkuvõte	34

Kasutatud Kirjandus	35
Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks	37
Lisa 2 – Eraldatud tunnuste loetelu	38
Lisa 3 – Näited Rakenduse vaadetest	39
Lisa 4 – Kõne tunnuste analüüsimise rakenduse tagasiside küsimustik	44

Jooniste loetelu

1	Rakenduse põhivaade ja selle osad.	26
2	Rakenduse salvestuste haldus.	27
3	Joondiagrammi näide ühe sõna ja tunnuse visualiseerimisel.	28
4	PCA koosinussarnasuse ja klastrite visualiseerimine hajuvusdiagrammil. .	29
5	Kõneveebi Audiofailide akustiliste omaduste võrdlemise rakenduse põhi- vaade.	30

Tabelite loetelu

1	Kasutusjuht GeMAPS tunnuste eraldamine kõnefailidest.	13
2	Salvestuste ja töödeldud andmete kustutamine.	14
3	Tunuste visualiseerimine.	15
4	Sarnaste kõnelejate leidmine.	15
5	Helisalvestuse esitamine ja helilaine visualiseerimine.	16
6	Andmete ja visualisatsioonide eksportimine.	16
7	Näide helifailist eraldatud tunnuste DataFrame'ist.	18
8	Salvestuste kolleksiooni väljad, nende andmetüübid ja semantika.	20
9	Sõnade kolleksiooni väljad, nende andmetüübid ja semantika.	20
10	Foneemide kolleksiooni väljad, nende andmetüübid ja semantika.	21

1 Sissejuhatus

Foneetikauuringutes kasutatakse kõnekorpusi, mis sisaldavad sadade inimeste kõnenäiteid. Korpuste analüüsil leitakse erinevaid akustilisi tunnuseid, nagu põhitoon (F0), formandid, spektrid ja muud hääle kvaliteeti kirjeldavad tunnused. Nende tunnuste abil saab tuvastada erinevaid kõnestiile, võõrkeele aktsente, emotsionaalseid ja tervislikke seisundeid.

Üheks laialt levinud tunnustekomplektiks foneetikauuringutes on GeMAPS (The Geneva Minimalistic Acoustic Parameter Set) [1]. See komplekt sisaldab mitmeid olulisi akustilisi tunnuseid, mida saab kasutada mitmesugustes uuringutes. Paraku ei ole hetkel kuigi palju tööriistu, mis võimaldaksid GeMAPS-tunnuste automaatset helisalvestisest eraldamist ja põhjalikumalt visualiseerimist ühes rakenduses.

Korpuste analüüsi tööprotsessi osad on sageli järgmised tegevused: akustiliste tunnuste eraldamine kõnefailidest, akustiliste tunnuste visualiseerimine, uurija poolt etteantud akustiliste omadustega kõnenäidete leidmine. Kuigi on olemas mitmeid tööriistu, mis suudavad neid ülesandeid eraldi täita, puudub hetkel terviklik rakendus, mis võimaldaks täita kõiki neid kõnekorpusete analüüsiga seotud ülesandeid.

Selle bakalaureusetöö eesmärk on luua tööriist, mis võimaldab:

- GeMAPS tunnuste eraldamist kõnefailidest,
- tunnuste visualiseerimist erinevatel diagrammitüüpidel: joondiagramm ajateljel, vokaalikaart, radar, karpdiagramm ja histogramm,
- sarnaste kõnelejate otsingut ja kõnekorpusetest.

Loodud rakendus on abiks uurijatele, kes soovivad kasutada GeMAPS-tunnuseid foneetika- või hääleanalüüsis. Rakendus koondab mitmed peamised uurimisprotsessi funktsioonid nagu: tunnuste eraldamine, andmete visualiseerimine ja sarnaste salvestiste leidmine.

2 Olemasolevad rakendused

Järgnevalt antakse lühike ülevaade sarnase funktsionaalsusega olemasolevatest kõneanalüüsi tööriistadest.

Kõneveebi Audiofailide akustiliste omaduste võrdlus [2] on veebirakendus, mis kasutab OpenSMILE eGeMAPS tunnuste komplekti kõnesalvestiste akustiliseks analüüsiks.

Rakenduse võimalused:

- .wav-formaadis pakendatud audiofailide üleslaadimine, maksimaalne lubatud failide suurus 900MB
- Võimalik määrata klassifikaatoreid CSV-vormingus.
- Visualiseeritakse valitud akustiliste tunnuste väärtuste jaotus joondiagrammina
- Arvutatud väärtused ja üleslaetud failid kustutatakse peale brauseri sulgemist.

Praat [3] on Amsterdami Ülikoolis arendatud vabavaraline kõneanalüüsi rakendus ning väga levinud foneetikauuringutes. Praat toetab Windows, macOS ja Linux operatsioonisüsteeme.

Rakenduse võimalused:

- Akustiline analüüs ja erinevate tunnuste arvutamine (F0, formandid, intensiivsus, jitter, shimmer jpm)
- Visualiseerimisvõimalused: spektrogrammid, põhitooni ja intensiivsuse graafikud ajateljel, formantide trajektoorid
- Kõnesüntees: põhitooni ja formantide põhjal on võimalik luua modifitseeritud või sünteesitud kõne
- Skriptide loomise võimalus: Praatil on oma skriptikeel, mis teeb rakenduse paindlikuks, ning võimaldab automatiseerida korduvaid analüüse

VoiceSauce [4] on MATLAB-i keskkonnas arendatud kõneanalüüsi rakendus. Rakendus on mõeldud teaduslikuks kõneanalüüsiks. Toetatud on Windows ja macOS operatsioonisüsteemid.

Rakenduse võimalused:

- Erinevate akustiliste hääle tunnuste visualiseerimine ja analüüs.
- Mõned parameetrid sõltuvad tugevalt arvatud F0 väärtustest, mistõttu võib analüüs olla ebatäpne müra rikaste helide korral
- Pikkade heliklippide analüüs võib tekitada ressursipuudust, kuna MATLABi keskkond võib intensiivsel analüüsil väga suurt osa mälu tarbida.
- Nõuab MATLABi kasutamist

Wasp (Windows Tool for Speech Analysis) [5] on tasuta kõneanalüüsi rakendus. Toetatud on operatsioonisüsteem Windows.

Rakenduse võimalused:

- Helifailide salvestamine ja esitamine, salvestatud faile on võimalik analüüsida
- Võimaldab luua annotatsioone
- Akustiliste tunnuste visualiseerimine
- Lihtne ja kasutajasõbralik kasutajaliides ning saadaval on veebiversioon rakendusega tutvumiseks

Kõik välja toodud rakendused (Kõneveebi veebirakendus, Praat, VoiceSauce ja Wasp) on võimelised kõnesalvestusi töötlemise ning pakuvad akustiliste tunnuste arvutamist ja visualiseerimisvõimalusi. Neist kõige rohkem võimalusi ja funktsionaalsust on Praatil, kuid sellel ei ole spetsiaalset funktsionaalsust GeMAPS tunnuste eraldamiseks. Samuti on erinevate visualisatsioonide ja analüüsides jaoks tavaliselt vajalik skriptide kirjutamine.

Selgelt erinev teistest on Kõneveebi lahendus, sest see on veebipõhine ning mõeldud ainult eGeMAPS-tunnuste analüüsiks. Selle funktsionaalsus on piiratud: rakenduses on küll võimalik akustilisi tunnuseid joondiagrammil visualiseerida, kuid puuduvad erinevad diagrammitüübid ning analüüsimeetodid sarnasuse otsinguks. Lisaks esineb failisuuruse piirang ja tulemusi ei salvestata.

Seega, ükski rakendus ei kata ühes keskkonnas kõiki käesolevas töös esile toodud analüüsiülesandeid: GeMAPS tunnuse arvutamine, erinevad visualiseerimis võimalused ning sarnasuse analüüsi.

3 Rakenduse nõuded

Järgnevalt kirjeldatakse rakenduse funktsionaalsete nõuete määratlemise protsessi ja esitatakse nende põhjal koostatud kasutusjuhud.

3.1 Funktsionaalsed nõuded

Rakenduse arenduse alguses lepitakse juhendajaga kokku funktsionaalsed nõuded, et määratleda rakenduse fookus ja võimalused.

Funktsionaalsed nõuded:

- GeMAPS tunnuste eraldamine kõnefailidest. Rakendus peab suutma kõnefailidest eraldada GeMAPS akustilisi tunnuseid. Selleks tuleb töödelda salvestused ja nendega seotud TextGrid failid.
- Tunnuste visualiseerimine erinevatel diagrammitüüpidel, nagu ajatelje joondiagramm, histogramm, radardiagramm, karpdiagramm, vokaalikaart.
- Sarnaste kõnelejate leidmine kõnekorpusel.
- Võimalus eksportida analüüsi tulemusi: pildifaile ja andmeid.

3.2 Funktsionaalsete nõuete kasutusjuhud

Arenduse käigus koostati täpsemad nõuded, mida täiendati jooksvalt, kui juhendajaga koos rakenduse testides tekkis uusi ideid. Järgnevalt esitatakse kokkulepitud funktsionaalsed nõuded kasutusjuhtudena tabelites 1-7.

Tabel 1. Kasutusjuht GeMAPS tunnuste eraldamine kõnefailidest.

Nimi	GeMAPS tunnuste eraldamine kõnefailidest
Kirjeldus	Kasutaja saab importida kõnesalvestused ja nendega seotud TextGrid failid, et eraldada ja salvestada GeMAPS akustilisi tunnuseid.
Eeltingimused	Kasutajal on kõnesalvestuste samanimelised .wav ja TextGrid failid.

Nimi	GeMAPS tunnuste eraldamine kõnefailidest
Põhiline sündmuste käik	<ol style="list-style-type: none"> 1. Kasutaja avab kõnefailide importimise akna. 2. Süsteem kuvab kõnefailide ja TextGridide üleslaadimise vormi. 3. Kasutaja valib kõnefailid ja vastavad TextGridid oma seadmest. 4. Süsteemis käivitub GeMAPS tunnuste eraldamise protsess 5. Süsteem töötleb failid, eraldab GeMAPS tunnused ja salvestab andmed andmebaasi. 6. Kasutajale kuvatakse õnnestumise teade.
Alternatiivne sündmuste käik	<ul style="list-style-type: none"> - Kui kasutaja ei vali faile, kuvatakse veateade ja protsessi ei alustata. - Kui failide töötlemine ebaõnnestub, kuvatakse veateade ja protsess peatub.
Lõpptulemused	<ul style="list-style-type: none"> - Eduka stsenaariumi korral GeMAPS tunnused salvestatakse andmebaasi edasiseks analüüsiks ja visualiseerimiseks. - Ebaõnnestunud stsenaariumi korral töötlust ei sooritata ning kasutajale kuvatakse teavitust.

Tabel 2. Salvestuste ja töödeldud andmete kustutamine.

Nimi	Salvestuste ja töödeldud andmete kustutamine
Kirjeldus	Kasutaja saab hallata salvestusi ja nendega seotud töödeldud andmeid, sealhulgas kustutada neid andmebaasist.
Eeltingimused	Kõnesalvestused ja töödeldud andmed on olemas andmebaasis.
Põhiline sündmuste käik	<ol style="list-style-type: none"> 1. Kasutaja avab salvestuste haldamise akna. 2. Süsteem kuvab salvestuste nimekirja. 3. Kasutaja valib salvestused, mida ta soovib kustutada. 4. Süsteem kuvab dialoogi, milles küsitakse kustutamise kinnitust. 5. Kasutaja kinnitab kustutamise. 6. Süsteem eemaldab valitud salvestuse andmed andmebaasist ning kuvab kinnitusteate.
Alternatiivne sündmuste käik	<ul style="list-style-type: none"> - Kui kasutaja tühistab kustutamise kinnituse, jäävad andmed muutmata.
Lõpptulemused	<ul style="list-style-type: none"> - Edukas stsenaariumi korral valitud salvestuse andmed kustutatakse andmebaasist. - Ebaõnnestunud stsenaariumi korral andmed ei muutu, kuvatakse vastav teavitust.

Tabel 3. Tunuste visualiseerimine.

Nimi	Tunnuste visualiseerimine
Kirjeldus	Kasutaja saab visualiseerida GeMAPS tunnuseid erinevatel diagrammitüüpidel, sealhulgas ajateljel, radardiagrammil, histogrammil, karpdiagrammil ja vokaalikaardil.
Eeltingimused	GeMAPS tunnused on eraldatud ja salvestatud andmebaasi.
Põhiline sündmuste käik	<ol style="list-style-type: none"> 1. Kasutaja avab tunnuste visualiseerimise menüü. 2. Süsteem kuvab valikud tunnuste, salvestuste ja graafikutüüpide valimiseks. 3. Kasutaja valib salvestuse, tunnused ja graafikutüübi. 4. Kasutaja vajutab visualiseerimise nupule. 5. Süsteem koostab ja kuvab valitud parameetrite alusel visualisatsiooni.
Alternatiivne sündmuste käik	- Kui valikud ei vasta nõuetele (nt pole valitud tunnuseid), kuvatakse veateade.
Lõpptulemused	<ul style="list-style-type: none"> - Eduka stsenaariumi korral kuvatakse valitud visualisatsioon ning kasutaja saab selle eksportida pildifailina või JSON-andmetena. - Ebaõnnestunud stsenaariumi korral visualisatsiooni ei looda ning kuvatakse teavitus.

Tabel 4. Sarnaste kõnelejate leidmine.

Nimi	Sarnaste kõnelejate leidmine
Kirjeldus	Kasutaja saab leida kõnekorpusest salvestused, mis on kõige sarnasemad valitud sihtsalvestusega
Eeltingimused	GeMAPS tunnused on eraldatud ja salvestatud andmebaasi.
Põhiline sündmuste käik	<ol style="list-style-type: none"> 1. Kasutaja avab sarnasuse analüüsi valiku juhtpaneelis. 2. Süsteem kuvab sihtsalvestuse ja teiste salvestuste valikuvõimalused. 3. Kasutaja valib sihtsalvestuse, millele sarnaseid otsitakse ja määrab sarnasuse arvutamise valikud (nt meetod, tagastavate salvestuste arv tulemuses). 4. Süsteem arvutab valitud meetodil sarnasused ja kuvab tulemused graafikuna.
Alternatiivne sündmuste käik	- Kui kasutaja ei määra sihtsalvestust, kuvatakse veateade.

Nimi	Sarnaste kõnelejate leidmine
Lõpptulemused	- Eduka stsenaariumi korral kuvatakse sarnaste salvestuste analüüs ja visualisatsioon. - Ebaõnnestunud stsenaariumi korral analüüsi ei sooritata, kuvatakse teavitus.

Tabel 5. Helisalvestuse esitamine ja helilaine visualiseerimine.

Nimi	Helisalvestuse esitamine ja helilaine visualiseerimine
Kirjeldus	Kasutaja saab kuulata kõnesalvestusi ning kuvatakse helilaine visualisatsioon.
Eeltingimused	Kasutaja on lisanud soovitud kõnesalvestuste failid rakenduse kausta nimega <i>data</i>
Põhiline sündmuste käik	1. Kasutaja avab rakenduse. 2. Süsteem kuvab saadaval olevate salvestuste nimekirja. 3. Kasutaja valib salvestuse ja käivitab selle esitamise. 4. Süsteem esitab helisalvestuse ja kuvab helilaine visualisatsiooni.
Alternatiivne sündmuste käik	- Kui kasutaja valitud salvestus ei ole mingil põhjusel saadaval, kuvatakse veateade.
Lõpptulemused	- Eduka stsenaariumi korral salvestus esitatakse ja helilaine kuvatakse visuaalselt. - Ebaõnnestunud stsenaariumi korral esitamist ei toimu ning kuvatakse teavitus.

Tabel 6. Andmete ja visualisatsioonide eksportimine.

Nimi	Andmete ja visualisatsioonide eksportimine
Kirjeldus	Kasutaja saab eksportida visualiseeritud andmeid ja graafikuid JSON- või pildifailidena.
Eeltingimused	Visualiseerimine või sarnasuse analüüs on edukalt läbi viidud.
Põhiline sündmuste käik	1. Kasutaja vajutab JSON faili ekspordi nupule juhtpaneelis või pildifaili ekspordi ikoonile visualisatsiooni vaate paremas nurgas. 2. Kasutaja valib failitüübi ja salvestuskoha ning kinnitab ekspordi. 4. Süsteem salvestab valitud failid kasutaja määratud asukohta.
Alternatiivne sündmuste käik	- Kui kasutaja ei määra salvestuskoha, ekspordi protsess katkestatakse ja kuvatakse teavitus.

Nimi	Andmete ja visualisatsioonide eksportimine
Lõpptulemused	- Eduka stsenaariumi korral eksporditud fail salvestatakse määratud asukohta. - Ebaõnnestunud stsenaariumi korral ekspordi protsess katkestatakse ja kuvatakse teavitus.

4 Rakenduse arendus

Selles peatükis antakse ülevaade rakenduse arenduse protsessist, mille etapid olid: akustiliste tunnuste eraldamine ja TextGrid-failide töötlemine, andmete salvestamine, kasutajaliidese arendus, visualisatsioonide loomine, kõnesalvestuste sarnasuse analüüs ning rakenduse testimine.

4.1 Tunnuste eraldamine ja TextGridide töötlus

Järgnevalt kirjeldatakse akustiliste tunnuste helifailidest eraldamise ning TextGridide töötlemise protsessi ja tehnoloogiaid.

4.1.1 OpenSMILE

Tunnuste eraldamiseks kasutati OpenSMILE [6] raamistikku, mille töötamise funktsioon võtab sisendiks helifaili ning annab väljundiks pandas DataFrame'i, kus iga rida akustilise tunnuse väärtus 10 millisekundilise ajaraami kohta.

Tabel 7. Näide helifailist eraldatud tunnuste DataFrame'ist.

Start (s)	Loudness_sma3	alphaRatio_sma3	slope0-500_sma3
0.00	0.120393	-19.566683	-0.073501
0.01	0.112910	-16.829172	-0.061550
0.02	0.103573	-14.812015	-0.051186
0.03	0.104770	-16.948393	-0.075717
0.04	0.105062	-19.221598	-0.086498
⋮	⋮	⋮	⋮
9.68	0.107817	-21.088533	-0.076632
9.69	0.110813	-16.712238	-0.043914
9.70	0.113438	-17.034071	-0.041286
9.71	0.112685	-16.894552	-0.032873
9.72	0.113219	-18.397005	-0.041675

OpenSMILE töötlemise funktsiooni parameetrikts määratakse eraldatav tunnuste komplekt. Neid on mitmeid erinevaid, selles rakenduses kasutatakse kõige väiksemat eGeMAPS v02 akustiliste tunnuste komplekti, kus on 88 tunnust, mis jagunevad:

1. Madala taseme deskriptorid (*Low Level Descriptors*), mis on otse helisignaalist tuletatud tunnused.
2. Funktsionaalsed tunnused (*functionals*): statistilised näitajad, mis arvutatakse madala taseme tunnustest, näiteks keskmised väärtused ja standardhälve.

Nendest omakorda valiti ainult madala taseme deskriptorid (kokku 25 tunnust), et tunnuste hulk oleks kasutajale hoomatav, sest visualiseerimisel keskendutakse üksikute tunnuste visualiseerimisele. Eraldatud tunnuste loetelu on esitatud Lisas 2.

4.1.2 TextGridide töötlemine

TextGrid [7] on failiformaat, mida kasutatakse kõnesalvestuse märgendamiseks. Lastekõne korpuse puhul, mida rakenduse arenduses nädisandmetena kasutati, on iga salvestuse TextGridis märgendatud IntervalTiers kiht, mis võib jaguneda veel näiteks: HMM-words, HMM-phonemes, cv (konsonant või vokaal) ja foot. Kihid, mida rakenduses TextGridide töötlusel kasutati on HMM-words ja HMM-phonemes. Nendes kihtides on andmed selle kohta, millal sõna või foneem algab ja lõpeb. Need andmed tuli salvestada, et võimaldada kindlate sõnade ja foneemide visualiseerimine.

Sõnade ja foneemide info töötlemiseks kasutati praatio teeki [8]. Sellega oli võimalik lihtsasti eraldada TextGridist kihtide infot. Praatio on Pythoni teek, mis lihtsustab TextGridide lugemist.

4.2 Andmete salvestamine MongoDB andmebaasis

Rakenduse andmebaasi valikul kaaluti võimalustest relatsioonilisi andmebaase või dokumendipõhiseid lahendusi nagu MongoDB. MongoDB [9] on NoSQL andmebaas, kus andmed salvestatakse BSON-formaadis dokumentidena. MongoDB on paindlik, sest see võimaldab paindlikku andmebaasi skeemi ning ühes kollektsioonis hoida erinevate väljade ja andmetüüpidega dokumente [10]. Kuna rakenduse arenduse alguses ei olnud veel selge, kuidas ja milliseid andmeid on vaja salvestada, tundus see valikul oluline. Samuti oli MongoDB seadistamine lihtne ja kiire.

Andmete salvestamise protsess: Peale tunnuste eraldamist ja TextGridide töötlemist salvestatakse andmed kolme kollektsiooni: *Recordings*, *Words* ja *Phonemes*.

Tabel 8. Salvestuste kolleksiooni väljad, nende andmetüübid ja semantika.

Välja nimi	Andmetüüp	Semantika
_id	ObjectId	Unikaalne identifikaator, mille MongoDB genereerib automaatselt.
recording_id	string	Salvestuse nimi (nt 148_10001148_26).
start	double	Salvestuse algusaeg sekundites.
end	double	Salvestuse lõppaeg sekundites.
duration	double	Salvestise kogukestus sekundites.
features	Object	Objekt, mis hoiab salvestise akustilisi tunnuseid. Sellel on kaks alamobjekti: <ul style="list-style-type: none"> ■ mean – akustiliste tunnuste keskmised väärtused kujul tunnuse nimi: tunnuse keskmine väärtus (double). ■ frame_values – massiiv 10 ms ajaraamide tunnuste väärtustest (double).

Ka sõnade ja foneemide puhul salvestatakse iga sõna või foneemi dokument vastavasse kolleksiooni, kuid tunnuste puhul salvestatakse ainult keskmised väärtused sõna või foneemi ajavahemiku kohta. Enne salvestamist filtreeritakse välja mitte-kõne segmentid, nagu vaikused ja müra, mis on vastavate siltidega (nt .noise või sil) TextGridides märgistatud. Nende segmentide tunnuste väärtuseid ei ole vaja andmebaasi salvestada, sest neid ei ole vaja eraldi visualiseerida, kuna vajalik on segmentidest ainult kindlaid sõnu või foneeme visualiseerida.

Tabel 9. Sõnade kolleksiooni väljad, nende andmetüübid ja semantika.

Välja nimi	Andmetüüp	Semantika
_id	ObjectId	Unikaalne identifikaator, mille MongoDB genereerib automaatselt.
duration	double	Sõna kestus sekundites.
start	double	Sõna algamise ajahetk sekundites.
end	double	Sõna lõppemise ajahetk sekundites.
text	string	Sõna esitus tekstina (nt linna+tulede).
recording_id	string	Salvestuse nimi (nt 148_10001148_26).
parent_id	ObjectId	Salvestuse dokumendi ID, millesse sõna kuulub.

Välja nimi	Andmetüüp	Semantika
features	Object	Objekt, mis hoiab sõna akustilisi tunnuseid: <ul style="list-style-type: none"> ■ mean – alamdokumentis on erinevate tunnuste keskmised väärtused kujul tunnuse nimi: double.

Tabel 10. Foneemide kollektsiooni väljad, nende andmetüübid ja semantika.

Välja nimi	Andmetüüp	Semantika
_id	ObjectId	Unikaalne identifikaator, mille MongoDB genereerib automaatselt.
duration	double	Foneemi kestus sekundites.
start	double	Foneemi alguse ajahetk sekundites.
end	double	Foneemi lõppemise ajahetk sekundites.
text	string	Foneemi esitus tekstina (nt "t").
parent_id	ObjectId	Viide selle sõna dokumendi ID-le, millesse foneem kuulub.
word_text	string	Sõna tekst, millesse foneem kuulub (nt linna+tulede).
recording_id	string	Salvestuse nimi (nt 148_10001148_26).
features	Object	Objekt, mis hoiab foneemi akustilisi tunnuseid: <ul style="list-style-type: none"> ■ mean – alamdokumentis on erinevate tunnuste keskmised väärtused kujul tunnuse nimi: double.

4.3 Kasutajaliidese loomine PyQt raamistikuga

PyQt [11] on Pythoni graafiliste kasutajaliideste arendamise raamistik. Selles on erinevad valmiskomponendid, näiteks nupud, aknad, menüüd ja otsingukastid. PyQt raamistik toetab Windows, macOS ja Linux operatsioonisüsteeme.

Peale PyQt kaaluti veel teisi GUI loomise raamistikke nagu Tkinter [12], mis on samuti Pythoni raamistik. Kuid selle funktsionaalsus on piiratud võrreldes PyQt-ga ning valmiskomponentide valik väiksem [13].

4.4 Visualisatsioonide loomine

Tunnuste visualiseerimise funktsionaalsuse eesmärk oli võimaldada kasutajal kuvada akustilisi tunnuseid mitmel erineval moel interaktiivsetel graafikutel, et andmeid oleks mugav uurida.

Rakenduse varasemates versioonides loodi graafikute visuaalid Matplotlib [14] teegiga, mis on tuntud ja usaldusväärne teek andmete visualiseerimiseks, kuid kuna see pakub vaikumisi vaid staatilisi jooniseid, siis hakati otsima teisi võimalusi, mis pakuksid rohkem valmis võimalusi graafikute interaktiivsuse ja disaini poolest.

Katsetati ka PyQt enda graafiku loomise komponente, kuid need olid väga sarnased Matplotlibi graafikutele, interaktiivsust seadistada oli keerulisem ning need ei pakkunud piisavalt sisseehitatud funktsionaalsust.

Lõpuks valiti graafikute loomise tehnoloogiaks Plotly [15], sest sellega on graafikud vaikumisi interaktiivsed, näiteks on võimalik suumimine, kerimine, andmepunktide (*tooltip*) lisainformatsiooni kuvamine ja pildifailide eksport.

Igale graafikutüübile tehti eraldi meetod, mis loob pandas DataFrame'ist Plotly graafikuobjekti. Graafikuobjekt teisendatakse HTML koodiks, mille PyQt QWebView komponent rakenduse põhivaate aknas renderdab. Kõigile graafikutele konfigureeriti ühtne legendi, telgede ja värvide stiil. Visualiseerimise graafikutüüpidest valiti realiseerimiseks:

- **Ajatelje diagramm**, mis kuvab valitud tunnuste väärtused ajas. Võimalik on nii terve salvestuse kui ka eraldi sõna ja foneemi visualiseerimine, näiteks saab visualiseerida, kuidas F0 põhisedus läbi kahe erineva valitud sõna muutub.
- **Histogramm** võimaldab visualiseerida, kuidas üksiku tunnuse väärtused valitud salvestuses, sõnas või foneemis väärtuste vahemikes jaotuvad. Väärtuste vahemike tulpade arv leitakse Sturgesi valemiga.
- **Karpdiagramm** võimaldab visualiseerida valitud tunnuste põhilisi statistilisi näitajaid nagu: miinimum, maksimum, mediaan ja kvartiilid.
- **Radardiagramm** võimaldab mitme salvestuse, sõna või foneemi valitud tunnuste võrdlemise. Tunnused normaliseeritakse enne kuvamist min-max meetodiga (tunnuste väärtused viiakse vahemikku 0-1).
- **Vokaalikaart** kujutab vokaalide paiknemist F1 ja F2 sagedusruumis. Rakendatakse Lobanovi normaliseerimist, mis teisendab F1 ja F2 väärtused z-skoorideks, et erinevate kõnelejate vokaalid oleksid paremini võrreldavad [16]. Graafikul kuvatakse iga vokaalipunkt koos häälikumärgiga.

- **Salvestuste klastrid ja koosinuskaugus hajuvusdiagrammil** on sarnasuse analüüsi esimene visualiseerimismeetod, mis kuvab valitud salvestused eri värvi klastritena PC1 ja PC2 tasandil hajuvusdiagrammil. Eraldi sümbolitega on klastrites märgitud salvestus, millele sarnaseid otsiti ning sellele leitud kõige sarnasemad salvestused.
- **Salvestuste koosinussarnasused tulpdigrammil** visualiseerib valitud salvestuste arvatud koosinussarnasuse väärtused tulpadena.
- **Andmete visualiseerimine tabelites:** lisaks graafiliste kujutistele on rakenduses võimalus kuvada iga visualisatsiooni DataFrame tabelina, mis kuvatakse visualisatsiooni all.

4.5 Sarnasuse analüüs

Sarnasuse analüüs võimaldab kasutajal tunnuste põhjal leida, millised salvestused on kõige sarnasemad, ning esitada arvatud tulemusi loeteluna tulpades või klastrite visualisatsioonina hajuvusdiagrammil. Järgnevalt kirjeldatakse meetodeid, mida rakenduse arenduses kasutati sarnasuse hindamiseks: klasterdamine ning koosinussarnasuse arvutamine. Enne sarnasuse arvutamist koostatati pandas DataFrame kõigi tunnuste salvestuste keskmiste väärtustega. See oli sisendiks kõigi rakendavate meetodite puhul.

4.5.1 Klasterdamine

Sisendiks on kõikide salvestiste eraldatud tunnuste keskmised väärtused. Iga tunnuse veerud skaleeritakse Scikit-learn StandardScaleri [17] abil, mille tulemusena on nende keskmine väärtus 0 ja standardhälve 1. Skaleeritud tunnuste hulk teisendatakse PCA meetodil vähendatud dimensioonidega ruumi. Saadud PCA ruumis jagatakse salvestused nelja klastrisse, kasutades Scikit-learn KMeans algoritmi. Nelja klastrit kasutatakse siinses töös näitejuhuna, tegelikkuses peaks kasutama optimaalse klastrite arvu leidmiseks kindlaid meetodeid.

4.5.2 Koosinussarnasuse arvutamine

Tunnused skaleeritakse StandardScaleriga samal viisil nagu klasterdamise puhul. Peale skaleerimist arvutatakse koosinuskaugus kõigile valitud salvestustele sihtsalvestuse suhtes. Koosinuskaugus valiti, sest see sobib kõnefailide analüüsiks. Plaan oli realiseerida ka teiste kauguste arvutamise kasutamine, nagu Manhattani, Eukleidese kaugus, kuid need jäid tegemata. Lõpuks valitakse kasutaja poolt määratud arv kõige sarnasemaid salvestusi ja visualiseeritakse tulpadena

4.5.3 Koosinussarnasuse arvutamine PCA-ga

Koosinussarnasuse arvutamist koos põhikomponentide analüüsiga kasutatatakse kahe visualisatsiooni koostamisel: hajuvusdiagrammil sarnaste punktide märkimisel tulpdiaagrammil sarnasuse väärtuste visualiseerimisel.

Peale tunnuste skaleerimist StandardScaleriga arvutatakse iga salvestise vahel koosinussarnasus PCA ruumis - sarnasust hinnatakse PCA-komponentide telgedel. PCA jätab kõrvale tunnused, mille varieeruvus on väike. Seega, kui on palju tunnuseid ja osa neist ei kannu kasulikku infot, siis need eemaldatakse.

Lõpuks valitakse kasutaja määratud arv kõige väiksema koosinuskaugusega ehk kõige sarnasemad salvestused ja kuvatakse nende koosinussarnasus tulpdiaagrammil või hajuvusdiagrammil eraldi sümbolitega.

4.6 Testimine

Rakendust testiti mitmetel kordadel koos juhendajaga. Rakenduse arenduse lõppedes tehti üks põhjalikum testimine kasutades rakenduse funktsionaalsuse kasutusjuhte. Testimise eesmärgid olid:

- Kasutatavuse kontrollimine
- Funktsionaalsuse kohta tagasiside saamine
- Tehniliste probleemide avastamine

4.6.1 Testitav rakendus

Tagasiside küsimiseks ning kasutajatele testimiseks loodi rakendusest PyInstalleriga [18] pakendatud ja käivitav versioon. Kuna andmebaasi ei seadistatud serverisse, siis oleksid kasutajad pidanud rakenduse kasutamiseks oma arvutis MongoDB andmebaasi seadistama. Seadistamis vaeva vähendamiseks kasutab pakendatud versioon MongoMock [19] tööriista, mis on mälu-põhine andmebaasiteek ja ei vaja eraldi paigaldust. MongoMock simuleerib MongoDB päringuid ja töötab näidisandmetega, mis testrakendusele lisati. See võimaldab kasutada kõiki rakenduse funktsioone samamoodi nagu päris MongoDB andmebaasiga. Ainuke erinevus on, et kui uusi andmeid importida, siis need on ajutised ja kaovad kui programm lõpetatakse. Testrakendusele lisati kaasa JSON formaadis näidisandmed ja .wav failid koos näidis helisalvestustega.

5 Tulemused ja analüüs

Selles peatükis tutvustatakse töö tulemusena valminud rakenduse põhilist funktsionaalsust ning hinnatakse, kuidas see täidab töö alguses püstitatud eesmärged. Lisaks võrreldakse saadud tulemusi ühe olemasoleva lahendusega, analüüsitakse rakendust testinud foneetikauuri tagasisidet ning kirjeldatakse võimalusi edasiarenduseks.

Valminud rakendus on avalikult kättesaadav GitHubi repositooriumis [20], kus on võimalik rakendus installeerida ja tutvuda koodiga.

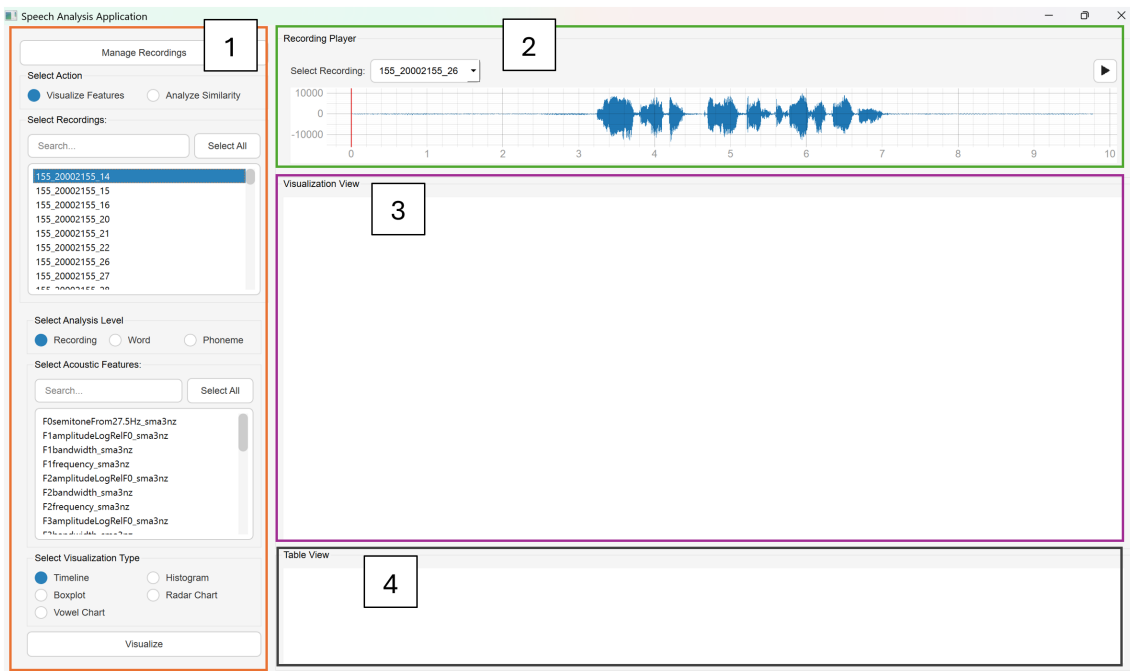
5.1 Rakenduse funktsionaalsus ja vaated

Järgnevalt analüüsitakse, kuidas rakenduse vaated ja funktsionaalsus täidavad töö alguses püstitatud kasutusjuhte. Kasutusjuhtude juures on lisatud rakenduse kuvatõmmised, et visualiseerida rakenduse toimimist.

5.1.1 Rakenduse kasutajaliidese tutvustus

Joonisel 1 on esitatud rakenduse põhivaade ja selle komponendid:

1. **Juhtpaneel** (*Control Panel*), kus kasutajal on võimalik salvestusi hallata, valida visualiseerimise tegevus (tunnuste visualiseerimine või sarnasuse visualiseerimine) ning peale tulemuste visualiseerimist andmeid eksportida.
2. **Helisalvestuste esitaja** (*Recording Player*) võimaldab helifaili esitamist ning kuvab selle amplituudi visualisatsiooni.
3. **Visualiseerimise vaade** (*Visualisation View*): vaates kuvatakse juhtpaneelis tehtud valikute põhjal genereeritud visualisatsioon.
4. **Tabelivaade** (*Table View*): vaates kuvatakse visualisatsiooni andmed tabeli kujul.



Joonis 1. Rakenduse põhivaade ja selle osad.

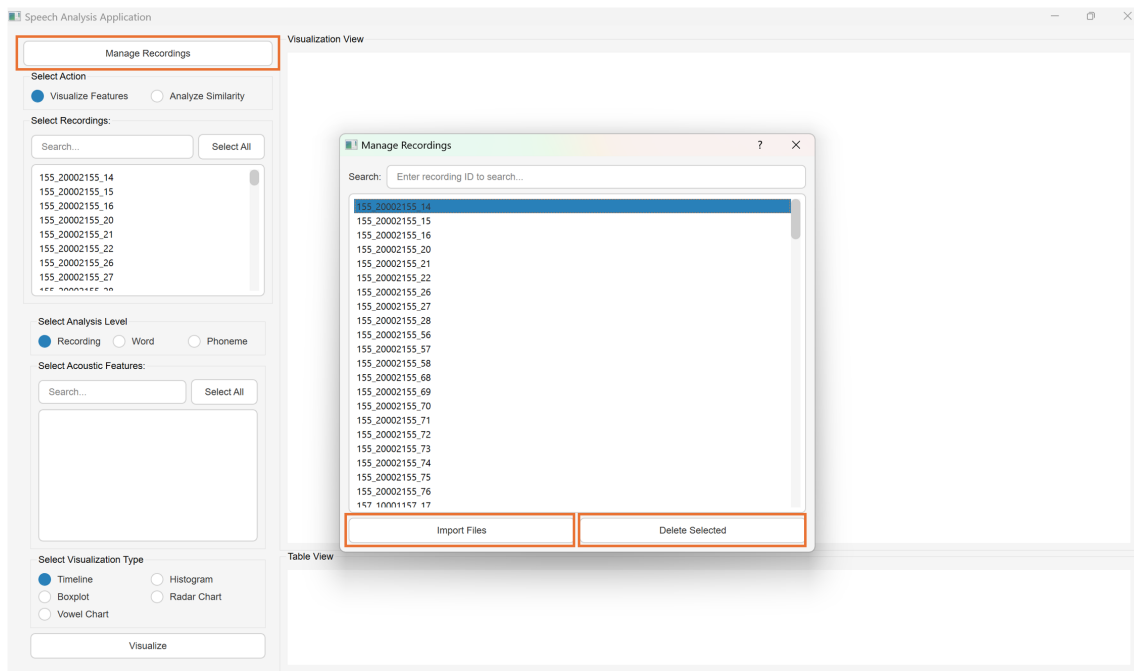
5.1.2 Kasutusjuhtude analüüs

Järgnevalt kirjeldatakse, kuidas loodud rakendus täidab funktsionaalsete nõuete kasutusjuhte.

GeMAPS tunnuste eraldamine kõnefailidest ja salvestuste töödeldud andmete kustutamine.

- Juhtpaneelis asuv nupp *Manage Recordings* avab salvestuste haldamise akna.
- Kasutaja saab *Import Files* nupu abil valida .wav ja .TextGrid faile, mille järel OpenSMILE'i abil eraldatakse GeMAPS tunnused ja salvestatakse MongoDB andmebaasi.

Joonisel 2 on esitatud salvestuste halduse aken, kus on näha töödeldud salvestuste nimekiri. Tulemuseks on, et Kasutaja saab töödelda ja hallata salvestusi. Rakendus kuvab nii õnnestumise- kui ka veateateid, mis on esitatud Lisa 3 joonistel 6 ja 7.

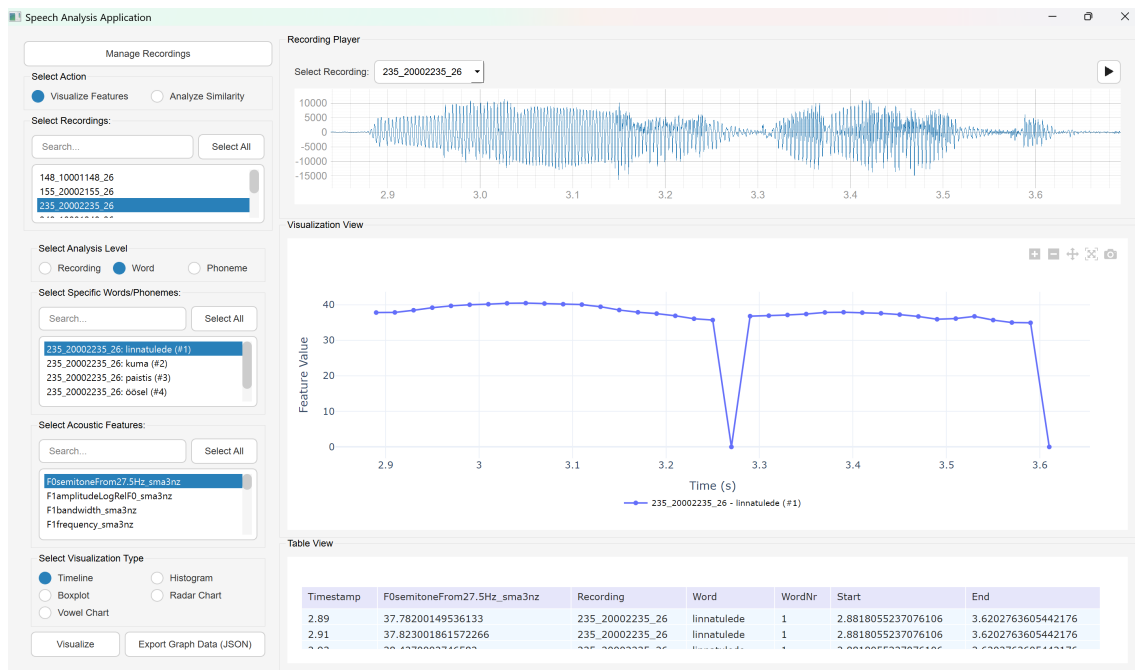


Joonis 2. Rakenduse salvestuste haldus.

Tunnuste visualiseerimine

- Visualisatsioonide kuvamiseks kasutaja valib *Visualize Features* valikutest sobivad tunnused.
- Valib salvestused (*Select Recordings*)
- Määrab, kas soovitakse vaadata tunnuseid salvestuse, sõna või foneemi tasandil (*Select Analysis Level*).
- Kui valiti foneemi või sõna tasand, tuleb valida ka sõnad või foneemid.
- Seejärel valib kasutaja konkreetsed GeMAPS tunnused (*Select Acoustic Features*) ning sobiva graafikutüübi (nt ajatelg, karp-, histogramm, radar, vokaalikaart).
- Pärast nupule *Visualize* vajutamist, kuvatakse *Visualization View*, kus on: interaktiivne graafik, mille paremas ülanurgas tööriistad (suurendus, pildi eksport jm).
- Visualisatsiooni vaate all kuvatakse tabelivaade (*Table View*).

Joonisel 3 on näha tunnuse visualiseerimine sõna kohta.

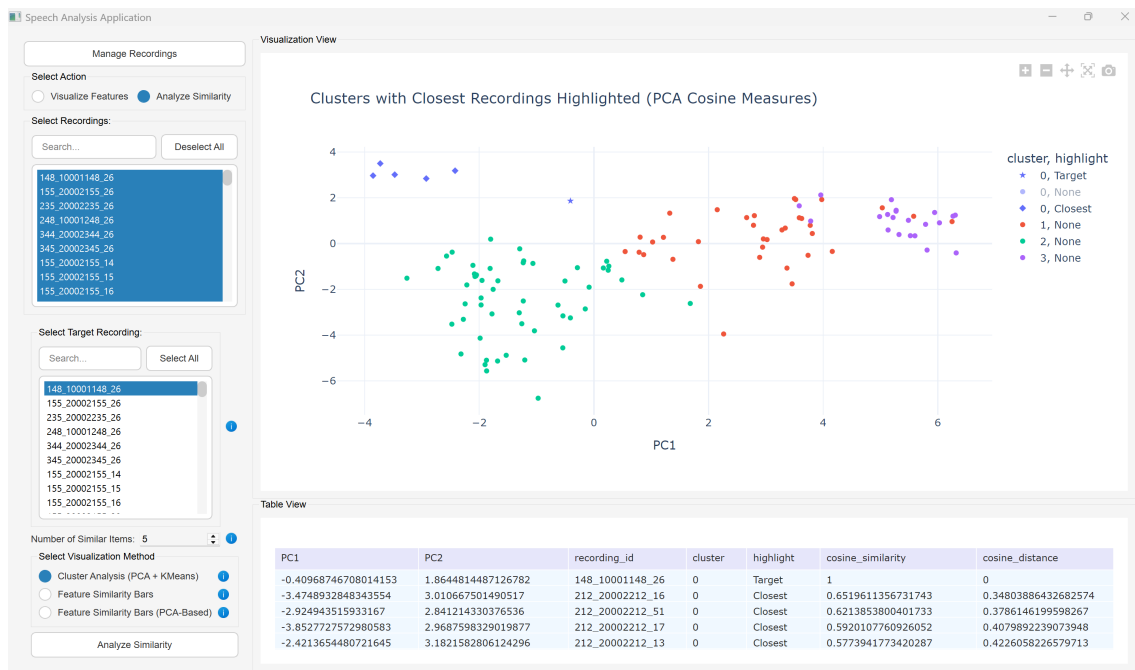


Joonis 3. Joondiagrammi näide ühe sõna ja tunnuse visualiseerimisel.

Sarnase kõneleja leidmine

- Kasutaja valib *Analyze Similarity* valiku, määrab salvestused, mida omavahel võrrelda (*Select Recordings*).
- See järel Valib ühe (*Target Recording*) salvestuse, millega teisi võrrelda ning sisestab arvu, mitu kõige sarnasemat salvestust soovitakse tulemusena saada (*Number of Similar Items*).
- Seejärel valib kasutaja sarnasuse visualiseerimise meetodi: *Cluster Analysis (PCA + KMeans)*, *Feature Similarity Bars* (koosinussarnasus) või *PCA-Based Similarity Bars* (koosinussarnasus PCA ruumis).
- Pärast nupuvajutust arvutab rakendus sarnasuse tulemused (nt klasterdamise või koosinussarnasuse), ning kuvatakse *Visualisation View*, kus on: tulpdiaagramm, mis näitab salvestuste sarnasuse väärtust või hajuvusdiagramm (klastrite korral), mis eristab klastreid värvidega ning märgib sihtsalvestuse ja kõige sarnasemad salvestused eri sümbolitega.

Joonisel 4 on esitatud näide sarnasuse visualiseerimisest PCA ja koosinussarnasuse meetodiga



Joonis 4. PCA koosinussarnasuse ja klastrite visualiseerimine hajuvusdiagrammil.

Helisalvestuse esitamine ja helilaine visualiseerimine

- *Recording Player* sektsioonis kuvatakse rakenduse *data* kaustas asuvad salvestused.
- Kasutaja valib faili ning vajutab *Play* nuppu.
- Rakendus mängib helisalvestust ning samal ajal kuvatakse helilaine. Mõlemad on kuvatud Joonisel 1.

Andmete ja visualisatsioonide eksportimine

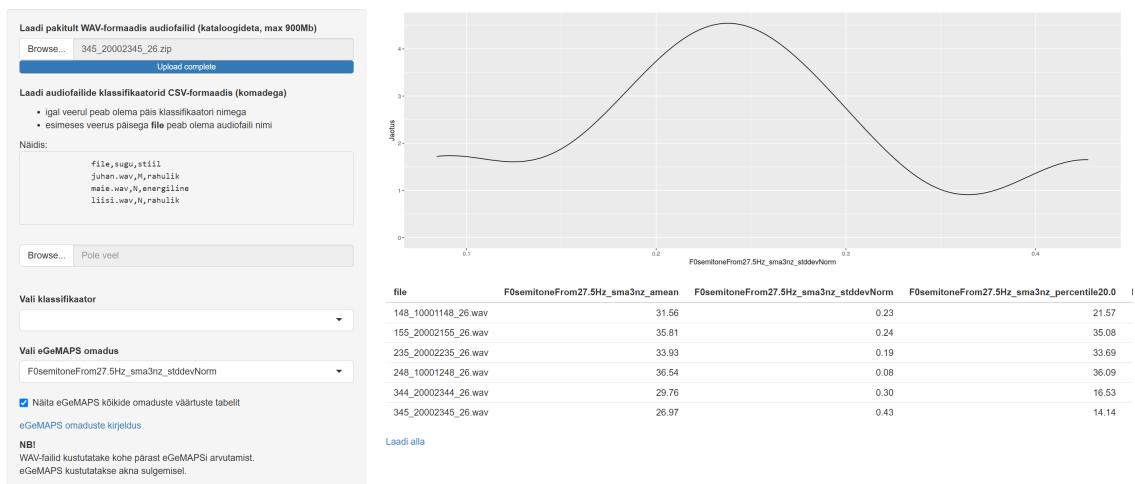
- Iga valminud graafiku kohal on kaamera märgiga nupp (Plotly sisseehitatud funktsioon). See võimaldab kasutajal salvestada staatilise pildifaili.
- Lisaks on juhtpaneelis *Export Graph Data (JSON)* nupp, mis salvestab kasutatud andmed JSON-vormingus. Nupud on nähtaval Joonisel 3.

5.2 Tulemuse võrdlus olemasoleva lahendusega

Järgnevalt võrreldakse valminud rakenduse funktsionaalsust Kõneveebi Audiofailide akustiliste omaduste võrdlemise lahendusega (Joonis 5). Kuna see tööriist on alternatiivsetest lahendusest kõige sarnasem oma fookuse, eesmärgi ja funktsionaalsuse poolest. Kuigi Praat või VoiceSauce on palju laiemalt kasutatud akustiliste tunnuste analüüsimisel, on nende fookus ja funktsionaalsus palju laiem ning nad ei keskendu kitsalt eGeMAPS-tunnuste

komplektile, nagu Kõneveebi lahendus ja ka loodud rakendus. Seega on sobivam rakendust võrrelda Kõneveebi lahendusega.

Audiofailide akustiliste omaduste võrdlus



Joonis 5. Kõneveebi Audiofailide akustiliste omaduste võrdlemise rakenduse põhivaade.

Omadus	Kõneveeb	Loodud rakendus
Tunnuste komplekt	Võimaldab analüüsida tervet 88 tunnusega eGeMAPS tunnuste-komplekti.	Võimaldab analüüsida ainult eGeMAPS v02 LLD komplekti, mis sisaldab üksnes madala taseme deskriptoreid (LLD).
Rakenduse tüüp	Veebirakendus.	Installitav töölauarakendus, töötab lokaalselt.
Kasutajaliides	Sarnane põhivaate ülesehitusele nagu loodud rakendusel: juhtpaneel valikutega vasakul ning paremal visualiseerimise ja andmetabeli sektsioonid.	Sarnane põhivaate ülesehitusele nagu kõneveebil: juhtpaneel valikutega vasakul ning paremal visualiseerimise ja andmetabeli sektsioonid.
Erinevate graafikute valik	Joondiagramm.	Pakub laiemat valikut tunnuste graafikuid, sealhulgas ajateljediagramm, histogramm, karpdiagramm, radardiagramm, vokaalikaart. Sarnasuse analüüsil on võimalik kasutada hajuv- ja tulpdiagramme. Samuti on graafikud interaktiivsed.

Omadus	Kõneveeb	Loodud rakendus
TextGridide info kasutamine	Ei paku võimalust lugeda ega töödelda TextGrid-failides sisalduvaid ajaintervalle (sõnu, foneeme). Analüüs piirdub salvestuse tasemega.	Võimaldab lugeda TextGrid-faili, seostada konkreetsete tunnuste väärtusi sõna või foneemi ajaintervalliga ning seejärel kuvada eraldi vaateid/visualiseeringuid valitud segmentidele.
Sarnaste salvestuste leidmine	Puudub.	Võimaldab mitut sarnasuse arvutamise meetodit (klasterdamine, koosinussarnasus, PCA-järgne koosinussarnasus). Visualiseerib tulemused kas tulpdiaagrammina (näiteks kõige sarnasemad salvestised) või hajuvusdiagrammil, kus sihtsalvestis ja sarnased salvestised on esile tõstatatud.
Andmete eksport	On napp arvatud tunnuste andmete allalaadimiseks, kuid see ei toiminud antud töö kirjutamise ajal.	Toetab andmete eksporti JSON faili ja pildifaili kujul.
Andmete haldus	Üleslaetud failid ja arvatud tunnused kustutatakse seansi lõpus. Failide maht on piiratud (kuni 900 MB) ja tulemusi ei salvestata.	Kasutab MongoDB andmebaasi analüüsi andmete hoidmiseks.

Kokkuvõttes on Kõneveebi rakendusel ja loodud rakendusel sarnane põhifookus, kuid veidi erinev ulatus. Kõneveeb on kindlasti mugav ja kiirem, kuna see ei vaja eraldi installimist ja seadistamist. Kuid käesolev rakendus on laiemas funktsionaalsusega ning sobib paremini mahukamaks analüüsiks, kus on vaja erinevaid visualiseerimise meetodeid, analüüsitulemuste salvestamist ja sarnasuse analüüsi.

5.3 Tagasiside

Rakenduse valideerimiseks küsiti tagasisidet ühelt foneetikaurijalt, kellele saadeti rakenduse installeritav versioon koos kasutusjuhendiga ja tagasiside küsimustega, mis on esitatud Lisas 4.

Rakenduse graafikute arusaadavus ja kasulikkus. Testija tagasid oli, et rakenduses loodud graafikud olid üldiselt arusaadavad ning märgiti, et need võiksid olla mugavad kiireks analüüsiks. Kõige kasulikumaks visualisatsiooniks peeti vokaalikaarti.

Tehnilised probleemid ja tõrked. Tagasiside põhjal ei märgatud rakenduse kasutamise ajal tehnilisi probleeme ega tõrkeid

Kasutajasõbralikkuse hinnang. Kasutajasõbralikkuse ja mugavuse osas anti rakendusele hinnang 3 skaalal 1-5, kus 1="väga mugav" ja 5="väga ebamugav". Kuigi rakendus ei olnud kasutaja arvates väga ebamugav, mainiti, et mõned klikkimisega seotud iseärasused vajasisid harjumist.

Näiteks tekitas segadust, et histogrammi visualisatsiooni valik muutub mitte-aktiivseks, kui on valitud korruga üle kahe tunnuse ja salvestuse korruga. See oli tahtlik piirang, et vältida olukordi, kus liiga paljude tunnuste ja salvestuste samaaegne kuvamine ühel graafikul muudaks tulemused raskesti loetavaks. Selle piirangu oleks pidanud kasutajale arusaadavamaks tegema näiteks teavituste või infonuppude abil, nagu tehti paljude rakenduse teiste funktsioonide puhul, kuid kahjuks jäid need selle piirangu jaoks lisamata.

5.4 Võimalused edasiarenduseks

Rakendusel on palju võimalusi edasiarendusteks. Järgmiselt tuuakse välja mõned võimalused, mis tulid esile peale arenduse lõppu:

- Võimaldada erinevate GeMAPS tunnustekomplektide arvutamine
- Võimaldada erinevate normaliseerimismeetodite rakendamine ja mitte rakendamine, et kasutajal oleks rohkem kontrolli andmete visualiseerimisel. Näiteks vokaalikaardi puhul normaliseeritakse väärtused Lobanov meetodiga ja ei ole võimalust teisi meetodeid kasutada või väärtused normaliseerimata jätta.
- Kasutajaliidese mugavdamine: näiteks menüüs võiks olla rohkem ruumi tunnuste ja salvestuste valimiseks
- Sarnasuse analüüsi valideerimine, rohkem meetodeid sarnaste salvestuste leidmiseks
- Andmete mahu probleemi lahendamine: kui andmete hulk kasvab, muutub rakendus aeglasemaks
- Muuta rakendus lihtsamini installeeritavaks, et kasutaja ei peaks MongoDB andmebaasi seadistama.

Lisaks tõi foneetikauuriija tagasisides esile järgmised soovitusid:

- Lisada funktsioon helifailide lõikamiseks.
- Luua võimalus esitada helifaile otse failide valiku menüüst.
- Kuvada rakenduses rohkem arvandmeid.

6 Kokkuvõte

Käesoleva bakalaureusetöö eesmärk oli luua kõneanalüüsi rakendus, mis võimaldab: GeMAPS akustiliste tunnuste eraldamist helisalvestustest, tunnuste visualiseerimist mitmesugustel diagrammitüüpidel (ajatelje-, histogrammi-, karp-, radari- ja vokaalikaardi diagrammid) ning sarnaste salvestiste leidmist.

Helifailidest eraldati GeMAPS madala taseme deskriptorid (LLD), kasutades OpenSMILE'i ning töödeldi kõnesalvestuste TextGrid faile märgendusinfo saamise jaoks. Tunnuste väärtused ja salvestuste andmed salvestati dokumentidena MongoDB andmebaasis. Rakenduse kasutajaliides loodi PyQt raamistikuga. Interaktiivsete diagrammide kuvamiseks (nt ajatelg, histogramm, karpdiagramm, radar ja vokaalikaart) kasutati Plotly teeki. Sarnaste salvestiste leidmiseks rakendatakse klasterdamist (KMeans) ja koosinussarnasuse arvutamist PCA-ga vähendatud dimensiooniruumis kui ka mitte vähendatud dimensioonidega tunnuste arvutamist. Sarnasusanalüüsi tulemused visualiseeritakse nii tulp- kui hajuvdiagrammidel.

Loodud rakendus vastas algselt püstitatud funktsionaalsetele nõuetele. Tagasiside ja testimise põhjal hinnati rakendust kasulikuks tunnuste kiireks visualiseerimiseks ning toodi esile võimalused kasutajaliidese mugavamaks muutmiseks ning funktsionaalsuse laiendamiseks.

Kasutatud kirjandus

- [1] Florian Eyben *et al.* *The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing*. [Kasutatud: 01.01.2025]. 2016. DOI: 10.1109/TAFFC.2015.2457417. URL: <https://ieeexplore.ieee.org/document/7160715>.
- [2] Kõneveeb. *Kõneveeb: Audiofailide akustiliste omaduste võrdlus*. Kasutatud: 01.01.2025. URL: <https://koneveeb.ee/akustika/>.
- [3] Praat. *Praat: doing phonetics by computer*. GitHub. [Kasutatud: 01.01.2025]. URL: <https://github.com/praat/praat>.
- [4] Y.-L. Shue *et al.* *VoiceSauce: A program for voice analysis*. UCLA Phonetics Laboratory. Kasutatud: 01.01.2025. URL: <https://phonetics.ucla.edu/voicesauce/>.
- [5] Mark Huckvale. *WASP (Windows Tool for Speech Analysis)*. Kasutatud: 01.01.2025. 2020. URL: <https://www.phon.ucl.ac.uk/resource/sfs/wasp/>.
- [6] OpenSMILE. *OpenSMILE: The Munich open-source large-scale multimedia feature extractor*. GitHub. [Kasutatud: 01.01.2025]. URL: <https://github.com/audeering/opensmile>.
- [7] Praat. *TextGrid file formats*. [Kasutatud: 15.01.2025]. URL: https://www.fon.hum.uva.nl/praat/manual/TextGrid_file_formats.html.
- [8] Praatio. *Praatio: A library for working with Praat TextGrid files in Python*. GitHub. Kasutatud: 01.01.2025. URL: <https://github.com/timmahrt/praatio>.
- [9] MongoDB Inc. *MongoDB: The developer data platform*. Official Website. Kasutatud: 01.01.2025. URL: <https://www.mongodb.com/>.
- [10] MongoDB Inc. *MongoDB Manual: Data Modeling*. Official Documentation. Kasutatud: 01.01.2025. URL: <https://www.mongodb.com/docs/manual/data-modeling/>.
- [11] PyQt5. *PyQt5: Python bindings for the Qt application framework*. Official Documentation. Kasutatud: 01.01.2025. URL: <https://www.riverbankcomputing.com/software/pyqt/intro>.
- [12] Tkinter. *Tkinter: The standard Python interface to the Tk GUI toolkit*. Python Documentation. Kasutatud: 01.01.2025. URL: <https://docs.python.org/3/library/tk.html>.

- [13] Martin Fitzpatrick. *PyQt vs. Tkinter — Which Should You Choose for Your Next GUI Project?* [Kasutatud: 15.01.2025]. URL: <https://www.pythonguis.com/faq/pyqt-vs-tkinter/>.
- [14] Matplotlib. *Matplotlib: Visualization with Python*. Kasutatud: 01.01.2025. URL: <https://matplotlib.org/stable/>.
- [15] Plotly. *Plotly: The interactive graphing library for Python, R, and other programming languages*. Kasutatud: 01.01.2025. URL: <https://plotly.com/>.
- [16] Erik R. Thomas ja Tyler Kendall. *NORM's Vowel Normalization Methods (v. 1.1)*. Kasutatud: 15.01.2025. URL: http://lingtools.uoregon.edu/norm/norm1_methods.php.
- [17] Scikit-learn. *sklearn.preprocessing.StandardScaler*. Kasutatud: 2025-01-15. URL: <https://scikit-learn.org/dev/modules/generated/sklearn.preprocessing.StandardScaler.html>.
- [18] David Cortesi, Giovanni Bajo ja William Caban. *PyInstaller Manual*. Kasutatud: 2025-01-15. URL: <https://pyinstaller.org/en/stable/>.
- [19] MongoMock. *MongoMock: Mocking library for MongoDB in Python*. GitHub. Kasutatud: 01.01.2025. URL: <https://github.com/mongomock/mongomock>.
- [20] Hanna Raudsepp. *Acoustic Feature Analysis*. Kasutatud: 2025-01-15. 2025. URL: <https://github.com/hrauds/acoustic-feature-analysis>.

Lisa 1 – Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks¹

Mina, Hanna Raudsepp

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose “Hääle akustiliste tunnuste visualiseerimise rakendus”, 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

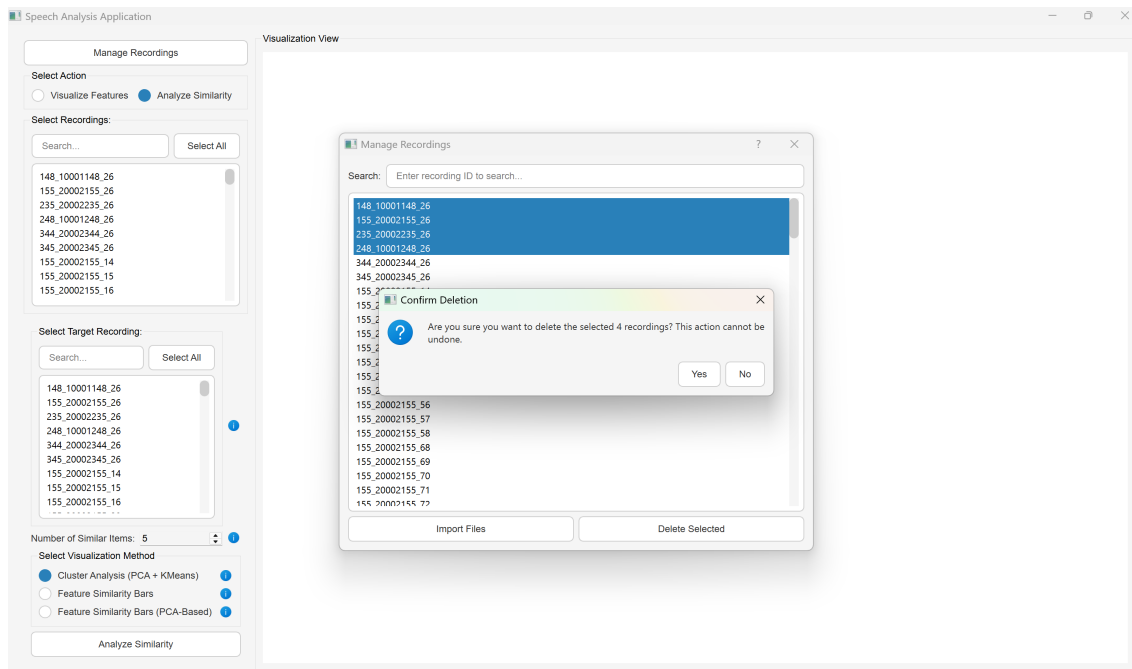
16.01.2025

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

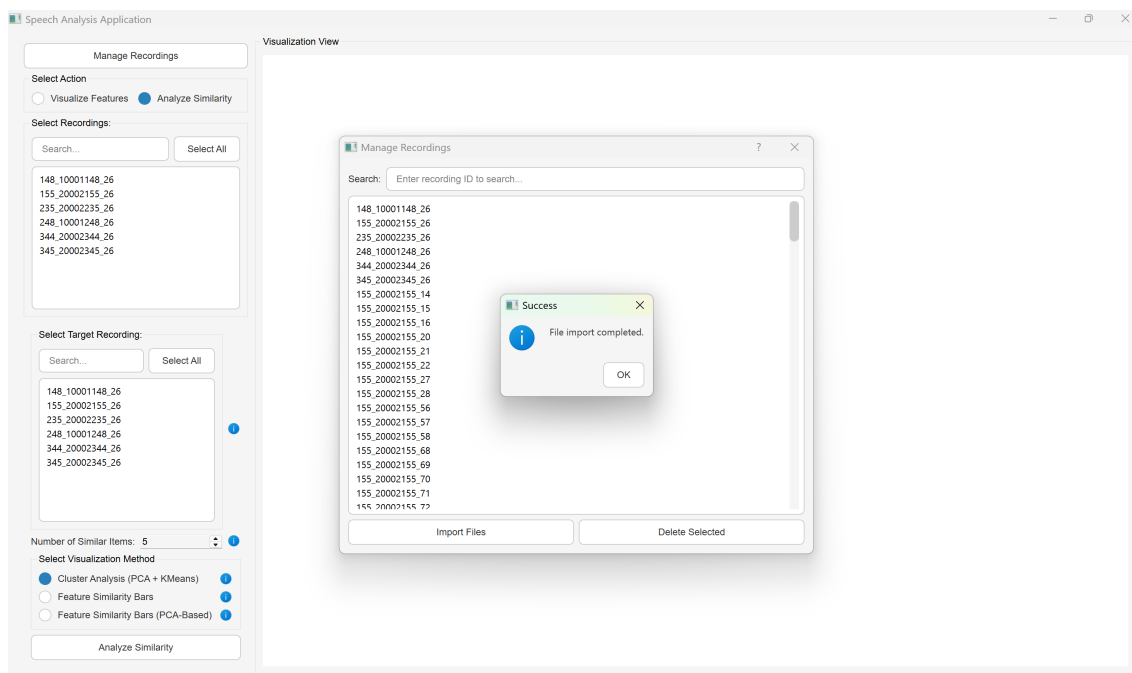
Lisa 2 - Eraldatud tunnuste loetelu

- Loudness_sma3
- alphaRatio_sma3
- hammarbergIndex_sma3
- slope0–500_sma3
- slope500–1500_sma3
- spectralFlux_sma3
- mfcc1_sma3
- mfcc2_sma3
- mfcc3_sma3
- mfcc4_sma3
- F0semitoneFrom27.5Hz_sma3nz18
- jitterLocal_sma3nz
- shimmerLocaldB_sma3nz
- HNRdBACF_sma3nz
- logRelF0-H1-H2_sma3nz
- logRelF0-H1-A3_sma3nz
- F1frequency_sma3nz
- F1bandwidth_sma3nz
- F1amplitudeLogRelF0_sma3nz
- F2frequency_sma3nz
- F2bandwidth_sma3nz
- F2amplitudeLogRelF0_sma3nz
- F3frequency_sma3nz
- F3bandwidth_sma3nz
- F3amplitudeLogRelF0_sma3nz

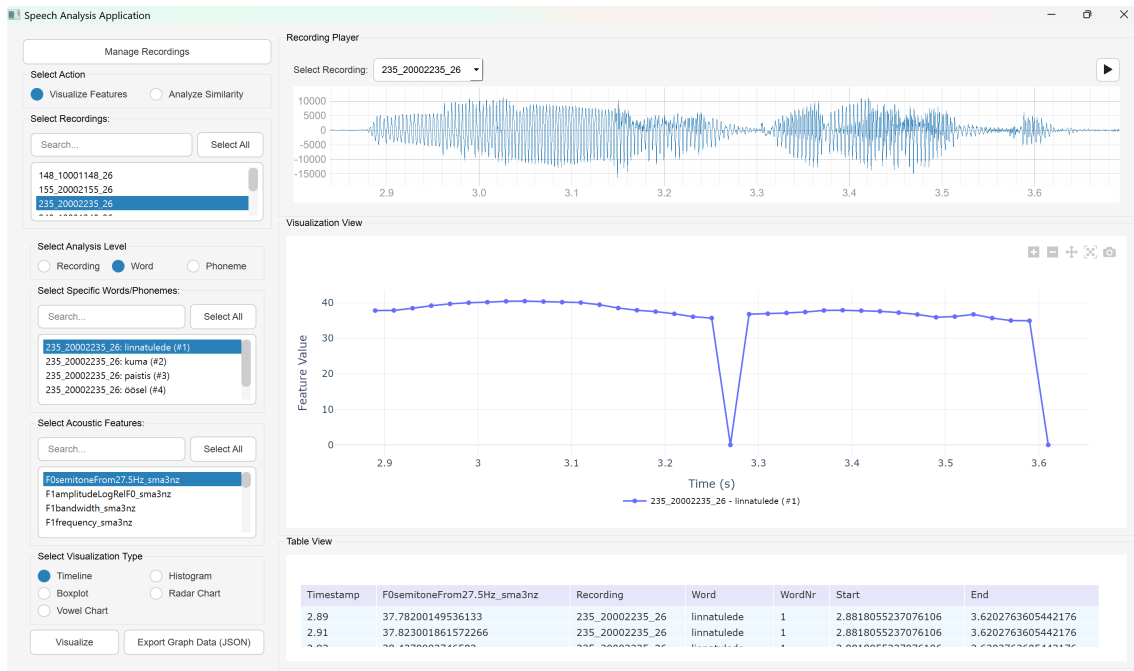
Lisa 3 - Näited rakenduse vaadetest



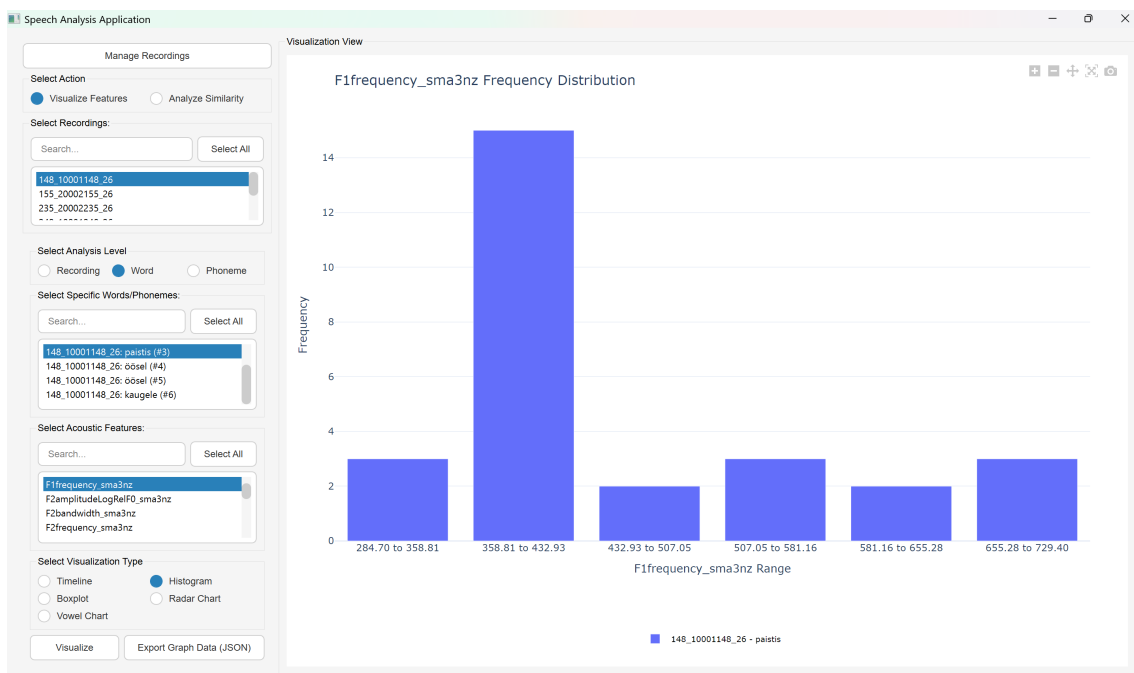
Joonis 6. Salvestuste andmete haldamise vaate kinnituse küsimise teavitus impordil.



Joonis 7. Salvestuste andmete haldamise vaate õnnestumise teavitus impordil.



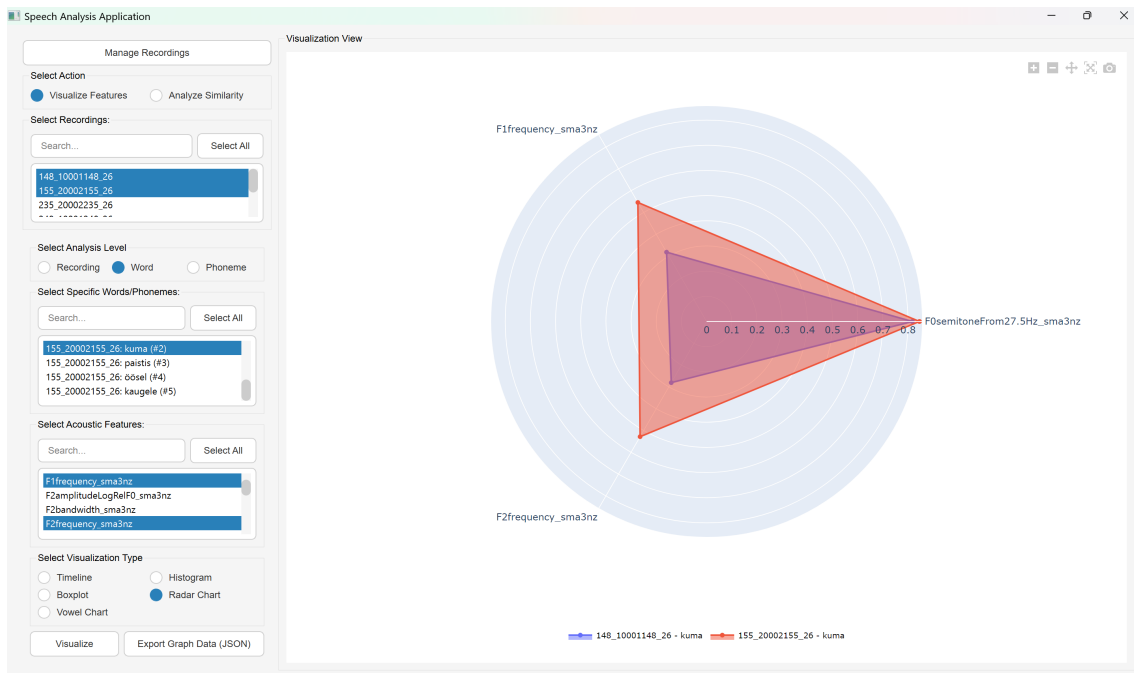
Joonis 8. Joondiagrammi näide ühe sõna ja tunnuse visualiseerimisel.



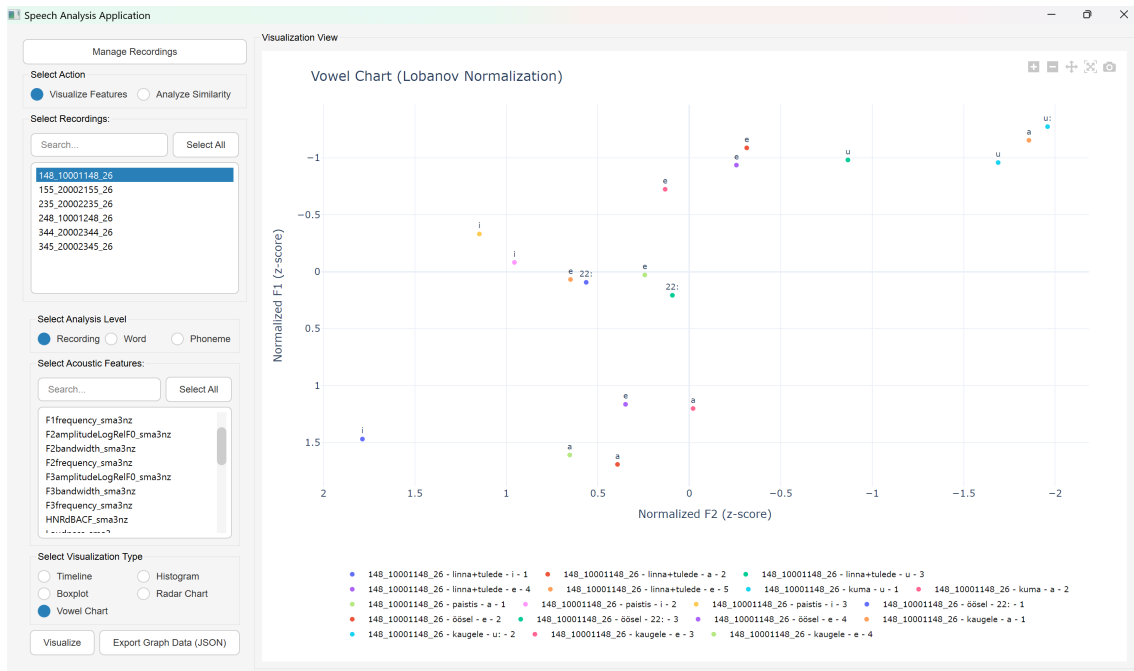
Joonis 9. Histogrammi näide ühe sõna ja tunnuse visualiseerimisel.



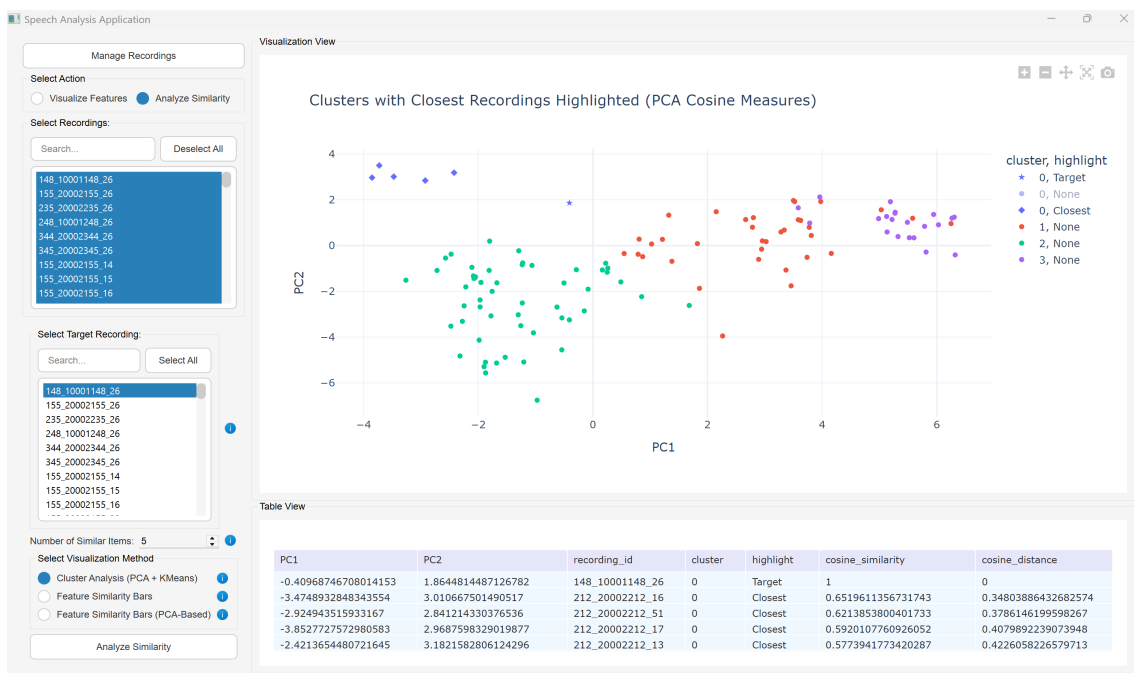
Joonis 10. Karpdiagrammi näide mitme sõna ja ühe tunnuse visualiseerimisel.



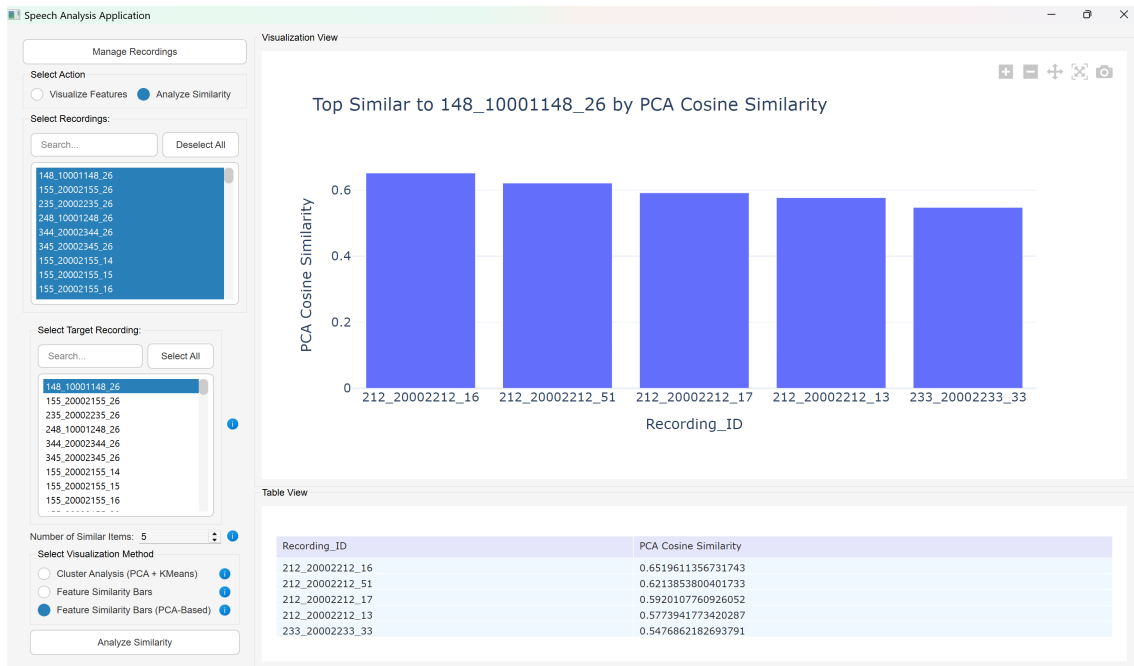
Joonis 11. Radardiagrammi näide ühe mitme sõna ja tunnuse visualiseerimisel.



Joonis 12. Vokaalikaardi näide ühe salvestuse visualiseerimisel.



Joonis 13. PCA koosinussarnasuse ja klastrite visualiseerimine hajuvusdiagrammil.



Joonis 14. PCA koosinussarnasuse visualiseerimine tulpdiagrammil.



Joonis 15. Koosinussarnasuse visualiseerimine tulpdiagrammil.

Lisa 4 - Tagasiside küsimused

Tere! Olen Hanna Raudsepp, Tallinna Tehnikaülikooli informaatika eriala bakalaureuseõppe tudeng. Minu lõputööks on "Hääle akustiliste tunnuste visualiseerimise rakendus" ning selle eesmärk on luua tööriist, mis võimaldab kõnesalvestuste akustiliste tunnuste analüüsi ja visuaaliseerimist.

Küsimustik on koostatud, et koguda tagasisidet rakenduse funktsionaalsuse ja kasutajamugavuse kohta.

Küsimused:

1. Kas tegelete foneetika või kõne uurimisega?
2. Kas rakenduse loodud graafikud olid arusaadavad ja kasulikud? (Rakenduse graafikud: ajagraafik, histogramm, karpdiagram, radar, vokaalikaart, sarnaste salvestuste klaster, sarnaste salvestuste tulpdiagrammid)
3. Kas märkasite rakenduse kasutamisel tehnilisi probleeme või tõrkeid? Kui jah, siis milliseid?
4. Hinnake skaalal 1–5, kui kasutajasõbralik ja mugav oli rakenduse kasutamine teie arvates? 1 = „väga ebamugav“, 5 = „väga mugav“. Selgitage oma hinnangut.
5. Millised rakenduse funktsioonid tundusid teie jaoks kõige kasulikumad?
6. Milliseid täiustusi või lisafunktsioone soovitaksite rakendusele lisada?
7. Muud kommentaarid