



Kauno technologijos universitetas

Informatikos fakultetas

**Akustinis 3D lokalizavimas naudojant plečiamą bevielį
mikrofonų masyvą**

Baigiamasis magistro studijų projektas

Ramūnas Purtokas

Projekto autorius

Doc. Armantas Ostreika

Vadovas

Kaunas, 2022



Kauno technologijos universitetas

Informatikos fakultetas

Akustinis 3D lokalizavimas naudojant plečiamą bevielį mikrofonų masyvą

Baigiamasis magistro studijų projektas

Dirbtinio intelekto informatika (6211BX007)

Ramūnas Purtokas

Projekto autorius

Doc. Armantas Ostreika

Vadovas

Prof. Rytis Maskeliūnas

Recenzentas

Kaunas, 2022



Kauno technologijos universitetas

Informatikos fakultetas

Ramūnas Purtokas

Akustinis 3D lokalizavimas naudojant plečiamą bevielį mikrofonų masyvą

Akademinio sąžiningumo deklaracija

Patvirtinu, kad mano, Ramūno Purtoko, baigiamasis projektas tema „Akustinis 3D lokalizavimas naudojant plečiamą bevielį mikrofonų masyvą“ yra parašytas visiškai savarankiškai ir visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

(vardą ir pavardę įrašyti ranka)

(parašas)

Purtokas Ramūnas. Akustinis 3D lokalizavimas naudojant plečiamą bevielį mikrofonų masyvą. Magistro studijų baigiamasis projektas / vadovas doc. Armantas Ostreika; Kauno technologijos universitetas, informatikos fakultetas.

Studijų kryptis ir sritis (studijų kryptčių grupė): Informatika (Informatikos mokslai).

Reikšminiai žodžiai: garso šaltinio lokalizavimas, neuroniniai tinklai, bevielis akustinių jutiklių tinklas, mikrofonų masyvas, mašininis mokymasis.

Kaunas, 2022. 55 p.

Santrauka

Garso šaltinio lokalizavimo metodai yra taikomi daugumoje akustinių įvesties įrenginių krypties ir atstumo įvertinimui, vienu garsų sustiprinimui, kitų slopinimui. Vis tikslesnio garso šaltinio vietos nustatymo reikia technologijų automatizavimo srityse: nuotolinių konferencijų metu, išmaniųjų namų valdyme, saugos bei stebėjimo sistemose ir kitur. Garso atvykimo krypties nustatymas, naudojant glaudžiai sujungtą mikrofonų masyvą, yra dažniausiai tiriama sritis. Tačiau paskirstytų sistemų tobulėjimas leidžia realizuoti bevelius akustinių jutiklių tinklus, kurie turi pranašumą dėl didesnių atstumų tarp mikrofonų.

Šiame darbe bevelių akustinių jutiklių tinklą siūloma naudoti garso šaltinio koordinacių nustatymui 3D erdvėje. Kai tokios sistemos mikrofonai gali judėti, turime naudoti ne tik garso įrašus, bet ir akustinių jutiklių koordinates. Tai nesudėtinga padaryti lėtai veikiančiuose lokalizavimo algoritmuose, bet iki šiol sukurtų mašininio mokymosi modelių struktūros netinka. Todėl siekiama įvertinti garso šaltinio vietos nustatymo galimybes naudojant dinamišką akustinių jutiklių sistemą.

Analizuojant mokslinę literatūrą išskiriami lokalizavimo metodai ir įvairių sukurtų neuroninių tinklų modelių pritaikymai garso apdorojimo užduotims. Parinkus tinkamą bevelių akustinių jutiklių sistemos struktūrą, realizuojama garso įvesties sistema duomenims realiomis sąlygomis rinkti. Naudojant geometrinio lokalizavimo ir tinklelio paieškos algoritmus, įvertinamas garso šaltinio vietos nustatymas. Bandoma nustatyti, kokios mašininio mokymosi technikos tinka labiausiai. Šiame darbe tiriamas konvoliucinių, rekurentinių, liekamosios vertės ir dėmesio mechanizmu pagrįstų sluoksnių tinkamumas garso šaltinio koordinacių nustatymui. Modifikuotų dažniausiai naudojamų neuroninių tinklų sėkmingas pritaikymas lokalizavimui leis indukuoti atitinkamų neuroninių sluoksnių tinkamumą garso analizei ir objekto koordinacių nustatymui.

Sukurta bevielė akustinių jutiklių tinklo sistema geba realiu laiku įrašinėti ir siųsti duomenis į centrinį valdiklį, kuriame atliekamas lokalizavimas. Tačiau garso įrašuose esantis statinis triukšmas ir galimai nepakankamai tikslus sistemos mazgų laiko sinchronizavimas, neleidžia pasiekti tikslių garso šaltinio lokalizavimo rezultatų. Tyrimo metu nustatyta, kad garso spektrogramos ir konvoliuciniai, rekurentiniai neuroniniai tinklai labiausiai tinka lokalizavimui, nes išskiria garsą nuo triukšmo ir sumažina nereikalingos informacijos kiekį. Dėl to modelis veikia greičiau ir tiksliau nei aritmetiniai algoritmai.

Purtokas Ramūnas. Acoustic 3D localization using a scalable array of microphones. Master's Final Degree Project / supervisor doc., Armantas Ostreika; Faculty of informatics, Kaunas University of Technology.

Study field and area (study field group): Informatics (Computing).

Keywords: sound source localization, neural networks, wireless acoustic sensor network, microphone array, machine learning.

Kaunas, 2022. 55.

Summary

Sound source localization methods are used in most acoustic input devices to estimate the direction and distance, to amplify some sounds, and to attenuate others. Increasingly accurate location of the audio source is needed in technology automation: remote conferencing, smart home management, security and surveillance systems, and more. Determining the direction of sound entry using a tightly coupled array of microphones is the most studied area. However, the development of distributed systems makes it possible to realize wireless networks of acoustic sensors, which have the advantage of greater distances between microphones.

In this work, it is proposed to use a network of wireless acoustic sensors to determine the coordinates of an audio source in 3D space. When the microphones of such a system can move, we must use not only the audio recordings but also the coordinates of the acoustic sensors. This is not difficult to do with slow-running localization algorithms, but the structures of the machine learning models developed so far are not appropriate. Therefore, the aim is to evaluate the location of the sound source using a dynamic system of acoustic sensors.

Analyzing the scientific literature, localization methods and applications of various developed neural network models to sound processing tasks are distinguished. After selecting the appropriate structure of the wireless acoustic sensor system, an audio input system for data collection under real conditions is implemented. The location of the audio source is evaluated using geometric localization and grid search algorithms. Attempts are made to determine which machine learning techniques are most appropriate. In this work, the suitability of convolutional, recurrent, residual and attention-based layers for the determination of sound source coordinates is investigated. The successful application of modified commonly used neural networks for localization will induce the suitability of the respective neural layers for sound analysis and object coordination.

The developed wireless acoustic sensor network system can record and send data in real time to a central controller where localization is performed. However, the static noise in the audio recordings and the time synchronization of the system nodes may not be accurate enough to achieve accurate localization of the audio source. The study found that sound spectrograms and convolutional recurrent neural networks are best suited for localization because they emit noise from sound and reduce the amount of unnecessary information. As a result, the model works faster and more accurately than arithmetic algorithms.

Turinys

| | |
|---|-----------|
| Lentelių sąrašas | 8 |
| Paveikslų sąrašas | 9 |
| Santrumpų ir terminų sąrašas | 10 |
| Įvadas..... | 11 |
| 1. Garso šaltinio lokalizavimo literatūros analizė | 13 |
| 1.1. Garso savybių naudojimas lokalizavimui..... | 13 |
| 1.1.1. Garso energija..... | 13 |
| 1.1.2. Atvykimo laikas..... | 13 |
| 1.1.3. Atvykimo kryptis..... | 14 |
| 1.1.4. Spindulio formavimas | 15 |
| 1.1.5. Dažnių spektrograma..... | 16 |
| 1.1.6. Kryžminės koreliacija..... | 16 |
| 1.2. Garso įvesties kanalų kiekis | 16 |
| 1.2.1. Monofoninis garsas | 16 |
| 1.2.2. Binauralinis garsas..... | 17 |
| 1.2.3. Trijų ir daugiau kanalų garsas | 17 |
| 1.3. Garso įvesties jutiklių išdėstymas | 17 |
| 1.3.1. Plokščiosios struktūros | 17 |
| 1.3.2. Erdvinės struktūros | 18 |
| 1.3.3. Dinamiškos struktūros | 18 |
| 1.4. Lokalizavimui naudojami modeliai | 19 |
| 1.4.1. Geometrinis lokalizavimas | 19 |
| 1.4.2. K-vidurkių metodas | 19 |
| 1.4.3. Pilnai sujungti neuroniniai tinklai | 20 |
| 1.4.4. Konvoliuciniai neuroniniai tinklai..... | 20 |
| 1.4.5. Rekurentiniai neuroniniai tinklai..... | 20 |
| 1.4.6. Liekamosios vertės neuroniniai tinklai..... | 20 |
| 1.4.7. Kodavimo-dekodavimo neuroniniai tinklai..... | 21 |
| 1.4.8. Dėmesiu pagrįsti neuroniniai tinklai | 21 |
| 1.5. Literatūros analizės išvados..... | 21 |
| 2. Garso šaltinio lokalizavimo sistemos projektas | 23 |
| 2.1. Reikalavimų specifikacija | 23 |
| 2.1.1. Projektuojamos priemonės paskirtis..... | 23 |
| 2.1.2. Projekto kūrimo pagrindimas | 23 |
| 2.1.3. Galimi sprendimo vartotojai..... | 23 |
| 2.1.4. Funkciniai reikalavimai | 24 |
| 2.1.5. Nefunkciniai reikalavimai | 25 |
| 2.2. Pasirinkti įrankiai ir technologijos..... | 25 |
| 2.3. Sistemos architektūra..... | 26 |
| 2.3.1. Sistemos įrangos struktūra..... | 26 |
| 2.3.2. Programinė sistemos struktūra | 27 |
| 2.3.3. Mazgų sinchronizavimas | 29 |
| 2.4. Duomenų specifikacija | 30 |
| 2.4.1. Modelio duomenų įvestis, išvestis..... | 30 |

| | |
|---|-----------|
| 2.4.2. Garso įrašų specifikacija..... | 31 |
| 2.4.3. Garso įrašų duomenų papildymas | 31 |
| 2.5. Vertinimo kriterijus | 31 |
| 2.6. Geometrinių modelių specifikacija..... | 32 |
| 2.6.1. Garso sklidimo skirtumo nustatymas | 32 |
| 2.6.2. Geometrinė lokalizacija..... | 32 |
| 2.7. Neuroninių modelių specifikacija..... | 32 |
| 2.7.1. Pilnai sujungtas neuroninis tinklas | 33 |
| 2.7.2. Konvoliuciniai neuroniniai tinklai..... | 33 |
| 2.7.3. Rekurentinis neuroninis tinklas | 35 |
| 2.7.4. Kodavimo-dekodavimo neuroniniai modeliai..... | 36 |
| 2.7.5. Liekamosios vertės neuroniniai modeliai..... | 38 |
| 2.8. Sistemos projektavimo išvados | 38 |
| 3. Garso šaltinių lokalizavimo tyrimas | 40 |
| 3.1. Akustinių jutiklių įvertinimas..... | 40 |
| 3.2. Sistemos sinchronizavimo įvertinimas | 41 |
| 3.3. Bevielių akustinių jutiklių tinklo įvertinimas | 41 |
| 3.4. Garso įrašinėjimo sąlygos..... | 42 |
| 3.5. Geometrinio lokalizavimo įvertinimas | 43 |
| 3.6. Neuroninių modelių įvertinimas..... | 45 |
| 3.6.1. Pirminis duomenų apdorojimas..... | 45 |
| 3.6.2. Mašininio mokymo metodologija..... | 46 |
| 3.6.3. Neuroninių modelių rezultatų palyginimas | 46 |
| 3.7. Tyrimo proceso apibendrinimas | 48 |
| Išvados | 49 |
| Literatūros sąrašas | 50 |

Lentelių sąrašas

| | |
|--|----|
| 1 lentelė. Energijos skirtumu grįsto geometrinio lokalizavimo algoritmo <i>Power</i> rezultatai | 43 |
| 2 lentelė. Atvykimo laikų skirtumu grįsto geometrinio lokalizavimo algoritmo <i>TDOA</i> rezultatai .. | 43 |
| 3 lentelė. Atvykimo laikų skirtumu grįsto geometrinio lokalizavimo modifikuoto algoritmo <i>TDOA_multi</i> rezultatai..... | 44 |
| 4 lentelė. Atvykimo laikų skirtumu grįsto geometrinio lokalizavimo k-vidurkių algoritmo <i>TDOA_kmeans</i> rezultatai | 44 |
| 5 lentelė. Geometrinių lokalizavimo algoritmų rezultatai..... | 45 |
| 6 lentelė. Dinamiškų garso šaltinio lokalizavimo algoritmų rezultatai | 46 |
| 7 lentelė. Visų garso šaltinio lokalizavimo algoritmų rezultatai | 47 |

Paveikslų sąrašas

| | |
|---|----|
| 1 pav. Garso atvykimo laiku pagrįsta vietos nustatymo sistema [1]..... | 14 |
| 2 pav. Garso atvykimo kryptimi pagrįstas lokalizavimo modelis vienam šaltiniui [1] | 15 |
| 3 pav. Garso atvykimo kryptimi pagrįstas lokalizavimo modelis keliems šaltiniams [1] | 15 |
| 4 pav. Sistemos panaudos atvejų diagrama..... | 24 |
| 5 pav. Sistemos diegimo diagrama | 26 |
| 6 pav. Sistemos tinklo diagrama | 27 |
| 7 pav. Garso apdorojimo veiklos diagrama | 28 |
| 8 pav. Garso įrašymo veiklos diagrama | 29 |
| 9 pav. Pilnai sujungtas neuroninis tinklas <i>NN_TDOA</i> | 33 |
| 10 pav. Konvoliucinis neuroninis tinklas su garso įvestimi <i>CNN_raw_pool</i> | 34 |
| 11 pav. Konvoliucinis neuroninis tinklas su garso įvestimi <i>CNN_raw_big</i> | 34 |
| 12 pav. Konvoliucinis neuroninis tinklas su spektrogramos įvestimi <i>CNN_spectro_single</i> | 35 |
| 13 pav. Konvoliucinis neuroninis tinklas su spektrogramų įvestimi <i>CNN_spectro_multi</i> | 35 |
| 14 pav. Rekurentinis neuroninis tinklas su spektrogramų įvestimi <i>CRNN_big</i> | 36 |
| 15 pav. Rekurentinis neuroninis tinklas su spektrogramų įvestimi <i>CRNN_small</i> | 36 |
| 16 pav. U-Net modelis su spektrogramų įvestimi <i>U-Net_spectro</i> | 37 |
| 17 pav. Dėmesio mechanizmu pagrįstas neuroninis modelis <i>Attention_NN</i> | 37 |
| 18 pav. Liekamosios vertės neuroninis modelis su garso įvestimi <i>Resid_NN</i> | 38 |
| 19 pav. Liekamosios vertės neuroninis modelis <i>ResNet-50</i> | 38 |
| 20 pav. Garso įrašas su periodiniais trikdžiais | 40 |
| 21 pav. Garso įrašas sumažinus trikdžius | 40 |
| 22 pav. Garso įrašas su energijos šaltinio keliamu triukšmu | 41 |

Santrumpų ir terminų sąrašas

Santrumpos:

ADC. – analoginis-skaitmeninis keitiklis (angl. Analogue to Digital Converter);

DMA. – tiesioginė prieiga prie atminties (angl. Direct Memory Access);

GŠL. – garso šaltinio lokalizavimas;

NTP. – tinklo laiko protokolas (angl. Network Time Protocol);

RMSE. – kvadratinės šaknies vidurkio paklaidos (angl. Root Mean Square Error);

ReLU. – „lygintuvo“ funkcija (angl. Rectified Linear Unit);

SRP-PHAT. – nukreipto atsako galios fazės transformacija (angl. Steered-Response Power Phase Transform);

UDP. – vartotojo datagramos protokolas (angl. User Datagram Protocol).

Įvadas

Žmonės naudojami viena iš pagrindinių juslių – klausa orientuotis aplinkoje, todėl jau nuo senų laikų binauralinio girdėjimo ypatumai buvo svarbi sritis. Bet tik 1989 metais pradėtas tyrinėti garso šaltinio lokalizavimo (GŠL) pritaikymas robotų kūrimui. Naudodami du mikrofonus ir garso atvykimo iki jų laiko skirtumą tyrėjai sukonstravo primityvų ir lėtai veikiančią robotuką. 1993 metai Japonijos mokslininkai taip pat susidomėjo šia sritimi ir nuo to laiko buvo išleista įvairių mokslinių darbų, nagrinėjančių garso bangos savybes ir jų pritaikymą. Dažniausiai naudojamos garso stiprio, atvykimo laiko ir krypties savybės GŠL sprendimuose. Tačiau pastebėta, kad ir garso deformacijos, dėl išorinės ausies formos, žmogui leidžia nustatyti sklindančio garso kryptį. Atlikta daug tyrimų siekiant išsiaiškinti klausos ypatumus ir kaip šias žinias pritaikyti, automatizuoti.

Tobulėjant kompiuteriams ir garso jutiklių tikslumui, naudojamos įvairios kombinacijos garso bangų savybių ir mikrofonų išdėstymo bei jų kiekio. GŠL probleminė sritis turi daug parametrų. Reikia įvertinti kiek ir kokių garso šaltinių gali būti aplinkoje, kokia įrašinėjimo sistemos struktūra, ar objektai juda aplinkoje, ar gali būti atspindžių, aido, kokia aplinkos temperatūra. Dažniausiai tiriama sujungtų mikrofonų rinkinys, kuris parenkamas atskaitos tašku. Naudojant garso jutiklių masyvą išdėstytą tam tikroje glaudžioje konstrukcijoje, galime iš anksto numatyti atstumus tarp mikrofonų bei centrinį tašką. Tačiau maži atstumai tarp jutiklių padidina jų garso įrašinėjimo kokybės ir tikslaus laiko sinchronizavimo svarbą. Dėl tokio sistemos mobilumo, paprastumo ir pritaikomumo robotams, tai plačiausiai ištirta GŠL sritis. Tyrimuose stengiamasi apskaičiuoti kryptį, iš kur atsklido garsas. Nustatinėjamas vienas arba du kampai atitinkamai dvimatėje arba trimatėje erdvėje. Tam naudojami algebriniai skaičiavimai, tačiau vis dažniau pritaikomi ir mašininio mokymosi metodai. Šias technologijas jau pritaikė „Siemens“ organizacija sukurdamą kamerą, kuri nedideliu atstumu parodo, iš kur sklinda garsas. Dėl glaudžios struktūros reikalinga brangesnė ir tikslesnė garso įrašinėjimo technika. Vis tiek pakankamai geri garso krypties nustatymo rezultatai pasiekiami tik įrašinėjant iš arti. Nors atlikta daug tyrimų, tačiau labai maža jų dalis stengiasi nustatyti objektų buvimo vietos koordinatas trijų dimensijų erdvėje. Iki šiol pasiūlytos metodikos sprendžia tik dalį GŠL probleminės srities.

Vis dažniau yra tiriamos ir kuriamos bevielės akustinių jutiklių sistemos. Nors jas sunkiau sinchronizuoti ir perduoti didelius kiekius duomenų realiu laiku, bet tobulėjančios technologijos leidžia realizuoti pakankamai tikslus ir pigius sprendimus. Ištyrus tokios sistemos galimybes, naudojant mašininio mokymosi metodus, būtų sukuriama universaliai plečiama ir plačiai taikoma GŠL metodika. Objekto vietos nustatymas garsu būtų naudingas gamyklose, išmaniuosiuose namuose, ar vaizdo konferencijų metu, kai nenaudojama brangi, daug vietos užimanti ir vartotojo sąveikos nereikalaujanti sistema. Bevielės akustinių jutiklių tinklas leidžia stebėti visą patalpos erdvę vienu metu. Todėl GŠL galimas didesniu atstumu ir nereikia fiziškai judinti garso įrašinėjimo įrangos. Dėl sistemos dinamiškumo svarbu žinoti mikrofonų koordinatas arba gebėti jas nustatyti automatiškai. Kadangi garsas įrašomas iš kelių skirtingų vietų, tokia sistema turėtų būti tinkamiausia garso šaltinio vietos, o ne tik krypties nustatymui.

Šio tiriamojo darbo tikslas yra įvertinti garso šaltinio vietos trimatėje erdvėje nustatymo galimybes, naudojant bevielę akustinių jutiklių tinklą. Siekiant tai įgyvendinti, iškelti tokie uždaviniai:

1. nustatyti literatūroje pateikiamas tinkamiausias garso savybes bei metodus, kuriais galima rasti garso šaltinio vietą 3D erdvėje;

2. įvertinti suprojektuotą dinamišką, pigią ir lengvai plečiama garso įvesties bei šaltinio vietos nustatymo realiu laiku trimatėje aplinkoje sistemą;
3. pritaikyti neuroninių modelių struktūras 3D koordinatinių nustatymui naudojant garso įrašus ir mikrofonų koordinates;
4. įvertinti geometrinių ir mašininio mokymosi modelių lokalizavimo tikslumą ir greitį.

Šiame dokumente pateikiama GŠL mokslinės literatūros analizės skyrius, kuriame įvertinamos garso savybės, lokalizavimo sistemų struktūros ir taikomos metodikos, algoritmai. Projektavimo skyriuje nurodoma siūloma garso įrašinėjos sistemos struktūra, duomenų rinkimo metodika ir GŠL algoritmų bei modelių struktūros. Lokalizavimo metodų tyrimo skyriuje aptariami eksperimentiniu būdu surinkti algoritmų rezultatai. Geometrinio lokalizavimo algoritmų rezultatų priklausomybės nuo įvesties duomenų formato dėsningumai analogiškai atsispindi ir neuroninių tinklų tyrime. Šiame darbe indukcijos būdu nustatoma konvoliucinių, rekurentinių, liekamosios vertės ir dėmesio mechanizmu grįstų sluoksnių tinkamumas garso šaltinio koordinatinių nustatymui.

1. Garso šaltinio lokalizavimo literatūros analizė

Šiame skyriuje analizuojami įvairūs GŠL probleminės srities aspektai, pateikti moksliniuose tyrimuose. Dėmesys skiriamas garso bangų savybėms, kurios leidžia nustatyti informaciją apie objektų buvimo vietą. Analizuojama įvesties jutiklių kiekio svarba ir kokius duomenis galima nustatyti tam tikromis aplinkybėmis. Aptariamos mikrofonų išdėstymo galimybės ir kokiomis savybėmis pasižymi kiekviena konfigūracija. Apžvelgiamos bevielių akustinių tinklų sistemos ir jų sinchronizavimo metodai. Analizuojami GŠL algebriniai ir mašininio mokymosi metodai. Įvertinus ir susisteminius pateiktų ir viešai prieinamų mokslinių straipsnių rezultatus bei išvalgas, apibendrinama probleminės srities analizė.

1.1. Garso savybių naudojimas lokalizavimui

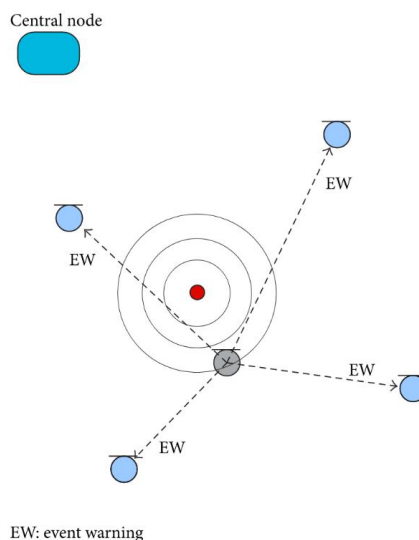
Siekiant nustatyti garso šaltinio vietą erdvėje, reikia išanalizuoti, kokios savybės leidžia tai padaryti. Garsas yra banga sklindanti įvairiomis medžiagomis, kuriomis keliaudama išsklaido savo energiją į aplinką. Energijos, sklidimo laiko, krypties pokytis aptinkamas stebėtojo, nutolusio nuo šaltinio, leidžia įvertinti bangos sklidimo kelią. Tuo tarpu garso dažnis leidžia atskirti skirtingus garsus. Įvairios metodikos, aprašytos straipsniuose [1, 2, 3], taiko garso savybes sprendžiant GŠL problematiką.

1.1.1. Garso energija

Kaip pateikiama straipsniuose [1, 4, 5], garso energijos duomenis galima panaudoti lokalizavimui. Garsas toldamas nuo šaltinio silpsta atvirkščiai proporcingai atstumo kvadratui, todėl matuojant garso stiprumą, galima įvertinti atstumą iki šaltinio. Naudojant du garso jutiklius galima aptikti energijos skirtumą, kuris kartu su mikrofonų buvimo vietomis gali nusakyti hipersferą, kurioje yra šaltinis. Kai naudojama daugiau mazgų, kiekviena jų pora leidžia apskaičiuoti po hipersferą, kurios susikirsdamos viename taške nurodo tikslią garsą skleidžiančio objekto buvimo vietą. Tokiu principu sukurta sistema gali veikti net ir mažai energijos vartojančiame paskirstytame jutiklių tinkle [4]. Ši metodika puikiai veikia, kai aplinkoje mažai triukšmo ir kai norime nustatyti vieno garso šaltinio buvimo vietą. Keli mikrofonai, esantys mažu atstumu ar viename mazge, nepatiria didelio energijos išsklaidymo skirtumo. Tinkamesnės paskirstytos garso jutiklių sistemos, kurias realizuoti paprasčiau dėl nedidelių įrangos tikslumo reikalavimų. Mazgų tarpusavio sinchronizavimas taip pat nėra itin svarbus aspektas, lyginant su kitais GŠL būdais. Galimi trumpesni garso įrašai bei įvesties vertimas į stiprumo reikšmes sumažina duomenų ir skaičiavimų kiekį. Dėl šių priežasčių garso energija yra viena iš palankiausių savybių vietos nustatymui.

1.1.2. Atvykimo laikas

Kitas būdas taikomas GŠL naudoja garso bangos atvykimo laiką iki akustinių jutiklių. Vienas iš rečiau taikomų energiją taupančių sprendimų pavaizduotas **1 pav.** aprašomas [1].

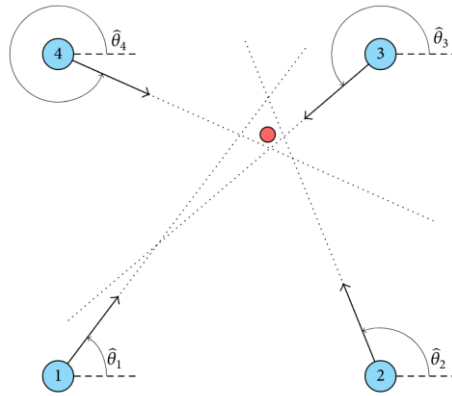


1 pav. Garso atvykimo laiku pagrįsta vietos nustatymo sistema [1]

Šioje sistemoje yra jutiklis esantis arti klausomo objekto, kuris užfiksuoja kada buvo išgirstas garsas ir įjungia kitus mikrofonus. Aptikus garsą kituose jutikliuose apskaičiuojamas atstumas, kurį nukeliavo banga ir randama jutiklio bei objekto vieta. Tokią sistemą galima pritaikyti pakankamai retais atvejais, nes dažniausiai nepatogu montuoti papildomą įrenginį šalia garso šaltinio arba tiesiog tas objektas nežinomas. Todėl daug populiaresnis atvykimo laikų skirtumo metodas, jis yra dažnai taikomas [6, 7, 8, 9, 10, 11], arba naudojamas palyginimui [4, 7, 12, 13]. Iš dviejų mikrofonų įrašų nustatoma kiek laiko atsilieka garsas. Pagal šį skirtumą apskaičiuojamas atstumas papildomai reikalingas pasiekti antrąjį jutiklį. Žinant mikrofonų koordinates galima nubrėžti lanką ar hipersferą atitinkamai dvimatėje ar trimatėje erdvėje, kuri nurodo galimas garso šaltinio vietas. Turint tris ir daugiau jutiklių gaunamas susikirtimo taškas, kuris nurodo tikslią objekto vietą. Svarbu atkreipti dėmesį, kad realiomis sąlygomis būna papildomo triukšmo, aido ir dažniausiai kreivės ar plokštumos nesusikerta viename taške. Taip pat itin svarbus garso jutiklių laiko sinchronizavimas. Dėl šių priežasčių garso atvykimo laikų skirtumo metodika reikalauja pakankamai daug erdvinių skaičiavimų ir papildomų aproksimacijos algoritmų.

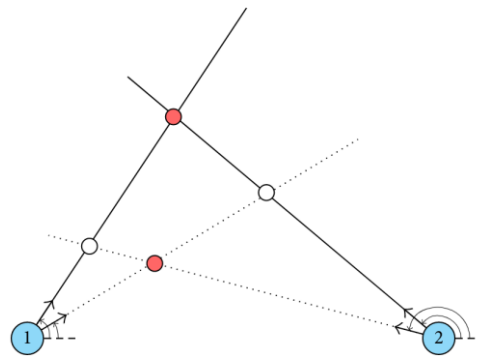
1.1.3. Atvykimo kryptis

Garso atvykimo krypties metodas yra aukštesnio lygio algoritmas nurodantis, kaip atlikti GŠL naudojant paskirstytą jutiklių tinklą. Kaip pateikta straipsnyje [1], sistemos mazgai individualiai nustato garso šaltinio kryptį, tada brėžiamos kryptinės kreivės, kurių susikirtimo taške yra garso šaltinis. Kaip pavaizduota **2 pav.**, spinduliai dėl netikslumų gali nesusikirsti viename taške, tačiau galima statistiškai išskaičiuoti objekto vietą.



2 pav. Garso atvykimo kryptimi pagrįstas lokalizavimo modelis vienam šaltiniui [1]

Kiekvienas mazgas gali nustatyti kryptį įvairiai, bet dažniausiai naudojami sujungti mikrofonų masyvai, kuriems taikomas garso energijos ar atvykimo laikų skirtumo metodikos. Tokia sistema sumažina komunikacijos kiekį tinkle ir tinka mažo ryšio pralaidumo sistemose. Kai objektai juda lėčiau nei apdorojama informacija, laiko sinchronizavimas tarp atskirų mazgų tampa mažiau svarbus. Algoritmas geba nustatyti ir vieno, ir kelių šaltinių buvimo vietas. Tačiau esant papildomiems šaltiniams, kryptinės rodyklės susikerta daugiau kartų, todėl nustatomi ir objektai vaiduokliai **3 pav.**



3 pav. Garso atvykimo kryptimi pagrįstas lokalizavimo modelis keliems šaltiniams [1]

Siekiant atskirti tikruosius objektus, reikalinga papildoma informacija apie kiekviena kryptimi aptiko garso savybes. Tuomet galima taikyti klasifikavimo algoritmus ir atrinkti realias objektų vietas. GŠL naudojant garso atvykimo kryptį sumažina komunikaciją tinkle ir nereikalauja tikslaus sinchronizavimo, tačiau kiekvienas mazgas atlieka daugiau skaičiavimų ir naudoja kitas metodikas krypties radimui.

1.1.4. Spindulio formavimas

Spindulio formavimas (angl. beamforming) yra signalo apdorojimo technika, kryptiniam signalo siuntimui ir priėmimui. Kombinuojant jutiklių masyvo duomenis, galima tam tikru kampu ateinančias bangas sustiprinti dėl konstruktyvios interferencijos, kol visas kitas slopinti – destruktuvia interferencija. Ši technika yra taikoma, tiriama ir GŠL srityje [14, 15, 16, 17]. Nukreipto atsako galios (angl. *Steered-Response Power*) algoritmas, pagrįstas spindulio formavimu, yra labai paklausus dėl savo gebėjimo tiksliai atlikti GŠL triukšmingose ir aidinčiose aplinkose. Labiausiai išpopuliarėjusi atmaina SRP-PHAT naudoja bangų fazių transformacijas [15, 18, 19, 20, 21, 22, 23]. Garso atvykimo

laikų skirtumo duomenys naudojami apskaičiuoti aplinkos žemėlapi, kuriame stipriausios galios taškai vaizduoja garso šaltinio buvimo vietą. Šis metodas sunkiai pritaikomas sistemose, veikiančiose realiu laiku, nes reikia atlikti daug skaičiavimų – išbandyti sustiprinti garso bangas visomis galimomis kryptimis ir sudaryti bendrą visų mazgų žemėlapi, iš kurio toliau atrenkami garso šaltiniai. Tyrimuose [23, 24, 25] siūlomos algoritmo modifikacijos, kurios pagreitina algoritmą vykdydamos dalinę paiešką. Taip pat šis algoritmas naudojamas kartu su neuroniniais tinklais [26], kad būtų pasiektas didesnis lokalizavimo tikslumas. Visuose tyrimuose naudojami glaudžiai sumontuoti mikrofonai, garso šaltinis yra atokiau nuo tokios konstrukcijos ir algoritmas grąžina kryptį nuo stebėtojo. Vadinasi šis GŠL metodas yra nepritaikytas paskirstytoms akustinių jutiklių sistemoms. Didelė tyrimų gausa rodo, kad spindulio formavimu pagrįstos metodikos yra plačiai paplitusios ir teikia pakankamai gerus rezultatus.

1.1.5. Dažnių spektrograma

Garso įrašuose užfiksuojamos įvairaus dažnio garso bangos gali būti atvaizduojamos spektrograma. Atskyrus skirtingo ilgio bangas, galima detaliau analizuoti įrašo sandarą ir lengviau atpažinti skirtingų garso šaltinių bangas. Tyrimuose [27, 28, 29, 30, 31] spektrogramos dažnai naudojamos garso savybių pateikimui į neuroninius modelius. Šis garso apdorojimo metodas sumažina duomenų kiekį ir aiškiai atvaizduoja bangos dažnių kitimą laiko intervalais. Nors taip ir prarandama dalis informacijos, bet tyrimai rodo, kad jos užtenka GŠL. Spektrogramos palengvina skirtingus garsus skleidžiančių objektų lokalizavimą, bei puikiai tinka garso savybių išskyrimui.

1.1.6. Kryžminės koreliacija

Siekiant įvertinti dviejų garso signalų prasislinkimą, taikoma kryžminė koreliacija. Ji nurodo kada garso įrašai labiausiai sutampa ir kiek vienas uždelstas kito atžvilgiu. Šis atsilikimas parodo kiek laiko reikėjo garsui papildomai nukeliauti iki antrojo jutiklio. Kryžminė koreliacija yra vienas iš pagrindinių metodų garso atvykimo laikų skirtumui nustatyti. Tyrimuose [10, 16, 26, 32], jis dažnai naudojamas kartu su kitomis GŠL metodikomis. Kryžminė koreliacija geriausiai veikia, kai pašalinis triukšmas yra silpnesnis už tiriamą garsą. Taip pat abu įrašai turėtų būti normalizuoti ir bangų vidutinis amplitudės poslinkis nuo nulio minimalus. Šis metodas nėra atsparus aidui ir kelių panašių garso šaltinių atskyrimui. Nepaisant trūkumų, kryžminė koreliacija yra viena iš paprastesnių ir plačiai taikomų GŠL parametru.

1.2. Garso įvesties kanalų kiekis

GŠL sprendimuose didelę įtaką turi garso jutiklių kiekis bei jų išdėstymas. Analizuojant kiekį – galime išskirti tris pagrindines grupes: kai turime vieną, du arba tris ir daugiau jutiklių. Kuriant lokalizavimo sprendimą, reikia pasirinkti mikrofonų kiekį ir jų išdėstymą. Sistemos struktūra dažnai lemia vietos nustatymo tikslumą, stebimos erdvės dimensijas ir dydį ar net apriboja galimas nustatyti garso bangas.

1.2.1. Monofoninis garsas

Sistema su vieno garso jutiklio įvestimi yra vadinama monofonine. Bendru atveju su vienu mikrofonu galima nustatyti tik garso stiprumą ir jo atvykimo laiką. Šią informaciją galima panaudoti keliais būdais: aptinkant triukšmą, perduodant garso pranešimus. Tai yra labiausiai paplitęs ir primityviausias garso įvesties būdas. Tačiau, buvo atlikti tyrimai ir eksperimentai [33] rodantys, kad

naudojant tik vieną mikrofoną galima nustatyti garso sklidimo kampą. Šis tyrimas buvo pagrįstas žiniomis apie žmogaus ausies biologiją ir bandymus, keičiant ausies kaušelio formą iš naujo apmokyti žmogų skirti garso kryptį. Kadangi garso bangos keliaudamos į ausies kanalą yra iškreipiamos išorinės ausies dalies, teigiama, kad ši garso modifikacija gali būti panaudota šaltinio krypties nustatymui. Atlikus bandymus su dirbtiniais modeliais, buvo pasiekti pakankamai geri rezultatai, leidžiantys apytiksliai lokalizuoti garso šaltinius. Šis tyrimas įrodo, kad kryptiniai monofoniniai garso įvesties jutikliai gali aptikti pakitimus garso signale pasikeitus klausymosi kampui.

1.2.2. Binauralinis garsas

Sistema naudojanti du garso įvesties jutiklius yra vadinama binauraline. Tai žmogui labiausiai pažįstamas garso suvokimo būdas. Gebėjimą lokalizuoti garsą šiuo metodu ir daugumą klausos savybių aptaria straipsnis [34]. Turint du įvesties jutiklius galima nustatyti horizontalų kampą iki šaltinio, skleidžiančio garsą. Tuo tarpu vertikalų kampą teigiama, kad daugiau padeda nustatyti monofoninis girdėjimas. Tam naudojamas ne tik išorinės ausies deformuojamos garso bangos, bet ir galvos, kūno judesiai - pasukimas. Pakeitus klausymosi poziciją pasikeičia atstumai ir kampai, todėl žmogus gali dar tiksliau įvertinti šaltinio buvimo vietą. Binauralinė klausa taip pat leidžia nustatyti ir atstumą iki garso šaltinio. Tai vyksta tokiu pačiu principu, kaip ir garso atvykimo laikų skirtumo nustatymas. Nors žmogaus organizmas yra prisitaikęs GŠL problemų sprendimui ir binauraliniai įrenginiai buvo taikomi seniau [3] ir net šiomis dienomis [12], bet mechaninės sistemos dažniausiai naudoja paprastesnius komponentus ir siekia tikslesnių rezultatų. Dviejų mikrofonų statinėse sistemose neužtenka, nes egzistuoja kelios veidrodiniu atžvilgiu vienodos vietos gebančios formuoti panašias įvestis garso jutikliuose. Atkartoti žmogaus klausos dar negalime, bet binauralinės įrašų savybės yra plačiai naudojamos GŠL srityje.

1.2.3. Trijų ir daugiau kanalų garsas

Kai sistema turi tris ir daugiau garso jutiklių, galima taikyti ne tik monofoninio ar binauralinio įrašų savybes, bet ir daug sudėtingesnius struktūrinius sprendimus, algoritmus. Didesnis įrašų kiekis užtikrina garso aptikimo patikimumą, klausomo ploto padidinimą bei leidžia lengviau išskirti garsus. Paskirstytų akustinių jutiklių sistemos [1, 4, 35], ir pavieniai įrašinėjimo įrenginiai [13, 22, 36, 37] skirti sudėtingesnėms GŠL problemoms dažniausiai naudoja daugiau nei du garso jutiklius. Didesniam duomenų kiekiui apdoroti ir perduoti reikalingi papildomi skaičiavimo resursai. Tačiau tobulėjančios technologijos leidžia projektuoti vis didesnes ir spartesnes garso įrašinėjimo sistemas.

1.3. Garso įvesties jutiklių išdėstymas

Populiarėjančios akustinių jutiklių masyvų sistemos lemia didžiulę struktūrų įvairovę. Galima išskirti du pagrindinius jutiklių išdėstymo būdus: statinį ir dinaminį. Kai sistemos mazgai nekeičia savo pozicijos vienas kito atžvilgiu, turime statinę struktūrą, kuri gali judėti ir sąveikauti su aplinka. Atitinkamai dinaminėje sistemoje garso jutikliai neturi iš anksto nustatyto išdėstymo arba gali judėti atskirai nuo kitų ir dažniausiai mazgai paskirstomi erdvėje atsitiktinai. Daugumoje tyrimų nagrinėjamos statinės sistemos išdėstytos pagal plokščiasias ir erdvinės geometrinės figūras.

1.3.1. Plokščiosios struktūros

Paprasčiausia garso įrašinėjimo sistemos struktūra yra plokščia, kai mazgai išdėstomi vienoje plokštumoje. Dažniausiai pasirenkami taisyklingųjų geometrinių figūrų, pavyzdžiui, kvadrato,

taisyklingojo trikampio, apskritimo taškai. Tai paprastina skaičiavimus, o suprojektavus vienodus atstumus tarp mikrofonų galima taikyti supaprastintus algoritmus ar net iš anksto apskaičiuoti tam tikras reikšmes. Su plokščiosiomis struktūromis tyrimuose [9, 12, 15, 16, 22] dažniausiai sprendžiamos 2D arba vieno, dviejų garso atvykimo kampų nustatymo 3D erdvėje [14, 17, 21, 26, 36, 38] lokalizavimo problemos. Tokias sistemas galima pritaikyti ir koordinacių nustatymui trimatėje aplinkoje, tačiau tai rečiau atliekami tyrimai [13, 39]. Atstumo nustatymas plokščiomis struktūromis trečiaja dimensija yra apsunkintas ir reikalauja didesnio tikslumo mažesniems garso pokyčiams nei lokalizavimui kitomis kryptimis. Siekiant pagerinti GŠL rezultatus naudojamos grupės plokščiųjų struktūrų. Tyrimuose [4, 8, 35] siūlomos paskirstytos sistemos, kurių mazgus sudaro mikrofonai išdėstomi vienoje plokštumoje. Platus šių struktūrų tyrimų spektras, lėmė ir pirmųjų GŠL produktų konstrukcijas. „Siemens“ sukurta kamera, nustatanti garso sklaidimo kryptį ir atvaizduojanti rezultatus nuotraukoje, naudoja apskritimu sumontuotus akustinius jutiklius [40]. Plokščiosios struktūros yra populiarios dėl savo paprastumo ir pritaikomumo įvairiose GŠL srityse.

1.3.2. Erdvinės struktūros

Erdvinės struktūros yra taikomos GŠL srities sprendimams trimatėje erdvėje. Sistemos mazgų išdėstymas dažnai parenkamas pagal taisyklingąsias erdvines geometrines figūras, tokias kaip kubas ir sfera. Skaičiavimai tokiose sistemose būna paprastesni, o tolygus garso įrašėjimas visomis trimis kryptimis užtikrina vienodą bangų pokyčių aptikimo jautrumą. Tyrime [19] naudojami mikrofonai išdėstyti kubo viršūnėse. Nesudėtinga struktūra leidžia aptikti garsus visomis kryptimis naudojant tik 8 jutiklius. Dėl to sistema gali veikti realiu laiku ir gali būti sumontuota ant roboto korpuso. Kitame tyrime [37] buvo naudoti 32 mikrofonai išdėstyti sferos paviršiuje. Tokia struktūra leidžia grupuoti ir aptikti garsus visomis kryptimis. Tačiau pasirinktomis tyrimo sąlygomis buvo apskaičiuota viršutinė ir apatinė atskiriamų garso dažnių ribos. Šie apribojimai neleidžia stebėti ir nustatinėti visų žmogui girdimų garsų. Svarbu pastebėti, kad įrašams su dideliu mikrofonų masyvu daryti reikalinga speciali techninė įranga. Taip pat didesni atstumai tarp akustinių jutiklių lemia ryškesnius garso bangos įrašų pokyčius. Erdvinės struktūros, skirtos GŠL, taikomos rečiau nei plokščiosios, tačiau labiau pritaikytos 3D problemoms spręsti.

1.3.3. Dinamiškos struktūros

Mažiau tirtos dėl savo kompleksiško ir sudėtingos realizacijos yra dinamiškos garso jutiklių sistemos. Tokios sistemos suskirstytos į mazgus, kuriuose yra vienas arba keli mikrofonai [4, 8, 35, 41]. O patys mazgai dažniausiai bevieliu tinklu jungiami į bendrą sistemą. Visuose lokalizavimo algoritmuose yra reikalinga garso įvesčių buvimo vieta, tai didžiausias sistemos realizacijos trūkumas. Mikrofonų koordinatės galima išmatuoti, tačiau automatinis sistemos kalibravimas būtų naudingesnis. Straipsnyje [1] aprašomi galimi jutiklių vietų nustatymo metodai. Vienas jų mazguose siūlo įmontuoti papildomus ne garsinius jutiklius, leidžiančius lokalizuoti mikrofonus. Pritaikius papildomą kalibravimo etapą ir naudojant žinomas garso šaltinio koordinatės, būtų įmanoma apskaičiuoti sistemos mazgų vietas.

Mikrofonų vietų nustatymas yra mažai ištirta sritis, tačiau plačiau išanalizuota yra mazgų laiko sinchronizavimo problema. Kadangi sistema dažniausiai jungiama bevieliu būdu, skirtingi įrenginiai turi žinoti tikslų ir vienodą laiką. Yra du pagrindiniai sinchronizavimo būdai, kai sistema turi centrinį valdiklį ir kai neturi. Plačiausiai taikomas kompiuterinėse technologijose yra UDP paketus naudojantis tinklo laiko protokolas NTP. Atliktame tyrime [42] nustatyta, kad idealiomis sąlygomis

šis algoritmas pasiekia didesnę nei 1 ms tikslumą. Atsižvelgiant į garso greitį ore 343 m/s, sinchronizuota sistema gali pasiekti mažesnę nei 34 cm tikslumą. Tokio tikslumo sinchronizavimas yra pakankamai geras, bet naudojant vietinius sinchronizavimo metodus galima rezultatus pagerinti. Yra įvairių algoritmų [43, 44], kurie naudoja vietinius maršrutizatorius ar kitus centrinius valdiklius prietaisų sinchronizavimui. Daug patvaresnės sistemos, kuriose mazgai patys sinchronizuojasi tarpusavyje [45, 46]. Visi šie įvairūs sprendimai nustato laiką siųsdami tinklo paketus. Tyrime [47] pateikiamas ir kitoks sinchronizavimo metodas, kuris naudoja lempuotę ir vaizdo kamerą. Norint pasiekti 1 ms tikslumą reikia brangios bent 1000 kadrų per sekundę padarančios kameros. Todėl tinkamiausią metodiką reikia pasirinkti pagal sprendžiamą GŠL problemą.

Nepaisant sudėtingos dinamiškos struktūros realizacijos, tokio tipo GŠL sistemos gali būti sumontuojamos įvairiose erdvėse. Akustiniai jutikliai išdėstomi atsitiktinai, bet stengiantis paskirstyti klausomoje zonoje. Garso šaltinių vietos nustatymas šia sistema lengvesnis, nes mikrofonai objektus supa iš visų pusių ir atstumai tarp jutiklių yra pakankamai dideli. Dinamiškos struktūros turi dar mažai ištirtą potencialą GŠL problemoms spręsti.

1.4. Lokalizavimui naudojami modeliai

Garso šaltinio vietos nustatymo modeliai vis tobulinami ir kuriami nauji. Pirmieji lokalizavimo sprendimai buvo aritmetiniai ir geometriniai algoritmai. Vėliau atsirado mašininio mokymosi metodikos ir galiausiai išpopuliarėjo neuroninių tinklų modeliai. Dažnai kuriami sudėtiniai sprendimai, apjungiantys įvairias technikas. Didžioji dalis jų aptariami apžvalginiuose straipsniuose [1, 48].

1.4.1. Geometrinis lokalizavimas

Vienas iš pirmųjų GŠL metodų, naudojančių garso atvykimo laikų skirtumą, yra geometrinis [6]. Šis algoritmas apskaičiuoja galimų objekto vietų buvimo hipersferas kiekvienai mikrofonų porai. Siekiant atrasti šių erdvinų figūrų susikirtimo tašką galima vykdyti pilną tinklelio paieškos (angl. *grid-search*) algoritmą. Kaip daugumoje tyrimų [18, 23, 49, 50, 51, 52] teigiama, pilna visos erdvės paieška naudoja didelius skaičiavimo resursus, todėl yra sunkiai pritaikoma realaus laiko sistemose. Tačiau pateikiami įvairūs optimizavimo sprendimai, kurie sumažina paieškos erdvę arba naudoja nepilnos paieškos algoritmus. Dažniausiai tokiuose tyrimuose skaičiuojama garso atvykimo kryptis, nors galima nustatinėti ir 3D koordinates. Tinklelio paieškos metodas ir jo modifikacijos dažnai naudojamos kartu su 1.1.4 poskyryje aptartu SRP-PHAT algoritmu. GŠL geometrinio būdu yra lengvai suvokiamas, bet daug skaičiavimų reikalaujantis metodas.

1.4.2. K-vidurkių metodas

Mašininis mokymasis yra naujesnė sritis palyginus su algebra, bet jau kuris laikas įvairūs metodai, kaip k-vidurkių klasifikavimas, yra taikomi sprendžiant GŠL problemas. Apžvalgoje [1] šis algoritmas siūlomas naudoti kelių garso šaltinių nustatymui. Straipsnyje [53] realizuota ir ištestuota k-vidurkių klasifikavimo modifikacija. Visais atvejais norima suskirstyti duomenis į grupes pagal jų panašumą. Algoritmas pasirenka k centrų ir cikliška keičia jų koordinates taip, kad atstumai iki artimiausių duomenų taškų būtų mažiausi. Šio metodo naudojimas kartu su kitais GŠL sprendimais padeda atskirti skirtingus garso šaltinius.

1.4.3. Pilnai sujungti neuroniniai tinklai

Didėjantys kompiuteriniai resursai leidžia kurti įvairesnius ir didesnius neuroninių tinklų modelius [48]. Ši mašininio mokymosi metodika gerai sprendžia netiesiškos priklausomybės problemas. Patys modeliai yra sudaryti iš neuronų išdėstytų sluoksniais, per kuriuos duomenys eina nuosekliai viena kryptimi. Neuroninius modelius nuo aritmetinių algoritmų skiria taikomos netiesinės aktyvavimo funkcijos. Dažniausiai naudojama ReLU funkcija, pavyzdžiui, tyrimuose [38, 54, 55]. GŠL srityje šie modeliai pritaikyti tiesioginiam garso šaltinio krypties nustatymui arba garso informacijos išskyrimui, kurią naudoji kiti algoritmai. Neuroniniams tinklams pateikiami dažniausiai jau apdoroti garso duomenys, pavyzdžiui, SRP-PHAT algoritmu. Kelis sluoksnius turintys modeliai, negali atlikti sudėtingų užduočių, todėl taikomi gilesni, daugiau sluoksnių turintys sprendimai. Tačiau tokių modelių mokymas yra sudėtingesnis. Jie per greitai prisitaiko prie mokymui skirtų duomenų, o tikrinant su testavimui skirtu rinkiniu gražina blogesnius rezultatus. Šiai problemai spręsti naudojami neuronų išmetimo (angl. *dropout*) ir duomenų partijos normalizavimo (angl. *batch normalization*) sluoksniai. Pilnai sujungti neuronų sluoksniai dažniausiai naudojami kartu su kitais naujesniais neuroninių tinklų sluoksniais.

1.4.4. Konvoliuciniai neuroniniai tinklai

Konvoliuciniai neuroninių tinklų sluoksniai leidžia išskirti duomenų savybes ir dėsningumus. Matematinė konvoliucijos operacija gražina dviejų funkcijų persidengimo kiekį. Todėl parinkus norimos formos filtrą galima aptikti vietas, kuriose pasikartojo ieškomos savybės. Toks filtravimo metodas pritaikytas ir neuroniniams modeliams, kurie plačiai tiriami ir naudojami [7, 26, 31, 38, 56]. GŠL srityje bangų savybių išskyrimas leidžia lengviau įvertinti įrašų prasislinkimą ar specifinių garsų klasifikavimą. Dėl šių savybių konvoliuciniai sluoksniai projektuojami modelių pradžioje. Taip ne tik sumažinami įvesties duomenų kiekiai ir tolimesnių sluoksnių dydžiai, bet ir pašalinama tikslui pasiekti mažiau reikalinga informacija. Konvoliucijos taikymas yra vienas iš populiariausių ir gerus rezultatus duodančių patobulinimų mašininio mokymosi srityje.

1.4.5. Rekurentiniai neuroniniai tinklai

Dar viena neuroninių modelių modifikacija yra orientuota į laikinių sekų apdorojimą. Rekurentiniai (angl. *recurrent*) sluoksniai ir jų pritaikymas pateiktas apžvalgoje [48]. Tyrimuose [57, 58] šie sluoksniai naudojami dėl būsenos išlaikymo savybės, kuri padeda apdoroti nuo laiko priklausomus duomenis. Paprasto rekurentinio modelio mokymo metu gradientai pakankamai greitai išauga arba sumažėja iki nulio. Apžvalgoje pateikti patobulinimai LSTM ir GRU šią problemą sumažina. Abu atmintį išlaikantys sprendimai yra taikomi GŠL srityje ir teikia pakankamai gerus rezultatus. LSTM realizacija tiriama [59, 60] straipsniuose, o GRU naudojamas [28, 29, 30]. Didžiausias šių sluoksnių skirtumas yra parametrų skaičius ir mokymo sudėtingumas. Todėl modelius naudojančius GRU galima apmokyti sparčiau. Atitinkamai LSTM yra tinkamas sprendimas turint didelį duomenį rinkinį. Rekurentinius neuroninius tinklus populiariau naudoti laikinių sekų apdorojimui.

1.4.6. Liekamosios vertės neuroniniai tinklai

Liekamosios vertės (angl. *residual*) neuroniniai tinklai naudojami siekiant pagerinti gilesnių modelių tikslumą. Tai padaroma papildomai sujungiant kelis ne iš eilės einančius sluoksnius. Tokie modeliai taikomi ir tiriami GŠL srityje [27, 61, 62]. Liekamosios vertės modeliai gauna geresnius rezultatus nei modeliai nenaudojantys sluoksnių praleidimo technikos. Informacijos išsaugojimas tolimesniems

sluoksniams sumažina nykstančio gradiento problemą ir leidžia įvertinti informacijos pokytį. Dėl šių savybių liekamosios vertės sluoksnių modifikacijos taikomos modelių tobulinimui.

1.4.7. Kodavimo-dekodavimo neuroniniai tinklai

Kodavimo-dekodavimo (angl. *encoder-decoder*) neuroniniai tinklai naudojami informacijos suspaudimui ir išplėtimui. Šio tipo modeliai dažnai aptariami mašininio mokymosi srityje ir gali būti pritaikyti net garso lokalizavimo užduotims [48]. Kodavimas leidžia sumažinti duomenų kiekį neprarandant informacijos, kurią vėl galima atstatyti dekoduojant. Šiuo principu projektuojami neuroniniai tinklai. U-Net yra konvoliucinis neuroninis tinklas, kuris naudoja keturias poras tarpusavyje sujungtų kodavimo ir dekodavimo bloką. Šis modelis palaipsniui performuoja įvestį į ypatybių žemėlapius ir iš jų vėl suformatuoja reikalingą informaciją. U-Net neuroninio tinklo pritaikymas tirtas ir garso apdorojimo srityje [63]. Šis modelis sėkmingai panaudotas kelių šaltinių krypties nustatymui. Papildomas nenuoseklus sluoksnių jungimas sumažina nykstančio gradiento problemą. Kodavimo-dekodavimo neuroniniai tinklai ir jų modifikacijos turi tinkamas savybes GŠL.

1.4.8. Dėmesiu pagrįsti neuroniniai tinklai

Dėmesiu pagrįsti neuroniniai tinklai (angl. *attention-based neural network*) naudojami svarbesnės informacijos išskyrimui. Apžvalgoje [48] aptariama, kad tokie tinklai leidžia atrinkti ieškomas duomenų savybes bei kad jie pritaikomi laikinoms sekoms. „DCASE“ tinklalapyje kelis praėjusius metus yra pateikiami su garso lokalizavimu susiję iššūkiai. Dažnai pasitaiko sprendimų naudojančių dėmesiu pagrįstus neuroninius tinklus [64]. Pagal šiuos modelius buvo sukurta transformerių neuroninių tinklų struktūra, kuri yra sudaryti iš kodavimo ir dekodavimo dalių. Jos išskiria svarbiausias duomenų savybes ir tuomet iš suspaustos informacijos suformuoja rezultatus. Pasiūlytas *PILOT* modelis [65] turi pritaikytą transformerio struktūrą ir naudoja kodavimo dalį garso įvykių lokalizavimui. Šis sprendimas turi pilnai sujungtų neuronų sluoksnius skirtus įrašų pirminiam apdorojimui. Tai yra vienas pirmųjų ir naujausias transformerio modelio pritaikymas garso įvykių lokalizavimui. Dėmesiu pagrįstos struktūros sėkmingai naudojamos ir garso savybių išskyrimui.

1.5. Literatūros analizės išvados

Garso bangos nueito kelio skirtumą iki dviejų erdvės taškų galima apskaičiuodami pagal laiko ir energijos pokyčius. Šie pasikeitimai nustatomi taikant kryžminę koreliaciją, kuri aptinka dviejų įrašų sutaptį. Spindulio formavimo algoritmas, kaip ir atvykimo laikų ar energijos skirtumai, leidžia nustatyti garso atvykimo kryptį ir ieškomo objekto vietą. Kai turime kelis garso šaltinius, jų savybes išskirti ir suklasifikuoti padeda dažnių spektrograma.

Akustinių jutiklių kiekis nurodo galimus taikyti GŠL algoritmus ir reikalavimus sistemai. Iš monofoninio įrašo galima nustatyti garso bangos stiprį, pokyčius einant laikui, aidą. Tinkamomis sąlygomis įmanoma apytiksliai lokalizuoti garsus. Jei turime du akustinius jutiklius, galime lyginti informaciją ir nustatinėti garso atvykimo kryptį, kartais ir atstumą. Didesnis mikrofonų kiekis leidžia atlikti daugiau palyginimų, tiksliau nustatinėti objektų vietas ir taikyti įvairius algoritmus.

Nuo akustinių jutiklių išdėstymo priklauso lokalizavimo algoritmai, garso pokyčio dydis ir sistemos tikslumas reikalingas GŠL. Mikrofonai į vieną mazgą dažnai jungiami pagal taisyklingąsias plokščiąsias arba erdvines geometrines figūras, dėl skaičiavimų ir išdėstymo paprastumo. Mažesni

atstumai tarp jutiklių reikalauja didesnio jautrumo nedideliems garso pokyčiams. Mazgus paskirsčius aplinkoje, gauname sudėtingesnę sistemą, bet reikalingas mažesnis tikslumas.

Pirmiausi pasiūlyti GŠL metodai yra pagrįsti geometrija. Prireikus atskirti kelis garsus naudojami klasifikavimo algoritmai, pavyzdžiui, k-vidurkių. Gerus lokalizavimo rezultatus parodė ir neuroniniai tinklai. Konvoliucija padeda išskirti garso savybes, o rekurentiniai sluoksniai saugo atmintį apdorojant laikines sekas. Siekiant išvengti persimokymo (angl. *overfitting*), naudojami išmetimo, duomenų normalizavimo ir liekamosios vertės sluoksniai. Šias technikas apjungia sudėtingesni U-Net ir transformerių neuroniniai tinklai, kurie yra pritaikyti ir tirti GŠL srityje. Dažniausiai sprendimai ieško vieno ar kelių garsų atvykimo krypčių, bet kartais ieškoma ir atstumo iki garso šaltinio arba tiesiog jo koordinatų.

2. Garso šaltinio lokalizavimo sistemos projektas

Šiame skyriuje pateikiami GŠL sistemos paskirtis ir funkciniai, nefunkciniai reikalavimai. Aprašomos realizavimui pasirinktos technologijos, įrankiai ir nurodomi struktūros bei veikimo principai. Specifikuojami pasirinktos GŠL srities parametrai ir rezultatų įvertinimo kriterijus. Pateikiami tyrimo metu naudoti geometriniai algoritmai ir plačiai taikomų neuroninių modelių struktūrų modifikacijos, pasirinktos srities sprendimui.

2.1. Reikalavimų specifikacija

Šiame poskyryje nurodomi sistemai keliami reikalavimai. Pateikiama bevielų akustinių jutiklių tinklo paskirtis ir naujų GŠL metodų kūrimo priežastis. Įvertinami galimi sukurto sprendimo vartotojai ir panaudojimo sritys. Taip pat pateikiami tyrimui naudojamos sistemos funkciniai ir nefunkciniai reikalavimai.

2.1.1. Projektuojamos priemonės paskirtis

Bevielė kelių kanalų garso įrašymo sistema yra skirta GŠL. Ši priemonė surenka duomenis, reikalingus algebriniams ir mašininio mokymosi metodams apmokyti ir vykdyti. Sukūrus ir pritaikius tinkamus algoritmus, sistema realiu laiku įrašinėdama garsą pateikia apskaičiuotą garso šaltinio buvimo vietą trimatėje erdvėje. Siūlomos sistemos maži bevieliai mazgai skirti patogiam diegimui įvairiose aplinkose. Ši priemonė pritaikyta montuoti stacionariai ir mobiliai. Toks universalumas numatytas, dėl galimybės integruoti siūlomą sprendimą į kitas sistemas. Nors pritaikyti bevielį akustinių jutiklių tinklą galima ir daug apstatytoje ar suskaidytoje erdvėje, bet šio tyrimo metu atliekami eksperimentai pakankamai atviroje, vientisoje erdvėje. Sistema skirta garsą skleidžiančio objekto vietos trimatėje erdvėje nustatymui ir šios funkcijos pritaikymui GŠL problemoms spręsti.

2.1.2. Projekto kūrimo pagrindimas

Bevielų akustinių jutiklių sistemų pritaikymas GŠL srityje yra mažai ištirtas. Tokių sistemų realizacijos ir tinkamų algoritmų apžvalga pateikta [1]. Taip pat kuriami, tiriami ir siūlomi techniniai sprendimai [4, 8, 41], leidžiantys realizuoti bevielų akustinių jutiklių tinklą. Šiuose tyrimuose pritaikomi algebriniai GŠL algoritmai ir įrodomas tokios sistemos tinkamumas. Mašininio mokymosi metodai dar nėra pritaikyti ir ištirti besikeičiančioms jutiklių vietoms, todėl ši sritis yra nauja. Tai įrodo ir vienas iš didžiausių proveržių, GŠL skirta „Siemens“ sukurta kamera [40]. Šis produktas leidžia nustatyti iš kur sklinda garsai ir tai atvaizduoti nuotraukoje. Rezultatai būna pakankamai tikslūs kamerą laikant tik kelių dešimčių centimetrų atstumu. Taip pat nustatymas vyksta tik viena kryptimi. Toks produktas turi savo panaudojimo paskirtį, tačiau jis netinkamas didelės erdvės pastoviam stebėjimui. Šiuos trūkumus išsprendžia bevielės akustinių jutiklių sistemos. Jų privalumas – didesni atstumai tarp mikrofonų – leidžia didesnes paklaidas nustatinėjant garso šaltinio vietą. Taip pat vartotojui nereikia fiziškai sąveikauti su sistema. Tobulėjančios technologijos palengvina siūlomo sprendimo kūrimą ir mazgų sinchronizavimą. Todėl plečiamumas yra pakankamai pigus, o pritaikomumas įvairioms GŠL problemoms – platus.

2.1.3. Galimi sprendimo vartotojai

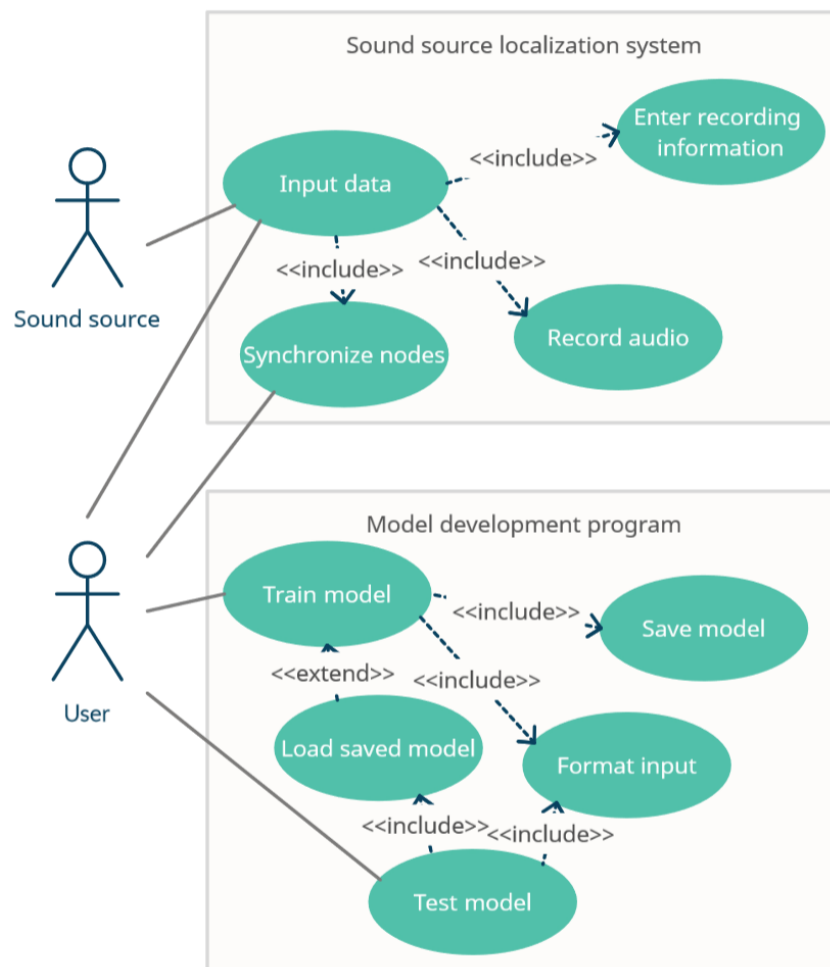
Projektuojamos sistemos galimi vartotojai:

- įvairių sistemų ir produktų kūrėjai – integruos projektuojamą priemonę, metodologiją, modelius į savo produktus GŠL problemoms spręsti;

- duomenų analitikai ir mašininio mokymosi sprendimų kūrėjai – panaudos pigią sistemą kelių kanalų garso įvesties duomenų kaupimui ir įvairių su GŠL problema susijusių modelių kūrimui, tobulinimui;
- GŠL sistemų kūrėjai – praplės ir papildys sukurtą sistemą sinchronizavimo, kelių šaltinių nustatymo funkcijomis.

2.1.4. Funkciniai reikalavimai

Šio tyrimo metu buvo sukurta bevielė akustinių jutiklių sistema ir algebrinių bei neuroninių modelių kūrimo, apmokymo ir testavimo programa. Su sistema sąveikauja vartotojas ir garso šaltinis, kaip pateikta panaudos UML diagramoje **4 pav.**



4 pav. Sistemos panaudos atvejų diagrama

Paveikslėlyje matome pagrindines sistemos ir programos funkcijas, kurios sudarytos atsižvelgiant į funkcinis reikalavimus:

- sistema turi automatiškai sinchronizuoti mazgus prieš įrašinėjimo sesiją. Vartotojas taip pat gali inicijuoti sinchronizavimą;
- sistema turi įrašyti garsą, kurį skleidžia garso šaltinis. Garso įrašymas turi būti atliktas su keliais garso įvesties įrenginiais, kad gautume kelių kanalų įrašą;
- garso įrašams reikalinga papildoma tekstinė informacija apie garso šaltinį, mikrofonų buvimo vietą. Vartotojas šiuos duomenis turi įvesti atlikus vieną įrašymo sesiją;

- išsaugoti duomenys turi būti atrenkami, suformatuojami į tinkamo dydžio laikines kelių kanalų garso įrašų sekas kartu su mikrofono bei garso šaltinio koordinatėmis;
- vartotojas gali apmokyti naują modelį arba jau anksčiau išsaugotą;
- programa turi išsaugoti apmokytus neuroninius tinklus ir gebėti juos panaudoti;
- vartotojas gali patikrinti modelio veikimo tikslumą arba gauti apskaičiuotas garso šaltinio koordinates.

2.1.5. Nefunkciniai reikalavimai

Projektuojamai GŠL sistemai nurodomi šie nefunkciniai reikalavimai:

- sistema turi nustatyti garso šaltinio buvimo koordinates trimatėje erdvėje;
- sistema naudodama neuroninius tinklus turi nustatyti objekto buvimo vietą tiksliau nei su geometriniu sprendimu;
- sprendimas turi teikti rezultatus greičiau nei per 2 sekundes;
- sistema turi apdoroti garso įrašus mažesnėmis nei 500 milisekundžių grupėmis;
- sistemos mazgai turi būti sinchronizuoti bent 1 milisekundės tikslumu;
- mazgai su akustiniais jutikliais išdėstomi atviroje apstatytoje patalpoje;
- garso įrašai turi būti nenutrūkstantys ir bent 16KHz.

2.2. Pasirinkti įrankiai ir technologijos

Šiame projekte naudojamos įvairios mašininio mokymosi srityje populiarios technologijos, programavimo kalbos, platformos, vykdymo aplinkos. Pasirinkti įrankiai ir technologijos leidžiančios greičiau ir paprasčiau sukurti bei įvertinti GŠL sistemą, kurią būtų galima lengvai plėsti, pritaikyti kitų problemų sprendimui:

- *Python* – programavimo kalba plačiai naudojama mašininio mokymosi projektuose. Ji yra palaikoma įvairių programų ir turi didelę bibliotekų bazę. Šia kalba bus aprašyti GŠL modeliai ir kodas jų paleidimui, vykdymui, rezultatų atvaizdavimui ir palyginimui;
- *Anaconda* – tai platforma ir paketų tvarkyklė skirta vykdymo aplinkai paruošti, bibliotekoms parsisiųsti ir atnaujinti, ji naudojama darbui su duomenų mokslu ir mašininio mokymusi;
- *Jupyter Notebook* – naršyklės programa skirta interaktyviems dokumentams rašyti ir programiniam kodui vykdyti;
- *Tensorflow* – platforma palengvinanti mašininio mokymosi modelių kūrimą ir testavimą.
- *C++* – programavimo kalba mikroprocesorių programavimui. Šia kalba aprašomi ESP32 mikroprocesoriai – mazgai su mikrofonais;
- *PlatformIO* – platforma įterptiniam programavimui ir projektų valdymui, vykdymui, bibliotekų parsisiuntimui ir integravimui;
- *Visual Studio Code* – universali programavimo grafinė sąsaja tinkanti tiek įterptiniam, tiek ir grafinės sąsajos, bei serverio programavimui;
- *PHP* – bendros paskirties programavimo kalba. Kuria aprašytas sistemos serveris, duomenų įvedimo grafinė sąsaja;
- *Apache2* – HTTP serveris leidžiantis talpinti įvairaus tipo internetinius puslapius, kuriuos galima pasiekti iš kitų įrenginių;
- *Laravel* – internetinių programų karkasas, kurį naudojant sukurta sistemos serverio dalis, komunikuojanti su sistemos mazgais ir duomenų baze.

Šiam projektui taip pat naudojama papildoma kompiuterinė įranga:

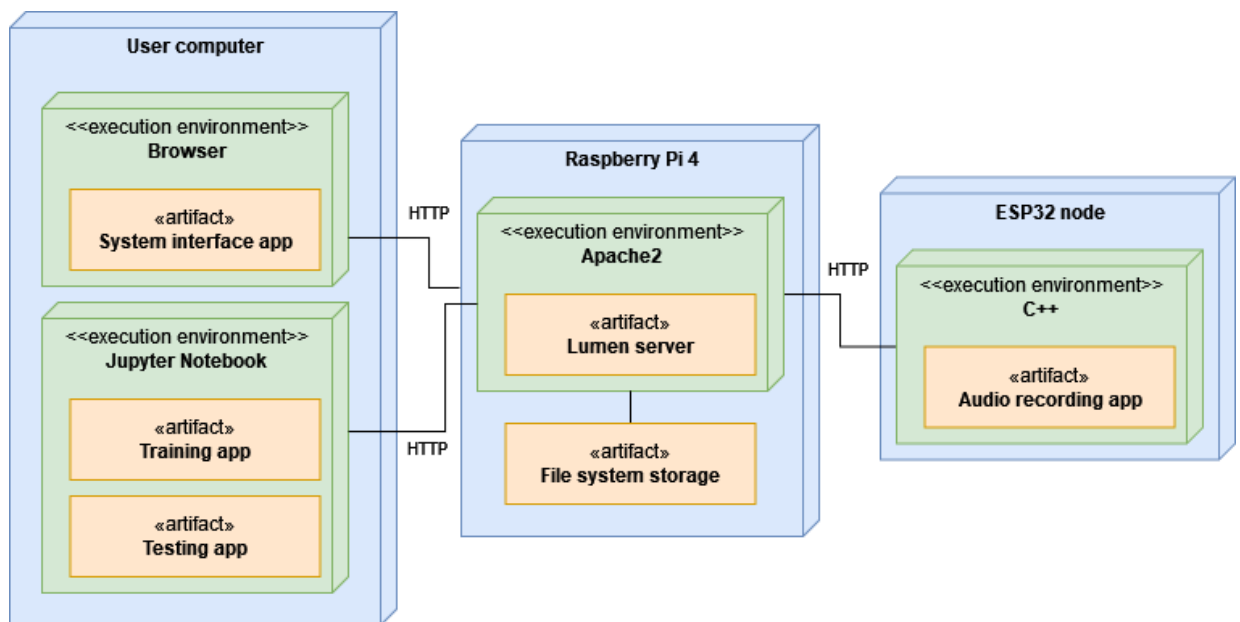
- *Raspberry Pi 4* – pigus ir pakankamai galingas kompiuteris, kuris naudojamas, kaip pagrindinis sistemos serveris.
- *ESP32* – pigus mažai energijos vartojantis mikroprocesorius, kuris naudojamas mazgams, nes turi WiFi funkciją.
- *MAX9814* – akustinis sensorius, mikrofonas naudojamas mazguose garso įrašymui.

2.3. Sistemos architektūra

Siūlomos bevielės akustinių jutiklių sistemos architektūra pateikiama fizinės įrangos ir programinių sąryšių aspektais. Nurodomos pasirinkimų priežastys ir pagrindiniai veikimo principai, sistemos mazgų sinchronizavimo algoritmas.

2.3.1. Sistemos įrangos struktūra

Bevielė akustinių jutiklių sistema susideda iš fiziškai tarpusavyje nesujungtų įrenginių. Yra trys pagrindinės jų rūšys: vartotojo kompiuteris, centrinis valdiklis ir garso įrašinėjimo mazgai. Šie įrenginiai ir jų sąsajos pavaizduotos diagramoje **5 pav.**



5 pav. Sistemos diegimo diagrama

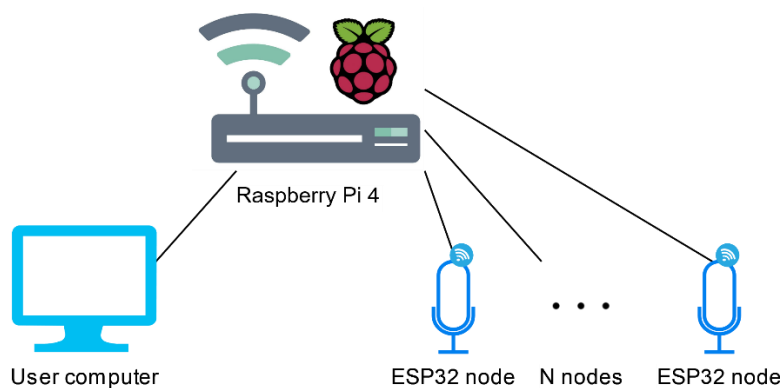
Vartotojo kompiuteris yra skirtas GŠL modelių kūrimui ir bevielės akustinių jutiklių tinklo valdymui. Kompiuterio naršyklė leidžia naudoti sistemos valdymo sąsają, kurioje yra pagrindinės sinchronizavimo ir įrašinėjimo funkcijos. Ši sistemos dalis nereikalauja didelių kompiuterinių resursų. Tuo tarpu *Jupyter Notebook* užrašinėje vykdomos neuroninių tinklų mokymo ir visų modelių tikrinimo programos atlieka didelius skaičiavimus. Čia apdorojami dideli kiekiai duomenų, todėl reikalingas galingas procesorius arba vaizdo plokštė. Užrašinė skirta modelio kūrimui ir naudojama viso šio tiriamojo projekto metu. Sukurtą modelį vėliau galima perkelti į centrinį valdiklį, kuriame ir bus atliekamas GŠL.

Abi vartotojo kompiuteryje veikiančios programos komunikuoja su *Raspberry Pi 4* valdiklyje esančiu pagrindiniu sistemos serveriu. Šis tinklu bendrauja su akustinių jutiklių mazgais, kurie teikia garso įrašus. Centrinis valdiklis saugo šiuos įrašus įrenginio failų sistemoje, kad vėliau juos būtų galima

panaudoti GŠL modelių kūrimui. Įkėlus garso lokalizavimo algoritmą, saugoti duomenų nebereikėtų, nes informacija būtų apdorojama realiu laiku ir serveris grąžintų objekto vietos koordinatas. Tačiau tyrimo tikslais visi garso įrašai yra saugomi. Pasirinktas *Raspberry Pi 4* įrenginys yra pakankamai galingas ir mažai energijos naudojantis kompiuteris, kurio operatyviosios atminties ir kitų resursų tikrai užtenka įvairių modelių vykdymui.

Bevielių akustinių jutiklių mazgams naudojami *ESP32* mikroprocesoriai. Jie yra pakankamai galingi ir turi du ADC, kurių vienas naudojamas WiFi ryšiui, o kitas garso įrašinėjimui. Norint viename mazge turėti daugiau mikrofonų, reikalinga specifiskai tam pritaikyta ir dažniausiai brangesnė įranga. Tai įvertinus ir atsižvelgus, kad monofoniniai garso įrašai taip pat turi pakankamai GŠL reikalingos informacijos, nuspręsta naudoti kryptinius akustinius jutiklius *MAX9814*. Šie pasirinkimai tyrimo metu leidžia įvertinti paprasčiausios ir vienos pigiausių bevielių akustinių jutiklių sistemos pritaikomumą GŠL problemoms.

Kaip visi įrenginiai susijungia į bendrą tinklą galime matyti **6 pav.**

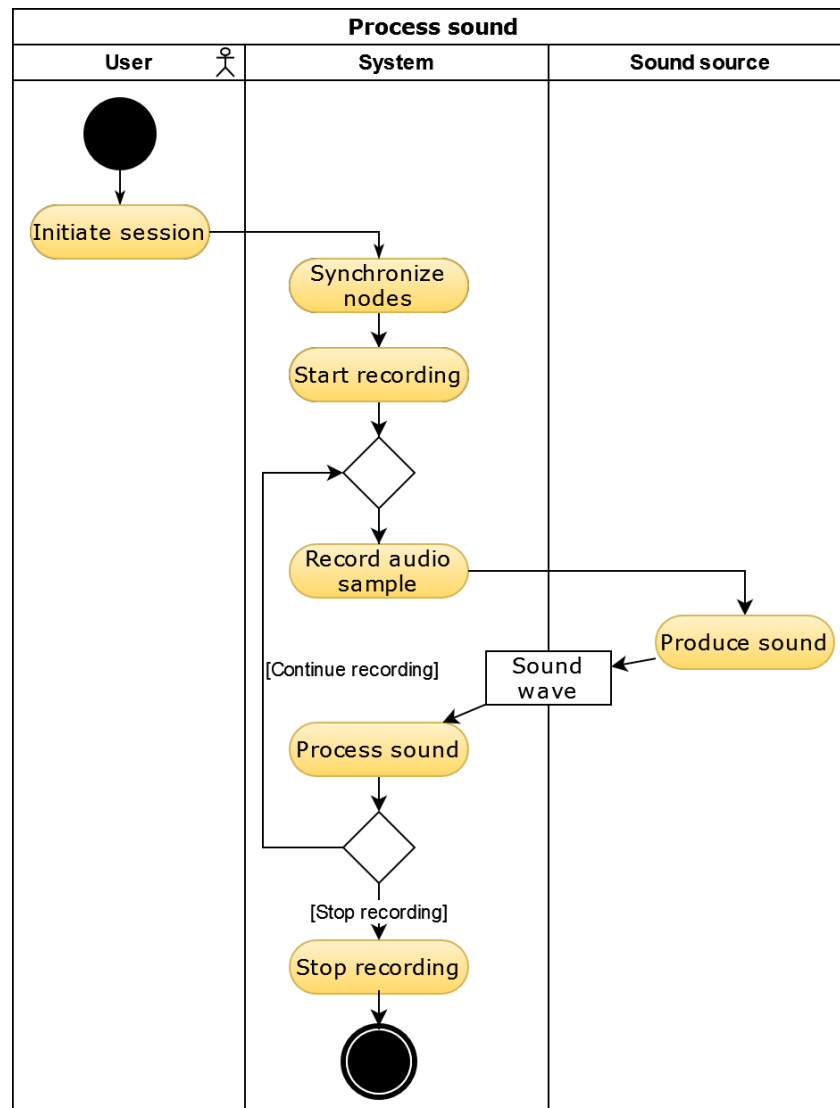


6 pav. Sistemos tinklo diagrama

Centrinis valdiklis *Raspberry Pi 4* atlieka maršrutizatoriaus vaidmenį. Prie jo jungiasi vartotojo kompiuteris ir visi bevieliai sistemos įrašinėjimo mazgai. Mikrofonų kiekį riboja tik tinklo konfigūracija ir centrinio valdiklio gebėjimas aptarnauti visus įrenginius. Šiame tyrime pateikiama ir analizuojami pati paprasčiausia tinklo struktūra. Tačiau galima lengvai ją praplėsti ir pritaikyti įvairaus dydžio GŠL problemoms spręsti.

2.3.2. Programinė sistemos struktūra

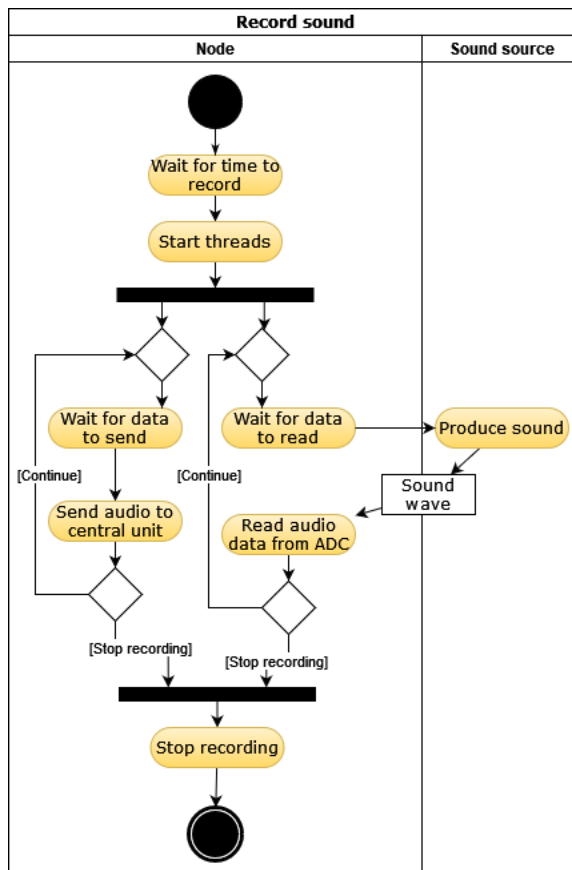
Bevielių akustinių jutiklių sistemos pagrindą sudaro centriniam valdiklyje veikiantis serveris. Jis atsakingas už sistemos valdymo grafines sąsajas pateikimą vartotojui, mazgų valdymą, jų sinchronizavimą ir garso įrašų saugojimą failų sistemoje arba GŠL, jeigu įkeliamas tam skirtas modelis. Sistemos mazgai įrašinėja garsą ir siunčia jį į centrinį valdiklį. Pagrindinis siūlomo sprendimo veikimo principas pateiktas **7 pav.**



7 pav. Garso apdorojimo veiklos diagrama

Vartotojas pirmiausia inicijuoja įrašinėjimo sesiją. Tada sistemos centrinis valdiklis sinchronizuoja laikus prisijungusiuose mazguose ir nurodo laiką įrašinėjimo pradžiai. Kiekvienas mazgas cikliškai kaupia garso bangos duomenis, kuriuos aptinka mikrofonu, ir siunčia į centrinį valdiklį apdorojimui. Įrašinėjimas vyksta nenutrūkstančiai iki kol pasiekama pabaigos sąlyga, jei tokia nustatyta. Norint gauti tam tikro ilgio garso bangos duomenis, apribojamas ciklų skaičius.

Beveiliai akustiniai sistemos mazgai prisijungia prie centrinio valdiklio ir vykdo jo siunčiamus nurodymus. Sinchronizavimas ir garso įrašinėjimas yra dvi pagrindinės mikrovaldiklių funkcijos. Pastarosios vykdymo eiga pateikiama veiklos diagrama 8 pav.



8 pav. Garso įrašymo veiklos diagrama

Įrašo darymo komanda nurodo tikslų laiką, kada reikia visiems mazgams pradėti. Tuomet mikrovaldikliai sukuria du lygiagrečius procesus ir inicijuoja garso įrašinį, kuris vykdomas nenaudojant procesoriaus. Tai padaryta įrašant duomenis su DMA – tiesiogine prieiga prie atminties. Viena iš procesoriuje vykdomų užduočių laukia, kol atsiras garso duomenų, o juos nuskaičius ir apdorojus išsaugo cikliniame buferyje. Kita užduotis ima garso bangos įrašus iš buferio formatuoja ir siunčia centriniam valdikliui. Tokiu principu veikianti sistema vienu metu gali įrašinėti ir išsiųsti garsus. Ciklinio buferio paskirtis yra užtikrinti, kad visi garso įrašų segmentai būtų perduoti, jeigu kai kurių tinklo paketų siuntimas užsidelsia. Įrašinėjimas vykdomas tol, kol baigiasi nurodytas ciklų skaičius arba gali tęstis be pabaigos.

2.3.3. Mazgų sinchronizavimas

Sistemos mazgų tarpusavio sinchronizavimas yra reikalingas siekiant tiksliai nustatyti garso atvykimo laiką iki mikrofonų skirtumą. Mikrovaldikliai nėra sujungti laidais, kurie leistų nykstamai mažomis paklaidomis įrašinėti lygiagrečius garso kanalus. Taip pat mazgai nėra paleidžiami vienu metu, todėl neturi bendro atskaitos taško laiko skaičiavimui. Tačiau bevielių sistemų sinchronizavimas yra pakankamai ištirtas. Dažnai kompiuterinėse sistemose taikomas NTP protokolas ar kiti laiko nustatymo ir suvienodinimo algoritmai naudoja UDP paketus. Kadangi šio tyrimo metu vykdomi tik trumpi garso įrašai, nereikia pastovaus mazgų sinchronizavimo ir galime pasirinkti paprasčiausią laiko nustatymo vietiniame tinkle būdą.

Sinchronizavimą inicijuoja serveris, kuris praneša mazgams pasiruošti priimti tinklu siunčiamus duomenis. Tai užtikrina, kad mikrovaldikliai nebūtų užsiėmę kitomis užduotimis. Atėjus UDP

paketui mazgas įsirašo procesoriaus laiką milisekundėmis ir laukia kito paketo arba baigia sinchronizavimo procedūrą, jei visi paketai jau gauti. Serveris dešimt UDP paketų siunčia pusės sekundės intervalais. Gavus visus paketus arba kai naujas paketas neateina per 5 sekundes, sinchronizavimo procesas yra stabdomas ir surinktos laiko žymės siunčiamos į serverį. Tuomet apskaičiuojamas kiekvieno mazgo procesoriaus darbo laiko skirtumas, lyginant su vėliausiai įsijungusiu mazgu, ir rezultatai išsiunčiami į mikrovaldiklius. Visi mazgai turi tokią pačią programinę ir fizinę įrangą, todėl apdoroti vienodą UDP paketą turėtų užtrukti panašiai tiek pat. Tokia sinchronizavimo procedūra nustato visuose mazguose laiko skaičiavimo atskaitos tašką, kurio tikslumas priklauso tik nuo Wi-Fi bangų sklidimo iki mazgų laiko skirtumo. Atsižvelgiant, kad radijo bangos sklidimo greitis laisvoje erdvėje yra apie 300 m/μs, o tyrimo patalpų dydis yra santykinai mažas, sinchronizavimas 1 μs tikslumu teoriškai yra įmanomas.

Vykdam sinchronizavimą laiko žymėms buvo naudojama *millis* funkcija, kurios mažiausias laiko vienetas yra 1 ms. Akivaizdu, kad tokiu tikslumu ir vykdomas sinchronizavimas. Išmatuoti laiko atskaitos taško nustatymo tikslumą skirtinguose mazguose būtų galima naudojant laidinę jungtį. Tačiau šio tyrimo modeliai turėtų nustatyti garso šaltinio vietą net ir su netiksliai sinchronizuota sistema. 1 ms paklaida 20 laipsnių temperatūros kambaryje, kuriame garso greitis yra apie 342.6 m/s, reiškia 34 cm paklaidą, nustatant garso šaltinio vietą. Tokio tikslumo GŠL modelis būtų labai didelis progresas tiriamoje srityje.

2.4. Duomenų specifikacija

Garso šaltinių išskyrimas ir objekto vietos nustatymas naudojant bevielį akustinių jutiklių tinklą yra dvi sudėtingos, nors ir glaudžiai susijusios problemos. Šio tyrimo metu pasirinkta analizuoti vieno išskirto šaltinio kelių kanalų garso įrašą. Poskyryje pateikiamas numatytas modelių įvesčių ir išvesčių formatas. Taip pat aprašoma, kokie yra renkamų duomenų raktiniai kriterijai, siekiant įvairovės. Nurodoma, kaip saugojami ir formatuojami garso įrašai bei būdas padidinti turimos informacijos mokymui kiekį.

2.4.1. Modelio duomenų įvestis, išvestis

Modelio įvestis priklauso nuo įrašo kanalų skaičiaus ir trukmės. Siūlomos sistemos mazgų skaičius apriboja maksimalų garso kanalų kiekį. Taip pat įvertinama, kad reikalingi mažiausiai trys įrašai, siekiant pakankamai tiksliai lokalizuoti objektą trimatėje erdvėje. Tyrimo metu dėmesys skiriamas GŠL algoritmams, kurie naudoja 3 ir 7 kanalų garso įrašus, tačiau paliekama galimybė keisti ši skaičių. Kai lokalizavimo modelio vidiniai parametrai yra dinamiški, taikomi 50, 150 ir 300 milisekundžių trukmės duomenų rinkiniai. Tačiau tam tikriems sprendimams realizuoti reikalinga specifinio formato informacija. Garso įrašymas vykdomas 16 KHz dažniu, vadinasi bangos duomenų ilgiai pateikiami modeliams yra: 800, 2400 ir 4800. Duomenų reikšmių amplitudė yra normalizuojama ir nustatoma intervale [0, 1].

Dar vienas įvesties parametras yra mikrofonų vietos trijų dimensijų erdvėje nurodančios koordinatės. Šios yra normalizuojamos iki skaičių skalės [0, 1]. Garso bangos ir jutiklių poziciniai duomenys pateikiami atskirai. O gaunamas modelio rezultatas palyginamas su taip pat normalizuotomis garso šaltinio koordinatėmis. Vadinasi siūlomi sprendimai turi nustatyti objekto vietą, naudojant garso įrašus ir sistemos mazgų pozicijas.

2.4.2. Garso įrašų specifikacija

Tyrimui ir GŠL problemos modelio kūrimui reikalingi duomenys įrašomi naudojant sukurtą sistemą. Siekiant užtikrinti pastoviai gerą siūlomo sprendimo veikimą realiomis sąlygomis, reikalingas pokyčiams atsparus lokalizavimo algoritmas. Tyrimo metu turi būti surinkta duomenų įvairovė. Tai užtikrinti galime keisdami šiuos parametrus:

- mikrofonų išdėstymas – mikrofonai patalpoje visada paskirstomi tolygiai patalpoje ir kuo įvairesnėmis galimų pozicijų kombinacijomis. Mazgai montuojami atvirose vietose;
- patalpos apstatymas ir aido susidarymas – priklausomai nuo patalpos ir jos apstatymo susidaro skirtingos galimybės atsirasti aidui. Todėl įrašai vykdomi bent 3-ijose skirtingose patalpose;
- objekto skleidžiamas garsas – naudojami skirtingų dažnių ir jų kombinacijų garso šaltiniai: vyro kalba, moterų balsai, muzika, keli pastovūs dažniai, šuns lojimas, tiksjėjimas;
- garso stipris – skleidžiami garsai taip pat gali būti tylūs, vidutiniai arba garsūs;
- garso šaltinio koordinatės sistemos atžvilgiu – tyrimo metu garso šaltinio buvimo vieta atsitiktinai keičiama patalpoje taip, kad objektas išliktų bevielių mazgų sistemos erdvėje arba netoli jos.

Garso įrašymo metu reikia atsižvelgti į šiuos kriterijus ir užtikrinti renkamų duomenų įvairovę. Taip sukurto modelio universalumas prie įvairių sąlygų bus didesnis ir įvertinimas teisingesnis.

2.4.3. Garso įrašų duomenų papildymas

Duomenų papildymas (angl. *data augmentation*) yra naudojamas, kai turimas apmokymams skirtas informacijos kiekis yra nedidelis. Ši duomenų analizės technika papildo surinktą duomenų rinkinį tiek modifikuotomis kopijomis. Tačiau patys pakeitimai unikaliai priklauso nuo informacijos. Papildymo technika padeda apmokyti neuroninius tinklus, nes sumažina persimokymo riziką. Garso įrašai turintys daugiau kanalų nei reikia, gali būti panaudoti kelis kartus. Skirtingų kanalų kombinacijos duoda unikalius duomenų rinkinius, nes mikrofonai išdėstyti dideliais atstumais turi skirtingas koordinates ir užfiksuoja kitokius garsus. Net sukeitus tų pačio garso įrašo kanalus vietomis gausime naują duomenų rinkinį. Toks duomenų papildymas leidžia labai daug kartų padidinti informacijos kiekį. Tačiau persistengus galima vėl pradėti patirti persimokymo problemą. Todėl šio tyrimo tikslais nuspręsta duomenų rinkinį padidinti 10 kartų.

2.5. Vertinimo kriterijus

Pagal siūlomą garso šaltinio vietos nustatymo metodiką, įvesties duomenys yra garso įrašai ir mikrofonų buvimo vietos, o išvestis yra garso šaltinio koordinatės trimatėje erdvėje. Žinodami tikrąsias garso šaltinio koordinates $p_T = (p_{Tx}, p_{Ty}, p_{Tz})$ ir modelio apskaičiuojamas koordinates $p = (p_x, p_y, p_z)$ galime įvertinti modelio tikslumą. Tai apskaičiuojame su kvadratinės šaknies vidurkio paklaidos formule (1) pateikta straipsnyje [66].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n |p_{Ti} - p_i|^2}{n}} \quad (1)$$

čia n – garso pavyzdžių kiekis, kuriam yra skaičiuojama paklaida; p_{Ti} – i -tojo duomenų elemento tikroji garso šaltinio buvimo vieta; p_i – modelio apskaičiuotos i -tajam duomenų elementui buvimo

vietos trimatėje erdvėje koordinatės. Ši formulė pritaikyta ir naudojama skaičiuojant net kelių objektų vietų aibės nustatymo tikslumą [39] ar atliekant kitas GŠL užduotis [4, 63].

2.6. Geometrinių modelių specifikacija

Vienas paprasčiausių GŠL metodų yra geometrinis. Jis susideda iš dviejų esminių dalių: garso kelio iki mikrofonų porų skirtumo nustatymo ir geometrinio erdvinių paviršių susikirtimo taško paieškos. Šio tyrimo metu naudojami keli skirtingi algoritmai kiekviename etape.

2.6.1. Garso sklidimo skirtumo nustatymas

Pirmojo GŠL etapo metu nustatoma kiek toliau garsui reikia keliauti iki skirtingų mikrofonų porų. Vienas iš paprasčiausių būdų naudoja garso energijos pokytį, kuris atvirkščiai proporcingas nueito kelio kvadratui. Garso banga svyruoja tarp neigiamų ir teigiamų reikšmių, todėl norėdami surasti vieno įrašo garso lygį skaičiuojame reikšmių kvadratų sumą. Tokiu būdu sumažinamas duomenų kiekis iki vieno skaičiaus kiekvienam garso segmentui. Reikia pastebėti, kad naudoti ilgus įrašus netikslinga, nes, esant vienam tuo metu nejudančiam garso šaltiniui, energijos skirtumai nuo trukmės nepriklauso. Apskaičiavus garso stiprį, galima kiekvienai mikrofonų porai nustatyti garso sklidimo kelio skirtumus.

Kitas dažnai GŠL srityje naudojamas metodas yra laikų atvykimo skirtumo nustatymas. Jį realizuoti galima aptinkant didžiausią dviejų įrašų sutaptį juos praslenkant vienas kito atžvilgiu. Kryžminė koreliacija atlieka garso aptikimą kitame ilgesniame įraše, tam reikia paimti dalį vieno įrašo ir apskaičiuoti sutaptį. Norint padidinti tikslumą ši procedūra taikoma du kartus sukeičiant įrašus vietomis, o sutapties sekos sudauginamos. Tuomet atrenkamas didžiausią reikšmę turintis taškas, kuris nurodo kiek viena duomenų seka yra prasislinkusi kitos atžvilgiu. Reikia pastebėti, kad triukšminguose įrašuose gali būti netinkamai randamas poslinkis. Todėl šio tyrimo metu siūlomi papildomi du algoritmai, kurie parenka nurodytą skaičių arba tam tikrą procentą didžiausių koreliacijos taškų. Tada grąžinamos kelios tikėtinos poslinkio reikšmės.

2.6.2. Geometrinė lokalizacija

Nustačius garso nueinamų atstumų skirtumus iki jutiklių ir žinant mikrofonų koordinates, galime apskaičiuoti tikėtiną objekto vietą. Kiekviena pora akustinių jutiklių nurodo po hipersferą, kurios paviršiaus taškuose gali būti garso šaltinis. Šių erdvinių figūrų susikirtimo paieška, ypač 3D erdvėje, reikalauja daug skaičiavimų. Lokalizavimui pasirinktas tinklelio paieškos algoritmas. Nors jis nėra greitas, bet jo paprasta realizacija leidžia pritaikyti įvairius susikirtimo vietos aproksimacijos algoritmus. Šio tyrimo metu realizuoti vidurkio ir medianos nustatymo algoritmai naudoja tinklelio paieška rastus susikirtimo taškus skaičiuojant tikėtiną garso šaltinio vietą. Aplinkoje gali atsirasti aidų ir gali būti numanomos kelios garsus skleidžiančios vietos. Todėl tyrimams pritaikytas ir k-vidurkių metodas, kuris galėtų atskirti triukšmus nuo ieškomos informacijos.

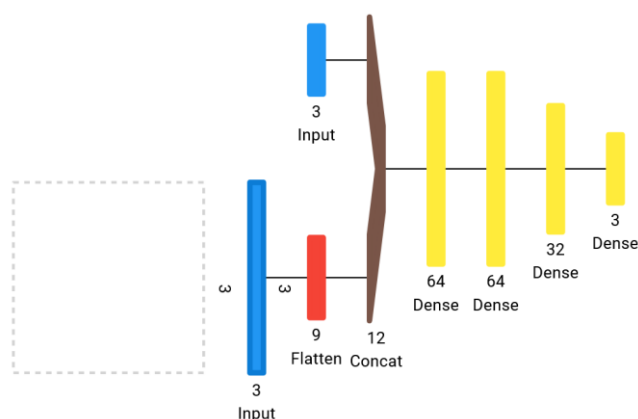
2.7. Neuroninių modelių specifikacija

Neuroniniai tinklai dažnai taikomi GŠL, bet pasirinktai probleminei sričiai, kai įvesčiai naudojamos ir mikrofonų koordinatės, modelių pritaikymų nėra. Šio tyrimo metu buvo pasirinkti keli populiarūs panašioms užduotims skirti neuroniniai tinklai ir modifikuoti sprendžiamai problemai spręsti. Vieni modeliai naudoja garso įrašą įvestims, kiti pirmiausiai apskaičiuoja spektrogramas arba laikų

atvykimo iki mikrofonų skirtumus. Dauguma siūlomų modifikacijų leidžia pasirinkti pradinių duomenų kanalų skaičių ir įrašo trukmę, bet kelios tiriamos struktūros nėra pakankamai lanksčios.

2.7.1. Pilnai sujungtas neuroninis tinklas

Pilnai sujungti neuroniniai tinklai yra seniai taikomi. Tokie modeliai yra patys primityviausi ir sprendžia pakankamai lengvas užduotis, nes šis neuroninis tinklas yra linkęs persimokyti. Vienas siūlomų sprendimų pavaizduotų **9 pav.** atlieka geometrinio modelio vaidmenį ir bando nustatyti garso šaltinio koordinatės iš atvykimo laikų iki mikrofonų skirtumų.

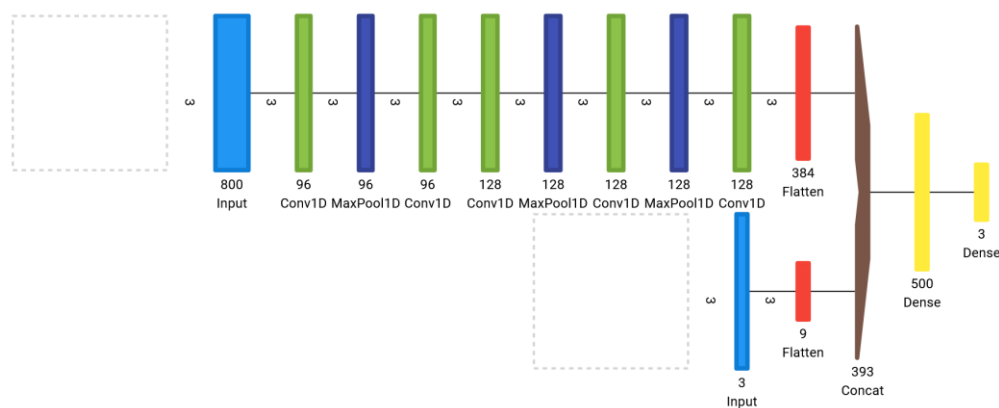


9 pav. Pilnai sujungtas neuroninis tinklas *NN_TDOA*

Čia pateikiamas neuroninis tinklas apdoroja trijų kanalų garso įrašus, tačiau tokią struktūrą lengvą plėsti ir galima naudoti su daugiau kanalų turinčiais įrašais. Šį neuroninį modelį darbe vadinsime *NN_TDOA*. Kaip matome paveikslėlyje mikrofonų koordinatės yra paverčiamos į vektorių ir sudedamos su normalizuotais atvykimo laikų skirtumų duomenimis. Tuomet išdėstyti keturi pilnai sujungti neuroniniai sluoksniai, kurie sumažėja iki trijų neuronų sluoksnio. Visų šiame darbe siūlomų neuroninių modelių paskutiniuose sluoksniuose naudojama *Softmax*, o visuose kituose ReLU aktyvavimo funkcijos. Taip pat reikia pastebėti, kad visų modelių pilnai sujungtiems neuroniniams sluoksniams taikomos išmetimo ir partijos normalizavimo technikos, kurios nebus pateikiamos neuroninių modelių struktūros schemose.

2.7.2. Konvoliuciniai neuroniniai tinklai

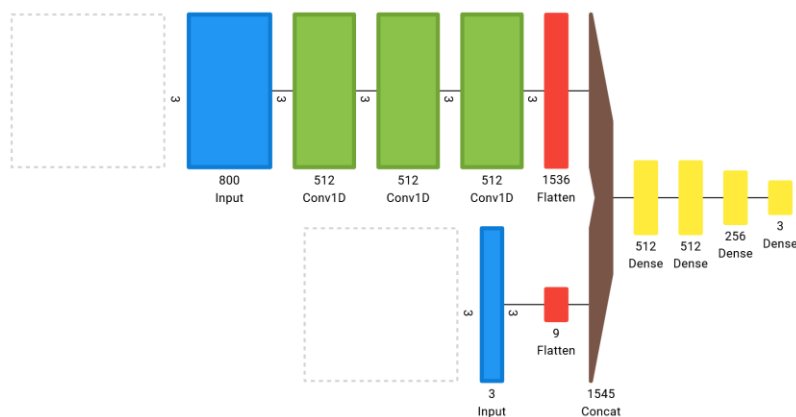
Konvoliuciniai neuroninių tinklų sluoksniai išskiria duomenų savybes, todėl jie puikiai tinka garso įrašų apdorojimui. Tyrime [39] yra siūlomas modelis iš karto nustatantis koordinatės, naudojant garso įrašus įvestims. Tokį neuroninį tinklą galima pritaikyti ir vieno objekto lokalizavimui. Taip pat įvesties įrašų ilgis ar kanalų skaičius gali būti lengvai pakeisti. Konvoliucinio modelio, kuris apdoroja 3 kanalų įvestį ir 50 ms ilgio įrašus, struktūra pateikta **10 pav.**



10 pav. Konvoliucinis neuroninis tinklas su garso įvestimi *CNN_raw_pool*

Šiam modeliui pakeisti įvesties ir išvesties sluoksniai, taip pat įterptas akustinių jutiklių koordinatų pridėjimas prieš pilnai sujungtus sluoksnius. Taip nuspręsta padaryti, nes konvoliuciniai sluoksniai gali iškraipyti ir prarasti svarbią vietos informaciją tarp sąlyginai mažiau svarbių garso įrašo duomenų. Tokį konvoliucinį modelį šiame darbe žymime *CNN_raw_pool*.

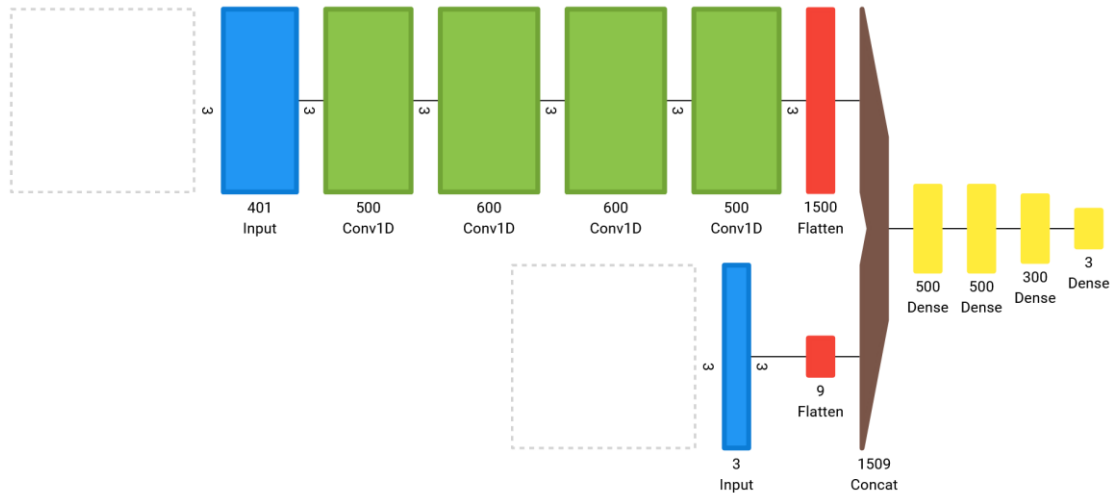
Šiame tyrime taip pat siūlomas iš daugiau pilnai sujungtų neuronų ir didesnių konvoliucinių sluoksnių sudarytas modelis *CNN_raw_big*, pateiktas **11 pav.**



11 pav. Konvoliucinis neuroninis tinklas su garso įvestimi *CNN_raw_big*

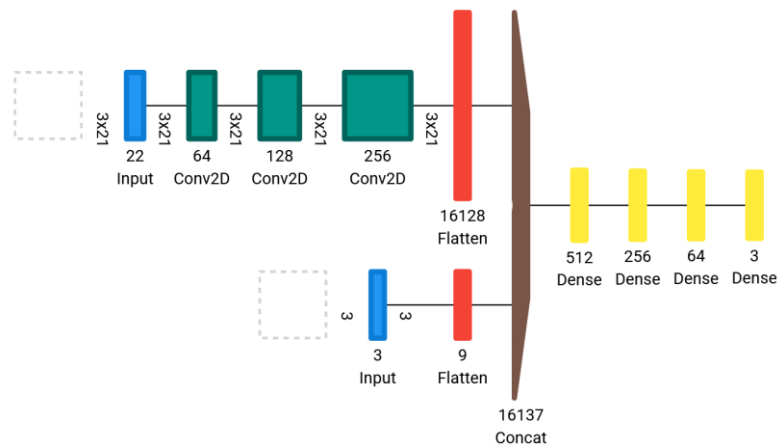
Šis neuroninis tinklas neturi *MaxPooling* sluoksnių, kaip pirmasis. Dažniausiai didesnę modelį apmokyti užtrunka daugiau laiko, todėl bus galima palyginti, kokia struktūra yra pranašesnė nustatant garso šaltinio koordinatės.

Konvoliuciniai neuroniniai tinklai geba apdoroti ne tik garso duomenis, bet ir spektrogramas. Geras pavyzdys yra [56] tyrime pateiktas krypties klasifikavimo modelis. Pakeitus įvesties ir rezultatų sluoksnius bei įterpus mikrofonų koordinatės, gauname **12 pav.** pavaizduotą *CNN_spectro_single* sprendimą.



12 pav. Konvoliucinis neuroninis tinklas su spektrogramos įvestimi *CNN_spectro_single*

Šis neuroninis tinklas įvestims naudoja vieno segmento visų garso kanalų spektrogramų sąjungą. Kadangi naudojama tik viena garso įrašo atkarpa, modelio dėmesys skiriamas garso dažnių skirtumų analizei. Patobulinta *CNN_spectro_multi* struktūra įvestims naudoja suskaidytą garso įrašą į smulkesnes dalis, kurių kiekvienai apskaičiuojama spektrograma. Neuroninis modelis naudoja dviejų dimensijų konvoliucinius filtrus, kaip parodyta **13 pav.**



13 pav. Konvoliucinis neuroninis tinklas su spektrogramų įvestimi *CNN_spectro_multi*

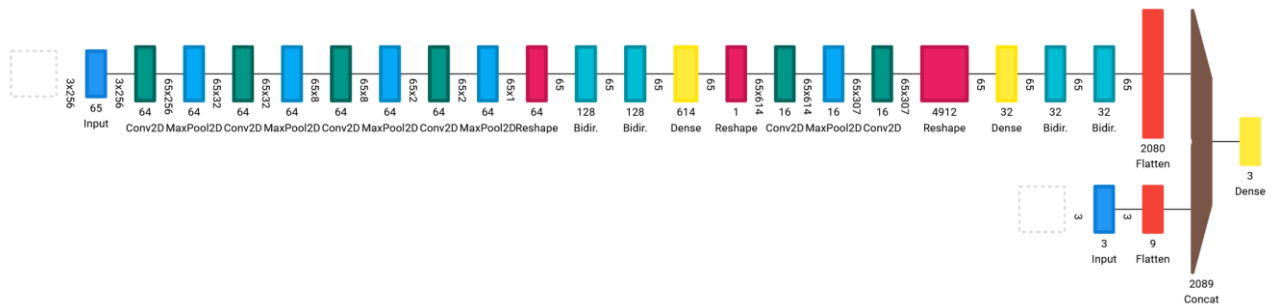
Šis neuroninis tinklas taip pat turi mažesnius sluoksnius, todėl jis turėtų apsimokyti ir veikti greičiau.

Visų pasiūlytų konvoliucinių modelių įvesčių struktūrą galima modifikuoti ir naudoti ilgesnius ar daugiau kanalų turinčius įrašus. Šio darbo metu įvertinamos kelios tokios pačios struktūros neuroninių tinklų versijos, turinčios skirtingą įvesties duomenų formatą.

2.7.3. Rekurentinis neuroninis tinklas

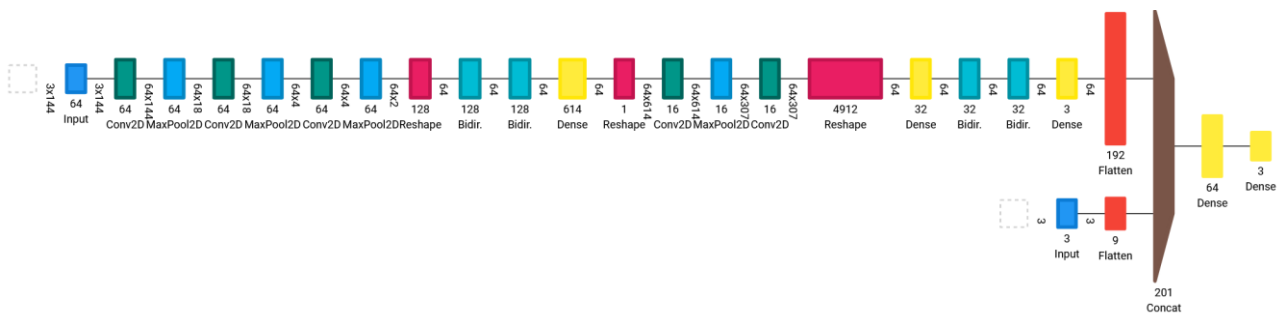
Rekurentiniai sluoksniai neuroninių tinklų modeliuose dažniausiai naudojami laikinėms sekoms apdoroti. Kadangi jie geba atsiminti informaciją ir ją naudodami interpretuoti naujus duomenis. Tokių modelių taikymas yra populiarus ir garso apdorojimo sistemose. Tyrimė [30] pateikiamas rekurentinis

neuroninis tinklas skirtas garso šaltinio krypties nustatymui. Šio modelio pritaikymas sprendžiamai GŠL problemai pavaizduotas **14 pav.**



14 pav. Rekurentinis neuroninis tinklas su spektrogramų įvestimi *CRNN_big*

Šio neuroninio tinklo garso įvesties ilgis yra nekeičiamas ir nustatytas projektavimo metu. Kadangi pateiktas modelis yra gilus, todėl tenka jam pateikti beveik dviejų sekundžių trukmės įrašą. Atsižvelgiant į šio tiriamojo darbo reikalavimus, toks sprendimas veiktų šiek tiek atsilikdamas, o ne realiu laiku. Siekiant *CRNN_big* pritaikyti pasirinktai projekto problemai, galima sumažinti sluoksnių skaičių. Tokio modelio šiek tiek mažesnė versija pateikiama **15 pav.**

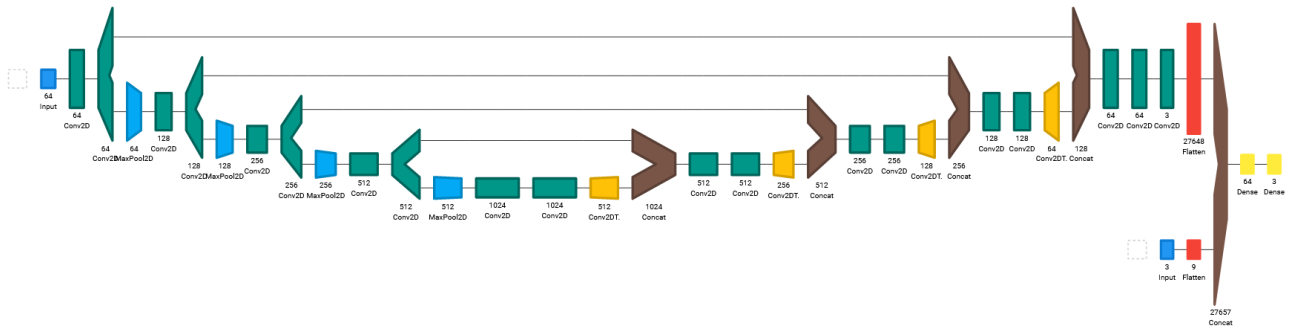


15 pav. Rekurentinis neuroninis tinklas su spektrogramų įvestimi *CRNN_small*

Atsisakius vieno konvoliucinio sluoksnio, modeliui *CRNN_small* galima paduoti mažesnės apimties duomenis. Papildomai pridėdamas pilnai sujungtas neuronų sluoksnius pabaigoje, turėtų leisti geriau įvertinti sąsajas su akustinių jutiklių koordinatėmis.

2.7.4. Kodavimo-dekodavimo neuroniniai modeliai

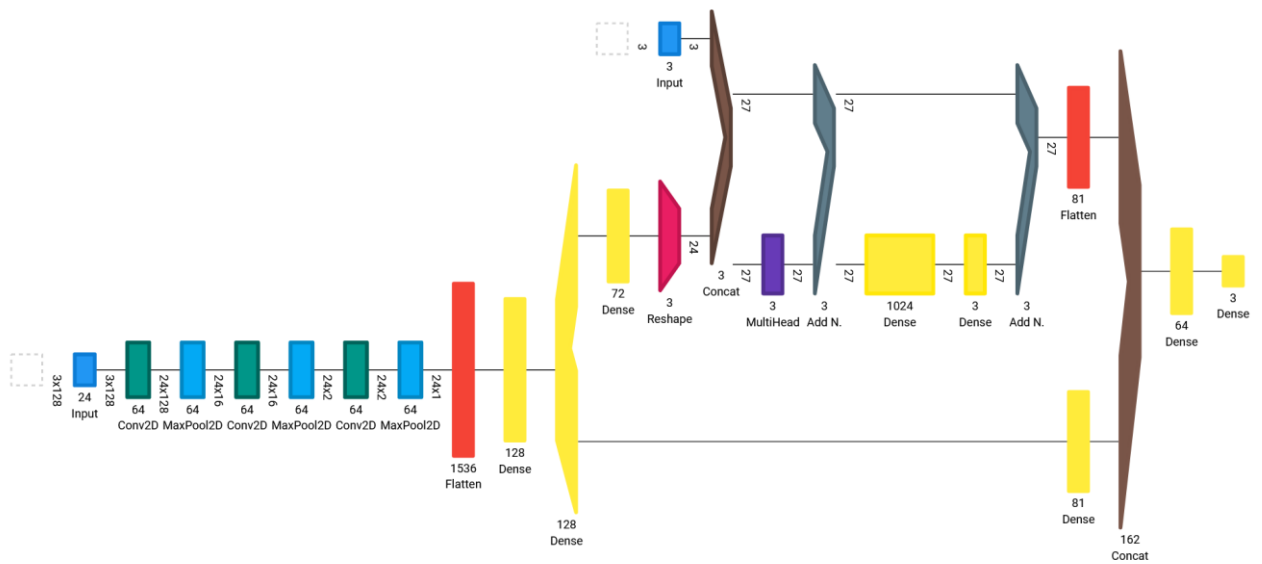
Šiame tiriamajame darbe taikomi du sudėtingesnės struktūros neuroniniai modeliai, kurių sudėtyje yra kodavimo-dekodavimo elementų. Vienas tokių modelių yra U-Net, kurio pritaikymą GŠL srityje galime rasti straipsnyje [63]. Šis modelis geba priimti ir grąžinti vienodai didelius duomenų masyvus. Tačiau pridėjus tinkamus sluoksnius modelio pabaigoje, galima iškart nustatinėti koordinates. Modifikuota struktūra pateikiama **16 pav.**



16 pav. U-Net modelis su spektrogramų įvestimi *U-Net_spectro*

Šiam *U-Net_spectro* modeliui reikalinga didelė duomenų imtis, maždaug vienos sekundės įrašas. Tai lemia didesnę delsą realaus laiko sistemose. Tačiau papildomos sluoksnių jungtys, padeda tą informaciją išsaugoti ir panaudoti tolimesniuose etapuose.

Vienas iš naujausių ir inovatyvesnių sprendimų yra *PILOT* modelis, pateiktas straipsnyje [65]. Tai transformerio modelio veikimu sukurtas neuroninis tinklas, kuris naudoja kodavimo elementą. Transformeriai dažniausiai naudojami nuosekliam duomenų sekos konvertavimui į kitokio tipo informaciją. Šio tipo modeliai naudoja dėmesiu pagrįstus sluoksnius, kurie padeda išskirti iš įvesčių svarbesnę informaciją. Pirmasis transformerio pritaikymas GŠL srityje, naudoja n kodavimo elemento kopijų, sluoksnių nustatinėdamas n garso įvykių viename įrašė. Tokį neuroninį tinklą galime pritaikyti ir šiame darbe tiriamai problemai, kaip pavaizduota 17 pav.

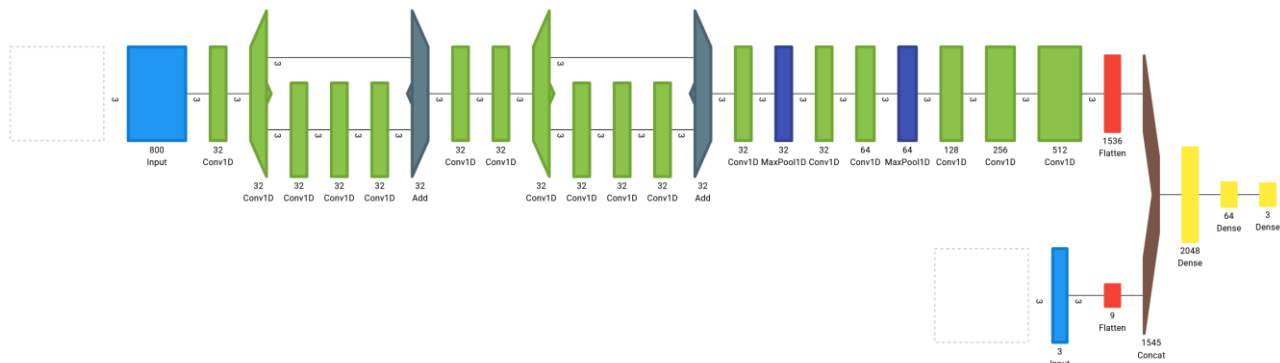


17 pav. Dėmesio mechanizmu pagrįstas neuroninis modelis *Attention_NN*

Vietoj pozicinių duomenų prijungiamų prieš pat kodavimo elementą, galima pateikti su kanalais susijusias mikrofonų koordinatas. Kodavimo elemento kopijos, skirtos kelių garsų lokalizavimui, šiuo atveju yra nereikalingos, todėl paliekamas tik vienas elementas. Apjungus išėjties duomenis pilnai sujungtais neuroniniais sluoksniais, gauname *Attention_NN* modelį, nuspėjantį garso šaltinio koordinatas.

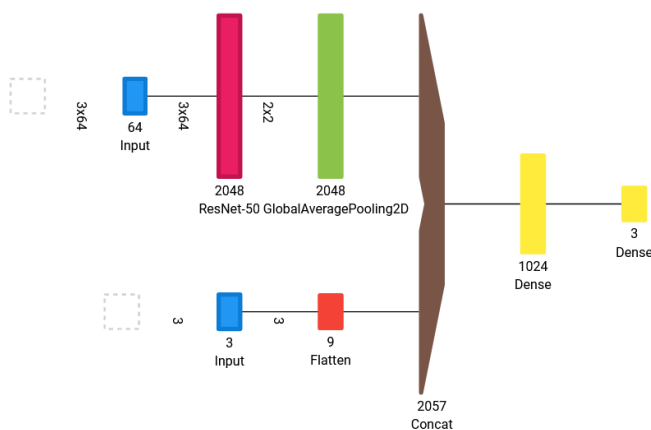
2.7.5. Liekamosios vertės neuroniniai modeliai

Šiame tyrime naudojami liekamosios vertės neuroniniai tinklai. Papildomas sluoksnių jungimas padeda išvengti persimokymo, todėl dažnai kuriami gilūs modeliai. Tyrime [27] siūlomas liekamosios vertės neuroninis tinklas geba nustatyti garso sklidimo kryptį. Šį modelį galima pritaikyti ir koordinacių nustatymui, kaip pateikta **18 pav.**



18 pav. Liekamosios vertės neuroninis modelis su garso įvestimi Resid_NN

Gilių modelių apmokymas yra ilgas procesas, todėl dažnai naudojami jau sukurti. Vienas iš populiarių liekamosios vertės modelių yra *ResNet-50* nuotraukų apdorojimo neuroninis tinklas. Atsižvelgiant, kad spektrograma galima vizualiai išreikšti garso sandarą, šį modelį verta išbandyti ir GŠL srityje. Didžiausia jau sukurtų modelių panaudojimo problema yra nepakeičiama struktūra. Galima modifikuoti ir iš naujo apmokyti įvesties ir išvesties sluoksnius, bet vidinės struktūros, neprarandant apmokytų parametrų, nepakeisi. Todėl *ResNet-50* modelį galime panaudoti tik trijų įvesties kanalų apdorojimui, kaip pavaizduota **19 pav.**



19 pav. Liekamosios vertės neuroninis modelis *ResNet-50*

Tyrimo metu šis modelis įvertinamas su apmokomais vidiniais parametrais ir nekaičiant apmokytų reikšmių. Šiuos eksperimentus atitinkamai žymi *ResNet-50_train* ir *ResNet-50_static*.

2.8. Sistemos projektavimo išvados

Kuriama bevielė akustinių jutiklių sistema pasižymi mobilumu, lengvu montavimu ir naudojimu. Siūloma priemonė gali būti taikoma kitų produktų kūrimui arba GŠL metodų tyrimams. Pagrindiniai

sistemai keliami reikalavimai yra: veikimo greitis, pakankamai tikslus automatinis laiko sinchronizavimas ir gebėjimas nustatyti garso šaltinio koordinatės.

Siūlomą sprendimą sudaro centrinis valdymo įrenginys, atsakingas už mazgų valdymą ir duomenų rinkimą. Dėl pasirinktų įrankių pagrindinis valdiklis gali realiu laiku vykdyti garso šaltinio lokalizavimą. Beveik akustiniai jutikliai yra pakankamai pigūs ir turi kryptinius mikrofonus. Todėl papildomų mazgų diegimas nereikalauja jokių sistemos pakeitimų. Sinchronizavimo procesas vyksta vietiniame tinkle. Pagal viešai prieinamus tyrimus, atliktus su NTP protokolu, beveik visuose įrenginiuose laiką galima nustatyti didesniu nei 1 ms tikslumu.

Tyrimo metu naudojamas akustinių jutiklių tinklas yra pritaikytas pastoviam garso įrašinėjimui. Tačiau GŠL modelių kūrimo tikslais duomenys saugomi sistemoje. Garso įrašai atliekami sesijomis, kurių metu formuojami kelių kanalų vienodos trukmės garso bangos failai. Kartu su failais saugoma ir garso šaltinių bei mikrofonų koordinatės.

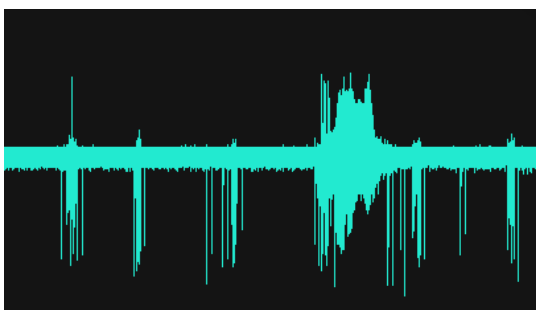
Nagrinėjami geometriniai paieškos algoritmai atlieka daug skaičiavimų, todėl bendru atveju yra pakankamai lėti. Naujausi mašininio mokymosi metodai, leidžia sukurti greitesnius modelius, kurie dažnai gali pasiekti geresnius rezultatus. Todėl tyrimo metu analizuojami konvoliuciniai, rekurentiniai ir liekamosios vertės neuroniniai sluoksniai. Taip pat įvertinami naujausi modeliai tokie, kaip U-Net, dėmesio mechanizmu grįstas transformerio kodavimo elementas. GŠL pritaikomas vaizdų apdorojimo neuroninis modelis *ResNet-50*.

3. Garso šaltinių lokalizavimo tyrimas

Šiame skyriuje pateikiami esminiai sistemos kūrimo etapai ir išvalgos, duomenų rinkimo ypatumai. Nurodomi tiriami geometriniai lokalizavimo algoritmai, palyginami jų rezultatai. Aptariamas duomenų rinkinio paruošimas ir naudojimas mašiniam apmokymui. Pateikiama neuroninių tinklų apmokymo eiga ir pasirinkti parametrai. Pateikiami panaudotų modelių rezultatai ir įvertinamas jų tikslumas bei greitaveika.

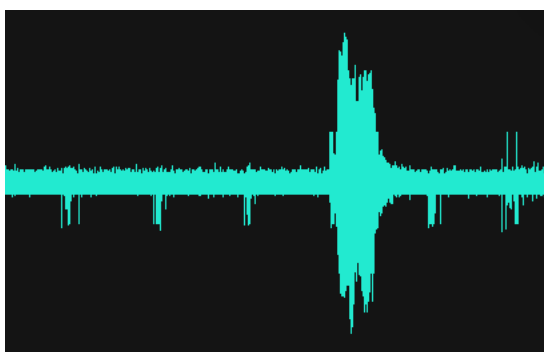
3.1. Akustinių jutiklių įvertinimas

Sistemos akustiniai jutikliai yra atsakingi už pastovų garso įrašinėjimą ir duomenų siuntimą centriniam valdikliui. Sujungus *ESP32* valdiklius ir *MAX9814* mikrofonus, pastariesiems buvo nustatytas mažiausias 40 dB garso stiprinimas. To užtenka, kad būtų girdimi pakankamai tylūs garsai, o statinis triukšmas būtų kiek galima mažesnis. Realizavus įrašinėjimo ir duomenų siuntimo procesus, gaunami garso įrašai turėjo ir pastovų nedidelį triukšmą, ir beveik periodiškai pasikartojančius bangų šuolius, pavaizduotus **20 pav.**



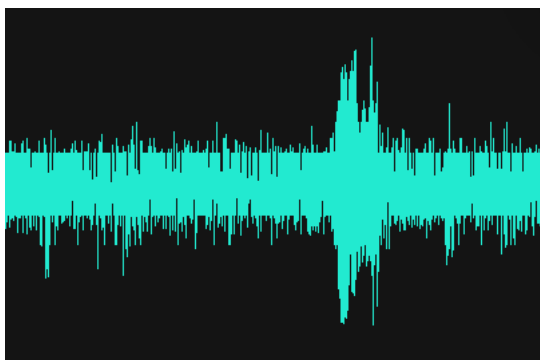
20 pav. Garso įrašas su periodiniais trikdžiais

Šie trikdžiai atsiranda, kai mikroprocesorius siunčia duomenis į pagrindinį valdiklį. Kondensatorius padėjo stabilizuoti srovę, tačiau trikdžių visiškai nepašalino. Kaip matome **21 pav.** įrašomas garsas palyginus su trikdžiais ženkliai išryškėjo.



21 pav. Garso įrašas sumažinus trikdžius

Garso įrašuose matomas ir pakankamai didelis statinis triukšmas, kurio nepašalina ir aukšto dažnio filtras. Taip pat kūrimo metu buvo pastebėta, kad ir energijos šaltinis gali sukelti tam tikrą triukšmą, kaip matome **22 pav.**



22 pav. Garso įrašas su energijos šaltinio keliamu triukšmu

Todėl buvo parinkti kiek įmanoma panašesni energijos šaltiniai. Visi mikrofonai ir valdikliai yra nauji ir kalibruoti gamykloje. Tačiau net įrašinėjant tomis pačiomis sąlygomis gaunami šiek tiek skirtingi garso įrašai. Pasiiekta garso įrašymo kokybė yra pakankama, tolimesniems darbams. Šio tyrimo tikslas yra įvertinti GŠL galimybes naudojant bevielį akustinių jutiklių tinklą, todėl tokie mikrofonai leidžia patikrinti siūlomo sprendimo veikimą realiomis, triukšmingomis sąlygomis.

Realizavus garso įrašinėjimo ir duomenų siuntimo procesus, svarbu parinkti tinkamą garso segmentų ilgį. Nuo to priklauso kiek dažnai bus siunčiami duomenys ir kokio dydžio jie bus. Garso įrašai turi būti pakankamai trumpi, kad dar užtektų laiko garso šaltinio lokalizavimui. Tai ypač aktualu realaus laiko sistemose. Eksperimentiniu būdu pastebėta, kad maždaug 200 ms trukmės įrašų siuntimas pradeda strigti, kai duomenis siunčia keli įrenginiai. Labai ilgų garso įrašų siųsti taip pat negalime, nes mikrovaldikliuose pritrūksta operatyvinės atminties informacijos saugojimui. Dėl šių priežasčių ir dėl paprastumo buvo nustatyta garso įrašus siųsti maždaug 1 sekundės segmentais.

3.2. Sistemos sinchronizavimo įvertinimas

Sistemos mazgų sinchronizavimas vykdomas prieš kiekvieną įrašymo sesiją. Nuo sinchronizavimo tikslumo priklauso garso nukeliauto atstumo nustatymo paklaida. Išskyrus atvejus, kai GŠL algoritmas naudoja garso energiją, atvykimo kryptį ar didesnio žingsnio spektrogramą. Kadangi bevielėse sistemose sinchronizavimas 1 ms tikslumu yra įmanomas, todėl galime teigti, kad įmanoma pasiekti ir 34 cm lokalizavimo tikslumą. Šios sistemos mazgų pasirinktas mažiausias laiko nustatymo vienetas taip pat yra 1 ms. Norint patikrinti, kaip tiksliai sinchronizuomi įrenginiai, reikia specialios įrangos. Todėl buvo pasirinktas paprastesnis įvertinimo būdas. Naudojant 60 kadru per sekundę greitaveikos kamerą, bevieliai mazgai filmuojami ir tikrinama ar visų lempučių nušinta tuo pačiu metu. Tokiu būdu galima nepastebėti 16 ms paklaidų. Tačiau stebėjimas atliktas 10 kartų ir nebuvo nustatyta netikslumų. Tai leidžia manyti, kad pasiektas užtektinas bevielių akustiniu jutiklių laiko sinchronizavimo tikslumas.

3.3. Bevielių akustinių jutiklių tinklo įvertinimas

Bevelis akustinių jutiklių tinklas turi dinamišką struktūrą, palyginus su kitais GŠL sprendimais. Sistemos mazgai gali būti specialiai garso įrašymui skirti įtaisai, kokie pateikiami šio tyrimo metu. Tačiau juos gali pakeisti ir kiti mikrofonus turintys įrenginiai, pavyzdžiui, telefonas, televizorius, išmaniųjų namų valdymo balsu įrenginys ir kiti. Dėl šios savybės siūlomas sprendimas yra pranašesnis už glaudžiai sujungtus specifinius GŠL įrenginius. Tačiau didžiausias dar neišspręstas

bevielių jutiklių tinklų trūkumas yra mazgų vietos nustatymas. Šio tyrimo metu mikrofonų koordinatės yra matuojamos, tačiau tai sunkus ir daug laiko atimantis procesas. Automatizavus mazgų lokalizavimą, būtų galima siūlomą sprendimą taikyti ir atvejais, kai sistemos mazgai juda aplinkoje. Be šios funkcijos sistemą lengviau pritaikyti nekintančiose aplinkose, kai mikrofonų koordinatės užtenka nustatyti vieną kartą diegimo metu.

Šiame tyrime analizuojamas sistemos veikimas vientisoje erdvėje, tačiau siūlomas sprendimas pritaikomas ir kitoms sąlygoms. Kelios patalpos ar viena, su garsą izoliuojančiomis konstrukcijomis atskiriančiomis erdvę į kelias atskiras, gali turėti sumontuotas kelias garso šaltinio vietą nustatančias sistemas. Atsižvelgiant į sistemos universalumą, mazgai gali būti išdėstomi skirtingose patalpose, bet jungiami prie vieno serverio, kuriame programiškai gali būti atskirti. Tuomet galima taikyti lokalizavimo algoritmą kiekvienai erdvei. Taip pat mazgų duomenys gali būti panaudojami kelių erdvių garso šaltinio lokalizavimui vienu metu. Didesnis garso kanalų skaičius leidžia algoritmui pasirinkti, kuriuos įrašus naudoti lokalizavimui. Taip pat galima vienu metu vykdyti kelis GŠL modelius su skirtingais kanalais ir taip patikslinti rezultatus. Siūlomo sprendimo mazgai jungiasi į vieną tinklą, todėl yra priklausomi nuo centrinio valdiklio. Yra galimybė sistemos sinchronizavimą vykdyti tarp mikrovaldiklių su akustiniais jutikliais, tačiau garso šaltinio lokalizavimui vis vien reikalingas atskiras įtaisas. Taip pat reikia atsižvelgti, kad mazgai valdikliui siunčia didelius kiekius duomenų, todėl tinklo pralaidumas sumažėja.

3.4. Garso įrašinėjimo sąlygos

Garso duomenų įrašymas vykdomas uždaroje patalpoje. Šio tyrimo metu buvo naudojamos atviros patalpos be tarpinių sienų ar didelių baldų sukuriančių garso barjerus. Įrašymai vykdomi trijose skirtingose patalpose $4,2\text{ m} \times 4,9\text{ m} \times 2,4\text{ m}$, $4,6\text{ m} \times 5,3\text{ m} \times 2,5\text{ m}$ ir $5,2\text{ m} \times 7,6\text{ m} \times 3,1\text{ m}$. Erdvių ir jų apstatymo baldais įvairovė lemia kintančias sąlygas. Nuo sienų ir daiktų priklauso aidų susidarymo galimybės bei garso sklidimo kelias. Taip užtikrinama, kad sukurtų sprendimų rezultatai išliktų stabilūs nepriklausomai nuo sistemos naudojimo aplinkos.

GŠL modeliai naudoja mikrofonų buvimo koordinatės trimatėje erdvėje nustatinėdami garso šaltinio buvimo vietą. Todėl duomenų sistemos mazgai yra atsitiktinai perkeliama kitur. Mikrofonai statomi atvirose vietose, kiek galima didesniais atstumais vienas nuo kito, bei skirtinguose aukščiuose. Mazgai išdėstomi taip, kad garso šaltinis būtų tarp mikrofonų, o ne šalia garso įvesties įrenginiais dengiamos erdvės. Kadangi pasirinkti mikrofonai yra kryptiniai, jie visada nukreipiami statmenai į viršų. Sumontavus sistemos mazgus patalpoje, išmatuojamos jų koordinatės. Atskaitos taškas, nuo kurio matuojama, nėra svarbus, nes koordinatės normalizuojamos į skaičių aibės [0-1] reikšmes.

Tyrimo metu naudojami 10 akustinių bevielių jutiklių sujungtų į bendrą sistemą. Garso įrašymai vykdomi sesijomis, kai suformuojami kiekvienam kanalui atskiri failai. Šiuos sudaro nesuspausti garso bangos įverčiai be papildomų metaduomenų. Nors toks formatas užima daugiau vietos atmintyje, bet informacija lieka nepakeista ir neapdorota. Sesijos metu pasirinkta formuoti 491520 duomenų taškų (30,72 sekundžių) garso įrašus. Tyrimui naudojama 16 KHz garso kokybė, leidžianti aiškiai suprasti garsus. Naudojant sukurtą bevielių akustinių jutiklių tinklą įrašyta 100 sesijų (3072 sekundžių) dešimties kanalų garso, tai sudaro apie 8,5 valandos monofoninio įrašo.

Visi duomenys saugomi *raw* formatu failų sistemoje. Tačiau garso bangos gali turėti vidutinio poslinkis nuo nulio (angl. *DC offset*) komponentę. Todėl visi įrašai filtruojami aukšto dažnio filtru, kuris naudojamas ir nuolatinės srovės poslinkiui pašalinti. Pašalinti statinio triukšmo, stipriai

nepaveikiant įvesties duomenų, pakankamai žemu filtru nepavyksta. Todėl taikome labai mažą 50 Hz dažnio filtrą, kuris mažiausiai paveikia įvesties duomenis.

3.5. Geometrinio lokalizavimo įvertinimas

Tiriant geometrinį GŠL, naudojamos keturios algoritmo versijos. Garso energijos skirtumų nustatymo būdu pagrįstas variantas šiame darbe žymimas *Power*. Šis metodas tiriamas su vidurkių ir medianos centrinių taškų nustatymo formulėmis taikant 3 ir 7 kanalų garso įrašus. Rezultatų duomenys pateikiami 1 lentelėje.

1 lentelė. Energijos skirtumu grįsto geometrinio lokalizavimo algoritmo *Power* rezultatai

| Įrašo trukmė, ms | Vidurkių metodas | | Medianos metodas | |
|------------------|------------------------|-------------------------|------------------|------------------|
| | 3 kanalai | 7 kanalai | 3 kanalai | 7 kanalai |
| 5 | 0.6638 21 ms | 0.5459 120 ms | 0.7498 22 ms | 0.5675 119 ms |
| 20 | 0.6641 21 ms | 0.5448 121 ms | 0.7499 21 ms | 0.5624 121 ms |
| 50 | 0.6645 23 ms | 0.5449 121 ms | 0.7499 22 ms | 0.5597 123 ms |
| 150 | 0.6631 24 ms | 0.5453 124 ms | 0.7497 24 ms | 0.5596 125 ms |
| 300 | 0.6632 26 ms | 0.5456 129 ms | 0.7497 26 ms | 0.5596 132 ms |

Lentelėje pateikiami RMSE rezultatai ir algoritmo vidutinė vykdymo trukmė. Galima pastebėti tendenciją, kad 7 kanalų nustatymas yra tikslesnis negu 3, tačiau apie šešis kartus ilgiau užtrunka. Didesnis duomenų kiekis leidžia tiksliau lokalizuoti, tačiau laiko kaina. Vidurkio aproksimacijos algoritmas visais atvejais grąžina geresnius rezultatus. Kaip buvo numatyta šiame darbe, energijos skirtumų nustatymo metodika leidžia vienodai gerai lokalizuoti su visokio ilgio garso įrašais. Todėl šį metodą geriausia naudoti trumpiems įrašams, nes juos greičiau apdoroja.

Kita geometrinio lokalizavimo metodika naudoja garso atvykimo laikų iki mikrofonų skirtumus, kurie randami kryžminės koreliacijos būdu, kai nustatomas vienas didžiausias sutapties įvertis. Šis modelis žymimas *TDOA* ir jo vykdymo rezultatai, pateikiami 2 lentelėje.

2 lentelė. Atvykimo laikų skirtumu grįsto geometrinio lokalizavimo algoritmo *TDOA* rezultatai

| Įrašo trukmė, ms | Vidurkių metodas | | Medianos metodas | |
|------------------|------------------------|-------------------------|------------------|------------------|
| | 3 kanalai | 7 kanalai | 3 kanalai | 7 kanalai |
| 50 | 0.6064 24 ms | 0.5451 126 ms | 0.6493 23 ms | 0.5644 125 ms |
| 150 | 0.5479 27 ms | 0.5451 142 ms | 0.5965 27 ms | 0.5619 143 ms |
| 300 | 0.5691 29 ms | 0.5343 142 ms | 0.5928 28 ms | 0.5442 143 ms |

Šiam geometriniam lokalizavimo algoritmui taip pat taikomi vidurkių ir medianos metodai, tačiau geresnius rezultatus duoda pirmasis. Čia taip pat galime teigti, jog didesnis garso kanalų skaičius

duoda geresnius rezultatus, bet per ilgesnį laiką. Garso įrašų trukmė turi šiek tiek įtakos rezultatams. Kai įrašas yra 50 ms ilgio, lokalizavimas yra prasčiausias. Vadinasi didesni garso įrašai padeda tiksliau lokalizuoti garsą.

Garso atvykimo laikų skirtumo geometrinio lokalizavimo algoritmą modifikavus taip, kad koreliacijos metu būtų nustatomi keli labiausiai tikėtini sutapties įverčiai, gauname *TDOA_multi* versiją. Ši metodika turėtų padėti išvengti atsitiktinių pavienių sutapties atvejų, kurie gali nutikti ir dėl mažos amplitudės triukšmo sutapties. Tyrimo rezultatai pateikiami 3 lentelėje.

3 lentelė. Atvykimo laikų skirtumu grįsto geometrinio lokalizavimo modifikuoto algoritmo *TDOA_multi* rezultatai

| Įrašo trukmė, ms | Vidurkių metodas | | Medianos metodas | |
|------------------|------------------------|-------------------------|------------------------|------------------|
| | 3 kanalai | 7 kanalai | 3 kanalai | 7 kanalai |
| 50 | 0.5667 74 ms | 0.5362 452 ms | 0.5954 74 ms | 0.5431 439 ms |
| 150 | 0.5404 74 ms | 0.5333 451 ms | 0.5616 75 ms | 0.5367 461 ms |
| 300 | 0.5396 79 ms | 0.5343 456 ms | 0.5571 76 ms | 0.5392 465 ms |

Ši algoritmo modifikacija pateikia panašius rezultatus, tik žymiai ilgiau užtrunka. Išlieka tendencija, kad iš 7 kanalų ir 150 ms arba 300 ms garso įrašų vidurkio metodu skaičiuojami rezultatai yra geriausi.

Taip pat buvo išbandytas k-vidurkių metodas skirtas klasifikuoti garso ir galimų triukšmų arba aido šaltinių vietoms. Šiam algoritmui nustatyta surinkti garso įrašų kryžminės koreliacijos įverčius didesnius nei 90 procentų didžiausios reikšmės ir surasti daugiausiai duomenų atstovaujantį k-vidurkių centrą. Šio tyrimo rezultatai pateikiami 4 lentelėje.

4 lentelė. Atvykimo laikų skirtumu grįsto geometrinio lokalizavimo k-vidurkių algoritmo *TDOA_kmeans* rezultatai

| Įrašo trukmė, ms | 1 grupės vidurkių paieška | | 2 grupių vidurkių paieška | |
|------------------|---------------------------|---------------------------|---------------------------|--------------------|
| | 3 kanalai | 7 kanalai | 3 kanalai | 7 kanalai |
| 50 | 0.5478 770 ms | 0.5337 2528 ms | 0.6393 1237 ms | 0.6068 3678 ms |
| 150 | 0.5297 2066 ms | 0.5312 8251 ms | 0.5936 2865 ms | 0.5809 9470 ms |
| 300 | 0.5269 2945 ms | 0.5301 12208 ms | 0.5355 3447 ms | 0.5571 13777 ms |

TDOA_kmeans algoritmas su vienos grupės paieška yra tiesiog vidurkių algoritmas, todėl naudojamas tik palyginimui. Šis modelis dėl didesnio duomenų apdorojimo kiekio veikia labai lėtai. Didesniam paieškos vidurkių skaičiui, vykdymo laikas labai išauga, o tikslumas tik blogėja. Vadinasi k-vidurkių algoritmas netinkamas vieno garso šaltinio lokalizavimui. Tačiau tai nereiškia, kad šis algoritmas blogai nustatinėtų kelis garso šaltinius.

Visų keturių algoritmų palyginimui parinkti rezultatai gaunami naudojant 150 ms garso įrašus ir vidurkių metodą, nes tada gaunami santykinai geri ir greiti rezultatai. 5 lentelėje pateikiamas geometrinio modelių lokalizavimo algoritmų palyginimas.

5 lentelė. Geometrinių lokalizavimo algoritmų rezultatai

| Algoritmas | 3 kanalų | 7 kanalų |
|--------------------|--------------------------|--------------------------|
| <i>Power</i> | 0.6631 24 ms | 0.5453 124 ms |
| <i>TDOA</i> | 0.5479 27 ms | 0.5451 142 ms |
| <i>TDOA_multi</i> | 0.5404 74 ms | 0.5333 451 ms |
| <i>TDOA_kmeans</i> | 0.5297 2066 ms | 0.5312 8251 ms |

Galima matyti, kad vieno garso šaltinio nustatymas būna tikslesnis su didesniu kanalų skaičiumi, tačiau trunka ilgiau. Nors k-vidurkių algoritmas duoda geriausius rezultatus, bet jis trunka per ilgai. Tikslumą padidina parenkamų kryžminės koreliacijos įverčių kiekis, tai galime matyti ir iš *TDOA_multi* geresnių rezultatų. Tačiau nedidelis tikslumo padidėjimas kainuoja dideliu vykdymo laiko pailgėjimu. Įvertinus visus modelius galima nuspręsti, kad *TDOA* algoritmas tiksliausiai geba nustatyti garso šaltinio vietą per pakankamai trumpą laiką. Todėl jį galima pritaikyti realaus laiko sistemose pastoviam lokalizavimui. Geometrinių algoritmų pasiekiamas tikslumas realiomis sąlygomis naudojamoje sistemoje yra gana prastas. To priežastis gali būti triukšmingi garso įrašai, kuriuose yra trikdžių, netikslus laiko sinchronizavimas. Naudojami vieni iš labiausiai paplitusių lokalizavimo metodų, be papildomų patobulinimų, nėra labai tinkami koordinatinių nustatymui, nors ir pakankamai tiksliai nustato garso atvykimo kryptį kituose tyrimuose. Šie rezultatai leis įvertinti siūlomų mašininio mokymosi metodų tikslumą.

3.6. Neuroninių modelių įvertinimas

Neuroniniai tinklai turi būti apmokyti, tam reikia paruošti ir tinkamai suformatuoti duomenis, kad šie atitiktų modelio įvesties parametrus. Šiame poskyryje aptariama neuroninių tinklų mokymo metodika ir naudojami parametrai. Sukurtų modelių rezultatai palyginami ir aptariami.

3.6.1. Pirminis duomenų apdorojimas

Mašininio mokymosi modeliai naudoja pasirinkto kanalų skaičiaus ir garso įrašo ilgio duomenis. Taip pat modeliams yra paduodamos mikrofonų koordinatės. Kadangi tyrimo metu naudojami 3 ir 7 kanalų ir 50 ms, 150 ms ir 300 ms garso įrašai, atitinkamai reikia paruošti šešis skirtingus garo įrašų rinkinius. Naudojant vieną įrašymo sesijos metu sukauptą įrašų rinkinį ir suformatuojant duomenis prieš visų mokymų vykdymą, gauname failų kurie užima apie 1 GB. Vadinasi išplėtus visus sukauptus duomenis gausime apie 100 GB informacijos. Tiek daug pasikartojančių garso įrašų saugoti neapsimoka. Todėl tyrimui sukurtas algoritmas, kuris reikalingą duomenų imtį sugeneruoja tik užklausus (angl. *lazy loading*). Dalis Neuroninių modelių naudoja spektrogramas ir garso atvykimo laikų skirtumus įvestims. Todėl sukuriamos duomenų bazės visiems reikalingiems formatams.

Duomenų skaitymas iš failų sistemos yra daug laiko trunkantis procesas. Jeigu norima gauti dalį garso įrašo iš failo *raw* formatu, pirmiausia jį reikia visą įkelti į operatyviają atmintį arba skaityti iki norimos vietos. Kad to nereikėtų daryti, visi pradiniai garso įrašai yra performatuojami į *numpy* failus. Iš jų galime skaityti duomenis tiesiai kaip iš masyvų. Šio pirminio duomenų apdorojimo metu yra panaudojamas ir aukšto dažnio filtras.

3.6.2. Mašininio mokymo metodologija

Neuroninius tinklus pirmiausia reikia apmokyti, net ir panaudojant sukurtą modelį jį reikia pritaikyti sprendžiamai problemai. Kadangi turime didelius duomenų kiekius, jų visų vienu metu apdoroti negalime, nes pritrūks kompiuterinių resursų. Visą rinkinį suskaidome į mažesnes grupes (angl. *batches*), kurias palaipsniui pateiksime modeliui. Tačiau ne visi duomenys naudojami mokymui, 80 % skiriama mokymui, o 20 % galutiniam įvertinimui. Kadangi vienos sesijos metu įrašomas tas pats garsas, esantis vienoje vietoje, todėl skirstoma sesijomis. Visi modeliai mokomi 100 epochų ir tada atliekamas tikslumo įvertinimas. Nustačius, kad mokymosi metu pasiektas tikslumas yra daug mažesnis negu tikrinimo, žinome, jog neuroninis tinklas persimokė. Tuomet didinamas išmetamų neuronų kiekis mokymo metu ir keičiama modelio struktūra, pridodant daugiau išmetimo, normalizavimo sluoksnių ar keičiami kiti parametrai. Šio tyrimo metu kuriami keli neuroniniai tinklai, todėl visų galimų optimizavimo būdų ir parametrų kiekvienam pritaikyti neverta. Tai galima daryti jau turint sukurtą modelį ir norint pagerinti jo rezultatus. Todėl visų neuroninių tinklų svoriams keisti buvo parinkta viena iš populiariausių *Adam* optimizavimo funkcija su 0,001 mokymosi greičiu. Šio tyrimo metu kuriami modeliai gali būti dar tobulinami ir gali pasiekti didesnę tikslumą, tačiau pasirinkta ištirti įvairių mašininio mokymosi technikų tinkamumą GŠL problemai.

3.6.3. Neuroninių modelių rezultatų palyginimas

Dalis neuroninių modelių turi pakankamai dinamišką struktūrą, todėl gali būti pritaikyti įvairiems įvesties formatams. Šie modeliai apmokyti su 50 ms, 150 ms ir 300 ms garso įrašų ilgio duomenimis. Lokalizavimo tikslumo, priklausomai nuo garso įrašo ilgio, nustatymo rezultatai pateikiami 6 lentelėje.

6 lentelė. Dinamiškų garso šaltinio lokalizavimo algoritmų rezultatai

| Įrašo trukmė, ms | 3 kanalai | | | 7 kanalai | | |
|---------------------------|----------------|----------------|----------------|----------------|-----------------|-----------------|
| | 50 ms | 150 ms | 300 ms | 50 ms | 150 ms | 300 ms |
| <i>NN_TDOA</i> | 0.4825 1 ms | 0.4928 5 ms | 0.5021 8 ms | 0.5144 8 ms | 0.5137 35 ms | 0.5234 61 ms |
| <i>CNN_raw_pool</i> | 0.4681 2 ms | 0.4215 4 ms | 0.5478 9 ms | 0.4431 4 ms | 0.4388 6 ms | 0.4809 12 ms |
| <i>CNN_raw_big</i> | 0.4651 1 ms | 0.4565 4 ms | 0.4473 8 ms | 0.4682 3 ms | 0.4782 10 ms | 0.4729 20 ms |
| <i>CNN_spectro_single</i> | 0.3893 1 ms | 0.4034 3 ms | 0.4089 6 ms | 0.4264 3 ms | 0.4022 10 ms | 0.4123 19 ms |
| <i>CNN_spectro_multi</i> | 0.2648 1 ms | 0.2567 3 ms | 0.3071 6 ms | 0.2714 3 ms | 0.2543 8 ms | 0.2593 17 ms |
| <i>Resid_NN</i> | 0.4011 1 ms | 0.3944 3 ms | 0.4258 5 ms | 0.4215 2 ms | 0.3901 6 ms | 0.3988 12 ms |

Galima įvertinti, kad didesnę duomenų kiekį apdorojantys modeliai veikia ilgiau, bet ne visada teikia geresnius rezultatus. Mažiausią paklaidą duoda konvoliucinis neuroninis tinklas, kuris įvestims naudoja spektrogramų masyvą. Iš gautų rezultatų sunku nustatyti kokio ilgio garso įrašai duoda geresnius rezultatus, nes skiriasi modelių struktūros ir sluoksnių dydžiai, tačiau atsižvelgiant į geometrinio lokalizavimo algoritmo rezultatų tendencijas pasirenkama toliau lyginti modelius, kurie naudoja 150 ms garso įrašus. Šie modeliai pakankamai greitai vykdo lokalizavimą ir dažniausiai

duoda gerus rezultatus. Toliau lyginami kiekvieno modelio rezultatai ir greitaveika 3 ir 7 kanalų duomenims pateikiama 7 lentelėje kartu su geriausiai veikiančio geometrinio lokalizavimo algoritmo atitinkamais įverčiais.

7 lentelė. Visų garso šaltinio lokalizavimo algoritmų rezultatai

| Algoritmas | 3 kanalų | 7 kanalų |
|---------------------------|-----------------|------------------|
| <i>TDOA</i> | 0.5479 27 ms | 0.5451 142 ms |
| <i>NN_TDOA</i> | 0.4928 5 ms | 0.5137 35 ms |
| <i>CNN_raw_pool</i> | 0.4215 4 ms | 0.4388 6 ms |
| <i>CNN_raw_big</i> | 0.4565 4 ms | 0.4782 10 ms |
| <i>CNN_spectro_single</i> | 0.4034 3 ms | 0.4022 10 ms |
| <i>CNN_spectro_multi</i> | 0.2567 3 ms | 0.2543 8 ms |
| <i>CRNN_big</i> | 0.3121 62 ms | 0.2943 93 ms |
| <i>CRNN_small</i> | 0.3305 25 ms | 0.3027 52 ms |
| <i>U-Net_spectro</i> | 0.2786 42 ms | 0.2645 72 ms |
| <i>Attention_NN</i> | 0.6645 13 ms | 0.7329 30 ms |
| <i>Resid_NN</i> | 0.3944 3 ms | 0.3901 6 ms |
| <i>ResNet-50_static</i> | 1.1090 10 ms | - |
| <i>ResNet-50_train</i> | 0.8913 10 ms | - |

Visi neuroninių tinklų modeliai išskyrus *U-Net_spectro* ir *CRNN_big* veikia greičiau nei geometrinis lokalizavimo algoritmas. Modelių veikimo laikas tiesiogiai priklauso nuo modelių dydžio, todėl didesni veikia lėčiau. Blogiausių rezultatus parodė *ResNet-50* modelis, tai lėmė faktas, kad šis modelis pritaikytas vaizdų apdorojimui, o gili jo struktūra neleidžia pakankamai greitai persimokyti. Panašu, kad dėmesio mechanizmu grįsto modelio modifikacija nebuvo tinkama ir gražinami prastesni rezultatai, nei geometrinio algoritmo. Geriausiai garso šaltinį lokalizuoja konvoliuciniai neuroniniai tinklai naudojantys spektrogramas. Rekurentiniai ir liekamosios vertės sluoksniai taip pat pasirodė geriau, tačiau tikėtina, kad dėl gilių struktūrų neužteko vykdyto apmokymo. Neuroniniai modeliai naudojantys garso įrašą įvestims turėjo geresnius rezultatus nei geometrinis nustatymo būdas, bet prastesnius nei kiti konvoliuciniai modeliai. To priežastis gali būti didelis nereikalingos informacijos kiekis paduodamas į modelį. Pilnai sujungtų neuronų tinklas, naudojantis garso atvykimo laikų skirtumą, veikė panašiai kaip ir geometrinis metodas. Aiškios tendencijos, ar lokalizavimas vyksta

tiksliu su 3 ar 7 kanalų įrašais, nematome. Tai lemia skirtingas modelių dydis ir naudojamų duomenų kiekis. Gaunamos lokalizavimo paklaidos yra pakankamai didelės, bet neuroninių tinklų sluoksnių palyginimui užtenka.

3.7. Tyrimo proceso apibendrinimas

Sukurtas bevielis akustinių jutiklių tinklas geba realiu laiku įrašinėti ir siųsti duomenis į centrinę valdiklį. Garso įrašuose išlieka nedidelis statinis triukšmas. Taip pat svarbu užtikrinti vienodą įrenginių maitinimą, tačiau net ir tada gaunami šiek tiek skirtingi garso įrašai. Nepaisant to, sistemos mazgų sinchronizavimas yra pakankamai tikslus ir GŠL gali būti vykdomas. Akustinio jutiklio struktūrą galima tobulinti elektroniniais sprendimais arba pakeičiant geresne įranga.

Siūlomo sprendimo architektūra yra labai lanksti ir pritaikoma įvairiose aplinkose. Šio tyrimo metu garsas įrašinėjamas atviroje patalpoje. Visi įrašai saugomi failų sistemoje. Sėkmingai realizuotas duomenų formatavimo sprendimas taupant sistemos resursus ir nesulėtinant modelių apmokymo proceso. Garso įrašai suskirstomi į 50 ms, 150 ms, 300 ms ir 3, 7 kanalų duomenų rinkinius, kurie gali būti naudojami ir spektrogramų ar garso atvykimo laikų skirtumų reikšmių formatu.

Tyrimo metu nustatyta, kad geometriniai lokalizavimo algoritmai geriau veikia su didesniu garso kanalų skaičiumi. Įvertinta, kad energijos stiprio pakitimo nustatymui, garso įrašo trukmė įtakos neturi. K-vidurkių algoritmo nesėkmingas taikymas parodė, kad geometrinio vieno objekto lokalizavo metu nesusidaro netikrų triukšmo ar aido šaltinių, kurie trukdytų algoritmo veikimui. Optimalus garso įrašo ilgis yra apie 150 ms. Tada geometrinis metodas veikia pakankamai greitai pritaikymui realiu laiku.

Siūlomų neuroninių tinklų įvertinimas leidžia nustatyti, kokio tipo modeliai labiausiai tinka GŠL. Šiame darbe pasiekiami tikslumo įvertinimai vis dar gali būti pagerinti, atliekant nedideles struktūros modifikacijas ar pakeičiant mokymo metodiką pasirinktam modeliui. Garso šaltinio lokalizavimui labiausiai tinka konvoliuciniai ir rekurentiniai neuroniniai sluoksniai ir garso spektrogramų naudojimas. Tyrimo metu gaunamos didelės lokalizavimo paklaidos galimai yra dėl netikslių įvesties įrenginių, paklaidų laiko sinchronizavime ir santykinai mažo duomenų rinkinio.

Išvados

1. Išanalizavus literatūrą, nustatyta, kad garso bangos energijos ir atvykimo laikų skirtumai yra pagrindinės lokalizavimui reikalingos savybės. Pastarajai apskaičiuoti naudojamas kryžminės koreliacijos įvertis. Garso bangas analizuoti ir atskirti padeda spektrogramos. O lokalizavimui taikomi geometriniai arba mašininio mokymosi metodai. Vieni iš sėkmingiausiai garso apdorojimo užduotis sprendžiančių modelių sudaryti iš konvoliucinių, rekurentinių, liekamosios vertės ir dėmesio mechanizmu pagrįstų neuroninių sluoksnių.
2. Suprojektavus ir realizavus dinamišką bevielę akustinių jutiklių sistemą, įsitikinta, kad pastovus garso įrašymas, siuntimas ir apdorojimas galimas net pigiomis, bet universaliomis sistemomis. Sukurtas įvesties įrenginys yra jautrus elektriniams triukšmams, kuriuos pavyko tik sumažinti. Todėl sunku tiksliau nustatyti garso šaltinio vietą. Sistemos mazgų laiko sinchronizavimo tikslumas yra iki 1 ms, tačiau teoriškai galima pasiekti dar mažesnę paklaidą. Siūlomo sprendimo architektūra yra gera, tačiau trūksta automatinės mazgų koordinatų nustatymo metodikos. Akustiniai mazgai ne vienodai įrašinėje garsus, todėl jiems reikalingas papildomas kalibravimas arba kitokie elektronikos tobulinimo sprendimai, kurie pašalintų triukšmus.
3. Konvoliucinių, rekurentinių ir liekamosios vertės modelių struktūros papildytos mikrofonų koordinatų įvestimi prieš paskutinius pilnai sujungtus neuroninius sluoksnius. Tai atlikta traktuojant, kad neuroninių tinklų modelis gebės išskirti garso bangų stiprio, laiko ar kitas savybes ir įvertinant mazgų koordinatas nustatys garso šaltinio vietą. Dėmesio mechanizmu grįstas *PILOT* modelis šiek tiek supaprastintas aptikti tik vieną garso šaltinį, o vietoje pozicinių duomenų pateikiama akustinių jutiklių vietos informacija.
4. Geometriniai lokalizavimo algoritmai tikslesnius rezultatus teikia su didesniu 7 kanalų ir ilgesniu garso įrašu. Tačiau vietos nustatymui užtrunkama šešis kartus ilgiau nei apdorojant 3 kanalų garsus. Iš visų tirtų geometrinių algoritmų, optimaliai gerus ir greitus rezultatus duoda atvykimo laikų skirtumą ir vidurkių aproksimaciją naudojantis metodas su 3 kanalų ir 50 ms ilgio garso įrašais. Neuroninių tinklų modeliai beveik visais atvejais atlieka lokalizavimą greičiau ir pasiekia geresnius rezultatus. Konvoliuciniai ir rekurentiniai sluoksniai kartu su spektrogramų naudojimu įvestims, leidžia tiksliau ir greičiau lokalizuoti garso šaltinius. Pasirinktas modelių *ResNet-50* ir *PILOT* pritaikymas garso šaltinių koordinatų nustatymui yra netinkamas.

Literatūros sąrašas

1. COBOS, M., et al. A Survey of Sound Source Localization Methods in Wireless Acoustic Sensor Networks. *Wireless Communications and Mobile Computing*. 2017, t. 2017 , 1-24. ISSN 1530-8669.
2. LIAQUAT, M.U., et al. *Localization of Sound Sources: A Systematic Review*. [interactyvus]. 2021, [žiūrėta 2022-03-12]. ISBN 1996-1073. doi: 10.3390/en14133910. Pieiga per: <https://search.proquest.com/docview/2549333629>.
3. RASCON, C. and MEZA, I. Localization of Sound Sources in Robotics: A Review. *Robotics and Autonomous Systems*. 2017, t. 96 , 184-210. ISSN 0921-8890.
4. DENG, F., et al. Energy-Based Sound Source Localization with Low Power Consumption in Wireless Sensor Networks. *IEEE Transactions on Industrial Electronics*. 2017, t. 64, nr. 6 , 4894-4902. ISSN 1557-9948.
5. LI, D. and HU, Y.H. Energy-Based Collaborative Source Localization using Acoustic Microsensor Array. *EURASIP Journal on Advances in Signal Processing*. 2003, t. 2003, nr. 4 , 985029. ISSN 1687-6180.
6. ALAMEDA-PINEDA, X. and HORAUD, R. A Geometric Approach to Sound Source Localization from Time-Delay Estimates. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2014, t. 22, nr. 6 , 1082-1095. ISSN 2329-9290.
7. BAI, M.R., LAN, S. and HUANG, J. Time Difference of Arrival (TDOA)-Based Acoustic Source Localization and Signal Extraction for Intelligent Audio Classification. [interactyvus]. 2018, [žiūrėta 2022-03-12]. ISBN 2151-870X. doi: 10.1109/SAM.2018.8448583. Pieiga per: <https://ieeexplore.ieee.org/document/8448583>.
8. VERREYCKEN, E., DAEMS, W. and STECKEL, J. *Passive Acoustic Sound Source Tracking in 3D using Distributed Microphone Arrays*. [interactyvus]. 2018, [žiūrėta 2022-03-12]. doi: 10.1109/IPIN.2018.8533679. Pieiga per: <https://ieeexplore.ieee.org/document/8533679>.
9. GRANDE, K.C., SCHNEIDER, N.B., SATO, G.Y. and SCHNEIDER, B. *Passive Acoustic Localization Based on Time of Arrival Trilateration*. XXVI Brazilian Congress on Biomedical [interactyvus]. 2019, [žiūrėta 2022-03-12]. ISBN 9789-811321191. Pieiga per: https://link.springer.com/chapter/10.1007/978-981-13-2119-1_80.
10. CHEN, L., LIU, Y., KONG, F. and HE, N. Acoustic Source Localization Based on Generalized Cross-Correlation Time-Delay Estimation. *Procedia Engineering; CEIS 2011*. 2011, t. 15 , 4912-4919. ISSN 1877-7058.
11. HOSSEINI, M.S., REZAIE, A. and ZANJIREH, Y. Time Difference of Arrival Estimation of Sound Source using Cross Correlation and Modified Maximum Likelihood Weighting Function. *Scientia Iranica*. 2017, t. 24, nr. 6 , 3268-3279. ISSN 2345-3605.
12. TAO, T., et al. *Sound Localization and Speech Enhancement Algorithm Based on Dual-Microphone*. [interactyvus]. 2022, [žiūrėta 2022-03-12]. ISBN 1424-8220. doi: 10.3390/s22030715. Pieiga per: <https://www.ncbi.nlm.nih.gov/pubmed/35161469>.
13. TELLAKULA, A.K. Acoustic Source Localization using Time Delay Estimation. Degree Thesis. Bangalore, India: Supercomputer Education and Research Centre Indian Institute of Science. 2007.

14. GOMBOTS, S., NOWAK, J. and KALTENBACHER, M. Sound Source Localization – State of the Art and New Inverse Scheme. *E & i Elektrotechnik Und Informationstechnik*. 2021, t. 138, nr. 3 , 229-243. ISSN 1613-7620.
15. BASAVA, S. *Microphone Array Wiener Beamforming with Modeling of SRP- PHAT for Speaker Localization*. [interactyvus]. 2012, [žiūrēta 2022-03-12]. Pieiga per: <http://bth.diva-portal.org/smash/get/diva2:833404/FULLTEXT01.pdf>
<http://urn.kb.se/resolve?urn=urn:nbn:se:bth-5988>.
16. RUIGROK, E.N., GIBBONS, S. and WAPENAAR, C.P.A. Cross-Correlation Beamforming. *Journal of Seismology*. 2017, t. 21, nr. 3 , 495-508. ISSN 1573-157X.
17. VALENCIA-PALMA, A. and CÓRDOVA-ESPARZA, D. *Sound Source Localization using Beamforming and its Representation in a Mixed Reality Embedded Device*. Pattern [interactyvus]. 2019, [žiūrēta 2022-03-12]. ISBN 9783-030210779. Pieiga per: https://link.springer.com/chapter/10.1007/978-3-030-21077-9_35.
18. Hoang Do and SILVERMAN, H.F. *SRP-PHAT Methods of Locating Simultaneous Multiple Talkers using a Frame of Microphone Array Data*. - 2010 IEEE International Conference on Acoustics, Speech and Signal Processing [interactyvus]. 2010, [žiūrēta 2022-03-12]. ISBN 1520-6149. doi: 10.1109/ICASSP.2010.5496133. Pieiga per: <https://ieeexplore.ieee.org/document/5496133>.
19. BADALI, A., VALIN, J.-., MICHAUD, F. and AARABI, P. *Evaluating Real-Time Audio Localization Algorithms for Artificial Audition in Robotics*. - 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems [interactyvus]. 2009, [žiūrēta 2022-03-12]. ISBN 2153-0866. doi: 10.1109/IROS.2009.5354308. Pieiga per: <https://ieeexplore.ieee.org/document/5354308>.
20. COBOS, M., MARTI, A. and LOPEZ, J.J. A Modified SRP-PHAT Functional for Robust Real-Time Sound Source Localization with Scalable Spatial Sampling. *IEEE Signal Processing Letters*. 2011, t. 18, nr. 1 , 71-74. ISSN 1558-2361.
21. ZHUO, D. and CAO, H. *Fast Sound Source Localization Based on SRP-PHAT using Density Peaks Clustering*. [interactyvus]. 2021, [žiūrēta 2022-03-12]. ISBN 2076-3417. doi: 10.3390/app11010445. Pieiga per: <https://search.proquest.com/docview/2559413962>.
22. SONG, T., et al. *A Sound Source Localization Algorithm using Microphone Array with Rigid Body*. 22nd International Congress on Acoustics;At: La Plate, Argentina [interactyvus]. 2016, [žiūrēta 2022-03-12]. Pieiga per: <http://www.ica2016.org.ar/ica2016proceedings/ica2016/ICA2016-0352.pdf>.
23. DEY, A.K. and SAHA, S. *Acoustic Beamforming : Design and Development of Steered Response Power with Phase Transformation (SRP-PHAT)*. [interactyvus]. 2011, [žiūrēta 2022-03-12]. Pieiga per: <http://urn.kb.se/resolve?urn=urn:nbn:se:bth-3226>.
24. Hoang Do, SILVERMAN, H.F. and Ying Yu. A Real-Time SRP-PHAT Source Location Implementation using Stochastic Region Contraction(SRC) on a Large-Aperture Microphone Array. [interactyvus]. 2007, [žiūrēta 2022-03-12]. ISBN 2379-190X. doi: 10.1109/ICASSP.2007.366631. Pieiga per: <https://ieeexplore.ieee.org/document/4217031>.
25. GRONDIN, F. and GLASS, J. SVD-PHAT: A Fast Sound Source Localization Method. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2019 , 4140-4144.

26. DIAZ-GUERRA, D., MIGUEL, A. and BELTRAN, J.R. Robust Sound Source Tracking using SRP-PHAT and 3D Convolutional Neural Networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2021, t. 29 , 300-311. ISSN 2329-9304.
27. YALTA, N., NAKADAI, K. and OGATA, T. Sound Source Localization using Deep Learning Models. *Journal of Robotics and Mechatronics*. 2017, t. 29, nr. 1 , 37-48. ISSN 0915-3942.
28. ADAVANNE, S., POLITIS, A., NIKUNEN, J. and VIRTANEN, T. Sound Event Localization and Detection of Overlapping Sources using Convolutional Recurrent Neural Networks. *IEEE/ Journal of Selected Topics in Signal Processing*. 2019, t. 13, nr. 1 , 34-48.
29. ADAVANNE, S., POLITIS, A. and VIRTANEN, T. *Localization, Detection and Tracking of Multiple Moving Sound Sources with a Convolutional Recurrent Neural Network*. [interactyvus]. 2019, [žiūrēta 2022-03-12]. doi: 10.48550/ARXIV.1904.12769. Pieiņa per: <https://arxiv.org/abs/1904.12769>.
30. ADAVANNE, S., POLITIS, A. and VIRTANEN, T. *Direction of Arrival Estimation for Multiple Sound Sources using Convolutional Recurrent Neural Network*. [interactyvus]. 2018, [žiūrēta 2022-03-12]. ISBN 2076-1465. doi: 10.23919/EUSIPCO.2018.8553182. Pieiņa per: <https://ieeexplore.ieee.org/document/8553182>.
31. TAN, T., LIN, Y., CHANG, Y. and ALKHALEEF AH, M. *Sound Source Localization using a Convolutional Neural Network and Regression Model*. [interactyvus]. 2021, [žiūrēta 2022-03-12]. ISBN 1424-8220. doi: 10.3390/s21238031. Pieiņa per: <https://search.proquest.com/docview/2608138153>.
32. LIU, N., CHEN, H., SONGGONG, K. and LI, Y. Deep Learning Assisted Sound Source Localization using Two Orthogonal First-Order Differential Microphone Arrays. *The Journal of the Acoustical Society of America*. 2021, t. 149, nr. 2 , 1069-1084. ISSN 0001-4966.
33. SAXENA, A. and NG, A.Y. *Learning Sound Location from a Single Microphone*. - 2009 IEEE International Conference on Robotics and Automation [interactyvus]. 2009, [žiūrēta 2022-03-12]. ISBN 1050-4729. doi: 10.1109/ROBOT.2009.5152861. Pieiņa per: <https://ieeexplore.ieee.org/document/5152861>.
34. RISAUD, M., et al. ScienceDirect. *Clinical Microbiology Newsletter*. 2018, t. 135, nr. 4 , 259-264. ISSN 0196-4399.
35. VERREYCKEN, E., et al. Bio-Acoustic Tracking and Localization using Heterogeneous, Scalable Microphone Arrays. *Communications Biology*. 2021, t. 4, nr. 1 , 1275. ISSN 2399-3642.
36. VERREYCKEN, E., DAEMS, W. and STECKEL, J. *Distributed Low-Cost Microphone Array for 3D Localization for Bio-Acoustic Applications*. [interactyvus]. 2017, [žiūrēta 2022-03-12]. ISBN 1930-0395. doi: 10.1109/ICSENS.2017.8234059. Pieiņa per: <https://ieeexplore.ieee.org/document/8234059>.
37. BRANDNER, M. and SONTACCHI, A. *3D Sound Source Localization using a Spherical Microphone Array*. [interactyvus]. January 1, 2014, [žiūrēta 2022-03-12]. Pieiņa per: https://www.researchgate.net/publication/326126874_3D_sound_source_localization_using_a_spherical_microphone_array.
38. HE, W., MOTLICEK, P. and ODOBEZ, J. *Deep Neural Networks for Multiple Speaker Detection and Localization*. [interactyvus]. 2018, [žiūrēta 2022-03-12]. ISBN 2577-087X. doi: 10.1109/ICRA.2018.8461267. Pieiņa per: <https://ieeexplore.ieee.org/document/8461267>.

39. VERA-DIAZ, J.M., PIZARRO, D. and MACIAS-GUARASA, J. Towards End-to-End Acoustic Localization using Deep Learning: From Audio Signal to Source Position Coordinates. *Sensors (Basel, Switzerland)*. 2018, t. 18, nr. 10 , 3418. ISSN 1424-8220.
40. *Sound Source Localization Systems | Siemens Software*. [žiūrēta 2022-02-20]. Prieiga per:<https://www.plm.automation.siemens.com/global/en/products/simulation-test/sound-source-localization-systems.html>.
41. ASTAPOV, S., BERDNIKOVA, J. and PREDEN, J. Optimized Acoustic Localization with SRP-PHAT for Monitoring in Distributed Sensor Networks. *International Journal of Electronics and Telecommunications*. 2015, t. 59, nr. 4 , 383-390.
42. VEITCH, D., RIDOUX, J. and KORADA, S.B. Robust Synchronization of Absolute and Difference Clocks Over Networks. *IEEE/ACM Transactions on Networking*. 2009, t. 17, nr. 2 , 417-430. ISSN 1558-2566.
43. RHEE, I., et al. *Clock Synchronization in Wireless Sensor Networks: An Overview*. [interactyvus]. 2009, [žiūrēta 2022-03-12]. ISBN 1424-8220. doi: 10.3390/s90100056. Prieiga per: <https://www.mdpi.com/1424-8220/9/1/56>.
44. MOCK, M., FRINGS, R., NETT, E. and TRIKALIOTIS, S. *Continuous Clock Synchronization in Wireless Real-Time Applications*. [interactyvus]. 2000, [žiūrēta 2022-03-12]. doi: 10.1109/RELDI.2000.885400. Prieiga per: <http://publica.fraunhofer.de/documents/2004925372.html>.
45. VERREYCKEN, W.E., LAURIJSSEN, D., DAEMS, W. and STECKEL, J. *Firefly Based Distributed Synchronization in Wireless Sensor Networks for Passive Acoustic Localization*. [interactyvus]. 2016, [žiūrēta 2022-03-12]. ISBN 2162-7347. doi: 10.1109/IPIN.2016.7743681. Prieiga per: <https://ieeexplore.ieee.org/document/7743681>.
46. PING, S. Delay Measurement Time Synchronization for Wireless Sensor Networks. *Intel Research Berkeley Lab*. 2003, t. 6 , 1-10.
47. LAURIJSSEN, D., et al. Low-Cost Synchronization of High-Speed Audio and Video Recordings in Bio-Acoustic Experiments. *The Journal of Experimental Biology*. 2017, t. 221, nr. Pt 4 , jeb173724. ISSN 0022-0949.
48. GRUMIAUX, P., KITIĆ, S., GIRIN, L. and GUÉRIN, A. *A Survey of Sound Source Localization with Deep Learning Methods*. [interactyvus]. 2021, [žiūrēta 2022-03-12]. doi: 10.48550/ARXIV.2109.03465. Prieiga per: <https://arxiv.org/abs/2109.03465>.
49. COBOS, M., MARTI, A. and LOPEZ, J.J. A Modified SRP-PHAT Functional for Robust Real-Time Sound Source Localization with Scalable Spatial Sampling. *IEEE Signal Processing Letters*. 2011, t. 18, nr. 1 , 71-74. ISSN 1558-2361.
50. YOOK, D., LEE, T. and CHO, Y. Fast Sound Source Localization using Two-Level Search Space Clustering. *IEEE Transactions on Cybernetics*. 2016, t. 46, nr. 1 , 20-26. ISSN 2168-2275.
51. CHO, Y., YOOK, D., CHANG, S. and KIM, H. Sound Source Localization for Robot Auditory Systems. *IEEE Transactions on Consumer Electronics*. 2009, t. 55, nr. 3 , 1663-1668. ISSN 1558-4127.
52. HAFEZI, S., MOORE, A.H. and NAYLOR, P.A. *3D Acoustic Source Localization in the Spherical Harmonic Domain Based on Optimized Grid Search*. [interactyvus]. 2016, [žiūrēta 2022-03-12]. ISBN 2379-190X. doi: 10.1109/ICASSP.2016.7471708. Prieiga per: <https://ieeexplore.ieee.org/document/7471708>.

53. Byoung-gi Lee and JongSuk Choi. *Multi-Source Sound Localization using the Competitive K-Means Clustering*. - 2010 IEEE 15th Conference on Emerging Technologies & Factory Automation (ETF A 2010) [interactyvus]. 2010, [žiūrēta 2022-03-12]. ISBN 1946-0740. doi: 10.1109/ETF A.2010.5641169. Pieiga per: <https://ieeexplore.ieee.org/document/5641169>.
54. OZANICH, E., GERSTOFT, P. and NIU, H. A Feedforward Neural Network for Direction-of-Arrival Estimation. *The Journal of the Acoustical Society of America*. 2020, t. 147, nr. 3 , 2035-2048. ISSN 0001-4966.
55. TAKEDA, R. and KOMATANI, K. *Sound Source Localization Based on Deep Neural Networks with Directional Activate Function Exploiting Phase Information*. - 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [interactyvus]. 2016, [žiūrēta 2022-03-12]. ISBN 2379-190X. doi: 10.1109/ICASSP.2016.7471706. Pieiga per: <https://ieeexplore.ieee.org/document/7471706>.
56. HIRVONEN, T. *Classification of Spatial Audio Location and Content using Convolutional Neural Networks*. [interactyvus]. May 6, 2015, [žiūrēta 2022-03-12]. Pieiga per: <http://www.aes.org/e-lib/inst/browse.cfm?elib=17718>.
57. JIN, L., et al. RNN for Solving Time-Variant Generalized Sylvester Equation with Applications to Robots and Acoustic Source Localization. *IEEE Transactions on Industrial Informatics*. 2020, t. 16, nr. 10 , 6359-6369. ISSN 1941-0050.
58. MURRAY, J.C., ERWIN, H.R. and WERMTER, S. Robotic Sound-Source Localisation Architecture using Cross-Correlation and Recurrent Neural Networks. *Neural Networks; what it Means to Communicate*. 2009, t. 22, nr. 2 , 173-189. ISSN 0893-6080.
59. AYUB, M.S., JIANFENG, C. and ZAMAN, A. Multiple Acoustic Source Localization using Deep Data Association. *Applied Acoustics*. 2022, t. 192 , 108731. ISSN 0003-682X.
60. QIN, D., TANG, J. and YAN, Z. *Underwater Acoustic Source Localization using LSTM Neural Network*. [interactyvus]. 2020, [žiūrēta 2022-03-12]. ISBN 1934-1768. doi: 10.23919/CCC50068.2020.9189504. Pieiga per: <https://ieeexplore.ieee.org/document/9189504>.
61. SUVOROV, D., DONG, G. and ZHUKOV, R. Deep Residual Network for Sound Source Localization in the Time Domain. 2018.
62. NARANJO-ALCAZAR, J., et al. *Sound Event Localization and Detection using Squeeze-Excitation Residual CNNs*. [interactyvus]. 2020, [žiūrēta 2022-03-12]. doi: 10.48550/ARXIV.2006.14436. Pieiga per: <https://arxiv.org/abs/2006.14436>.
63. SUDO, Y., ITOYAMA, K., NISHIDA, K. and NAKADAI, K. *Multi-Channel Environmental Sound Segmentation Utilizing Sound Source Localization and Separation U-Net*. [interactyvus]. 2021, [žiūrēta 2022-03-12]. ISBN 2474-2325. doi: 10.1109/IEEECONF49454.2021.9382730. Pieiga per: <https://ieeexplore.ieee.org/document/9382730>.
64. SUDARSANAM, P., POLITIS, A. and DROSSOS, K. *Assessment of Self-Attention on Learned Features for Sound Event Localization and Detection*. [interactyvus]. 2021, [žiūrēta 2022-03-12]. doi: 10.48550/ARXIV.2107.09388. Pieiga per: <https://arxiv.org/abs/2107.09388>.
65. SCHYMURA, C., et al. *PILOT: Introducing Transformers for Probabilistic Sound Event Localization*. [interactyvus]. Jun 7, 2021, [žiūrēta 2022-03-12]. Pieiga per: <https://arxiv.org/abs/2106.03903>.
66. BARNSTON, A.G. *Correspondence among the Correlation, RMSE, and Heidke Forecast Verification Measures; Refinement of the Heidke Score*. [interactyvus]. 1992, [žiūrēta 2022-03-

12]. doi: 10.1175/1520-0434(1992)0072.0.CO;2.
<https://ui.adsabs.harvard.edu/abs/1992WtFor...7..699B>.

Pieiga per: