



УНИВЕРЗИТЕТ У НОВОМ САДУ
ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА



Ненад Четић

**Модел бежичних акустичких сензора за
командовање гласом у паметним кућама**

– ДОКТОРСКА ДИСЕРТАЦИЈА –

Ментор:

проф. др Мирослав Поповић

Нови Сад, 2019

Велику захвалност дугујем свом ментору проф. др Мирославу Поповићу за велико стрпљење, мудрост и веру у мене.

Захваљујем се и свим члановима комисије на корисним саветима и помоћи.

Посебно се захваљујем колеги доц. др Миодрагу Ђукићу, за све корисне савете у току докторских студија.

Захваљујем се колегама Мирославу Стефановићу, Давору Додеру, Срђану Фурману, Александру Вуку Павловићу и Небојши Рајићу на колегијалности и сарадњи у току рада на овој дисертацији.

Захваљујем се доц. др Јелени Ковачевић и проф. др Николи Теслићу за велику подршку у оквиру колектива и рада у оквиру РТ-РК института.

Највећу захвалност дугујем мојој породици.



КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА

Редни број, РБР:	
Идентификациони број, ИБР:	
Тип документације, ТД:	Монографска документација
Тип записа, ТЗ:	Текстуални штампани материјал
Врста рада, ВР:	Докторски рад
Аутор, АУ:	Ненад Четић
Ментор, МН:	Проф. др Мирослав Поповић
Наслов рада, НР:	Модел бежичних акустичких сензора за командовање гласом у паметним кућама
Језик публикације, ЈП:	Српски
Језик извода, ЈИ:	Српски
Земља публикавања, ЗП:	Република Србија
Уже географско подручје, УГП:	Војводина
Година, ГО:	2019.
Издавач, ИЗ:	Ауторски репринт
Место и адреса, МА:	Нови Сад; Трг Доситеја Обрадовића 6
Физички опис рада, ФО: (поглавља/страна/ цитата/табела/слика/графика/прилога)	6 поглавља / 110 страна / 110 цитата / 31 табела / 35 слика / 3 прилога
Научна област, НО:	Електротехничко и рачунарско инжењерство
Научна дисциплина, НД:	Рачунарска техника и рачунарске комуникације
Предметна одредница/Кључне речи, ПО:	Бежичне сензорске мреже, реверберација, управљање гласом, аутоматско препознавање гласа, потрошачка електроника
УДК	
Чува се, ЧУ:	У библиотеци Факултета техничких наука, Нови Сад
Важна напомена, ВН:	
Извод, ИЗ:	Основни циљ истраживања у дисертацији је испитивање примене дистрибуираних акустичких сензора на проблем аутоматског препознавања говора. Услед реверберације говора у затвореном простору долази до проблема да снимљени сигнал садржи више одјека и шума него директног звука. Овај проблем се решава применом различитих алгоритама базираних на обради сигнала помоћу микрофонских низова и решетки. На основу аналитичког модела простора са реверберацијом реализован је систем дистрибуираних акустичких сензора за подршку говорне комуникације у паметним кућама. Добијени резултати указују да је једноставним дистрибуираним сензорима могуће остварити тачност препознавања говора попут тачности код комерцијалних система.
Датум прихватања теме, ДП:	22. децембар 2016
Датум одбране, ДО:	
Чланови комисије, КО:	Председник: Др Никола Теслић, ред. проф. Члан: Др Мило Томашевић, ред. проф. Члан: Др Владо Делић, ред. проф. Члан: Др Миодраг Ђукић, доцент Ментор: Др Мирослав Поповић, ред. проф.
	Потпис ментора



KEY WORDS DOCUMENTATION

Accession number, ANO :	
Identification number, INO :	
Document type, DT :	Monographic publication
Type of record, TR :	Textual printed material
Contents code, CC :	PhD Thesis
Author, AU :	Nenad Cetic
Mentor, MN :	Miroslav Popovic, PhD
Title, TI :	Model of wireless acoustic sensors for voice commands in smart homes
Language of text, LT :	Serbian
Language of abstract, LA :	Serbian
Country of publication, CP :	Republic of Serbia
Locality of publication, LP :	Vojvodina
Publication year, PY :	2019.
Publisher, PB :	Author's reprint
Publication place, PP :	Novi Sad, Dositeja Obradovica sq. 6
Physical description, PD : (chapters/pages/ref./tables/pictures/graphs/appendixes)	6 chapters / 110 pages/ 110 references / 31 tables / 35 pictures / 3 appendix
Scientific field, SF :	Electrical Engineering
Scientific discipline, SD :	Computer Engineering, Engineering of Computer Based Systems
Subject/Key words, S/KW :	Wireless sensor networks, reverberation, voice commands, automatic speech recognition, consumer electronics
UC	
Holding data, HD :	The Library of Faculty of Technical Sciences, Novi Sad, Serbia
Note, N :	
Abstract, AB :	<p>The main goal of the dissertation is to examine the application of distributed acoustic sensors to the problem of automatic speech recognition. Due to the reverberation of indoor speech, there is a problem that the recorded signal contains more echoes and noise than direct sound. This problem is solved by applying different algorithms based on signal from microphone arrays and microphone grids. Based on the analytical model of the reverberated space, a system of distributed acoustic sensors was realized to support voice communication in smart home environment. Results show that simple distributed sensors can achieve recognition performance similar to those found in state-of-the-art systems available on the market.</p>
Accepted by the Scientific Board on, ASB :	22. December 2016
Defended on, DE :	
Defended Board, DB :	
President:	Prof. Nikola Teslić, PhD
Member:	Prof. Milo Tomašević, PhD
Member:	Prof. Vlado Delić, PhD
Member:	Assistant prof. Miodrag Đukić, PhD
Mentor:	Prof. Miroslav Popović, PhD
	Mentor's sign

Садржај

ПОГЛАВЉЕ 1. УВОД	7
1.1 Предмет истраживања	8
1.2 Циљ истраживања	9
1.3 Научни допринос	9
1.4 Организација дисертације	10
ПОГЛАВЉЕ 2. СТАЊЕ У ОБЛАСТИ	12
2.1 Интернет технологије	14
2.1.1 Пренос гласа путем интернета	16
2.2 Појам паметне куће	17
2.3 Акустика у архитектури	19
2.3.1 Акустички параметри	22
2.3.2 Технике моделовања акустике у затвореном простору	24
2.4 Говор као спрега	27
2.4.1 Трендови у развоју и примени аутоматског препознавања говора	31
2.4.2 Командовање гласом у потрошачкој електроници	33
2.5 Бежични акустички сензори	36
2.5.1 Проблем синхронизације	37
2.5.2 Бежичне акустичке сензорске мреже	41
2.5.3 Модели мрежног саобраћаја	43
ПОГЛАВЉЕ 3. МОДЕЛ БЕЖИЧНИХ АКУСТИЧКИХ СЕНЗОРА	45

3.1	Модел простора са реверберацијом	46
3.2	Модел бежичног акустичког сензора	49
3.3	Модел акустичких сензора за командовање гласом	52
ПОГЛАВЉЕ 4. ЕКСПЕРИМЕНТАЛНИ РЕЗУЛТАТИ И ДИСКУСИЈА		54
4.1	Провера модела изворишног саобраћаја	55
4.2	Провера модела мреже бежичних акустичких сензора	58
4.3	Командовање гласом - студија случаја	64
ПОГЛАВЉЕ 5. ПОВЕЗАНА ИСТРАЖИВАЊА И РЕЗУЛТАТИ		67
5.1	Командовање гласом код Андроид апликација	68
5.2	Реализација система дистрибуираних акустичких сензора	72
5.3	Реализација говорне спреге за уређаје са ограниченим ресурсима	75
5.4	Реализација јефтиног бежичног акустичког сензора	78
5.5	Систем за испитивање гласовних команди у паметним кућама	81
ПОГЛАВЉЕ 6. ЗАКЉУЧАК		84
ЛИТЕРАТУРА		86
ПРИЛОГ А - ГРЕШКЕ ПРИ АУТОМАТСКОМ ПРЕПОЗНАВАЊУ ГОВОРА БЕЗ ИЗОБЛИЧЕЊА		98
ПРИЛОГ Б - МЕРЕЊЕ ВРЕМЕНА РЕВЕРБЕРАЦИЈЕ		100
ПРИЛОГ В - ИСПИТНИ СЛУЧАЈЕВИ		103

СПИСАК СЛИКА

ИЛУСТРАЦИЈА 1.1 ТИПИЧНА ГАРТНЕРОВА КРИВА ОЧЕКИВАЊА.....	14
ИЛУСТРАЦИЈА 2.3.1 ЈАЧИНА ЗВУКА У ПРОСТОРИЈИ.....	21
СЛИКА 2.3.2 ПРИМЕР ОМНИДИРЕКЦИОНОГ ИЗВОРА ЗВУКА.....	22
ИЛУСТРАЦИЈА 2.3.3 МЕРЕЊЕ ВРЕМЕНА РЕВЕРБЕРАЦИЈЕ - RT60	23
СЛИКА 2.4.1 ПРЕГЛЕД ГОВОРНИХ ТЕХНОЛОГИЈА (ПРЕУЗЕТО ИЗ [DELIC2]).....	28
СЛИКА 2.4.2 ПРИМЕР КОРЕКЦИЈЕ ТЕКСТА КОД АНДРОИД ПАМЕТНОГ ТЕЛЕФОНА.....	30
ИЛУСТРАЦИЈА 2.4.3 ПРИКАЗ МИКРОФОНСКЕ РЕШЕТКЕ.....	31
ИЛУСТРАЦИЈА 2.5.1 ОСНОВНИ ПРОБЛЕМ СИНХРОНИЗАЦИЈЕ ДИСТРИБУИРАНИХ ЧВОРОВА	37
ИЛУСТРАЦИЈА 2.5.2 ПРИМЕР СИНХРОНИЗАЦИЈЕ ПОД МРЕЖЕ.....	38
ИЛУСТРАЦИЈА 2.5.3 МЕРЕЊЕ И СИНХРОНИЗАЦИЈА ВРЕМЕНА У RTP	40
ИЛУСТРАЦИЈА 2.5.4 ДИСТРИБУИРАНИ АКУСТИЧКИ СЕНЗОРИ.....	42
ИЛУСТРАЦИЈА 3.1 ПРОЦЕС ИЗГРАДЊЕ МОДЕЛА БЕЖИЧНИХ АКУСТИЧКИХ СЕНЗОРА.....	45
ИЛУСТРАЦИЈА 3.1.1 ПРОСТИРАЊА ЗВУКА У ПРОСТОРУ БЕЗ РЕВЕРБЕРАЦИЈЕ	48
ИЛУСТРАЦИЈА 3.1.2 ПРОСТИРАЊЕ ЗВУКА У ПРОСТОРУ СА РЕВЕРБЕРАЦИЈОМ	48
ИЛУСТРАЦИЈА 3.1.3 МОДЕЛ ПРЕСЛИКАВАЊА ИЗВОРА ЗА НИЖЕ ФРЕКВЕНЦИЈЕ (125 HZ И 128 HZ) 49	
ИЛУСТРАЦИЈА 3.2.1 ПРИМЕР ИЗВОРИШНОГ САОБРАЋАЈА КОД МОДЕЛА WASN.....	51
СЛИКА 3.3.1 РАД ПОВЕЗАНИХ АКУСТИЧКИХ СЕНЗОРА	52
ИЛУСТРАЦИЈА 4.1 СПЕКТАР ИЗГОВОРЕНЕ КОМАНДЕ	55
СЛИКА 4.1.1. УТИЦАЈ РЕВЕРБЕРАЦИЈЕ НА ИЗВОРИШНИ САОБРАЋАЈ	57
ИЛУСТРАЦИЈА 4.2.1 ПРОСТОР У КОЈЕМ СУ СЕ ОДВИЈАЛИ ЕКСПЕРИМЕНТИ И МЕРЕЊА.....	59
СЛИКА 4.2.2 ФОТОГРАФИЈЕ ПРОСТОРА СА РЕВЕРБЕРАЦИЈОМ	60
СЛИКА 4.2.3 ПОЗИЦИЈЕ ГОВОРНИКА И СЕНЗОРА У СИМУЛАЦИЈИ	62
СЛИКА 4.2.3 ОДРЕЂИВАЊА БРОЈА ПОТРЕБНИХ МИКРОФОНА	63
ИЛУСТРАЦИЈА 5.1.1 ПРИМЕРИ АНДРОИД АПЛИКАЦИЈА	69
СЛИКА 5.2.1 ПРИМЕР ТОПОЛОГИЈЕ ЗВЕЗДЕ	73
СЛИКА 5.2.2. АРХИТЕКТУРА СИСТЕМА	74
СЛИКА 5.3.1 SMARTMP3 ПЛОЧА СА ИНТЕГРИСАНИМ ЗВУЧНИКОМ	76
СЛИКА 5.3.2 КОНЦЕПТ УРЕЂАЈА	76
СЛИКА 5.4.1 ПРОТОТИПОВИ ЈЕФТИНОГ АКУСТИЧНОГ СЕНЗОРА	80
СЛИКА 5.5.1 БЕЖИЧНИ СЕНЗОР ЗА ИСПИТИВАЊЕ ГЛАСОВНИХ КОМАНДИ	83
ИЛУСТРАЦИЈА В.1 РАСПОРЕД ГОВОРНИКА И СЕНЗОРА У ПРОСТОРИЈИ 1.....	104
ИЛУСТРАЦИЈА В.2 РАСПОРЕД ГОВОРНИКА И СЕНЗОРА У ПРОСТОРИЈИ 2.....	105
ИЛУСТРАЦИЈА В.3 РАСПОРЕД ГОВОРНИКА И СЕНЗОРА У ПРОСТОРИЈИ 3.....	105
ИЛУСТРАЦИЈА В.4 РАСПОРЕД ГОВОРНИКА И СЕНЗОРА У ПРОСТОРИЈИ 4.....	106
ИЛУСТРАЦИЈА В.5 РАСПОРЕД ГОВОРНИКА И СЕНЗОРА У ПРОСТОРИЈИ 5.....	106

СПИСАК ТАБЕЛА

ТАБЕЛА 3.1.1 ПАРАМЕТРИ СИМУЛАЦИЈЕ	47
ТАБЕЛА 4.1.1 ПАРАМЕТРИ ЕКСПЕРИМЕНТА ЗА ИСПИТИВАЊЕ ИЗВОРИШНОГ САОБРАЋАЈА	56
ТАБЕЛА 4.2.1 ВРЕМЕ РЕВЕРБЕРАЦИЈЕ У ПРОСТОРИЈАМА.....	60
ТАБЕЛА 4.2.2 ПАРАМЕТРИ СИМУЛАЦИЈЕ У ПРОСТОРИЈИ 1	61
ТАБЕЛА 4.2.3 ГРЕШКА ОДАБИРА МИКРОФОНА НА ОСНОВУ SNR	62
ТАБЕЛА 4.2.4 ГРЕШКЕ ПРЕПОЗНАВАЊА ГОВОРА У ИЗВЕДЕНИМ ЕКСПЕРИМЕНТИМА	64
ТАБЕЛА 4.3.1 ОДНОС ИСПИТНИХ СЛУЧАЈЕВА.....	65
ТАБЕЛА 4.3.2 ГРЕШКЕ ПРЕПОЗНАВАЊА ГОВОРА КОД СТУДИЈЕ СЛУЧАЈА	66
ТАБЕЛА 5.1.1 ЛЕВЕНШРАЈНОВО РАСТОЈАЊА КОД АПЛИКАЦИЈЕ ДАЉИНСКОГ УПРАВЉАЧА.....	70
ТАБЕЛА 5.1.2 ЛЕВЕНШРАЈНОВО РАСТОЈАЊЕ КОД АПЛИКАЦИЈЕ ЛОГИЧКЕ ИГРЕ	70
ТАБЕЛА 5.1.3 LSC РАСТОЈАЊА АПЛИКАЦИЈЕ ДАЉИНСКОГ УПРАВЉАЧА	71
ТАБЕЛА 5.1.4 LSC РАСТОЈАЊА КОД АПЛИКАЦИЈЕ ЛОГИЧКЕ ИГРЕ	71
ТАБЕЛА 5.1.5 ГРЕШКЕ ПРЕПОЗНАТИХ КОМАНДИ КОД АПЛИКАЦИЈЕ ДАЉИНСКОГ УПРАВЉАЧА	71
ТАБЕЛА 5.1.6 ГРЕШКЕ ПРЕПОЗНАТИХ КОМАНДИ КОД АПЛИКАЦИЈЕ ЛОГИЧКЕ ИГРЕ	71
ТАБЕЛА А.1 WER ГОВОРА БЕЗ ИЗОБЛИЧЕЊА	98
ТАБЕЛА А.1 WER ГОВОРА БЕЗ ИЗОБЛИЧЕЊА - НАСТАВАК	99
ТАБЕЛА Б.1 ВРЕМЕ РЕВЕРБЕРАЦИЈЕ У ПРОСТОРИЈИ 1	100
ТАБЕЛА Б.2 ВРЕМЕ РЕВЕРБЕРАЦИЈЕ У ПРОСТОРИЈИ 2	101
ТАБЕЛА Б.3 ВРЕМЕ РЕВЕРБЕРАЦИЈЕ У ПРОСТОРИЈИ 3	101
ТАБЕЛА Б.4 ВРЕМЕ РЕВЕРБЕРАЦИЈЕ У ПРОСТОРИЈИ 4	101
ТАБЕЛА Б.5 ВРЕМЕ РЕВЕРБЕРАЦИЈЕ У ПРОСТОРИЈИ 5	102
ТАБЕЛА В.1 ПАРАМЕТРИ СИМУЛАЦИЈЕ У ПРОСТОРИЈИ 1	107
ТАБЕЛА В.2 ПАРАМЕТРИ СИМУЛАЦИЈЕ У ПРОСТОРИЈИ 2	107
ТАБЕЛА В.3 ПАРАМЕТРИ СИМУЛАЦИЈЕ У ПРОСТОРИЈИ 3	108
ТАБЕЛА В.4 ПАРАМЕТРИ СИМУЛАЦИЈЕ У ПРОСТОРИЈИ 4	108
ТАБЕЛА В.5 ПАРАМЕТРИ СИМУЛАЦИЈЕ У ПРОСТОРИЈИ 5	109
ТАБЕЛА В.6 ГРЕШКЕ ПРЕПОЗНАВАЊА У ПРОСТОРИЈИ 1.....	109
ТАБЕЛА В.7 ГРЕШКЕ ПРЕПОЗНАВАЊА У ПРОСТОРИЈИ 2.....	109
ТАБЕЛА В.8 ГРЕШКЕ ПРЕПОЗНАВАЊА У ПРОСТОРИЈИ 3.....	110
ТАБЕЛА В.9 ГРЕШКЕ ПРЕПОЗНАВАЊА У ПРОСТОРИЈИ 4.....	110
ТАБЕЛА В.10 ГРЕШКЕ ПРЕПОЗНАВАЊА У ПРОСТОРИЈИ 5.....	110

СКРАЋЕНИЦЕ

ADC	Analog to Digital Converter
ADT	Android Development Tools
AEC	Acoustic Echo Canceller
AEC	Acoustic Echo Canceling
AIR	Aachen Impulse Response
ASR	Automatic Speech Recognition
AVS	Alexa Voice Service
DC	Direct Current
DFT	Discrete Furies Transformation
DNN	Deep Neural Network
DOA	Direction Of Arrival
DSP	Digital Signal Processing
DSP	Digital Signal Processor/Processing
FSM	Finite State Machine
GPIO	General Purpose Input/Output
HTTP	Hypertext Transfer Protocol
ID	Identification
IoT	Internet of Things
IRIG	Inter-Range Instrumentation Group
ISM	Image Source Modeling
ISM	Industrial Scientific and Medical
JDK	Java Development Kit
JSON	Java Script Object Notation
LCS	Longest common subsequence
MQTT	Message Queuing Telemetry Transport
NTP	Network Time Protocol
PCM	Pulse Code Modulation
PRC	Remote procedure call
PTP	Precision Time Protocol
PWM	Pulse Width Modulation
RIR	Room Impulse Response
RMS	Root Mean Square
RTP	Real-time Transport Protocol
RTSP	Real-time Streaming Protocol
RTT	Round Trip Time
SASwNG	Simple Acoustic Sensor with the Noise Gate
SIP	Session Initiation Protocol
SNR	Signal to Noise Ratio
SNTP	Simple Network Time Protocol
STT	Speech To Text
UDP	User Datagram Protocol
URL	Uniform resource locator
VAD	Voice Activity Detection
VAD	Voice Activity Detector
VI	Voice Interface

WAN	Wide Area Networks
WASN	Wireless Acoustic Sensor Network
WER	Word Error Rate
WLAN	Wireless Local Area Network
WMSN	Wireless Multimedia Sensor Networks
WSN	Wireless sensor network
TTS	Text To Speech

ПОГЛАВЉЕ 1.

Увод

Тренд који видимо у области потрошачке електронике је примена све више сензора у животном простору људи. Овакво окружење се често назива и „паметна кућа“. Оно представља сложен систем сачињен од мноштва сензора, актуатора и процесних јединица за скупљање података, њихову анализу и управљање електричним апаратима и уређајима. Употреба мноштва сензора има за циљ стварање што детаљније слике о животном простору. Чест циљ оваквог система је остварење боље енергетске ефикасности и унапређење квалитета живота. Овакви системи су засновани на уређајима са ограниченим ресурсима због своје ниске цене и мале потрошње енергије. Повезивање једноставних акустичких сензора мале процесне моћи бежичним путем носи са собом низ проблема. Пренос података путем конвенционалне интернет мреже не гарантује минимално кашњење и не штити од губитка пакета.

Људски говор је специфичан вид комуникације који одваја људе од осталих живих бића. Рачунарски склоп који има за циљ да разуме говор мора покрити све горе наведене процесе и решити сложен технички проблем. Процес препознавања говора почиње дигитализацијом говорног сигнала превођењем у његов временско дискретни квантизовани облик. На тачност препознавања говора у многоме утиче квалитет звучног записа, тачније, однос корисног сигнала и шума,

односно сметњи. У затвореном простору какав је животни простор људи тачност препознавања говора зависи од:

- акустичких параметара простора,
- положаја говорника
- положаја сензора.

У неким ситуацијама положај акустичких сензора у односу на говорника може резултовати неповољним односом сигнал-шум. У тим ситуацијама говоримо о акустичком проблему далеког дифузног поља (енгл. *far field problem*). Класичан приступ решавању овог проблема заснива се на алгоритмима обраде сигнала и примени микрофонских низова и решетки. Идеја предложена у овој дисертацији је решавање овог проблема дистрибуираним акустичким сензорима.

1.1 Предмет истраживања

Предмет истраживања је постизање интеракције између човека и рачунара ради командовања гласом у оквиру паметне куће. Спрега човека и рачунара се остварује помоћу сензора и сензорских мрежа (енгл. *Wireless sensor network - WSN*). Тачније, ради се о акустичким сензорима, па се ужа област истраживања ове докторске дисертације односи на мреже акустичких бежичних сензора (енгл. *Wireless acoustic sensor network - WASN*). Фокус ове докторске дисертације су акустички сензори повезани у систем за управљање, односно командовање гласом. Сензори могу бити самостални бежични микрофонски уређаји направљени искључиво за ову намену. Додатно, електрични апарати у кући могу бити опремљени сличним електронским склопом. Тако опремљени могу се користити као део система. Више акустичких сензора се повезује бежичним путем. У оквиру ове докторске дисертације дат је детаљан опис ограничења и захтева у погледу сложености, квалитета и могућности оваквог приступа.

Предложени систем бежичних акустичких сензора се може користити као систем за контролу гласом у окружењу паметних кућа. У овој дисертацији је сагледано тренутно стање и дата је анализа предности и мана предложеног приступа. Данас, велики број комерцијалних решења доносе говорне технологије уз помоћ уређаја повезаних на удаљене рачунаре. Због велике сложености сама област препознавања и синтезе говора, иако у константном развоју, и даље садржи

велики скуп проблема и интригира научну јавност. Изучавање природних језика је шира област од саме анализе и синтезе говора. Холистички, ово се мора употпунити свим оним што представљају говорне технологије. Ипак, главни фокус ове дисертације је сама анализа говора у затвореном простору какав је животни простор људи. Прецизније, фокус је моделовање система за прикупљање односно складиштење људског говора, његово бежично преношење и аутоматско препознавање. У том циљу било је потребно сагледати више аспеката проблема и развити модел који адресира те аспекте. Исто тако, примену предложеног модела је било потребно потврдити у различитим ситуацијама кроз испитивање предложеног решења у различитим акустичким амбијентима.

1.2 Циљ истраживања

Циљ истраживања је потврда идеје о примени дистрибуираних акустичких сензора на проблем аутоматског препознавања говора услед реверберације у затвореном простору.

Хипотеза ове дисертације је да је на основу модела пресликавања извора могуће реализовати симулационо окружење простора са реверберацијом, које довољно добро описује реалан простор, и реализовати систем једноставних дистрибуираних акустичких сензора за подршку говорне комуникације човек-машина у паметним кућама чија перформанса је упоредива са решењима на бази микрофонских низова и решетки.

1.3 Научни допринос

У раду је коришћен аналитички прорачун, симулација и експериментални метод. Након анализе проблема направљен је модел који описује акустички простор са реверберацијом. На њега је надовезан модел сензора који је у складу са анализом доступних техничких решења и технологије уграђених уређаја (енгл. *embedded device*). Уграђени уређаји су рачунарски склопови са ограниченим ресурсима. На основу ова два модела добијен је модел изворишног саобраћаја бежичних акустичких сензора у простору са реверберацијом [Cetic1]. Повезивањем више микрофона добијен је систем микрофона за командовање гласом. Коначни модел је затим коришћен као основа за истраживање система за командовање. Преглед додатних повезаних истраживања [Cetic3], [Stefanovic],

[Doder1], [Doder2], [Rajić] је дат у виду доприноса до којих се дошло радом на овој дисертацији.

1.4 Организација дисертације

Дисертација је организована у седам поглавља и садржи три прилога.

У уводном поглављу дат је кратак увод у проблем који се решава. Представљен је предмет истраживања, циљ истраживања и кратко је описан постигнут научни допринос.

Након тога, дат је преглед стања у области који садржи кратак преглед релевантних информација о Интернет технологијама са акцентом на пренос гласа путем интернета. Уведен је појам паметне куће као окружење којем је предложено решење намењено. Представљен је проблем акустике затворених простора односно дат је преглед стања у области истраживања акустике у архитектури. Затим је дат преглед техника моделовања акустике у затвореном простору и детаљније дат коришћени модел пресликавања извора. Дат је преглед трендова у развоју и примени техника аутоматског препознавања говора. На крају је дат преглед стања у области бежичних сензора са фокусом на бежичне акустичке сензоре.

У трећем поглављу је дат приказ развијеног модела акустичких сензора за командовање у паметним кућама. Дати су детаљи модела простора са реверберацијом и модел једног решења бежичног акустичког сензора. Затим је описано на који начин су бежични сензори повезани у систем за ефикасније прибављање говора у простору са реверберацијом. На крају поглавља је дат опис система за командовање гласом.

Четврто поглавље садржи приказ изведених експеримената и дискусију на основу добијених резултата. Урађена је провера модела изворишног саобраћаја са сензором постављеним у просторију са реверберацијом. Након тога приступљено је испитивању система бежичних акустичких сензора у приватној кући. Извршена су мерења времена реверберације у просторијама. Испитан је систем за командовање гласом односно успешност аутоматског препознавања гласа и дато је поређење са референтним системом.

Пето поглавље даје преглед повезаних истраживања на којима је аутор радио у току израде дисертације. Ово поглавље служи да употпуни разумевање неких одлука и ставова које су дате у поглављу три и четири.

Шесто поглавље даје преглед постигнутих доприноса и закључак. Такође приказана је кратка дискусија о предностима и недостацима развијеног система и дат је преглед даљих праваца истраживања.

Након тога дат је приказ коришћене литературе.

На крају је дат преглед мерених величина и додатних резултата у изведеним експериментима у виду три прилога. Прилози су издвојени као посебне целине ради читљивости и прегледности четвртог поглавља.

ПОГЛАВЉЕ 2.

СТАЊЕ У ОБЛАСТИ

Аутоматско препознавање говора (енгл. *Automatic Speech Recognition – ASR*) је веома широка област истраживања која је и даље актуелна са низом изазова укључујући и језичку зависност [Delic1]. Почети развоја везују се за Белове лабораторије (енгл. *Bell Laboratories*) и њихов предлог модела анализе и синтезе говора. Нешто касније, 1952. године научници су реализовали систем за препознавање изговорених бројева једног говорника [Denes] уз помоћ измерених фреквенција при изговору самогласника [Biing-Hwang].

Многи аутори сматрају да су клауд технологије (енгл. *Cloud technologies*) једне од кључних технологија за даљи развој пословних могућности и остварење зараде [Luftman]. Напредак потрошачке електронике у виду револуције код развоја паметних телефона, већа процесна моћ уграђених рачунара и стални напредак клауд базираних платформи и сервиса омогућио је да се говорне технологије базирају на нове методе машинског учења [Delic2]. Ово ствара нове могућности за другачију примену говорних технологија и доводи до тога да све више људи прихвата говорне технологије и примењује их у свом свакодневном животу [Kinsella2019]. У исто време корисници се ослобађају страха од повреде приватности услед система који стално ослушкују и почињу све више користити удаљене рачунарске сервисе занемарујући нарушавање приватности, сигурност личних података и безбедности на интернету. Комерцијална решења нуде конвенционалне видове заштите као што су шифровање самих мрежних порука

и класичне видове аутентификације. Пошто идентитет говорника није могуће поуздано препознати, примена оваквих система могућа је само код управљања активностима са малим ризиком. Следећи велики корак у погледу сигурности биће аутентификација корисника на основу биометријских обележја у изговореној фрази.

Очекује се да се све више људи окрене персоналним асистентима и интегрише их у своју свакодневницу. Како технологија напредује могућности асистената ће бити све већа. Моћнија интеграција између паметног телефона и доступних сервиса доводи до тога да се корисници ослањају на услуге асистента. Основна замисао је да се комуникација одвија путем природног језика налик разговору са правом особом. Асистент прима команде, пружа услуге, информације и обавља за човека досадне и заморне послове. Поред тога служи и као спрега са остатком система односно као посредник при контроли других рачунарских система. Комуникација гласом је природни начин комуникације са оваквим асистентом. Што се корисник више ослони на овакву интеракцију, његова очекивања доступности асистента су већа. Циљ је да се омогући удобна вербална комуникација која се одваја са било које позиције унутар затвореног озвученог простора попут паметне куће уз нормални ниво људског говора, без напрезања. Због природе простирања звука у затвореном простору ово је сложен акустички проблем. Потребно је обезбедити оптималну покривеност акустичким сензорима уз минималан утрошак материјала и енергије.

Да би смо овај проблем потпуно разумели потребно је разумети и контекст у ком је настала ова докторска дисертација која се простира кроз неколико области. Уређаји којима се остварује гласовна спрега су уређаји потрошачке електронике. Већина ставова долазе из рада у индустрији потрошачке електронике и сталног праћења њеног развоја. Идеја је да се акустички проблем реши постављањем јефтиних акустичних сензора и њихово повезивање у систем за препознавање команди. Ово доводи до формирања идеје о бежично повезаним микрофонима и њиховој могућој примени. У наставку је дат преглед стања области интернет технологија, уводи се појам паметне куће, даје преглед стања области акустике затворених простора попут животног простора људи, преглед

стања у области говорних технологија и на крају преглед области бежичних акустичких сензора којим се решава проблем реверберације.

2.1 Интернет технологије

Интернет технологије су саставни део и покретач модерног друштва. Могућност примене интернета потиче од његове децентрализоване архитектуре. Сама појава интернета се везује за решавање проблема међусобног повезивања великих рачунарских система распоређених на раздвојеним географским локацијама. Интернет је заправо скуп технологија за комуникацију различитих уређаја заснованих на рачунару. Глобалне комуникације на бази интернета довеле су до тога да данас говоримо о концептима као што су интернет ствари – IoT (енгл. *Internet of Things*) или интернету свега (енгл. *Internet of everything*). Ово свакако представља тренутни врхунац интернет технологије и његове примене где последњи извештаји говоре о 31×10^9 повезаних уређаја до краја 2018. године [Markit]. Са друге стране консултантска кућа Гартнер сматра да се IoT технологија тренутно налази близу такозваног врхунца инфлације очекивања [Gartner] и да се полако спушта у долину истине услед отрежњења као што је приказано на Илустрација 1 илустрацији 1.1.



Илустрација 1.1 Типична Гартнерова крива очекивања

Главна парадигма иза IoT јесте пружање услуга уз помоћ великог броја дистрибуираних уређаја са ограниченим ресурсима међусобно повезаних уз помоћ конвенционалне интернет глобалне мреже. Често се овакви уређаји називају и паметни уређаји (енгл. *Smart devices*) али у пракси то су често веома једноставни уређаји чија памет долази од саме повезаности у већи систем. Због различитих критеријума оптимизације попут економских или енергетских, овакви уређаји се физички ограничавају у погледу процесорске моћи, доступне меморије и могућности спрежних система. Електронски уређаји често бивају само мали додатак на постојећи физички предмет или електрични апарат те се на тај начин остварује праћење и управљање стварима на вишем нивоу апстракције. Ствари на овај начин добијају дигиталну репрезентацију и њима се манипулише помоћу дистрибуираног система за управљање. Иако није искључиво, често се овакви уређаји повезују на локални уређај-капију илити гејтвеј (енгл. *gateway*) који служи у повезивању уређаја у локалну мрежу, и њихово повезивање на глобалну мрежу али и као конвертер протокола. Локални уређаји не морају бити опремљени са подршком за интернет протокол. Они често користе неке од техника и протокола за бежичну комуникацију на малим растојањима [Gill].

Уграђени рачунари су темељ савремене рачунарске технологије. Сваки модерни електрични апарат има неку процесну јединицу, па је то систем заснован на рачунару. Рачунарска технологија прелази са уграђених уређаја на раширено или свеприсутно рачунање што значи да имамо све више рачунара који су уграђени свуда и у свему. Са напретком интернета у погледу покривености и брзине, такви уређаји често имају могућности повезивања и формирају оно што се данас популарно назива интернет ствари (IoT).

IoT парадигма претпоставља многе уређаје повезане преко конвенционалне интернет мреже. Овакви уређаји обично имају ограничене ресурсе, тако да се прибегава премештању дела имплементације у клауд инфраструктуру. С друге стране, повезивање са многим уређајима може бити веома захтевно. За употребу IoT у свакодневном животу, уређаји морају поштовати принципе ергономије. У овој дисертацији се испитује потенцијал људског гласа као спреге за један или више уређаја у IoT екосистему. Оваква спрега се користи за издавање команди и пријем информација путем гласовних порука.

Један од протокола који покушава да одговори на проблем ограничених ресурса је 6LoWPAN. То је протокол заснован на компресији заглавља интернет протокола верзије 6. Овај протокол користи компресију без памћења (енгл. *Stateless*). Протокол је пажљиво пројектован за уређаје са малом процесном моћи и веома ограниченим меморијским ресурсима. Ипак, пружа спрегу класичне утичнице (енгл. *Socket*) која повезује било које две тачке, чиме се уклања ограничење у погледу подржане топологије. На физичком нивоу се ослања на 802.15.4 протокол. Велики проблем губитка пакета и ре-трансмисије се овде ублажава максималном величином пакета од само 127 октета.

Веома важне особине IoT система су:

- Скалабилност – могућност лаког проширења уз очување својстава које су присутне када је систем мали.
- Поузданост – огледа се у отпорности система на промене комуникационог медијума или промене других физичких услова па и заузетости односно количине саобраћаја.
- Сигурност – отпорност на малициозне утицаје у погледу неовлашћеног приступа или ометања комуникације. Постиге се шифровањем порука и садржаја меморије на уређају али не треба занемарити ни физичко обезбеђење уређаја и приступних тачака.
- Приватност – дистрибуираност ових система ствара услове за повреду приватности услед продирања кроз сигурносне мере. Овде је важно заштитити личне податке корисника IoT система од стране нежељеног приступа. Често се ради о техникама којима се постиже анонимност и гарантује се правним актима потписаним од стране корисника.

2.1.1 Пренос гласа путем интернета

Пренос гласа путем интернета је подржан мрежним протоколима. Имплементација виртуалних позивних центара је базирана на TCP и интернет протоколу верзије 4 [Porovic]. Један од најважнијих протокола је протокол успоставе сесије (енгл. *Session Initiation Protocol - SIP*). Овај протокол регулише многе важне аспекте код преноса гласа преко интернет протокола [Basicevic1]. За

сигурност испоруке пакета се доста ослања на механизме познате из HTTP протокола попут аутентификације и шифровања. Говор се преноси компримовано уз помоћ кодерско декодерских система или краће кодека (енгл. *CODEC*). Компаративну анализу перформанси различитих кодека код преноса гласа путем интернета дат је у [Војовић].

Природа SIP протокола је погодна за успоставу сесије произвољног типа. У раду [Basicевић2] је приказана примена SIP протокола у контексту архитектуре за извршавање скупа задатака (енгл. *bag-of-tasks*). SIP протокол је употребљен за реализацију извршног управљачког сервиса. SIP заступник са чувањем стања (енгл. *Stateful proxy*) се користи као посредник захтева извршења задатака. Овај рад даје опис концепта и прототип система који је изграђен, као и прорачун процењеног нивоа перформанси и њен однос према максималном времену пута (енгл. *Roung trip time – RTT*) код мрежног система. Основна предност ове једноставне мрежне архитектуре је поновна употреба инфраструктуре. Дат је и кратак преглед неких приступа и математичких модела рачунарских мрежа.

Поред SIP протокола за сам пренос мрежних пакета може се користити протокол за транспорт у реалном времену (енгл. *RTP - Real-time Transport Protocol*). Док се контрола извора гласа може контролисати преко RTSP (енгл. *Real Time Streaming Protocol*) протокола. Пример оваквог преноса гласа дат је у поглављу 5.2.

2.2 Појам паметне куће

Појам паметне куће односно паметног животног простора није нова идеја. Још у току седамдесетих година најављиван је продор у овој области и демонстрирани су први системи [Brush]. Овај појам заправо представља сложен систем који углавном подразумева неки вид аутоматског управљања зарад постизања веће удобности [Evans], боље енергетске ефикасности [Jain], сигурности [Demirkol] и друго. Паметна кућа неретко поседује сложен софтверски и хардверски систем са под системима за управљање осветљењем [Milošević], управљање квалитетом ваздуха, управљање системом за грејање и хлађење, управљање кућним апаратима [Vrga], мултимедијалним уређајима [Stanivuk], праћење присуства људи [Mrazovac], њихових активности и томе

слично. Иако се чини да су сви потребни технички услови задовољени и даље су отворена многа питања попут начина инсталације и адаптације постојећих кућа, проблем подешавања и администрације, сврха паметних уређаја и увођење система вештачке интелигенције у животни простор [Rejovic]. Системи кућне аутоматизације углавном имају везу са интернетом и омогућавају даљински приступ и контролу уређаја у кући што изискује додатне технике заштите.

Уграђени рачунарски системи [Isikdag] су део наше модерне свакодневице. На тржишту потрошачке електронике све се више појављују уређаји повезани у локалну бежичну мрежу. Конвенционални уређаји често долазе са новим и напреднијим функционалностима за праћење њиховог стања или даљинско управљање. Све ово је омогућено напретком и развојем интернет технологија. Концептуално, ова глобална појава се назива интернет ствари (IoT), а њен продор у наше домове оплемењује стару идеју звану паметна кућа [Jie]. Уређаји са уграђеним рачунарима заједно повезаним на локалну или глобалну мрежу чине скуп интелигентних ствари које заједно померају границе комотности људског простора и пружају додатну удобност [Alaa].

Велики број уређаја у оваквим системима ствара потребу за униформном и лако спрегом са оваквим системом. Једно елегантно решење свакако представља природни говор. Овакав приступ је доступан корисницима паметних телефона дуги низ година [Stefanovic]. Многи аутори се баве анализом и синтезом говора на самим уређајима, али овакав приступ има низ мана. У нашем скорашњем раду смо дали предлог говорне спреге за уређаје са изразито ограниченим ресурсима [Cetic3].

Фокус ове дисертације је командовање гласом у паметним кућама. Овај проблем се дуго изучава и појављује се све више уређаја којима је могуће командовати гласом путем паметних звучника и сличних уређаја [Guamán]. Истраживања на простору Сједињених Америчких Држава показују драматичан скок код прихватања паметних звучника са мање од 1% одрасле популације на чак 19,7% у току 2018 [Kinsella2018] док нови подаци показују чак 26,2 % [Kinsella2019]. Ово прати и глобални тренд где се овакви системи појављују и ван енглеског говорног подручја [Giannoulis].

2.3 Акустика у архитектури

Значајан проблем у модерној архитектури је процена акустичких својстава пројектованог простора. Процена импулсног одзива затвореног простора даје увид у акустичке карактеристике тог простора. Овакве методе имају примену при:

- пројектовању и изградњи великих просторија попут позоришта, амфитеатра, концертних дворана, цркава и храмова, спортских дворана и слично.
- постављању и подешавању система за репродукцију звука. Овде се првенствено мисли на просторије слушаонице које су намењене за слушање музике [Kurtović].
- просторном планирању и уређењу радног и животног простора.

Акустика у затвореном простору фасцинира људе дуже од стотину година. За пионира модерне акустике у архитектури сматра се амерички физичар Волес Сабин (енгл. *Wallace Clement Sabine*) који је живео у раздобљу од 1868. до 1919. године [APSNArc]. Дотадашњи приступ акустици базирао се искључиво на геометријским моделима који су изучавали пропацију звучних фронтана. Сабин уводи новине узимајући у обзир дифузију енергије у затвореном простору и фокусира се на истраживања карактеристика разних материјала у погледу рефлексије и апсорпције. Узимајући у обзир број могућих параметара задатак није био нимало лак. Сабин и његови сарадници изводе многа мерења уз помоћ оргуља и штоперике и тиме дају емпиријску потврду теорије.

Даљи рођак Волеса Сабина, Пол Сабин (енгл. *Paul Sabine*) наставља његов рад и даје велики допринос у изучавању акустике у затвореним просторима. Он развија методу којом се простор акустички оцењује. Метода изгледа тако што се у просторији направи кратак звук, попут Дираковог импулса, након чега се звук снима или слуша. Примећено је да звук остаје у просторији услед одбијања о зидове док се потпуно не апсорбује. Овај феномен се назива реверберација. Сабин је овом методом и простим слушањем успео да измери време трајања са завидном прецизношћу.

Ефекат реверберације је посебно изражен у просторијама са глатким или сферним зидовима од чврстог материјала попут оних у црквама, великим халама

и ходницима. Ефекат реверберације у тим случајевима може трајати и неколико секунди. Такве просторије се називају веома живе просторије. Просторије у којима је ефекат реверберације траје мање од 0,3 секунде називамо анехоичне (глуве) собе или мртви простори.

Према закону одржања енергије снага долазећег звука P_d мора бити једнака збиру снага рефлектованог P_r и апсорбованог звука P_a [Kurtović].

$$P_d = P_a + P_r \quad (2.1)$$

Овим дефинишемо коефицијент апсорпције α :

$$\alpha = \frac{P_a}{P_d} \quad (2.2)$$

Аналогно овој дефиницији често се користи и однос интензитета апсорбованог звука J_a и интензитета директног звука J_d . Сам коефицијент апсорпције је својство материјала од кога се звук одбија. Није исти на свим фреквенцијама, односно потребно га је одредити на свим фреквенцијама од интереса. У литератури се користи и $\bar{\alpha}$ као ознака за средњу вредност или укупну вредност коефицијента.

У акустици се још дефинише и комплементарни коефицијент рефлексије r као однос снага P_r и P_d или интензитета звука J_r и J_d . Коефицијенти r и α на тај начин стоје у релацији:

$$1 = \alpha + r \quad (2.3)$$

Звук слаби услед механичког трења честица у медијуму кроз који се простире али је ово мали узрок слабљења. Главно слабљење интензитета звука се дешава услед ширења сферног таласног фронта. Уколико се посматра слободно простирање звука од извора може се сматрати да интензитет слаби са квадратом растојања по формули:

$$J_d = \frac{P_a}{4\pi r^2} \quad (2.4)$$

Ако посматрамо простирање звука у било ком затвореном простору примећујемо да се звук који допире до нас састоји од директног звука и додатних компоненти сачињених од рефлексија. У тачкама од интереса, односно у тачкама у којим анализирамо звук прво допире директан звук, затим прве рефлексије и након тога ослабљене касне рефлексије. У случају затвореног простора звук

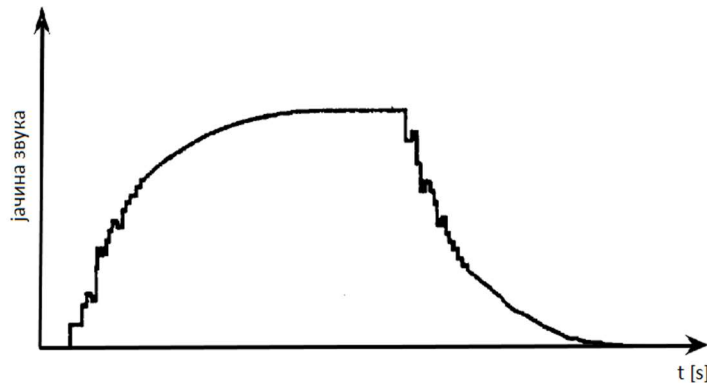
додатно слаби услед апсорпције при свакој рефлексији. Апсорпција A_0 просторије је збир свих утицаја површина од којих се звук одбија:

$$A_0 = \sum_{i=1}^N \alpha_i S_i \text{ [m}^2\text{]}. \quad (2.5)$$

Појединачни утицаји су сразмерни изложеној површини S_i помноженој са одговарајућим коефицијентом апсорпције α_i . За утицај људи у просторији се у просеку узима $A_c = 0.5 \text{ m}^2$. У случају када је у просторији или већој дворани присутно n људи имамо укупну апсорпцију у виду следећег израза:

$$A = A_0 + n \cdot A_c \quad (2.6)$$

Након што извор започне емитовање звука у просторији, јачина звука у свакој тачки експоненцијално расте док се не успостави равнотежа емитоване и апсорбоване звучне енергије у јединици времена. Када побуда нестане, јачина звука опада и на крају остаје само рефлектујући звук као што је приказано на илустрацији 2.3.1.



Илустрација 2.3.1 Јачина звука у просторији.

На довољној удаљености од извора формира се дифузно поље. Уколико је поље хомогено интензитет звука је представљен изразом:

$$J_{max} = \frac{4 \cdot P_a}{A} \quad (2.7)$$

У близини извора доминира директан звук. У литератури се раздаљина на којој се изједначава директан и реверберантни звук назива радијус просторије или критична раздаљина (енгл. *Critical distance*). Изједначавањем израза 2.4 и 2.7 добијамо једначину за радијус просторије:

$$r_h = \frac{1}{4} \sqrt{\frac{A}{\pi}} \quad (2.8)$$

Тачно одређивање радијуса просторије у произвољном простору није тривијалан задатак [Мијић1][Shalkouhi][Мијић2]. Техникама мерења интензитета звука се постижу мере одређеног нивоа тачности. Иако је радијус просторије веома важан са становишта одређивања позиције бежичног микрофона његово тачно одређивање није директно предмет ове дисертације.

2.3.1 Акустички параметри

У литератури постоји низ параметара који се користе ради испитивања акустичких својстава. Резултати мерења звука у некој тачки се користе за процену других величина као што су ниво звучног притиска или мерења звучне изолације [ISO 3382-2]. За мерења се користе прецизни инструменти који најближе могу имитирати тачкасти извор звука за стварање хомогеног звучног поља као што је приказан на слици 2.3.2. Генерална препорука гласи да се овакав извор звучног притиска поставља у средину просторије да би се тиме омогућило хомогено распрострањање таласа односно равномерна дистрибуција звучног притиска.



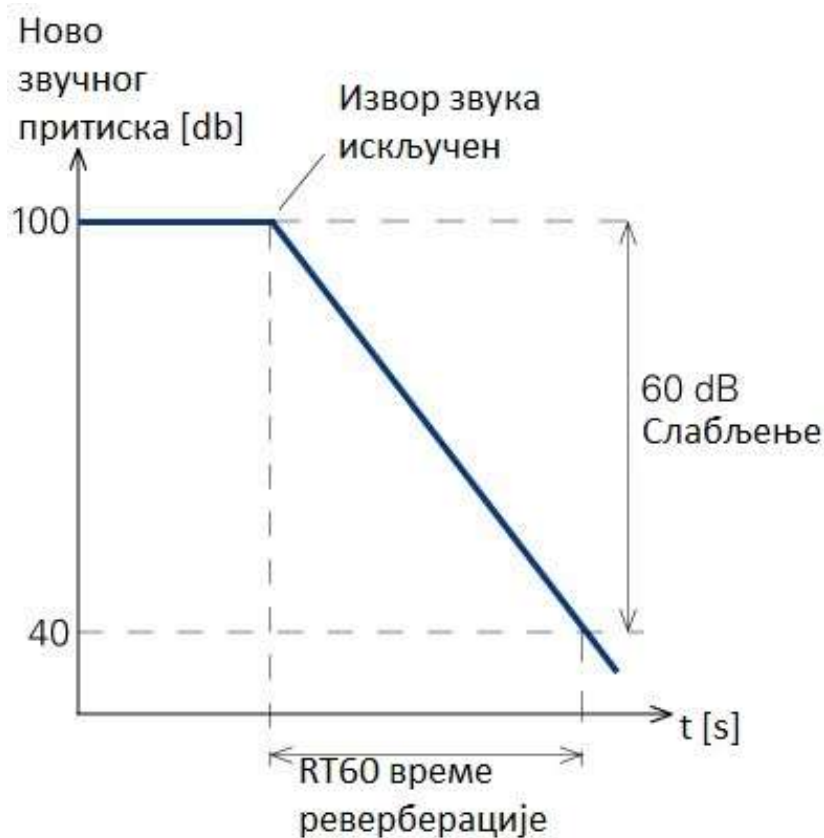
Слика 2.3.2 Пример омнидирекционог извора звука

Стандард ISO 354: 2003 прописује метод мерења коефицијента апсорпције одређеног материјала и њихов утицај на ефекат реверберације. Предмети као што

су намештај, особе или специјалне површине за апсорпцију звука (апсорбери) морају се узети у обзир приликом прецизног одређивања односно мерења акустичких параметара. Добијени резултати се могу користити за потребе поређења али и за даља израчунавања код просторног планирања, озвучавања и контроле буке [ISO 354:2003]. Иако су ове методе прописане стандардом некада их није једноставно спровести. Мерења која се добију нису довољног квалитета. Уколико се такви подаци користе при симулацији, модел може бити неприкладан и врло лако може довести до погрешних закључака [Christensen].

2.3.1.1 Време реверберације

Прописани стандард ISO3382-2: 2008 одређује методе за мерење времена одзива у обичним просторијама. Описује процедуру мерења, потребан апарат, потребан број мерних положаја и метод за процену података и презентацију извештаја о мерењу. Мера RT60 се назива време реверберације. Служи за процену колико је потребно времена да ниво звучног притиска опадне за 60 dB након искључивања звучног извора.



Илустрација 2.3.3 Мерење времена реверберације - RT60

Сабиновим обрасцем можемо израчунати приближно RT_{60} на основу познате укупне апсорпције A и димензија просторије, односно запремина просторије V :

$$RT_{60} \cong \frac{0.163 V}{A}. \quad (2.9)$$

Мерење времена реверберације RT_{60} дефинисано је стандардом ИСО 3382-2 за обичне просторије. RT_{60} се мери у секундама и често се наводи као једна вредност која се по правилу односи на фреквенцију од 1 kHz. Услед последице да је апсорпција материјала различита по фреквенцијама, потпуна информација о реверберацији је функција фреквенције па се у пракси врше мерења у опсегу од 50 Hz до 8 kHz у корацима од једне октаве или трећине октаве. Сваки материјал има различите способности апсорпције и расипања. Због тога се за побуду користе сегменти белог шума пропуштени кроз филтер са пропусним опсегом центрираним на карактеристичним фреквенцијама. Ова техника је примењена код одређивања RT_{60} у експериментима који су спровођени. Резултати мерења су дати у табелама у прилогу Б.

2.3.2 Технике моделовања акустике у затвореном простору

Како би се постигло оптимално решење приликом пројектовања новог грађевинског објекта често се користе технике моделовања и симулације. Помоћу оваквих метода дати критеријум можемо проверити и пре него што је пројектовани простор изграђен. На овај начин се постижу значајне уштеде у новцу и времену.

2.3.2.1 Модел пресликавања извора

За моделовање пропагације звука у простору са реверберацијом Ален и Беркли су предложили врло прецизан метод под називом модел пресликавања извора (енгл. *Image Source Modeling – ISM*) [Allen]. Цео процес се одвија унутар рачунара и поред датог описа методе дат је прилог у виду имплементације у програмском језику Фортран. Метода ISM се најчешће примењује код моделовања омотача облика квадрата са шест нормалних правоугаоних површина, али није ограничен само на такве омотаче. У теорији, помоћу ISM се може моделовати било који облик затвореног простора [Borish]. У публикацији [Öçal]

показује се метод праћења објеката у не-конвексним просторијама користећи обрнути ISM метод. У оквиру ове дисертације остаје се само при моделовању правоугаоних омотача, имајући у виду чињеницу да се већина животног простора релативно добро може свести на облик квадрата у прихватљивој тачности [Allen]. Ова метода спада у геометријске методе. Оригинални ISM метод је рачунски захтеван. У новијем раду [Lehmann1], Лиман и Јохансон су предложили побољшани метод у којем се користи процена опадања енергије.

Основна идеја се базира на израчунавању преносне функције, односно преносне карактеристике импулсног одзива собе (енгл. *Room Impulse Response - RIR*). У литератури је познат појам преносне функције која се одређује између било које две тачке у простору узимајући у обзир димензије и облик омотача, односно својства рефлексије и апсорпције зидова, пода и плафона. Овде је важно нагласити да се у симулацији занемарује утицај намештаја и људи у просторији. Након што се конструише RIR, конволуцијом се може добити сигнал у произвољној тачки у простору. На тај начин се добија сигнал акустичког сензора у којем се симулира ефекат реверберације.

Симулација започиње тако што се у простору одређују позиције извора звука и позиције тачака од интереса. У свакој тачки од интереса поставља се један акустички сензор. Посматра се Декартов координатни систем на следећи начин. Позиција извора сигнала обележава се са:

$$p_{source} = \begin{bmatrix} x_{source} \\ y_{source} \\ z_{source} \end{bmatrix}, \quad (2.10)$$

позиција од интереса (акустички сензор) је:

$$p_{sensor} = \begin{bmatrix} x_{sensor} \\ y_{sensor} \\ z_{sensor} \end{bmatrix}, \quad (2.11)$$

и на крају димензије омотача се дефинишу са:

$$d_{room} = \begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix}. \quad (2.12)$$

Преносна функција у временском домену $h(t)$ се израчунава за сваку тачку од интереса користећи следећу формулу:

$$h(t) = \sum_{u_{sum}=0}^1 \sum_{l_{sum}=-\infty}^{\infty} A(u_{sum}, l_{sum}) \cdot \delta(t - \tau(u_{sum}, l_{sum})), \quad (2.13)$$

где $\delta(\cdot)$ означава функцију Дираковог импулса. Са $u_{sum}(u, v, w)$ и $l_{sum}(l, m, n)$ су означене троструке суме над сваким триплетом интерних индекса. Триплети (u, v, w) и (l, m, n) су параметри индексирања слике извора по свим димензијама ширини, дужини и висини. Ознака $A(\cdot)$ означава фактор слабљења који је одређен изразом (2.14), док је кашњење $\tau(\cdot)$ услед пропагације звука брзином c дато једначином (2.15):

$$A(u_{sum}, l_{sum}) = \frac{\beta_{x,1}^{|l-u|} \beta_{x,2}^{|l|} \beta_{y,1}^{|n-v|} \beta_{y,2}^{|m|} \beta_{z,1}^{|n-w|} \beta_{z,2}^{|n|}}{4\pi \cdot d(u_{sum}, l_{sum})}, \quad (2.14)$$

$$\tau(u_{sum}, l_{sum}) = d(u_{sum}, l_{sum}) / c. \quad (2.15)$$

Ако узмемо у обзир раздаљине простирања звука у просторијама релативно малих димензија у односу на брзину звука $c = 343$ m/s, добијамо велики број рефлексција у одређеном времену. Овај број се назива густина рефлексција. Ефекат реверберације се добија сабирањем рефлектујућих сигнала са кашњењем τ услед пропагације сигнала. Сам ефекат рефлексције се моделује коефицијентима рефлексције $\beta_{i,j}$, који се изводе на основу коефицијената апсорпције $\alpha_{i,j}$:

$$\beta_{i,j} = \sqrt{1 - \alpha_{i,j}}. \quad (2.16)$$

Пошто се коефицијенти апсорпције и рефлексције у литератури дефинишу на различите начине, овде је битно напоменути да су коефицијенти $\beta_{i,j}$ посматрају онако како су дефинисани код [Allen] и треба их разликовати од коефицијента рефлексције r из израза (2.2) где говоримо о простом односу коефицијената код једног типа материјала.

Еуклидска удаљеност $d(u_{sum}, l_{sum})$ између сваке слике извора и места од интереса се рачуна множењем дијагоналне матрице $diag(\cdot)$ и параметара (u, v, w) и (l, m, n) , позиције извора p_{source} и димензије омотача d_{room} :

$$d(u_{sum}, l_{sum}) = \left\| \begin{array}{c} diag(2u-1, 2v-1, 2w-1) \cdot p_{source} + \\ p_{sensor} - diag(2l, 2m, 2n) \cdot d_{room} \end{array} \right\|, \quad (2.17)$$

Једначина (2.18) даје RIR $H(\omega)$ у фреквентном домену:

$$H(\omega) = \sum_{u_{sum}=0}^1 \sum_{l_{sum}=-\infty}^{\infty} A(u_{sum}, l_{sum}) \exp(-j\omega\tau(u_{sum}, l_{sum})), \quad (2.18)$$

Након тога, да би се одредио сигнал у тачки од интереса $y_{sensor,i}(t)$ означеним са i сигнал се преводи у фреквенцијски домен дискретном Фуријеовом трансформацијом (енгл. *Discrete Furie Transformatio - DFT*) над сигналом извора звука:

$$Y_{source}(\omega) = F(y_{source}(t)). \quad (2.19)$$

Над сигналом извора звука се ради конволуција у фреквентном домену за сваки $H_i(\omega)$ из добијене банке одзива. Затим се сигнал у временском домену добија инверзном Фуријеовом трансформацијом $F^{-1}\{\cdot\}$ за сваку тачку од интереса:

$$y_{sensor,i}(t) = F^{-1}\{Y_{source}(\omega) * H_i(\omega)\}. \quad (2.20)$$

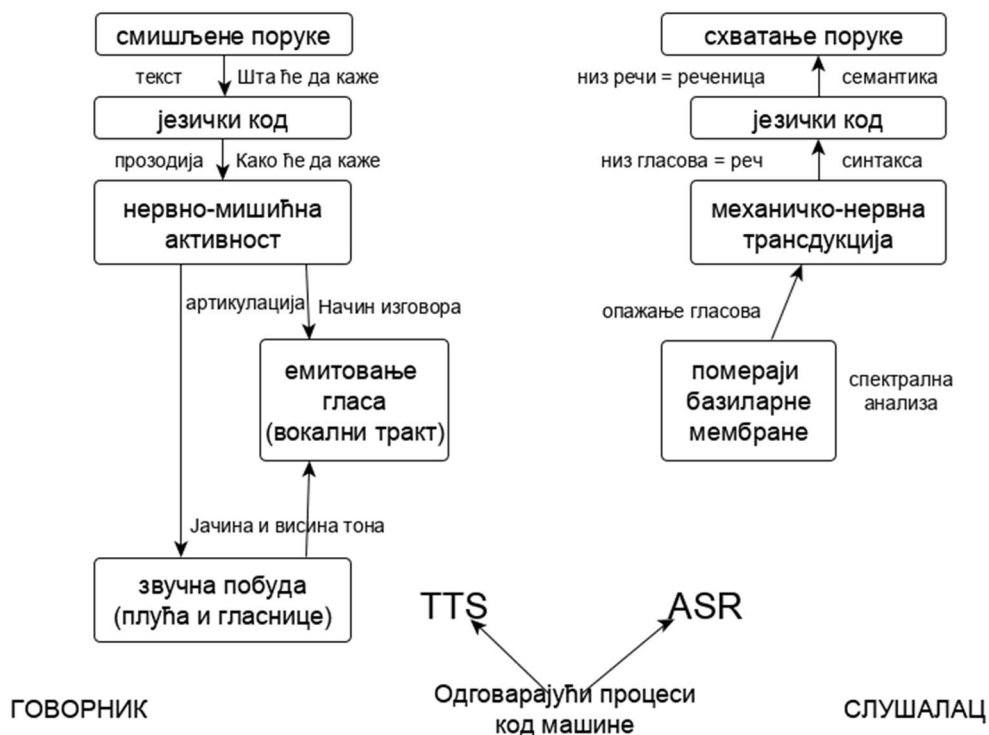
У оквиру симулација које користимо у [Cetic1] одређивање $\alpha_{i,j}$ и $\beta_{i,j}$ коефицијената се раде у складу са проценом опадања енергије односно кривом опадања енергије (енгл. *energy decay curve EDC*) [Lehmann1]. Као параметри модела користе се пропорцијални коефицијенти разлике у апсорпцији површина и време реверберације RT60 на основу којих се добијају стварни $\alpha_{i,j}$. Пропорцијални коефицијенти представљају односе апсорпција. Уколико се узме у обзир шесторка (1, 1, 1, 1, 1, 1) то значи да је апсорпција зидова, пода и плафона иста и можемо очекивати потпуно хомогену апсорпцију у таквој просторији. У другом случају ако узмемо [1 1 0,9 0,9 0,7 0,7] добићемо да бочни зидови имају 10% мању апсорпцију од предњег и задњег зида, а плафон и под имају 30% мањи коефицијент апсорпције [Lehmann2].

2.4 Говор као спрега

Примена препознавања говора која се појављује у последњих неколико деценија укључује гласовно бирање или гласовно позивање, усмеравање позива, једноставне податке попут уноса броја кредитних картица и претрага звучног записа који садржи одређене речи. Могуће је одредити место у звучном запису где су задате речи изговорене. Временом, уз напредак у развоју и повећањем тачности препознавања говора, апликације које користе ову технологију добијају све више на важности. Тако је ова технологија пронашла примену у медицини,

војној индустрији, телекомуникацијама, роботизи, домаћинству и разним уређајима потрошачке електронике.

Говорна комуникација је сложен процес који са стране говорника обухвата мисаоне, односно когнитивне процесе, неуро-мускуларну активност, стварање звука и његово емитовање уз помоћ плућа кроз глотис и вокални тракт. Са стране слушаоца посматрамо процес који код здравог човека почиње у уху. Говор заједно са осталим звуковима преко ушног канала долази до базиларне мембране. Даље се механичке вибрације путем неуралних трансдјусера преносе као импулси на основу којих се добијају обележја говора. На крају, у когнитивном процесу се од низа препознатих обележја долази до разумевања изговорених речи и реченица [Delic2].



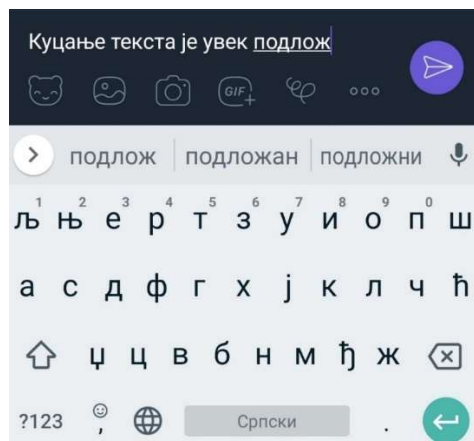
Слика 2.4.1 Преглед говорних технологија (преузето из [Delic2])

Данас се препознавање говора остварује на много начина. Ипак, ниједна метода још није доведена до врхунца. Многи аутори се баве анализом и синтезом говора на самим уређајима, али овакав приступ има своја ограничења. У [Cetic3] је дат предлог говорне спреге за уређаје са изразито ограниченим ресурсима базиране на снази удаљених рачунара. Овакав приступ је популаран код многих произвођача оригиналне опреме. Главни покретач овог феномена је стални

напредак у снази удаљених рачунарских система. Иако је у почетку постојао одређени отклон према самој идеји, корисници масовно прихватају решења која снимљени говор шаљу на анализу у системе ван своје контроле. Са становишта сигурности и приватности овакав приступ није повољан, нарочито код примене у медицини, судству и одбрани. Ипак, појединцима погодности које добијају могу бити од веће важности од њихове приватности и могу сматрати да им безбедност није угрожена.

Глас као корисничка спрега и даље са собом носи велики број проблема. Иако се многи аутори баве овом облашћу и даље има пуно простора за напредак у анализи и синтези говора. Системи за командовање гласом и сама примена аутоматског препознавања гласа у свим условима је и даље нерешен и комплексан проблем. Главни проблеми и даље представљају ситуације са неповољним односом сигнала и шума, варијабилност говорника и говора у погледу акцента и емоција, као и сложености контекста и говор у конверзацијама. Поред овога, специфичности различитих језика, њихова фонетска и граматичка различитост и даље је велика област истраживања где се модели и технике у једном језику некада тешко преносе и примењују у другом. Ова језичка зависност говорних технологија је издваја у односу на многе друге нове технологије.

Говорне технологије су уско везане за област обраде природног језика (енгл. *Natural language processing - NLP*). Шире говорећи овде се не ради само о обради звучног записа већ заправо о под области лингвистике и њеној примени у рачунарској техници и информационим технологијама. Пример примене NLP се види кроз разне услуге у свакодневници савременог човека. На слици 2.4.2 дат је приказ корекције текста при његовом уносу код паметних телефона.



Слика 2.4.2 Пример корекције текста код Андроид паметног телефона

Ради ефикасне обраде звучног сигнала развијени су дигитални сигнал процесори (енгл. *Digital Signal Processor* – DSP) који имају специфичну намену и посебно прилагођену архитектуру [AADSP1]. Посебно прилагођен скуп инструкција и проширене аритметичко логичке могућности су главна предност оваквих процесора. За њихово програмирање потребни су додатни напор и умеће. Модерне DSP архитектуре се ослањају на паралелизме на нивоу инструкције што изискује специфичне технике превођења помоћу наменских развијених преводиоца из виших програмских језика у машински код [Djukic]. Често чипови поседују више процесорских језгара ради постизања паралелне обраде на нижим тактовима. Ово изискује додатну пажњу при изради програмске подршке и пратећих библиотека [Cetic2]. DSP стално напредују у погледу процесорске моћи и подржаних могућности. Ради постизања енергетске ефикасности примењују се технике моделовања и процене потрошње енергије на основу анализе извршавања програма [Kronic].

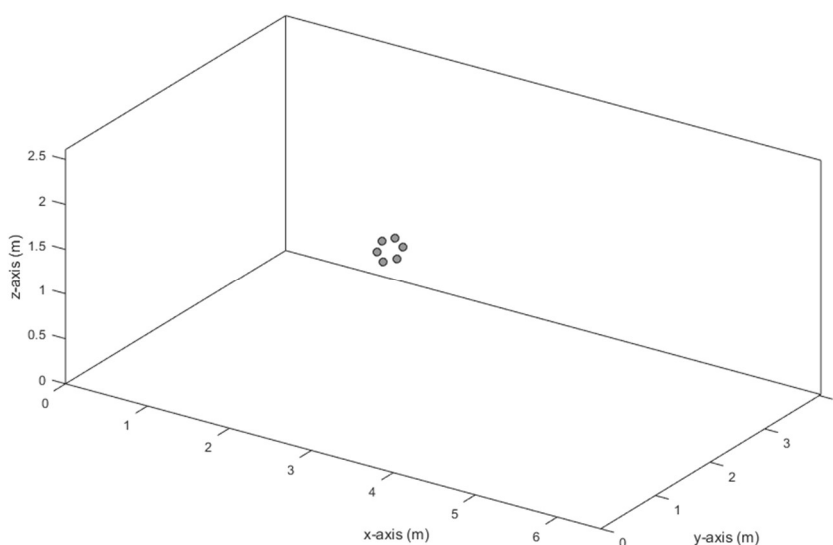
Звучни сигнал $y(t)$ се описује функцијом времена. Сигнал пролази кроз некакав медијум и трпи одређена изобличења. У општем случају сигнал од интереса се означава са $s(t)$ док се изобличења сматрају за буку или дисторзију $n(t)$.

$$y(t) = s(t) + n(t) \quad (2.21)$$

Процена шума се користи код техника попут отклањања акустичког еха (енгл. *Acoustic Echo Cancellation* – AEC). Традиционално, микрофонски низови (енгл. *Arrays*) или микрофонске решетке (енгл. *Grid*) се користе као алат за

екстракцију различитих звучних информација, локализацију извора звука и смањење шума [Zheng]. Често се користе технике процене угла из којих долази звук (енгл. *Direction of arrival – DOA*).

Проблем са оваквим системом је приказан на илустрацији 2.4.3 где видимо да су сви микрофони груписани на једном месту у просторији. Пошто микрофонски низови снимају само локално звучно поље и често са превелике удаљености, такав сигнал обично има низак ниво корисног сигнала у односу на шум [Bertrand].



Илустрација 2.4.3 Приказ микрофонске решетке

2.4.1 Трендови у развоју и примени аутоматског препознавања говора

Увидом у литературу може се уочити да је аутоматско препознавање говора и даље веома актуелно. За аутоматско препознавање говора се користи теорија статистичког моделовања говора и теорија скривених Марковљевих ланаца (енгл. *Hidden Markov model - HMM*) [Hwang]. Ово је статистички модел који производи низ симбола или ознака. Разлог због кога се HMM употребљава за препознавање је својство сигнала који се може посматрати по деловима као краткотрајан стационарни сигнал. У кратком временском периоду од око 10 милисекунди, говор може бити схваћен као стационаран процес. Други разлог због кога су HMM били популарни је чињеница да могу бити обучени аутоматски. Додатно, они су једноставни и практични за реализацију код система базираних на

рачунару. У овој области дуги низ година дешавала се стагнација без великог напретка. Чинило се да су се исцрпеле могућности статистичког приступа.

Ипак, недавно се појавио велики напредак применом дубоких неуронских мрежа (енгл. *Deep Neural Network* - DNN), односно продор уз помоћ техника машинског учења [Delic2]. DNN налазе примену и код препознавања српског језика. У [BPopović1] је приказано препознавање говора базирано на Калди алатима отвореног кода. Постигнут је изразито добар резултат са грешком препознавања речи мањим од 2% на реченицама од више стотина хиљада речи.

Код DNN за аутоматско препознавање говора јављају се два приступа при обучавању дубоких неуронских мрежа у зависности од коришћеног звучног материјала. Такозвани чисти акустички модели (енгл. *Clean*) су они модели код којих је скуп података прикупљен у једном сценарију односно просторној ситуацији без варирања акустичких параметара. Код другог приступа се користи већи скуп података и често се користи симулација где је могуће добити синтетизован звук са различитим односима сигнала и шума.

Многи велики ASR сервиси базирани на рачунарским подсистемима у клауду пружају услуге тек након аутентификације корисника. Готово увек користе доступан персонализован контекст, односно додатне информације да би побољшали препознавање. У раду [Raju1], група аутора из тима за машинско учење компаније Амазон је демонстрирала како се статистички модел језика може побољшати коришћењем познатог контекста за сваког корисничког агента у сценарију виртуалног персоналног асистента.

Постоје обимна истраживања у области аутоматског препознања говора. Истакнуте публикације у литератури показују велики број различитих приступа. Заједничко им је да се за испитивање перформанси увек формира велики скуп звучних записа са одређеним својствима. Овакав скуп записа се назива корпус и обично садржи аотиран велики скуп изговорених речи и реченица од стране више говорника. Поред аудио записа мора да садржи и транскрипције а често и фонетске и фонолошке аотације. Један од познатијих је [TIMIT] који је објављен још 1993. године који садржи прочитане речи 630 говорника у различитим дијалектима са Енглеског говорног подручја. Интересантан је и рад [Panayotov] где се корпус конструисао на основу аудио књига које су јавно доступне.

Последњих година научна јавност објављује конкурсе у којима истраживачи решавају релевантне проблеме у овој области. Конкурс СНiМЕ се одржава трећи пут [Barker1]. Има за циљ сепарацију говорника и препознавање на основу микрофона распоређених на ободу таблет уређаја. У недавном извештају [Barker2] су објављени резултати 26 учесника. У оквиру конкурса *REVERB challenge* је учествовало још више учесника [Kinoshita]. Састоји се из два задатка: побољшања сигнала путем уклањања ефеката реверберације и препознавања говора. Постоје и напори у домаћим научним круговима који праве корпус за српски језик [Pakoci]. Код ових истраживања скуп сигнала није јавно доступан. До њега је вероватно могуће доћи на захтев. Последњи конкурс узет у разматрање је [ACE]. Даје врло детаљне описе просторија у којима су вршени експерименти и пружа јавно доступни скуп снимљеног говора у глувој соби. Овај скуп је коришћен у практичним истраживањима у оквиру рада на овој дисертацији.

Грешка препознавања речи је позната мера код оцене ASR и означава се са WER (енгл. *Word Error Rate*). Ово је мера за разлику између W и W' секвенци речи. Заправо представља количник између укупног броја грешака и броја речи у референтној реченици. У литератури је познато неколико варијација код рачунања грешке. Израз (2.15) је мера грешке изражене у процентима која се користи у већини испитивања на које је аутор наишао.

Грешке код препознавања се деле на четири врсте грешака:

- C – број исправно препознатих речи (енгл. *correct*)
- S – број речи које су замењене (енгл. *substituted*)
- D – број речи које су обрисане (енгл. *deleted*)
- I – број речи које су додате (енгл. *inserted*)

$$WER = \frac{S + D + I}{C + S + D} \times 100\% \quad (2.22)$$

2.4.2 Командовање гласом у потрошачкој електроници

Многа истраживања доприносе овом пољу проналазећи нове апликације за гласовне контроле. У раду [Škraba] је представљен развој платформе за инвалидска колица. Пројекат отвореног кода WebKit користи се као услуга

препознавања гласа на бази клауд технологије. Било би занимљиво видети упоређивање перформанси са другим системима. У овом раду се користи само анализа говора. Аутори у [Wei] користе исти приступ који користи клауд инфраструктуру како би омогућио препознавање говора за преносне уређаје. У оба рада не спомиње се синтеза говора. У раду [Soda] аутори користе мрежу микрофона за изградњу гласовне спреге за систем паметне куће. Такође ови аутори истражују употребу виртуелних агената као веб сервиса како би обогатили корисничко искуство у паметној кући. Њихов рад је концентрисан на јапански језик. У оквиру публикације [Alumäe] аутори дају опис једне архитектуре за реализацију препознавања естонског говора за Андроид уређаје. Српски језик је такође изучаван од стране домаћих аутора [Pekar]. Између осталог, развијен је интерактивни говорни систем и разне друге апликације са применом говорних технологија.

Такође, данас су доступна и нека комерцијална решења. Компанија HouseLogic нуди два производа која се могу користити као говорна спрега за систем контроле у паметној кући. Прва је VoicePod мобилна апликација за паметне телефоне која је говорни асистент на природном језику. Нуди флексибилно командовање и тихи рад. Друго решење је VoicePod стони уређај који је осмишљен за употребу у просторијама. Ово решење је повезано са системом кућне аутоматизације.

Компанија Амазон је представила производ назван Амазон Ехо. Он служи као говорни асистент који је способан да успостави дијалог са корисником путем синтетизованог гласа. Аутор дисертације поседује мању верзију уређаја Ехо Дот која у погледу препознавања има сличне карактеристике као и већи уређај осим што је у њега уграђен звучник мање снаге. Додатно Амазон кроз своје клауд сервисе пружа аутоматско препознавање. Овај сервис је коришћен код истраживања у овој дисертацији и реализовани систем за испитивање је објављен у публикацији [Rajic]. Детаљи овог система су дати у петом поглављу.

Поред мноштва других произвођача увидом у тржиште потрошачке електронике издвајају се четири велике компаније: Гугл, Епл, Мајкрософт и Амазон. Свака од њих нуди своје сервисе на различитим платформама и сваки од ових сервиса је могуће надоградити са додатним могућностима:

- Гугл Асистент је виртуални персонални асистент развијен од стране компаније Гугл [Schalkwyk]. Примарно је намењен мобилним уређајима и уређајима попут паметних звучника. Корисник обично користи природан говор, али је подржан и унос текста преко тастатуре уколико се комуницира са асистентом путем мобилног телефона или веб претраживача. Тренутно долази са највећим бројем подржаних језика. Гугл Асистент омогућује кориснику да претражује по интернету користећи Гугл претраживач, да забележи различите догађаје и аларме у календару, промени поставке уређаја и прикаже информације са корисниковог Гугл профила. Компанија је најавила и скупљање визуалних информација путем камере, класификацију музичких записа и извођача, куповина и слање новца и слично.
- Алекса је виртуални асистент развијен од стране Амазона. Првенствено се појавила и користила у Амазон Ехо и Амазон Ехо Дот уређајима. Приликом коришћења ових уређаја, користи се природан говор, где уређај препознаје да треба да се активира кад корисник изговори кључну реч „Алекса“ на почетку реченице. Алекса може да ради и на другим платформама. Препознавање кључне речи није обавезно. Подржана је гласовна комуникација на притисак тастера (енгл. *push to talk*). Ово је послужило за испитивања у оквиру ове дисертације приликом студије случаја. У току истраживања подржани су енглески, немачки и јапански језик. Омогућује кориснику разговор, репродукцију музичких садржаја, прављење подсетника, подешавање аларма, добављање сервисних информација из света у реалном времену, коришћење интернет услуга и командовање и управљање паметним уређајима у кућама.
- Сири је интелигентни персонални асистент развијен од стране Епла. Она долази уз све оперативне системе компаније Епл. За коришћење система користи се природан говор и текстуалне поруке. Сири ће покушати да одговори на корисникова питања, понуди предлоге или одради задате команде. Програм константно учи и прилагођава се на

корисников говор, претраге и подешавања, тако да је сваки одговор специфичан за корисника.

- Компанија Мајкрософт је такође дала један од битнијих доприноса у овој области. Њихов виртуални асистент назива се Кортана. Долази уз *Windows 10* и оперативне системе за мобилне телефоне. Нешто касније се појавио и паметни звучник. Примарно се користи природан говор за интеракцију, али је подржан и унос преко тастатуре.

Константан развој ових система и сервиса је очевидан. У [ВРоровић2] је дат приказ реализације асистента за српски језик. У оквиру рада на овој дисертацији аутор је користио доступне сервисе компаније Гугл и Амазон и приказао могућност интеграције са уређајима у паметним кућама уз помоћ виртуалних асистената ове две компаније. Тренутни рад је ограничен само на енглески језик. Аутор има жељу да користи српски језик у свом будућем раду.

Примена говорних технологија види се у производима као што су бежични звучници, сатови, камере, роботи, мобилни телефони и таблет уређаји. Овакви уређаји пружају гласовну корисничку спрегу која лако налази примену у паметним кућама.

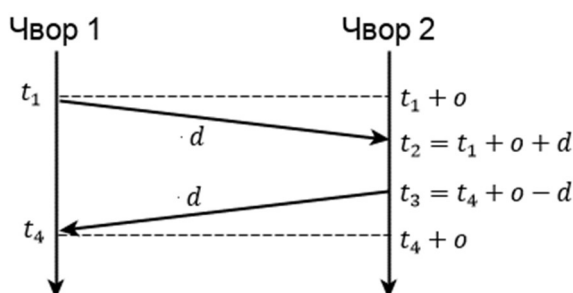
2.5 Бежични акустички сензори

У оквиру ове дисертације аутор се бави сензорима са ограниченим ресурсима бежично повезаним на процесорске јединице у локалној мрежи (*WLAN*), сензоре повезане са удаљеним рачунарским системима путем конвенционалне интернет мреже и сензоре повезане другим радио везама. Сензори повезани на овај начин пружају могућност различитих техника фузије података добијених са појединачних сензора. Да би се ушло у анализу оваквог система и дао предлог за његову реализацију и примену потребно је разумети следеће аспекте:

1. Достигнућа у области бежичних сензора.
2. Могућност примене и намене сензорских мрежа.
3. Преглед проблема и ограничења.

2.5.1 Проблем синхронизације

Ако је потребно омогућити корелацију сигнала или неку сложенију синтезу сензорских података неопходно је решити проблем синхронизације дистрибуираних чворова. Ово је један од кључних техничких проблема у мрежним комуникацијама. Нека од решења захтевају посебне проводнике попут IRIG (енгл. *Inter-Range Instrumentation Group*) система [Webb]. Главни проблем који је потребно решити је заправо проблем непредвидивости саобраћаја код чворова са уграђеним редом чекања. Овакви чворови уводе произвољно кашњење и уколико се о томе не води рачуна није могуће прецизно одредити време унутар дистрибуиране мреже. Основни проблем синхронизације се заправо решава мерењем кашњења услед пропагације порука кроз физички медијум као што је приказано на илустрацији 2.5.1.



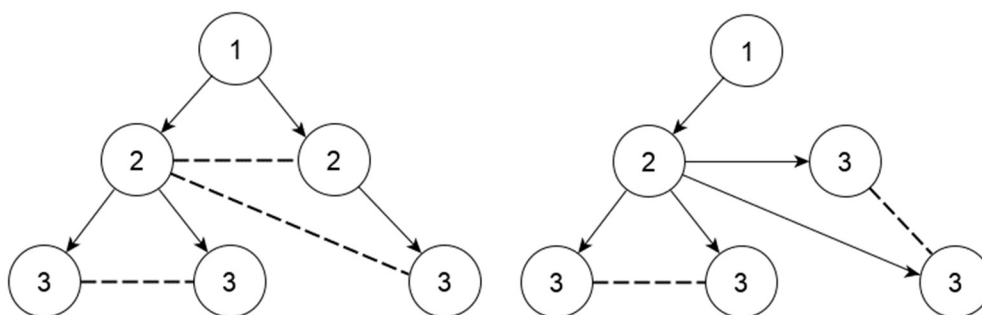
Илустрација 2.5.1 Основни проблем синхронизације дистрибуираних чворова

$$o = \frac{(t_2 - t_1 + t_3 - t_4)}{2} \quad (2.23)$$

Код интернет мрежа базираних на IP протоколу постоји неколико потенцијалних решења у зависности од постављеног критеријума синхронизације. Код мрежа широког подручја – WAN (енгл. *Wide Area Networks*) издвајају се два протокола која практично решавају дистрибуцију часовника. То су мрежни временски протокол NTP (енгл. *Network Time Protocol*) [RFC5905] и његов сродни једноставнији протокол SNTP (енгл. *Simple Network Time Protocol*) [RFC4330]. Оба протокола имају идентичну синтаксу мрежних пакета. Разликује се само начин на који систем реагује на примљене пакете ради синхронизације времена у чворовима. Каснијим допунама је протокол додатно растерећен увођењем могућности игнорисања неких поља из проширења [RFC7822]. У иницијалној верзији синтаксе пакета, информација о времену се размењује у

временском печату ширине 64 бита. У питању је број у непокретном зарезу где горњих 32 бита престављају целобројни број секунди који је протекао од 1. јануара 1900 године, док је разломљени део доњих 32 бита. Овим је постигнута резолуција од 2^{-32} што је приближно 232 пикосекунде. Касније у NTPv4 верзији протокола уведе се још два формата временског печата. Скраћени формат од 32 бита и проширени формат од 128 бита који се још зове и формат датума (енгл. *NTP date format*). И поред овако велике резолуције печата у пракси се уз помоћ NTP протокола углавном ради о синхронизацији на нивоу милисекунди.

SNTP протокол се препоручује само код примена где захтеви за синхронизацију нису велики. Ови протоколи су базирани на клијент-сервер архитектури, односно код ових протокола заинтересована страна шаље упит о времену. Са друге стране NTP протокол се сматра потпуним решењем код кога се клијент и сервер синхронизују са већом тачношћу [Mills]. NTP сервери обично базирају своје време на основу атомског часовника који користи стандардну резонантну фреквенцију атома као бројач времена. Синхронизација локалних часовника у чворовима се постиже математичким и статистичким методама којима се избацују неочекивани пакети и њихови временски печати и тиме се постиже фино кориговање локалног времена. Код платформи код којих је то могуће унутар NTP система узима се у обзир и појава дрифта и девијације периодичности часовника те се периодично раде корекције односно убрзања и успорења часовника. NTP систем се гради хијерархијски где се уводе нивои сервера времена који се зову стратуми (енгл. *stratums*) као што је приказано на илустрацији 2.5.2.



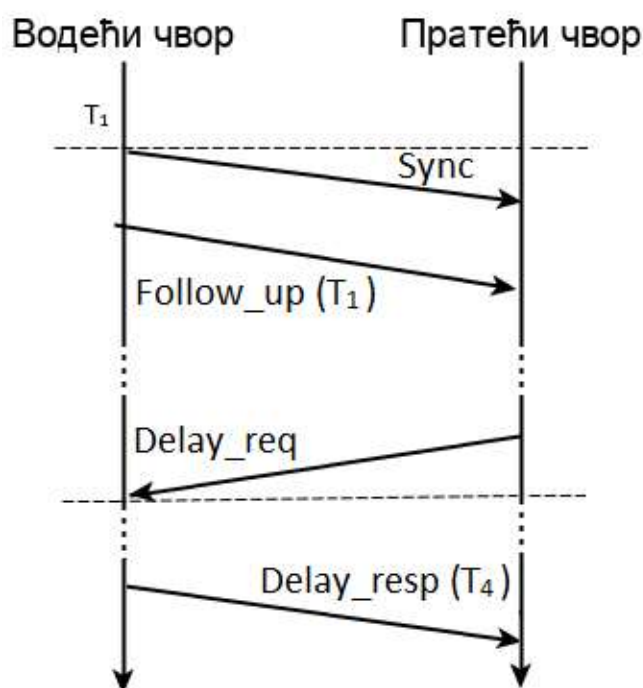
Илустрација 2.5.2 пример синхронизације под мреже.

SNTP протокол не подразумева чување стања у дугом временском периоду и стога је прикладнији код платформи са ограниченим ресурсима. Овим се

потпуно игнорише присуство дрифта и девијације и обично се имплементира само уз помоћ веома ограничених метода корекције часовника уз помоћ једноставног прескакања корака интерног часовника (енгл. *stepping*) или прецизнијег клизања (енгл. *slewing*). На овај начин се не постиже велика тачност али је ипак SNTP протокол и даље важан и користан уколико се примени уз јаке препоруке из [RFC4330]. На нижим слојевима (стратумима) се заправо користи NTP протокол док се SNTP користи само у екстремитетима под мреже односно на последњем слоју мреже. Ипак, примена SNTP система се јавља тамо где нема потребе за прецизном синхронизацијом односно где би примена NTP протокола била тешко оправдана. Детаљније, SNTP клијент би требало да се појави само у листовима стабла синхронизације мреже док се SNTP сервер може појавити само у корену под мреже која се синхронизује уз ограничење да ради у конфигурацији где не постоји ни један други извор синхронизације.

Оба протокола NTP и SNTP немају никакву потпору у хардверској имплементацији. По датим препорукама, потпуно се ослањају на софтверску имплементацију часовника приликом потписивања пакета али хардверска подршка се не забрањује односно не искључује се та могућност. NTP протокол је замишљен као децентрализован протокол синхронизације намењен глобалној мрежи и обично се налази на трећем нивоу Исо Оси модела.

Са друге стране имамо протокол прецизног времена (енгл. *Precision Time Protocol – PTP*) који је дат кроз IEEE 1588 стандард [IEEE1588]. Постоје две верзије стандарда које међусобно нису компатибилне, објављене 2002 и 2008 године. Принцип синхронизације кроз комуникацију водећег и пратећег чвора дат је на илустрацији 2.5.3.



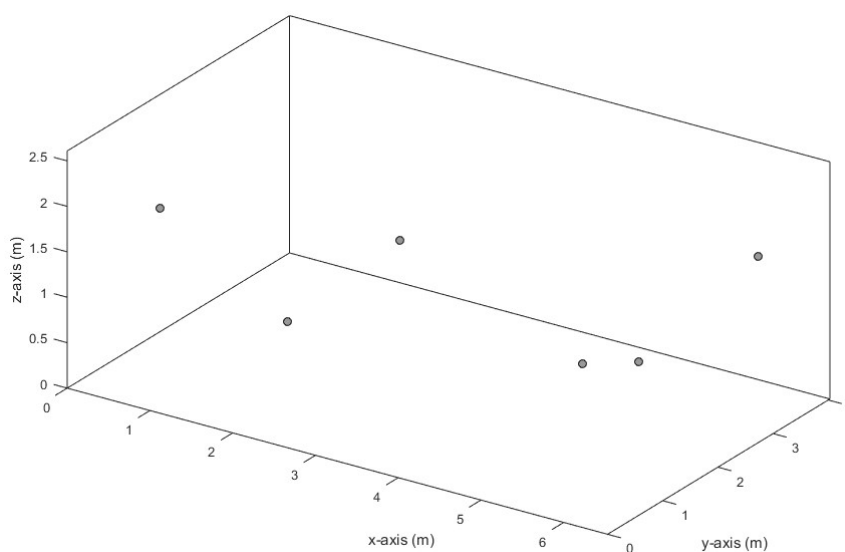
Илустрација 2.5.3 Мерење и синхронизација времена у RTP

Протокол за транспорт у реалном времену (енгл. *RTP - Real-time Transport Protocol*) у оквиру сваког пакета садржи и редни број и временски печат. Да би сваки пакет садржао валидну информацију у RTP пакетима потребно је синхронизовати сваки дистрибуирани чвор. Девиијације на нивоу целог система конституишу кумулативну грешку приликом комуникације са краја на крај. Чворови са ограниченим ресурсима могу постићи синхронизацију са резолуцијом испод микро секунда користећи механизам потписивања пакета прописан IEEE 1588 стандардом [Maу]. На овај начин могуће је постићи довољно добро поравнавање у времену и потенцијално користити технике типичне за класичне микрофонске низове. Ипак, за примену у ситуацијама са тврдим временским роковима, потенцијална препрека може бити додатно кашњење целог система које уносе дигитални кодери сигнала и сама пропација пакета кроз мрежу преко великог броја мрежних сегмената (енгл. *hops*). Стога се у пракси прибегава локалним системима док се удаљени рачунари повезани путем дељене везе користе када синхронизација није неопходна [Gomez].

2.5.2 Бежичне акустичке сензорске мреже

Акустички сензори су тема која се дуго изучава, али се традиционални приступ полако мења појавом све доступнијих мобилних уређаја и самим напретком телекомуникационих система. У литератури се овакви системи зову бежичне акустичке сензорске мреже (енгл. *WASN – Wireless Acoustical Sensor Networks*), али се некада помиње и термин бежичне мултимедијалне сензорске мреже (енгл. *WMSN – Wireless Multimedia Sensor Networks*). Да би се заобишао проблем са једним микрофоном или малим микрофонским низом, прибегава се примени WASN. На тај начин нема ограничења саме мреже и ствара се могућност произвољног постављања микрофона у простору [Bertrand]. Ова идеја је коришћена у развоју система за прикупљање акустичког материјала у животном простору путем бежичних сензора [Doder1].

Јефтине бежични акустички сензори који се разматрају у оквиру ове дисертације звук прикупљају помоћу омнидирекционих микрофона. Овакви микрофони се масовно користе у потрошачкој електроници због своје ниске цене. На тржишту су распрострањене веома јефтине микрофонске капсуле које је битно повезати на правилан начин и обезбедити конверзију аналогног сигнала. Предност у погледу ефикасности и цене доносе дигитални микрофони посебне израде. Они су базирани на микро електро механичком систему (енгл. *Microelectromechanical – MEMS*). Са њима се остварује комуникација дигиталним путем пошто је сама конверзија сигнала интегрисана у сам MEMS микрофон.



Илустрација 2.5.4 Дистрибуирани акустички сензори

Илустрација 2.5.4 приказује простор са распоређеним сензорима у просторији. Бежичне акустичке сензорске мреже (WASN) су релативно нова технологија за снимање и обраду звука. WASN користи чворове бежичног микрофона за снимање података звучног поља. Чворови бежичног микрофона који садрже један микрофон имају мале физичке димензије и могу се дистрибуирати по околини без физичких ограничења. Због бежичне везе, микрофони се могу постављати на места где је тешко поставити жичане микрофоне. Са овим предностима чворови микрофона могу физички да покрију много веће подручје интереса. Још једна предност акустичких сензора је та што могу да детектују активност иза препрека, јер не морају да буду у линији вида омогућавајући бољу покривеност са мање сензора [Hollosi2].

Постоје многе примене у којима се може интегрисати бежична мрежа акустичких сензора, као што су слушни апарати, телефонија без употребе руку, акустички надзор и амбијентална интелигенција [Bertrand]. У [Xiuhong] се усредсређује на акустичко надгледање унутрашње средине, као и на праћење отворених простора попут руралних подручја. У ту сврху користи се бежична звучна мрежа сензора са великим бројем бежичних акустичких сензора. Сензори који су узети у обзир у нашем истраживању су уређаји са ограниченим ресурсима. У систему акустичког надгледања често видимо да се подаци обрађују у централном делу система у виду апликативних сервера. Овај приступ омогућава истовремену примену сложенијих обрада и анализа на све прикупљене звучне

записе. У случају додавања нових алгоритама све промене се реализују на апликативном серверу, што омогућава лакши развој и проширење.

Истраживања показују веома добар потенцијал хибридних система попут *Auditeur* - мобилног клауда [Nirjon] где се користе паметни телефони. Такође демонстриран је и систем који пореди окружења акустички користећи мобилне уређаје [Jokic]. Ипак, овакав приступ је дискутабилан у погледу поузданости, енергетске ефикасности и цене. Велике инвестиције су омогућиле да читави градови послуже као тест полигон за разна истраживања [Pham1]. Омогућен је пренос звука у реалном времену кроз ад-хок мрежу у топологији решетке уз помоћ *XVee* бежичних чворова [Pham2]. Класификација звучних информација у градовима уз помоћ звучног надзора и различитих техника идентификације и локализације извора буке, отвара велики потенцијал за даљи развој и примену ради постизања удобнијег акустичког хоризонта и подизања квалитета животне средине [Ye]. Пројекат *EAR-IT* се бавио интелигентним системима базираним на акустичким сензорима и њиховом интеграцијом у велике IoT мрежне системе [Holloosi1]. Главна предност додавањем WASN је могућност комбинације информација о екстрахованим звучним догађајима са осталим подацима добијеним другим сензорима.

Код оваквих хибридних система у први план се истичу проблеми интегритета временских печата код дискретних одбирака сигнала стога су само неке примене могуће. Уколико се разматрају системи који користе конвенционалну, дељену интернет мрежу и при томе користе удаљене рачунаре и сервисе на њима, мора се узети у обзир и капацитет везе као потенцијално уско грло и извор произвољног кашњења. Класична имплементација комуникационих чворова базирана на редовима пакета не може гарантовати квалитет пружене услуге и поузданост.

2.5.3 Модели мрежног саобраћаја

Модели мрежног саобраћаја имају за циљ да опишу мрежне системе. Веома су корисни код испитивања мрежних перформанси. Често се моделују читаве мреже. Понашања мреже се потврђују симулацијама ради испитивања могућих загушења и квалитета услуге. Често се испитују губици пакета и примењују се

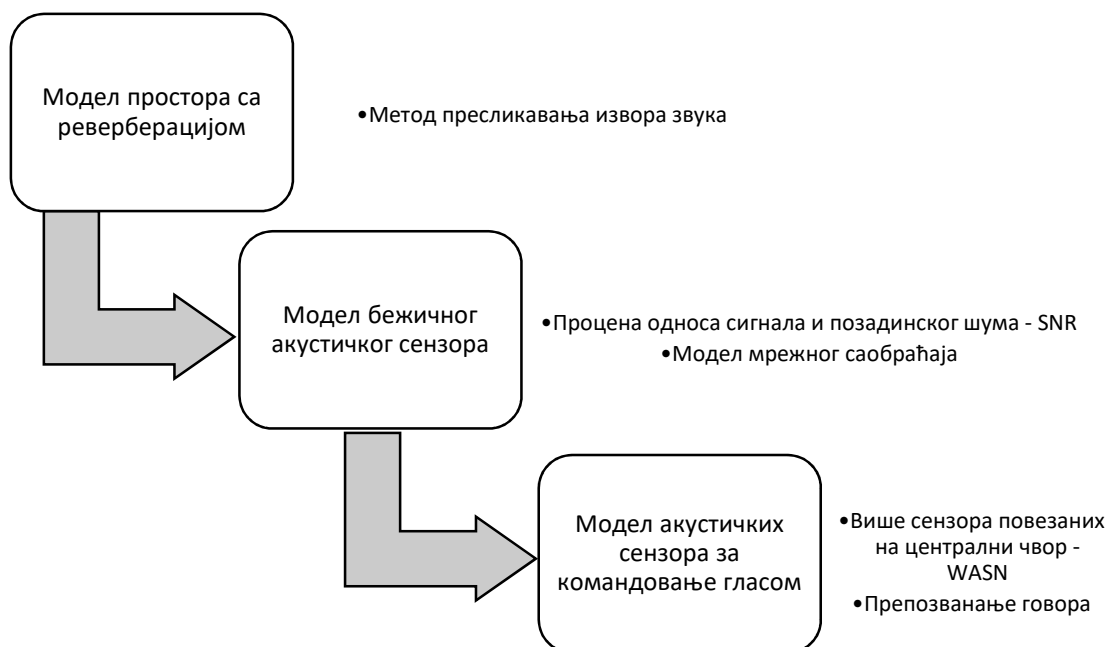
моделу редова чекања. Почетак моделовања креће од модела изворишног саобраћаја. Постојећа истраживања модела изворног мрежног саобраћаја нису примењива код *WASN*.

Модел саобраћаја података у *WSN* за детекцију уљеза базира се на геометријском распореду сензора дуж штићеног коридора [Demirkol]. У литератури се налазе истраживања која разматрају саобраћај података код сензора коришћених у медицини. Код [Messier] модел је изграђен на основу емпиријског скупа података. Оба модела су важни доприноси у овој области, али се они не могу директно користити за *WASN* који се разматра у овој дисертацији. Приступ описан у [Wang] се у одређеној мери може применити на *WASN* пошто је природа сигнала у систему за праћење слична природи звучног сигнала. Ипак, овај модел се мора даље прилагодити. Да би се добио прави модел мрежног саобраћаја у *WASN* потребно је узети у обзир простирање звука унутар простора са реверберацијом. За потребе испитивања примене мултимедијалних сензора или прецизније примену дистрибуираних акустичких сензора аутор је морао да конструише нови модел изворишног саобраћаја. Приказ развијеног модела је дат у наредном поглављу.

ПОГЛАВЉЕ 3.

МОДЕЛ БЕЖИЧНИХ АКУСТИЧКИХ СЕНЗОРА

У овом поглављу је дат преглед корака изградње модела бежичних сензора за управљање гласом. Саму мотивацију за конструисање овог модела као и разлоге за донете одлуке приликом моделовања дати су у уводу и прегледом стања у области. Модел бежичних акустичких сензора за командовање гласом у паметним кућама је настао у три корака која су приказана на илустрацији 3.1.



Илустрација 3.1 Процес изградње модела бежичних акустичких сензора

Након конструисања модела могуће је урадити низ експеримената којима се може упоредити овако представљен систем са референтним системом. У литератури која показује реализоване системе и уређаје у потрошачкој електроници постоје разни приступи овом проблему. Поређењем се може утврдити да ли је предложени модел валидан. Ово је детаљно описано у наредном поглављу 4. Поређењем предложеног система са системима познатим у пракси демонстрира се потенцијал примене оваквог система у паметним кућама ради остварења успешнијег командовања гласом.

3.1 Модел простора са реверберацијом

На основу увида у стање у области моделовања простора одабран је модел пресликавања извора као метод за моделовање простора са реверберацијом. Акустичка симулација собе је базирана на Лимановом раду који је доступан као Матлаб имплементација отвореног кода [Lehmann2]. У његовом раду је дата основа и неки примери простора са реверберацијом над којим је испитана симулација. Допринос у оквиру ове дисертације се огледа у унапређењу постојеће Матлаб имплементације [Cetic1]. Код оригиналне имплементације примећен је проблем са прекомерним коришћењем радне меморије због читавања читаве улазне датотеке. Улазна датотека може имати више канала који описују различите изворе звука и слике извора сабране у тачкама од интереса. Оригинална имплементација покушава да обради све то у једном проласку чувајући одбирке сигнала у радној меморији. Ово је у раду изазивало проблеме са недостатком меморије (енгл. *out of memory*). Да би се тај проблем решио имплементирана је блоковска обрада улазних, интерних и излазних одбирака сигнала.

Оригинални сигнал $s(t)$ се конволуцијом у фреквентном домену комбинује са симулираним импулсним одзивом собе $Hi(\omega)$. Овим се добија ефекат реверберације, односно ефекат да се у микрофону комбинује сигнал који директно долази од говорника и све рефлексije овог сигнала. На овако добијени сигнал се додаје позадински шум $n(t)$. Оваквим поступком могуће је варирати услове под којим ради препознавање говора. Метод приказан у [Cetic1] је

користан за испитивање великог броја симулираних просторија уз присуство различитог позадинског шума.

Извор сигнала смештен је у стационарну тачку. Сензори се постављају да покривају целу површину просторије у равни на висини извора звука од 1.3m. Детаљни параметри модела у симулацији дати су у табели 3.1.1. Да би се боље приказао ефекат реверберације урађене су симулације две просторије са различитим нивоима апсорпције, односно времена реверберације.

Табела 3.1.1 Параметри симулације

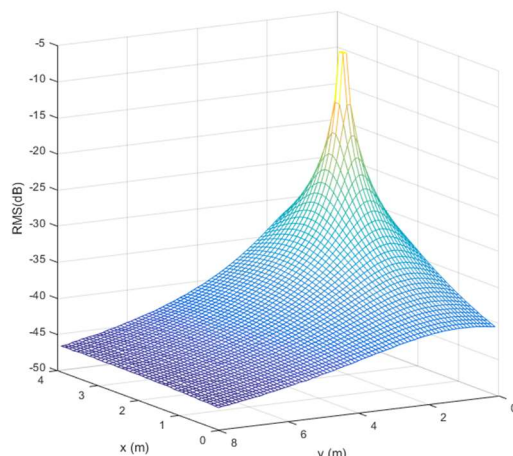
Параметар	Опис параметра	Вредност параметра
d_{room}	Димензије собе	[4 m , 8 m, 2,3 m] ^Г
p_{source}	Позиција извора звука	[3 m, 1 m, 1,3 m] ^Г
	Број постављених сензора	40 × 80 = 3200 сензора
$p_{sensor,k,j}$	Позиције сензора	k * 0,1 m; j * 0,1 m; 1,3 m
Пригушена просторија		
α	Коефицијенти апсорпције	[0,82 0,81 0,71 0,69 0,37 0,42]
RT_{60}	Време реверберације	200 ms
Уобичајени ентеријер		
α	Коефицијенти апсорпције	[0,23 0,24 0,21 0,21 0,11 0,13]
RT_{60}	Време реверберације	520 ms

Симулација се одвија тако што се на улаз система доводе различити сигнали. Ови сигнали представљају говорника и посматрани су као извор звука. У свакој тачки где се налази сензор израчунава се средња вредност квадратног корена (енгл. *Root mean square - RMS*) сигнала и изражена је у децибелима:

$$RMS_{dB} = 20 \log_{10}(RMS_y(t)), \quad (3.1)$$

где је 0 dB пун дозвољени опсег квантизованог сигнала. На илустрацијама је приказан различит ниво сигнала у зависности од релативне позиције у односу на извор звука.

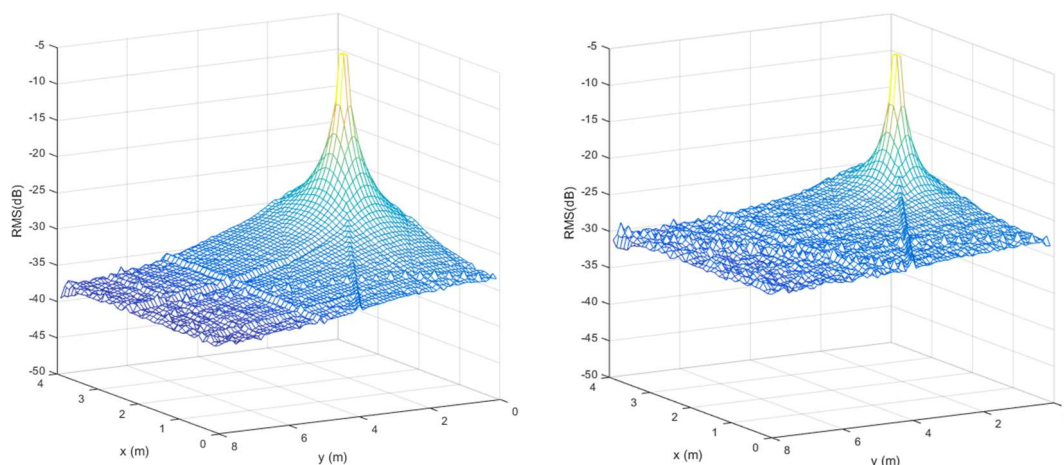
У првом случају посматра се простирање белог шума кроз простор. Илустрација 3.1.1 приказује ниво емитованог белог шума у случају када нема реверберације у просторији – симулација анехоичне собе (енгл. *Anechoic room*).



Илустрација 3.1.1 Простирања звука у простору без реверберације

Исто понашање постоји и у случају отвореног простора, односно ако не постоје препреке од којих се звучни талас одбија. На илустрацији 3.1.1. се види да ниво директног сигнала опада по закону обрнуте квадратне функције.

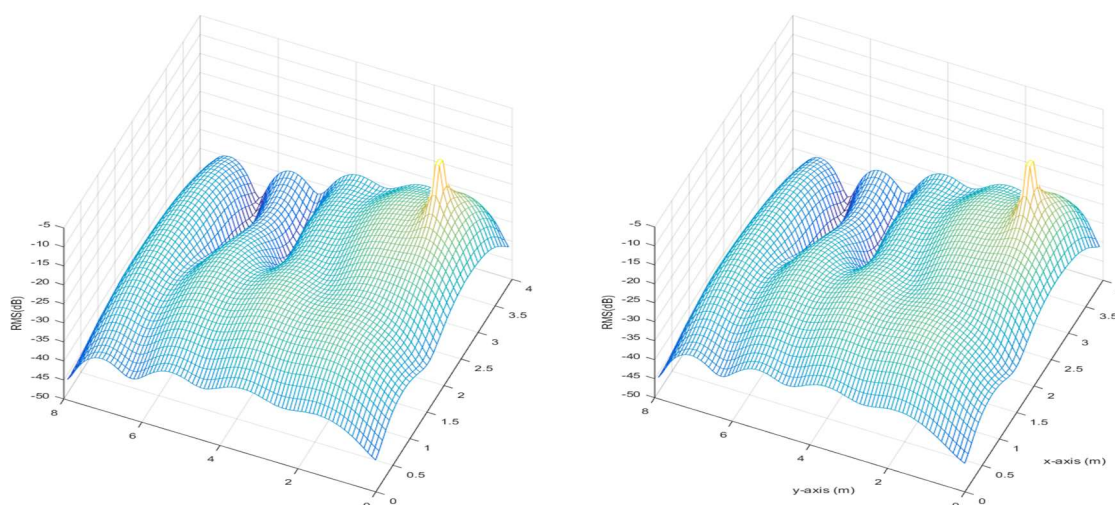
Два случаја са реверберацијом приказују симулације простирања белог шума услед ефеката реверберације са параметрима симулације датим у табели 3.1.1. Енергија шума опада спорије услед рефлексија од зидова просторије, што је приказано на илустрацији 3.1.2. Лево је приказан случај у пригушеној просторији. На слици десно је дат приказ уобичајеног ентеријера. Овим се потврђује основна исправност симулације.



Илустрација 3.1.2 Простирање звука у простору са реверберацијом

На следећим двема сликама се види како изгледа простирање звучне енергије у зависности од фреквенције сигнала на извору. Познато је да метода пресликавања извора није довољно добра за ниске фреквенције пошто не води

рачуна о дифракцији. На илустрацији 3.1.3 је приказано како изгледа простирање сигнала за синусоидни сигнал на 125 Hz (лево) и 128 Hz (десно).



Илустрација 3.1.3 Модел пресликавања извора за ниже фреквенције (125 Hz и 128 Hz)

На крају се мора нагласити да је за прецизно моделовање потребно узети у обзир и импулсне одзиве снимљене у правим просторијама - AIR. Постоје напори да се створе библиотеке ових одзива попут оног у [Szöke]. Поређење модела развијеног у овој дисертацији са доступним библиотекама може бити предмет даљег истраживања.

3.2 Модел бежичног акустичког сензора

У раду [Cetic1] је конструисан модел мрежног саобраћаја имајући у виду простирање звука у простору. У раду је претпостављено да сензорски чворови користе јефтине микрофоне слабе перформансе из економских и практичних разлога. Због саме несавршености оваквих сензора очекује се звучна слика мале тачности. Ова чињеница уводи ограничен динамички опсег сигнала и нека додатна изобличења у спектру. Сваки сензор мора да има прилагођавање његовим физичким карактеристикама, путем процене, при непрекидном снимању звука локално у сензорском чвору. Критеријум који се претпоставља је да се подаци шаљу само у присуству акустичних догађаја. Такође, императив је да *WASN* буде енергетски ефикасан. Због тога се предлаже једноставан акустични сензор са капијом буке (енгл. *Simple Acoustic Sensor with the Noise Gate - SASwNG*). Користи се алгоритам за обраду дигиталног сигнала сличан систему за надзор

путем акустичких сензора описан у [Goetze]. У овој дисертацији се користи мање рачунски-интензиван алгоритам за раздвајање акустичких догађаја од позадинског шума. Симулирајући окружење са реверберацијом испитана је процена перформансе овако добијене капије буке.

Процена снаге дискретног сигнала $P(y(t), k, N)$ над блоком од N одбирака креће над блоком на позицији k :

$$P(y(t), k, N) \approx \hat{P}(y(t), k, N) = \frac{1}{N} \sum_{t=k}^{k+N} |y(t)|^2. \quad (3.2)$$

На самом почетку рада SASwNG потребно је урадити процену позадинског шума. Процена се одвија над првих M одбирака ($t=0$; $k=0$). Овде је битно нагласити да оваква метода у правом уређају мора бити значајно сложенија пошто се у општем случају не може ослонити на чињеницу да у првих M одбирака неће бити присуства неког другог садржаја осим стационарног позадинског шума. Уколико постоји присуство нестационарног шума процена шума се ради непрекидно. Поједностављена процена изгледа овако:

$$P(n(t)) \approx \hat{P}(n(t), 0, M) = \hat{P}(y(t), 0, M). \quad (3.3)$$

Дакле, процена се ради уз претпоставку да је сигнал од интереса $s(t)$ једнак нули у трајању од M одбирака. Након тога се одређује доњи праг буке (*eng. Noise floor*) који у себи садржи и буку из окружења, али и несавршености самог микрофонског сензора. Над свим наредним блоковима може се урадити процена односа сигнала и шума (енгл. *signal-to-noise ratio - SNR*) која се рачуна у датом моменту k као:

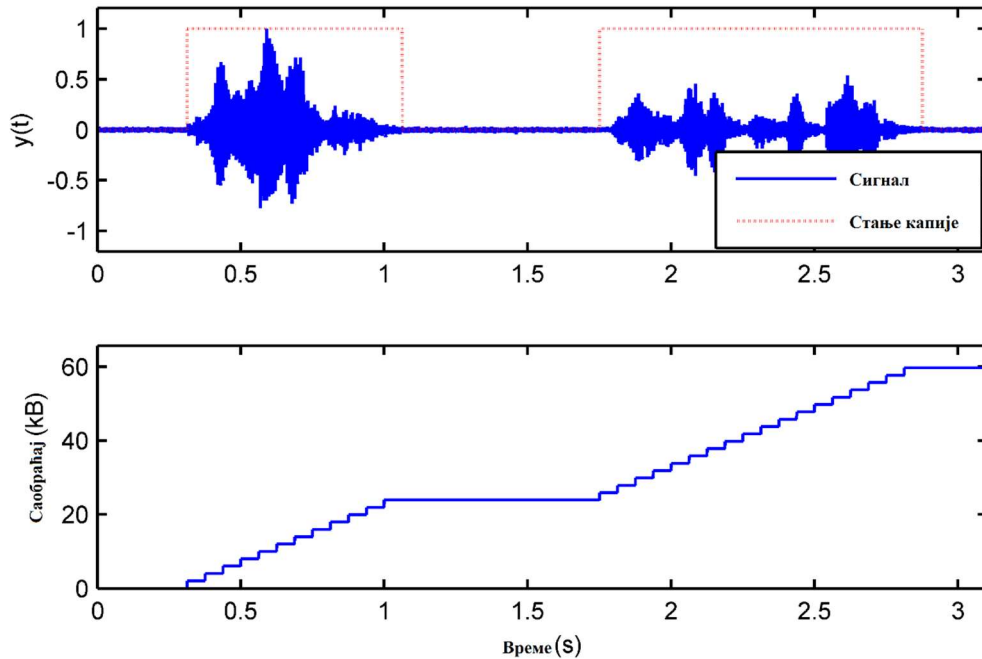
$$SNR(s(t), k, N) = 10 \log_{10} \left(\frac{\hat{P}(y(t), k, N)}{\hat{P}(n(t), 0, M)} \right). \quad (3.4)$$

Процена SNR се упоређује са параметром $SNR_{threshold}$ на основу чега се отвара и затвара капија буке. Алгоритам ради над блоковима од N одбирака који се у случају непостојања значајног акустичног догађаја мењају блоком тишине. Овде се уводе нови параметри у модел ради динамике сензора која се жели постићи. Време напада T_a (енгл. *attack time*) је време да капија буке започне пропуштање пакета. Слично томе, време отпуштања T_r (енгл. *release time*) је време за које

капија треба да заустави пропуштање пакета услед појаве сигнала чији ниво се налази испод $SNR_{threshold}$. У пракси, T_a се поставља на вредност која је по величини мања од T_r . Када су T_a и T_r постављени на нулу, алгоритам изгледа овако:

$$y(t) = \begin{cases} \forall y(t), k < t < kN, \\ y(t) & SNR(y(t), k, N) > SNR_{threshold} \\ -\infty \text{ dB} & \text{иначе} \end{cases} \quad (3.5)$$

Када се „чист“ сигнал без изобличења, односно са незнатним присуством одјека, уврсти у израз (2.20) па затим примени (2.20) добија се ефекат реверберације над тим сигналом и добија се ефекат слабљења звучног фронта. Тако добијен сигнал се уврштава у израз (3.5). Повезивањем два модела се сигнал са утицајем реверберације доводи на улаз капије буке и на тај начин се добија модел изворишног саобраћаја као што је приказано на илустрацији 3.2.1.



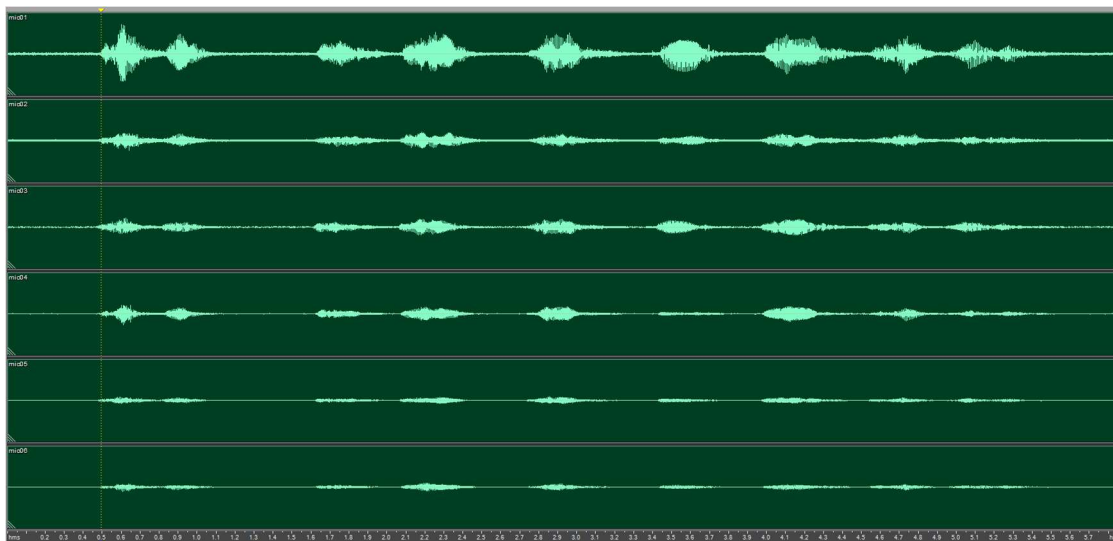
Илустрација 3.2.1 Пример изворишног саобраћаја код модела WASN

Главна идеја у овој дисертацији је употреба врло једноставног и јефтиног сензора мале процесорске моћи и мале потрошње електричне енергије. Ово се подудара са принципима који су присутни код приступа имплементације у IoT. Помоћу оваквог WASN могуће је лоцирати објекте [Holloosi2], измерити ниво

звука саобраћаја на раскрсници, препознати одређене ситуације у блиском окружењу [Kelly]. Остале важне предности бежичних звучних сензора су да се могу поставити на неприступачна места и да им не треба оптичка видљивост околине да би радила. Синхронизација акустичких сензора попут оне у [Atmoko] у случају дистрибуираних сензора није нимало лак задатак. Истраживања у овом правцу ће бити предмет даљих истраживања и она нису у оквирима ове дисертације.

3.3 Модел акустичких сензора за командовање гласом

У овом поглављу је описан модел система повезаних акустичких сензора. Рад повезаних сензора илустрован је сликом 3.3.1. На слици је приказана амплитуда сигнала шест акустичких сензора који се налазе у простору са реверберацијом. Енергије сигнала су зависне од позиције сензора у односу на говорника. Такође однос директног и рефлектованог звука зависи од раздаљине у односу на микрофон у складу са раније дефинисаним радијусом просторије. Очеvidно је да сензор који је ближе говорнику има повољнији SNR.



Слика 3.3.1 Рад повезаних акустичких сензора

Сензори раде у три режима:

- Процена позадинског шума
- Процена SNR у оквиру сензора
- Слање говорне команде

Како је приказано у поглављу 3.2 процена енергије позадинског шума се ради над блоковима након чега сензор прелази у стање процене енергије сигнала и пореди ове две енергије. Када се деси скок, односно када вредност пређе задати праг, добијени SNR се међусобно пореде. Након тога, тражи се максимум SNR и на основу њега бира се сензор. Сензор код кога је најповољнији однос прелази у стање слање команде, док се остали сензори враћају процени SNR.

У будућем раду уместо процене SNR одабир сензора је могуће радити на основу индекса сличности који се добија код препознавања кључне речи. Концепт кључне речи, или речи која буди уређај за командовање, је опште прихваћен метод код говорних спрега. Корисник се обраћа уређају тако што га ослови именом, и непосредно затим изговара реченицу коју систем за командовање треба да препозна.

ПОГЛАВЉЕ 4.

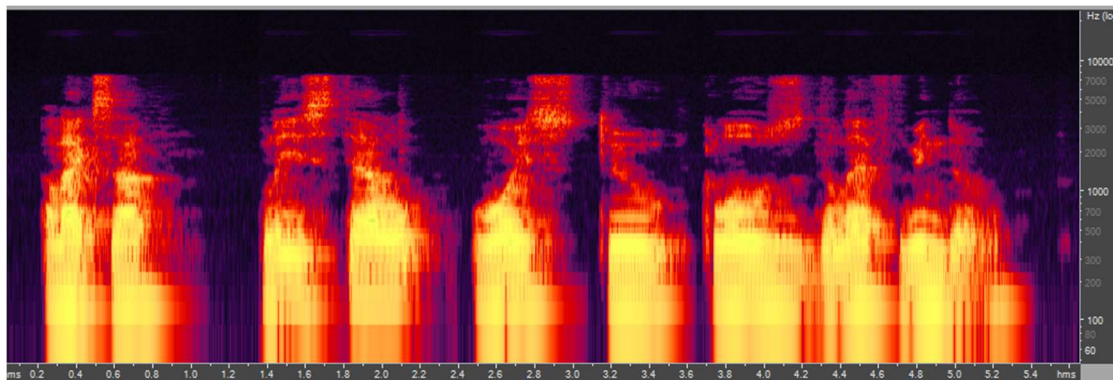
ЕКСПЕРИМЕНТАЛНИ РЕЗУЛТАТИ И ДИСКУСИЈА

У овом поглављу је приказан рад предложеног система путем симулација над изабраним узорцима дигиталног записа говора. Параметри симулације се бирају тако да описују познат животни простор. У истом простору изводе се и реална мерења чиме се потврђује исправност модела и симулације. Изграђени модел се затим користи као основа за истраживање система за командовање. Експериментални резултати су настали на основу:

- аналитичког прорачуна,
- моделовања познатог простора и симулације,
- експеримената у простору са реверберацијом.

Наш фокус је употреба сензора за препознавање гласа. Аутори у [Bouzid] у алгоритму на бази микрофонске решетке за потребе локализације извора звука предлажу учесталост одабирања од 44,1 kHz. Раније се веровало да је за потребе разумевања говора довољна учестаност од 8 kHz. Овакав сигнал се у литератури назива уско појасни (енгл. *Narrow band*). Напретком технологије обраде, преноса и складиштења говора се прешло на 16 kHz (енгл. *Wide band*). Алгоритам у овој дисертацији своди се само на процену енергије сигнала и он ради искључиво са сигналом говора. Испитивања су рађена само на учестаности 16 kHz. Пример спектралне слике једне изговорене гласовне команде дат је на слици 4.1. На слици

је приказан сигнал са учестаношћу одабирања од 48 kHz. Глас је узоркован у просторији са реверберацијом без значајног позадинског шума. Види се да фреквенција изнад 8 kHz скоро да нема. Снимак је начињен уз помоћ уграђеног микрофона у персонални рачунар.



Илустрација 4.1 Спектар изговорене команде

4.1 Провера модела изворишног саобраћаја

Модел изворишног саобраћаја се добијају спајањем модела акустичког простора са реверберацијом и модела акустичког сензора. Први експеримент се односи на проверу модела изворишног саобраћаја SASwNG. Мерена је просечна битска брзина тако што су различити сигнали довођени на улаз SASwNG, након чега су добијене величине послатих пакета. Коришћени сигнали описују три случаја:

1. чист говор без реверберације,
2. говор добијен у моделу акустичког простора,
3. говор у реалној просторији.

Овај експеримент је извођен у празној канцеларији без присуства људи са мало плочастог намештаја. Говорник и сензор су се налазили на дијаметрално супротним странама просторије да би ефекат реверберације био што израженији. На овај начин се у снимљеном сигналу види постојање дифузног поља. Параметри простора и уједно параметри симулације приказани су у табели 4.1.1.

Одређивање коефицијената апсорпције α се мора пажљиво радити. Потребно је узети у обзир различите материјале и врсту апсорпционе површине на основу којих се коефицијенти добијају. За ово испитивање је претпостављено да евентуална нетачност неће имати утицај пошто се позиција говорника и

микрофона не мењају. У том случају дозвољено је претпоставити да се ради о униформно распоређеном звуку услед истих ефеката рефлексije.

Табела 4.1.1 Параметри експеримента за испитивање изворишног саобраћаја

Параметар	Опис параметра	Вредност параметра
d_{room}	Димензије просторије	[5,3 m 2,5 m 2,67 m] ^T
p_{source}	Позиција извора звука	[5,0 m 2,2 m 1,3m] ^T
p_{sensor}	Позиција сензора	[0,2 m 0,2 m 1,3m] ^T
T_{60}	Време реверберације	899 ms
F_s	Учестаност одабирања	16000 Hz
Q	Величина одбирака	16 бита
M	Величина блока процене шума	5000 одбирака
N	Величина блока процене сигнала	од 100 до 5000 одбирака
H	Величина заглавља	40 октета = 320 бита
λ	Измерен однос корисног сигнала	0,45406915

Процена позадинског шума се обавља над првих M одбирака. Процена енергије корисног сигнала се ради над блоковима величине N одбирака. Овај параметар варира од 100 до 5000 одбирака. За сваку величину се рачуна просечна битска брзина. Ово је поновљено за сваки улазни сигнал.

Теоретски максимум битске брзине се добија за наивни случај када се сви блокови шаљу у континуитету. У генералном случају се види да на битску брзину $R(F_s, N, \lambda)$ утичу:

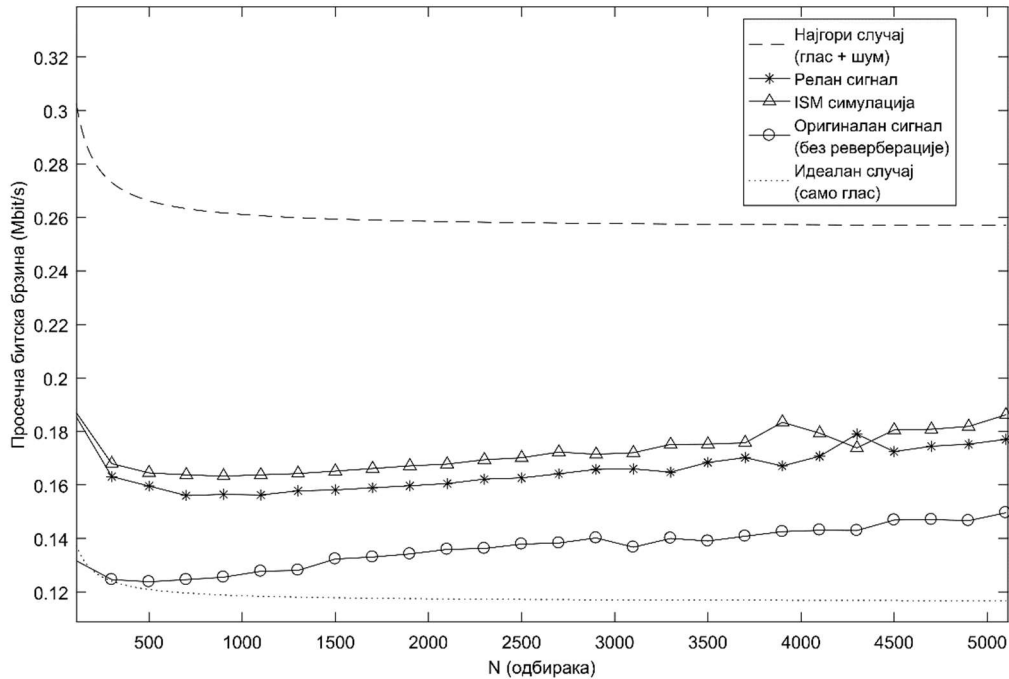
- F_s - учестаност одабирања
- N - величина блока обраде
- λ - однос дужина сигнала од интереса и позадинског шума
- H - величина заглавља датаграма
- Q – прецизност квантизације, односно ширина одбирака у битима.

Фактор λ је ручно одређен аотирањем звучног записа сваке секвенце говора која је коришћена за испитивање.

$$\frac{F_s}{N}(H + NQ)\lambda \leq R(F_s, N, \lambda) \leq \frac{F_s}{N}(H + NQ), \quad (4.1)$$

Види се да се битска брзина мора налазити између две границе што је дато изразом (4.1). Максимална битска брзина R_{max} се добија у наивном случају када се

звук шаље непрекидно, односно када су сви пакети послати. Минимална битска брзина R_{ideal} се добија уколико би SASwNG успео идеално да филтрира позадински шум и шаље само сигнал од интереса. R_{ideal} се рачуна тако што се R_{max} множи фактором λ . Величина заглавља H је у случају RTP датаграма фиксне дужине и индиректно утиче на промену брзине због варирања величине блока обраде. Резултати експеримента су приказани на слици 4.1.1.



Слика 4.1.1. Утицај реверберације на изворишни саобраћај

Резултати мерења потврђују симулирани модел у односу на сигнал добијен у правој просторији. Може се приметити да се брзина незнатно повећава са повећањем N . Узрок овог ефекта је немогућност поравнања блокова обраде са кључним временима у сигналу попут, почетка и завршетка гласних деоница. Овај недостатак се може потиснути уколико се процена ради на мањим блоковима и техникама прозорирања. Међутим, ово би алгоритам у SASwNG учинило сложенијим. За случај који је приказан, добијена је уштеда од 43,78% у односу на R_{max} . Експеримент такође показује како ефекат реверберације неповољно утиче на уштеду.

Увидом у литературу уочава се да је могуће унапредити перформансу алгоритма капије буке [Atmoko] али уз коришћење више микрофона. Пошто је

намера аутора била да прво испита могућност примене врло једноставних сензора даља унапређења су остављена за будући рад.

4.2 Провера модела мреже бежичних акустичких сензора

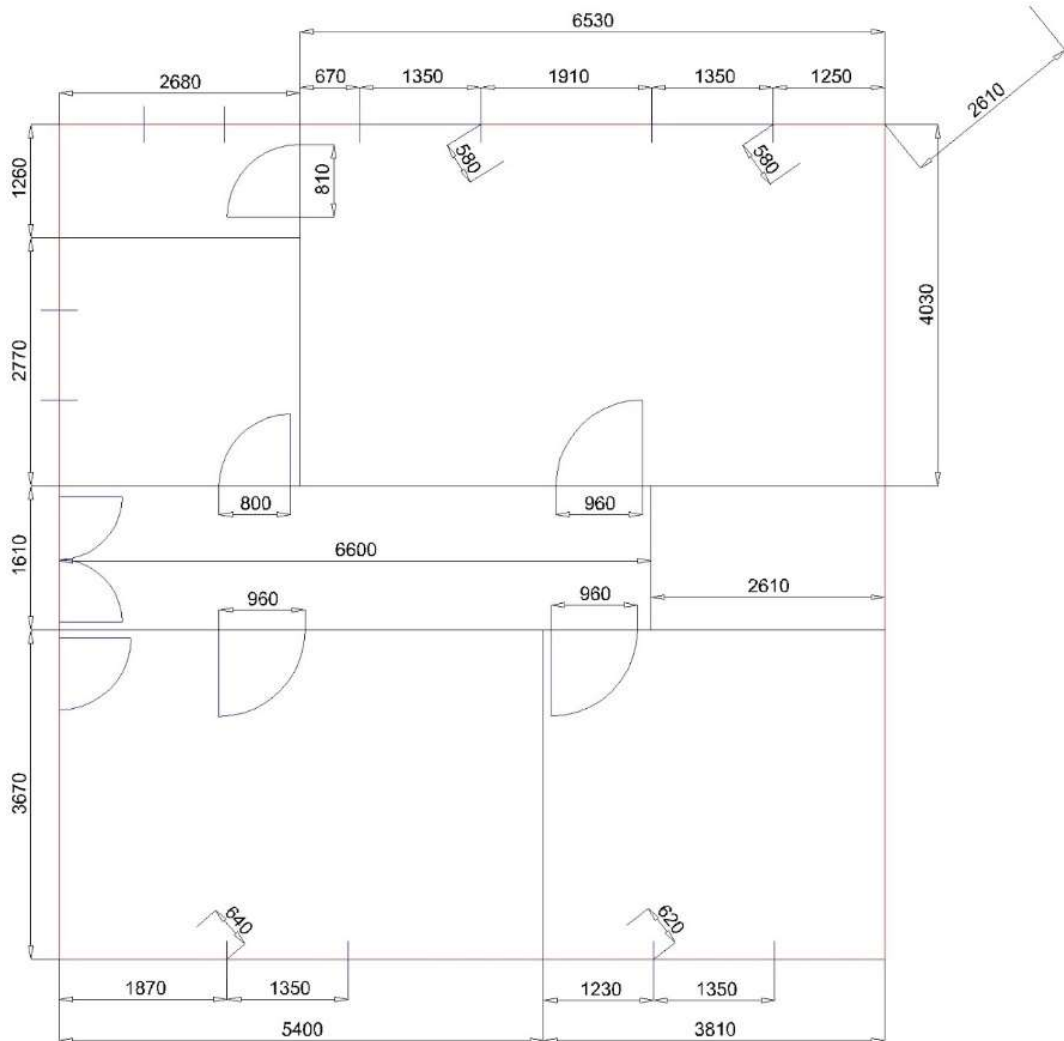
За испитивање мреже бежичних сензора потребно је одабрати над којим корпусом извести испитивање. Пошто је потребно испитати утицај реверберације за експерименте су потребни репрезентативни узорци снимљеног гласа у посебним условима. Претрагом јавно доступног садржаја одабран је скуп “чистог” говорног сигнала снимљеног у глувој соби из корпуса [ACE]. Састоји се из гласа пет мушких и пет женских говорника са различитим дужинама реченица.

Као референтни систем за аутоматско препознавање говора користили смо Амазон Транскрајб сервис и систем за испитивање гласовних команди [Rajić]. Прво су испитане грешке код препознавања „чистог” сигнала да би се омогућило поређење и утврдио утицај реверберације. Над 2912 речи добијена је грешка од 17,31%, односно 504 грешке. Детаљи препознавања појединачних датотека су дати у прилогу А.

Провера модела бежичних акустичких сензора је урађена у животном простору људи, односно стамбеном простору једне етажне вишеспратне породичне куће. Тлоцрт етажне је приказан на слици 4.2.1. За симулацију је одабрано пет просторија са различитим акустичким својствима. Мања помоћна просторија (горе лево) и степениште (средина десно) нису узете у разматрање.

За детаљан опис акустичког простора потребно је одредити односе апсорпционих коефицијената и уврстити их у коришћени модел [Cetic1]. Ово су заправо тежински фактори који се користе код описа зидова, пода и плафона. Аутор се ослонио на јавно доступна мерења апсорпције одређених материјала у [Davis]. Узете су у обзир димензије зидова, пода и плафона, односно њихова укупна површина. Од површине зидова се одузимају површине свих површина који су уграђени у зид попут прозора и врата или других плочастих материјала. Свакој идентификованој површини додељен је коефицијент апсорпције из [Davis]. Исто је урађено и за део непрекривене површине зида, плафона и пода. Овако добијени парови коефицијената и површина су уврштени у израз (2.5) и

тима је добијено шест параметара који описују акустички простор. Добијени коефицијенти апсорпције су дати у табелама у додатку В.



Илустрација 4.2.1 Простор у којем су се одвијали експерименти и мерења

Развијени модел има низ предности и даје могућност за обимна испитивања која би у реалном физичком простору било теже извести. Једна од предности дата је на примеру одређивања броја микрофона на основу симулације. Другим речима, одговара се на питање колико је сензора потребно поставити у одређену просторију да би се постигао оптимум у погледу озвучености простора ради успешног препознавања говорних команди. Дат је приказ експеримената у симулацији којим је број микрофона утврђен.



Слика 4.2.2 Фотографије простора са реверберацијом

У табели 4.2.2 су дати параметри симулације који су добијени на основу измерених величина физичког простора дневне собе приказане на слици 4.2.2 (горе). На слици доле лево је приказана просторија ходник онако како свакодневно изгледа. У просторијама које су биле део испитивања ово је била просторија са највише реверберације где је $RT60$ уносило 690 ms. Изглед просторије са уклоњеним дебелим кончаним тепихом је приказан доле десно. Након уклањања тепиха ефекти реверберације су били још израженији. Сва мерења су дата у прилогу Б. У табели 4.2.1 приказани су запремина просторије и $RT60$ на основу медијане десет поновљених мерења.

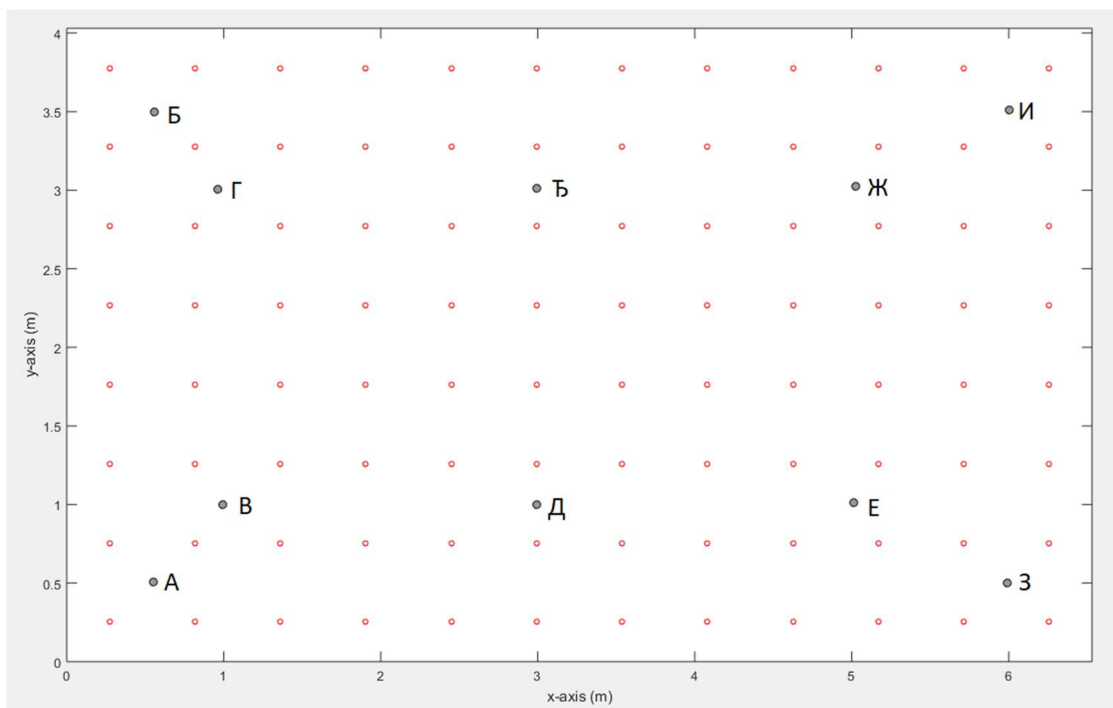
Табела 4.2.1 Време реверберације у просторијама

Р. Б.	Опис, намена	Запремина	$RT60$
1	Дневна соба	68,68 m ³	520 ms
2	Спаваћа соба	51,13 m ³	435 ms
3	Ходник	27,73 m ³	855 ms
4	Купатило	19,38 m ³	460 ms
5	Дечија соба	36,49 m ³	320 ms

Намера је испитивање перформансе система за издавање команди из било које тачке у просторији. Експеримент је испитао понашање система сензора са једним говорником чија се позиција мења тако да покрива целокупну површину просторије. Сензори су постављени тако да су сагледана физичка ограничења у реалној просторији која се моделује. На овај начин је добијен практичан распоред сензора који је могуће постићи без ометања укупана у њиховим нормалним активностима. Време реверберације је одређено на основу мерења у физичком простору. Распоред сензора и симулираних говорника је приказан на слици 4.2.3 док су тачне координате дате у табели 4.2.2.

Табела 4.2.2 Параметри симулације у просторији 1

Параметар	Опис параметра	Вредност параметра
d_{room}	Димензије собе	[6,53 m 4,03 m 2,61 m] ^T
	Број извора звука	12 × 8 = 96 позиција
$p_{source, i, j}$	Позиције извора звука	[0,27 + I * 0,54 m 0,25 + j * 0,5 m 1,3 m]
$p_{sensor, A}$	Позиције сензора	[0,5 m 0,5 m 1,3 m] ^T
$p_{sensor, Б}$	-	[3,5 m 0,5 m 1,3 m] ^T
$p_{sensor, B}$	-	[1 m 1 m 1,3 m] ^T
$p_{sensor, Г}$	-	[3 m 1 m 1,3 m] ^T
$p_{sensor, Д}$	-	[3 m 1 m 2,51 m] ^T
$p_{sensor, Ђ}$	-	[3 m 3 m 2,51 m] ^T
$p_{sensor, E}$	-	[5 m 1 m 1,3 m] ^T
$p_{sensor, Ж}$	-	[5 m 3 m 1,3 m] ^T
$p_{sensor, З}$	-	[6 m 0,5 m 1,3 m] ^T
$p_{sensor, И}$	-	[6 m 3,5 m 1,3 m] ^T
α	Коефицијенти апсорпције	[0,2417 0,3022 0,3022 0,3022 0,6715 0,3022]
RT_{60}	Време реверберације	520 ms



Слика 4.2.3 Позиције говорника и сензора у симулацији

Симулацијом простора са реверберацијом користећи параметре из табеле 4.2.2, за сваку улазну датотеку добијено је 960 излазних датотека. Појединачна датотека садржи другачије акустичке услове у погледу односа директног и дифузног звука. Након тога је за сваку датотеку урађено аутоматско препознавање говора и израчунати су WER.

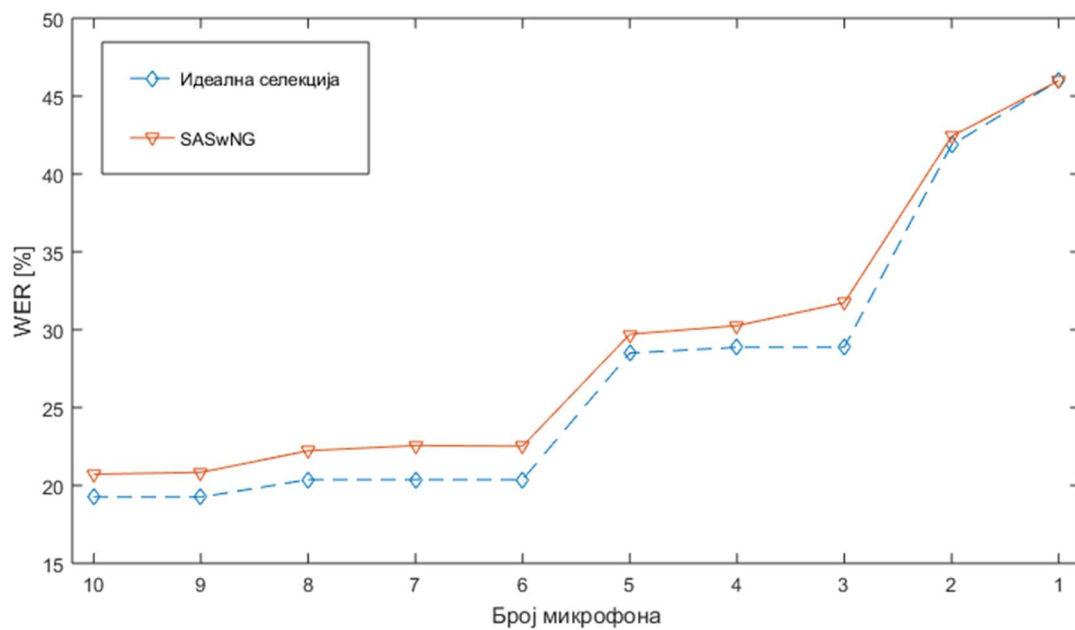
Алгоритам за процену SNR не даје увек идеалне резултате у односу на позицију и раздаљину говорника и сензора. У табели 4.2.3. су приказани одабири најповољнијег сензора на основу процене SNR.

Табела 4.2.3 Грешка одабира микрофона на основу SNR

i,j	1	2	3	4	5	6	7	8	9	10	11	12
1	Б	Б	БГ	ГЂБ	ЂГ	Ђ	ЂЖ	ЖЂИ	ЖИЂ	Ж	И	И
2	Б	БГ	ГБ	ГЂ	Ђ	Ђ	Ђ	ЖЂ	Ж	Ж	ИЖ	И
3	ГБ	Г	Г	ГЂ	Ђ	Ђ	Ђ	ЖЂ	Ж	Ж	Ж	ИЖ
4	ГБВ	ГБВ	ГВ	ГЂВД	ЂДГВ	ЂД	ЂДЖ	ЖЂДЕ	ЖЕ	ЖЕ	ЖЕИ	ИЖ
5	ВАГ	ВГАБ	ВГ	ВДГЂ	ДЂВГ	ДЂ	ДЂЕЖ	ЕДЖЂ	ЕЖ	ЕЖ	ЕЗЖ	ЗЕИЖ
6	ВА	В	В	ВД	Д	Д	Д	ЕД	Е	Е	ЕЗ	ЗЕ
7	АВ	ВА	ВА	ВД	Д	Д	Д	ЕД	Е	Е	З	З
8	А	А	ВА	ВДА	ДВА	Д	ДЕ	ЕД	ЕЗ	ЕЗ	З	З

На примеру експеримента за извор звука $p_{source, i, j}$ где је $i = 4$ и $j = 4$ види се да су бирани сензори „Г“, „Ђ“, „В“ и „Д“. Ово за последицу има да успешност препознавања говора у односу на одабрани микрофон не мора дати најповољнији могући резултат.

На основу ове симулације се може одредити колико је сензора потребно да би се добили добри резултати препознавања говора. Слика 4.2.3 приказује како грешка препознавања говора расте како се број сензора смањује. Априори одабир сензора на основу процене SNR се разликује од постериорних одабира на основу свих познатих грешака препознавања – случаја идеалне селекције. На слици се види разлика која долази из несавршености процене SNR а самим тим и одабира сензора. На примеру експеримената у просторији 1 WER за сваког сензора је испитан на више улазних сигнала за сваку од 96 позиција.



Слика 4.2.3 Одређивања броја потребних микрофона

У табели 4.2.4 су дати резултати одређивања броја микрофона код експеримената у пет просторија, намена просторија и грешке у изведеним експериментима. Детаљи експеримената дати су у прилогу В.

Табела 4.2.4 Грешке препознавања говора у изведеним експериментима

Просторија:	1	2	3	4	5
Намена:	Дневна соба	Спаваћа соба	Ходник	Купатило	Дечија соба
Број сензора:	6	5	5	3	3
Грешке препознавања – WER [%]					
Симулација:	20,91	24,79	42,84	30,02	22,35
Реални простор:	21,81	25,89	44,58	30,62	23,52

4.3 Командовање гласом - студија случаја

Након провере реализованог модела и система бежичних сензора, приступљено је испитивању система за командовање гласом у паметним кућама. Као пример референтног система који је комерцијално доступан на тржишту и сматра се за последњу реч технике (енгл. *state-of-the-art*) узет је паметни звучник, уређај компаније Амазон под називом Ехо Дот (енгл. *Echo Dot*). Подршка за развој уређаја коју нуди Амазон се показала веома погодно у току истраживања.

Експерименти су извођени у раније поменутој просторији 1, дневној соби у кући у којој су извођени сви експерименти. Ради директног поређења, паралелно је постављен Ехо Дот уређај и систем дистрибуираних микрофона - СДМ тако да говорник издаје команде и једном и другом систему истовремено.

СДМ је пријављен у систем као развојни уређај. Све интеракције са системом се чувају на серверима компаније Амазон и доступне су кориснику путем посебних веб сајтова. Корисник има контролу над њима и може затражити њихово брисање. Такође, за сваку изговорену команду дат је текстуални запис препознатог говора што је послужило за рачунање грешке код аутоматског препознавања.

На почетку је посматран рад референтног система. Корисник започиње интеракцију са виртуалним асистентом тако што изговара кључну реч, односно ословљава асистента именом. У нашем случају коришћено име асистента је женско име Алекса. Препознавање кључне речи се ради посебним техникама запажања речи и одвија се на самом уређају [Raju2]. Ова техника је од недавно доступна и за развојне уређаје. Запажање кључне речи је специфичан проблем и

носи са собом посебне проблеме у виду сметњи од стране телевизијског апарата, буке других уређаја у кући, репродукције музике на самом паметном звучнику и другог говора који није намењен систему за командовање. Веома је тешко реализовати систем који има низак ниво лажног односно нежељеног препознавања (енгл. *false alarm rates - FAR*) или одбацивања препознате речи (енгл. *false rejection rate - FRR*) [Raju2]. У оквиру студије случаја, у периоду од годину дана утврђен је FAR у износу од 16,25% у односу на укупан број исправно уочених кључних речи. Ово индицира да уочавање кључне речи користи другачији алгоритам од онога за препознавање команди. У случају FAR, референтни уређај је и даље у стању да исправи грешку и обустави даље препознавање говора. Пошто је препознавање кључне речи доступно само на референтном систему и није реализовано на СДМ, случајеви са FAR и FRR као и присуство кључне речи су изузети из поређења перформанси система.

Испитивања оба система су изведена на основу 6525 изговорених речи при издавању команди. У испитивању су учествовали аутор и његови укућани. Аутор је издао 68 % од укупних команди, односно толики проценат одлази на мушког одраслог говорника. Одрасла женска особа у укупном броју учествује са 11,49 % док остатак од 20,47 % чине команде деце различитог узраста. Све команде су издаване на енглеском језику. Гласовне команде су препознаване на два одвојена система са посебним корисничким налозима да би се лакше утврдиле разлике. Свака команда је класификована у три случаја у зависности од позиције говорника при издавању команде у односу на референтни уређај. Испитани су случајеви блиског говора (енгл. *close talk*), блиског поља (енгл. *near-field*) и далеког, дифузног поља (енгл. *far-field*). Мора се нагласити да број испитних случајева у ове три категорије није био равномерно расподељен. Предност је дата испитивањима проблема далеког дифузног поља. Однос испитних случајева је дат у табели 4.3.1 Резултати препознавања говора код оба система су приказани у табели 4.3.2.

Табела 4.3.1 Однос испитних случајева

Close talk	Near-field	Far-field
17,35 %	17,59 %	65,06 %

Табела 4.3.2 Грешке препознавања говора код студије случаја

	Close talk	Near-field	Far-field	Укупно
Ехо Дот	12,99 %	14,72 %	26,03 %	21,78 %
СДМ	21,47 %	22,47 %	21,92 %	21,96 %

На основу добијених резултата видимо да је постигнут укупни WER приближан оном код комерцијалног уређаја. Односно перформанса СДМ је у рангу комерцијалног решења када се посматрају сви испитни случајеви. Оно што се мора истаћи је да је систем остварио боље препознавање код случајева дифузног поља за 4,12 % чиме је потврђена постављена хипотеза.

Мора се нагласити да је у овом истраживању разматран случај са само појединачним говорником у једном тренутку. Проблем са више говорника, који истовремено говоре, је познат проблем у литератури. Дистрибуирани микрофони могу бити предност код решавања овог проблема. По сазнањима аутора овај приступ није довољно истражен те је даље разматрање остављено за неку другу прилику. Испитивани систем је далеко од готовог решења. Питања у погледу даље интеграције са мултимедијалним системом и остатком уређаја у паметној кући нису била предмет овог истраживања.

У литератури су познате многе технике побољшања снимљеног говора. Почевши од алгоритама за потискивање позадинског шума до оних који воде рачуна о разумљивости говора и укупном квалитету звучног записа. Примена ових техника код појединачних акустичких сензора може додатно побољшати резултате.

ПОГЛАВЉЕ 5. ПОВЕЗАНА ИСТРАЖИВАЊА И РЕЗУЛТАТИ

У току рада на овој дисертацији аутор је радио на неколико научно истраживачких пројеката и као запослени на институту РТ-РК. У овом поглављу ће бити приказана истраживања која су омогућена радом на овој дисертацији. Ови сегменти истраживања су градивни блокови којима се постепено граде елементи дисертације и даје практична потврда изнетој хипотези и објашњење изнетим ставовима.

У току рада на дисертацији аутор се бавио применом постојеће технологије код проблема командовања гласом. Први резултат у овој области примене препознавања говора је истраживање обављено 2012. године које обухвата два прототипа којима се командује гласом. Први прототип даљинског управљања телевизијским пријемником демонстрира контролну петљу у којој корисник може да бира канале, мења ниво звука и искључи пријемник. Други прототип успешно демонстрира интеракцију са видео игром. Детаљи ових експеримената су дати у овом поглављу. Код обе примене задатак аутоматског препознавања гласа се одвија уз помоћ клауд технологија односно у удаљеном рачунару у облаку (енгл. *cloud computing*). Након тога дат је приказ реализације система дистрибуираних акустичких сензора.

У наредним под поглављима дат је приказ публикација каје садрже делове истраживања везаних за саму тезу дисертације. Уз сваки објављени резултат свакако остаје простор за дискусију и даљи наставак истраживања.

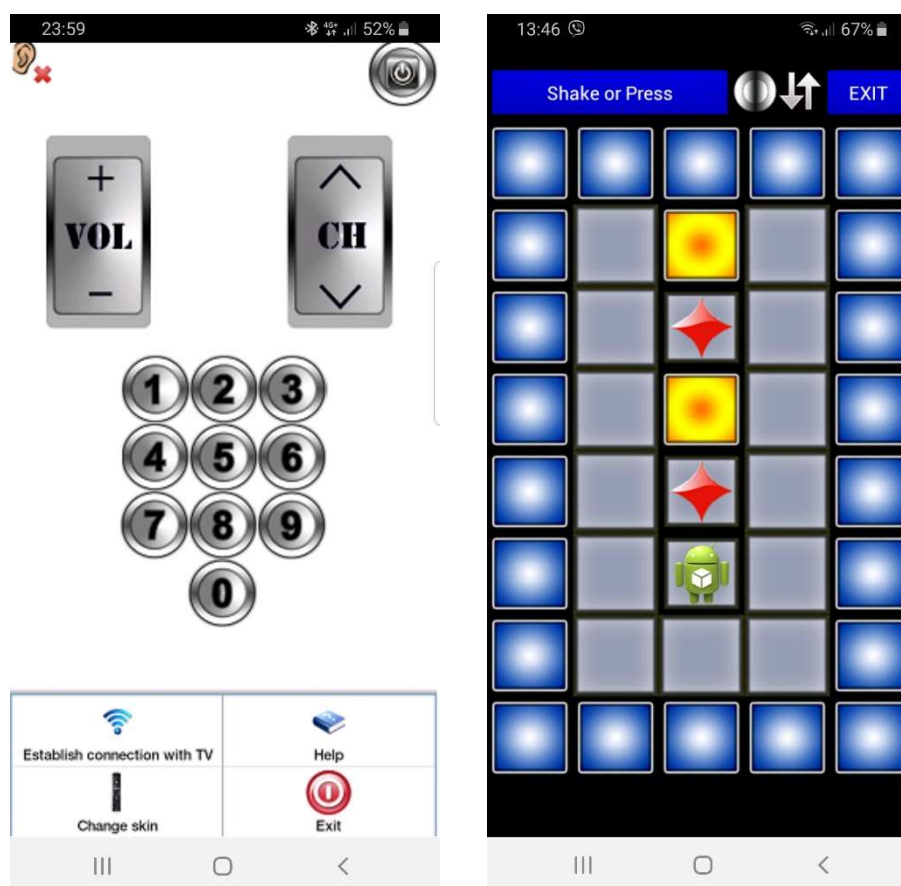
5.1 Командовање гласом код Андроид апликација

У оквиру публикације [Stefanovic] види се приказ реализације система за управљање гласом базиран на сервису за препознавање говора компаније Гугл. Ово је прва публикација у оквиру које се аутор бави применом говорних технологија. Примарни циљ је истраживање могућности интеграције и примене гласовне спреге за командовање. Командовање је реализовано помоћу аутоматског препознавања говора и његовог претварања у текст. Сервис за препознавање је иначе доступан за Андроид базирани системи али и на другим платформама путем веб спреге. Поред саме потврде примене дат је допринос кроз предлог побољшања препознавања на основу додатне обраде добијеног текста. Овим је дошло до бољих перформанси система за командовање гласом у погледу тачности односно разумевања изговорених команди.

Процес препознавања говора је сложен процес и захтева одговарајући хардвер. Захваљујући клауд технологијама [Sullivan] могуће је користити препознавање говора на уређајима који не поседују процесор високих перформанси нити сложене алгоритме обраде. Наиме, потпуна обрада снимљеног говора се извршава на удаљеном рачунару. Ово примењује компанија Гугл и нуди као услугу за претварање говора у текст. Приступ овој услузи омогућен је свим корисницима Андроид платформе и интернет претраживача путем *HTTP* протокола и *gRPC* спреге. Овај сервис могуће је користити и на платформи са веома ограниченим ресурсима како је описано у поглављу 5.3.

Поступак имплементације две Андроид апликације је описан у даљем тексту. Показало се да је реализација једноставна, брза и довољно поуздана. Ово је постигнуто тако што се захтевна обрада попут аутоматског препознавања говора одвија у клауд инфраструктури која је доступна. На Андроид уређају ипак остаје простор за даље побољшање препознавања команди што је уједно и главни допринос у [Stefanovic].

Развијене су две интерактивне апликације које имају циљ испитивања могућности командовања гласом. Прва Апликација симулира рад даљинског управљача телевизијског пријемника. Друга Андроид апликација је рачунарска игра у којој играч врши померање графичких елемената на екрану да би решио логички задатак. Апликације су написане у програмском језику Јава, применом објектно оријентисаног програмирања заснованог на коришћењу Андроид системске подршке. Коришћено је Еклипс развојно окружење помоћу ADT проширења за развој Андроид апликација са подршком за специфичне развојне алате намењене циљној платформи. На илустрацији 5.1.1 је приказана графичка спрега реализованих апликација.



Илустрација 5.1.1 Примери Андроид апликација

Други корак представља реализација модула побољшаног препознавања гласовних команди. Основна идеја је да се на основу претраге у врло ограниченом простору команди лоше препозната реч одреди на основу поређења добијеног текста са подржаним командама у апликацији. Добијени превод се састоји од низа

карактера који се захтева од STT сервиса. Сваки стринг тог низа представља једну хипотезу који је ASR препознао на основу изговорених речи или фразе. Да би се изабрала команда са највећим процентом поклапања у односу на оригинално изговорене речи неопходно је рашчланити добијени израз на низ речи, а затим их појединачно обрадити. У апликацији симулатора даљинског управљача команде су поређене Левенштајновим растојањем [Levenshtein] док се у апликацији са логичком игром, препознате команде пореде помоћу алгоритма за тражење најдужег заједничког под низа (енгл. *Longest common subsequence – LCS*). Сличан приступ се види у [Lukić] код проблема претраге садржаја телевизијског програма.

Ради илустрације простора претраге у следећим табелама је дат списак команди са израчунатим растојањима по ове две метрике.

Табела 5.1.1 Левенштајново растојања код апликације даљинског управљача

	channel down	channel up	turn off	mute	stop	volume down	volume up
channel down	0	4	9	11	11	6	10
channel up	4	0	9	9	9	10	6
turn off	9	9	0	7	7	8	8
mute	11	9	7	0	4	9	7
stop	11	9	7	4	0	10	8
volume down	6	10	8	9	10	0	4
volume up	10	6	8	7	8	4	0

Табела 5.1.2 Левенштајново растојање код апликације логичке игре

	down	up	left	right	stop	undo	reset	exit
down	0	4	4	5	4	4	5	4
up	4	0	4	5	3	3	5	4
left	4	4	0	4	4	4	3	3
right	5	5	4	0	5	5	3	4
stop	4	3	4	5	0	4	5	4
undo	4	3	4	5	4	0	5	4
reset	5	5	3	3	5	5	0	3
exit	4	4	3	4	4	4	3	0

Табела 5.1.3 LSC растојања апликације даљинског управљача

	channel down	channel up	turn off	mute	stop	volume down	volume up
channel down	12	8	3	1	1	6	2
channel up	8	10	2	1	1	2	4
turn off	3	2	8	1	2	3	2
mute	1	1	1	4	1	2	2
stop	1	1	2	1	4	1	2
volume down	6	2	3	2	1	11	7
volume up	2	4	2	2	2	7	9

Табела 5.1.4 LSC растојања код апликације логичке игре

	down	up	left	right	stop	undo	reset	exit
down	4	0	0	0	1	2	0	0
up	0	2	0	0	1	1	0	0
left	0	0	4	1	1	0	2	2
right	0	0	1	5	1	0	2	2
stop	1	1	1	1	4	1	2	1
undo	2	1	0	0	1	4	0	0
reset	0	0	2	2	2	0	5	2
exit	0	0	2	2	1	0	2	4

За потребе испитивања и верификације решења направљен је мали корпус са снимљеним гласовним командама. У испитивању је учествовало 20 особа. Оне су изговориле по 10 реченица са по 10 команди на енглеском језику. Команде можемо видети у табелама 5.1.1, 5.1.2. Овде треба напоменути да се ради о особама којима енглески није матерњи језик. Резултати успешности препознавања команди код обе апликације су приказани у табелама 5.1.5. и 5.1.6.

Табела 5.1.5 Грешке препознатих команди код апликације даљинског управљача

Команда:	channel down	channel up	turn off	mute	stop	volume down	volume up
Гугл	71,54%	58,00%	29,24%	71,54%	0,00%	29,24%	14,62%
ESASR	29,24%	0,00%	0,00%	14,62%	0,00%	0,00%	0,00%

Табела 5.1.6 Грешке препознатих команди код апликације логичке игре

Команда:	down	up	left	right	stop	undo	reset	exit
Гугл	71,43%	58,58%	28,58%	71,43%	0,00%	28,58%	14,29%	71,43%
ESASR	28,58%	14,29%	0,00%	14,29%	0,00%	0,00%	0,00%	28,58%

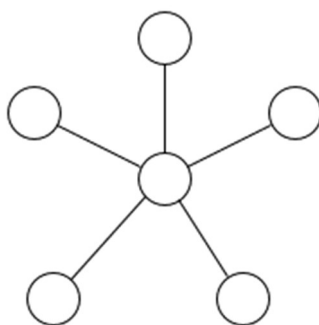
На основу резултата добијених додатном обрадом показано је да се овом техником добијају бољи резултати код препознавања команди. У зависности од позадинског шума, јачине говора, положаја микрофона и акустичких услова просторије, погрешни изговори речи или другачијег акцента команде неће бити препознате у потпуности.

Препознавање уз помоћ обраде говора у клауд сервису пружа лаку реализацију говорне спреге. Овај рад је потврдио концепт и испитао могућности примене клауд технологија. Савремени паметни мобилни телефон није добар пример уређаја са ограниченим ресурсима. Његова процесна моћ у многоме превазилази једноставне сензоре који су главни предмет ове дисертације. У литератури су познате реализације говорних асистената које целокупну обраду извршавају на доступним ресурсима модерних телефона [ВРовић2].

5.2 Реализација система дистрибуираних акустичких сензора

У раду [Doder1] дат је опис сваке компоненте у систему који садржи сервер и чвор бежичног микрофона. Сервер апликација темељи се на LIVE555 библиотеци за мрежни пренос мултимедије. LIVE555 библиотека већ подржава RTSP и RTP протокол и пренос различитих аудио формата. Ради верификације решења и испитивања перформанси, било је потребно направити бежични микрофон који ће деловати попут акустичког сензора.

WASN имплементација је веома изазован задатак, посебно када се користи у комбинацији са алгоритмима за обраду у реалном времену [Tozlu1]. Код ове примене коришћена је мрежа са топологијом звезде као на слици 5.2.1. Микрофонски чворови WASN преносе свој сигнал на уређај централног сервера где се обрађују сви сигнали.

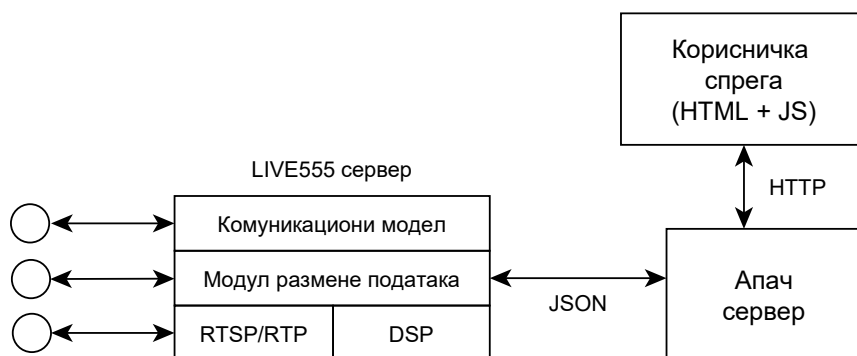


Слика 5.2.1 Пример топологије звезде

Урађена је анализа звучног кодека. Избор одговарајућег аудио формата је веома важан. Код [Toulouris] се види пример кодека за ZigBee мрежу. Компресија утиче на потенцијално загушење мреже и потрошњу енергије сензорског чвора [Tozlu2]. У [Doder1] комуникација се одвија путем 16 битних одбирака на 16 kHz учестаности одабирања. Користи се један канал јер сензорски чвор садржи један микрофон. Овим форматом, једном сензору је потребно 31,25 kbit/s пропусне брзине података за рад у нормалним условима. Компримовани формат је коришћен у каснијим фазама рада у оквиру дисертације.

Комуникација између сензорских чворова и централног сервера се одвија путем два протокола. Први је RTSP протокол на нивоу апликације за контролу испоруке података при комуникацији у реалном времену [Schulzrinne1]. Овај протокол се директно не бави преносом битских токова. У нашем случају се користи као протокол руковања између чворова бежичног микрофона и сервера. RTP протокол је коришћен за пренос података у реалном времену, као што је, у нашем случају, снимљени звук [Schulzrinne2]. RTP пружа услуге доставе са краја на крај и укључује идентификацију типа података, нумерисање редоследа и временско жигосање. RTP се обично ослања на UDP (енгл. *User Datagram Protocol*). Величина UDP датаграма може утицати на перформансе мреже. Употреба премалог пакета може довести до нежељених трошкова при преносу. Ово је детаљније објашњено у поглављу 4.

Реализовани сервер за акустичке сензоре део је укупног система као што је приказано на слици 5.2.2. Сервер се ослања на функционалност веб сервера Апач (енгл. *Apache*) како би путем корисничке спреге омогућили приступ обрађеним вредностима. На овај начин, користећи интернет претраживач, корисник је у могућности да надгледа активности на сензорској мрежи.



Слика 5.2.2. Архитектура система

Кориснички интерфејс је реализован у виду веб странице. На веб страници се налази приказ свих активних сензора са графиконом процене нивоа звука и додатним информацијама о сензорима. Подаци на графиконима се ажурирају у реалном времену. Предност ове реализације је могућност праћења од стране више корисника на удаљеним локацијама истовремено.

Намена Апач сервера је приступ веб страници путем које се могу пратити активности. Може се налазити на удаљеном рачунару независно од сервера за акустичке сензоре. Апач сервер се ослања на комуникациони модул од кога тражи информације о сензорима и тренутним читавањима сензора када је то неопходно.

Комуникациони модул служи за комуникацију са Апач сервером и испоручује тражене информације у JSON формату. За прикупљање података о акустичким сензорима овај се модул ослања на модул за размену података. Овај модул омогућава одвајање програмске подршке за приказ веб странице и обраду података са сензора.

Модул за размену података комбинује информације о сензорима које пружа RTSP модул и резултате обрађених сигнала које пружа DSP модул. Представља базу података повезаних сензора и чува резултате различитих алгоритама обраде за појединачни сензор. Посебан модул омогућава лакшу имплементацију могућих проширења.

Модул RTSP је одговоран за успостављање везе са сензорима помоћу RTSP протокола. RTSP модул је заправо посебан сервер који послушкује на дефинисаном порту и чека да сензори пошаљу захтев за повезивање. Након пријема захтева, сервер започиње даљу комуникацију до успостављања коначне

везе. Након успешне успоставе везе, започиње се пренос звучног сигнала. У овој имплементацији се користе фиксне адресе сервера. Касније је приступ унапређен тако што се сензори рекламирају помоћу емитованих датаграма.

DSP модул садржи обраду примљених сигнала. Имплементиран је и пример обраде. Обрада се врши над блоком одбирака. Величина блока може утицати на перформансе. Велики блокови могу изазвати спорији одговор корисничког интерфејса на активност сензора. Имплементирано је уклањање једносмерне компоненте уз помоћ БИО филтра са следећом функцијом преноса:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1-z^{-1}}{1-\gamma z^{-1}}, \gamma = 0,9921875 \quad (5.1)$$

У раду је дата процена перформанси система у погледу губитка пакета у два сценарија:

- локална бежична мрежа
- пренос путем интернета

У случају преноса преко интернета саобраћај мора бити преусмерен преко неколико мрежних чворова. У овом случају укупни губици повећани су у малом проценту, али то је и даље прихватљиво за многе потенцијалне апликације. У овом раду коришћена је само сензорска мрежа са топологијом у облику звезде, користећи сервер у центру мреже. Даља анализа протокола и различите топологије мреже могу бити предмет будућих истраживања.

5.3 Реализација говорне спреге за уређаје са ограниченим ресурсима

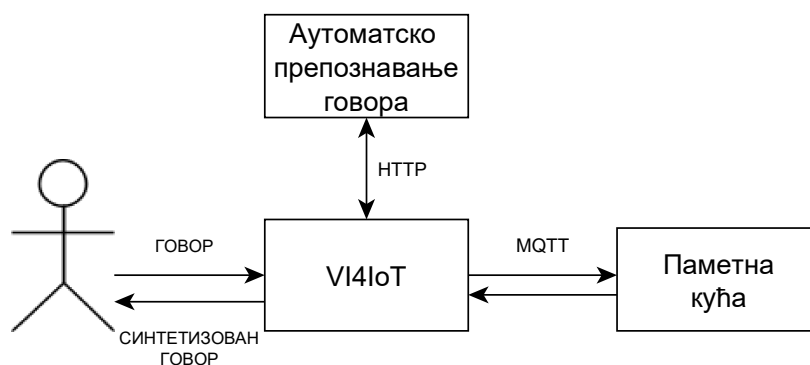
У [Cetic3] је показана реализација говорне спреге за уређај са веома ограниченим ресурсима. Дат је опис реализованог прототипа говорне спреге (енгл. *Voice interface* – VI) за IoT уређаје - VI4IoT. Као главни процесни чип VI4IoT користи 500MIPS xCORE вишејезгарни микроконтролер који има осам логичких језгара [May]. Повезана је интернетом помоћу TiWi-SL IEEE 802.11b/g Wi-Fi модула заснованог на Texas Instruments CC3000 чипу. За снимање и репродукцију звука коришћена је MikroElektronika SmartMP3 развојна плоча са VS1053 кодек чипом. Плоча је приказана на слици 5.3.1. На њој се види

интегрисани звучник мале снаге. Подржани формати за кодовање су ADPCM и MP3 формат за декодовање звучног записа синтетизованог говора.



Слика 5.3.1 SmartMP3 плоча са интегрисаним звучником

Концепт предложеног VI4IoT уређаја приказан је на слици 5.3.2. Корисник издаје гласовне команде. Глас се прикупља и шаље у сервис за аутоматско препознавање говора користећи HTTP POST метод. Као одговор, он добија информације у JSON формату. Доступне су алтернативе препознатог говора. Препознати говор се затим пореди са скупом подржаних команди користећи претходно развијени алгоритам [Stefanovic].



Слика 5.3.2 Концепт уређаја

Када се команда препозна, извршава се одређена рутина за руковање команде. Имплементиран је скуп команди. Покретање наредбе може резултирати са два могућа догађаја. Ако су актуатори директно повезани са VI4IoT уређајем, може се повезати излазно/улазну ножицу опште намене (GPIO) или неки интензитет помоћу модулације ширине импулса (PWM) као што је показно у [Vrga]. Уређај подржава и SPI и I2C магистрале па су могућа даља проширења. У

случају да контролисани IoT уређај није директно везан на VI4IoT могуће га је повезати путем протокола који се обично користе у IoT [Tang]. VI4IoT уређај може да комуницира и са веб сервисом попут портала за временску прогнозу и слично.

Активирање путем детекције гласа (енгл. *Voice Activation Detection – VAD*) је прво стање у коме VI4IoT детектује говорне активности. Користимо алгоритам за раздвајање предњег и позадинског сигнала заснованог на енергији слично као [Goetze]. Када се препозна гласовна активност, VI4IoT започиње HTTP POST методу на Гугл STT веб сервису. Док је присутна гласовна активност, уређај шаље податке. Када је слање завршено, уређај чека резултат аутоматског препознавања гласа. Резултат је примљен као JSON датотека.

Након резултата препознавања добијени текст се анализира. Анализа је објашњена у поглављу 5.1. Уколико је резултат претраге позитиван команда се извршава. VI4IoT је тада у стању чекања до потврде извршавања команде.

За неке команде је потребно дати гласовну потврду или је сама команда заправо упит за неко од доступних информација. У том случају се приступа синтези говора. Путем HTTP захтева синтетизована гласовна порука се затим репродукује уз помоћ MP3 декодерске плоче. По завршетку репродукције поруке, уређај се враћа у стање у коме чека следећу говорну активност.

Уређај је испитан у случајевима командовања кућним апаратима које је требало укључити или искључити. Затим провера температуре у просторији и порукама о тренутном временском стању и временској прогнози.

У [Cetic3] је дата кратка анализа кроз приказ потребне меморије и потрошње енергије. Такође, иста количина података се мора проследити и послати локалној IEEE 802.11 приступној тачки. У време када је уређај развијан доступна компресија је била само она без губитака Флак (енгл. *FLAC*) кодера. Ово је алгоритамски превише сложен кодек на уређају који је узет у разматрање. Из тог разлога одлучено је да се шаљу некомпримовани одбирци као што је назначено у [Zordan]. Интересантан податак је могућност уштеде енергије када процесорска јединица троши више енергије ради обрада сигнала од енергије потребне за бежични пренос података без компресије.

Имплементација аутоматског препознавања говора са великим речником није увек могућа на платформама са веома ограниченим ресурсима. Синтеза говора на уређају са ограниченом меморијом могућа је уз помоћ неких техника уз велике компромисе квалитета синтетизованог говора [Tokuda]. Додатно, када порука садржи нумеричку вредност попут тачног времена или температуре ваздуха, чување свих вредности у меморији није могуће. Синтеза говора се одвија на удаљеном рачунару па је у предложеном решењу потребно чувати само текстуални запис поруке.

Предложен уређај има добар потенцијал за употребу у домаћинству. Будући да систем користи услуге у клауду, могуће је препознати произвољне команде и дати одговор синтетизованим говором. Подршка за више језика омогућава комуникацију на матерњем језику. Кључна предност Гугл сервиса је што нуди подршку за српски језик.

5.4 Реализација јефтиног бежичног акустичког сензора

У раду [Doder2] је приказана имплементација јефтиног бежичног акустичког сензора. Прва верзија сензора је заснована на Espressif ESP8266 модулу за Wi-Fi (IEEE 802.11) комуникацију [ESP8266EX], PCM4201 аналогно-дигитални конвертор [SBAS342B] и додатну електронику за управљање микрофоном и напајањем. Створено је јефтино решење које се може користити у великом броју сензора. Овакви сензори се користе за прикупљање акустичних догађаја из окружења. Обрада и анализа прикупљених аудио података се одвија на чворовима за обраду.

Бежични акустички сензори користе се у многим IoT апликацијама, почевши од паметне куће, препознавања гласа, сигурносних система, надгледања животне средине и сличних апликација. Традиционални акустички сензори реализовани су тако да буду оптимизовани за специфичну примену. Како је приказано у [Cetic3] коришћење континуираног тока података звука од акустичког сензора до моћнијег чвора за обраду има пуно предности у односу на систем где податке израчунава и обрађује сам сензор. На страни сервера постоји приступ свим снимљеним звучним подацима који пружају бројне могућности за обраду сигнала [Doder1].

Будући да је потребан велики број сензора, укупни трошкови могу бити веома високи. Да би смањили трошкове система, сензори морају бити јефтине и приступачни, али и довољно флексибилни да омогуће различите анализе. Дизајн напајања је посебно важан јер сви коришћени делови захтевају стабилно напајање како би радили нормално. У раду су дати неки аспекти модула за напајање.

Са овом компонентом се постижу мале димензије сензора, па се сензор може интегрисати у окружење без потешкоћа. Цена комплетног сензора не прелази 10 долара. Прва верзија бежичног акустичког сензора састоји се од четири модула:

1. Мрежни модул.
2. Модул за аналогно дигиталну конверзију.
3. Микрофонски модул.
4. Модул напајања.

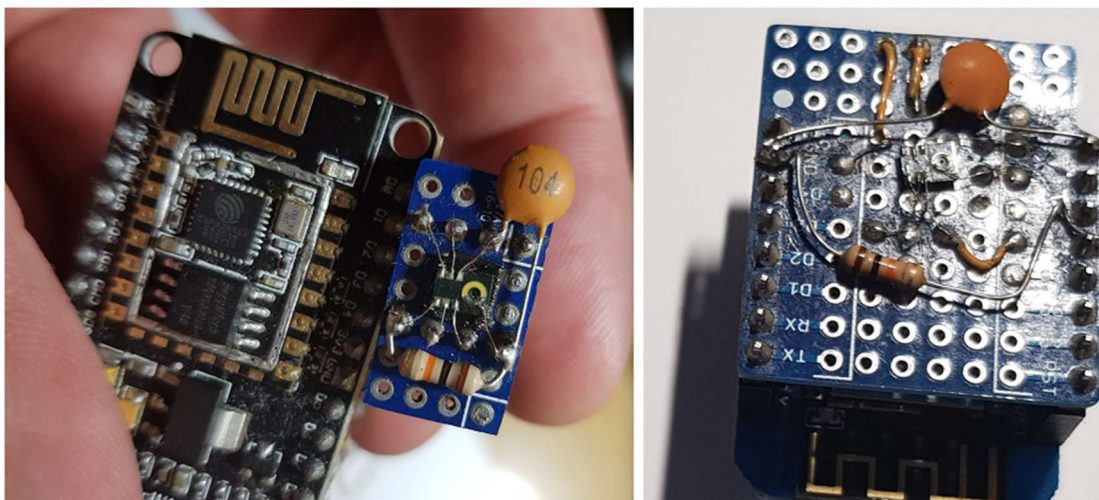
ESP8266 чип нуди потпуно и самостално решење за бежично умрежавање. Конкретан модул поседује интегрисану керамичку антену која показује солидне перформансе у пракси, али мрежни модул може бити додатно опремљен спољном антеном за добијање снаге сигнала.

Примљени звучни подаци са I2S интерфејса чувају се у меморијском низу и када имамо довољну количину PCM података, RTP пакет се конструише и шаље на сервер. Величина пакета може утицати на перформансе и потрошњу енергије. Употреба веће величине пакета ће Wi-Fi задржати у стању мировања и уштедети више енергије.

У раду је приказана анализа потрошње оваквог уређаја. Потребно је нагласити да је мерење потрошње урађено само на прототипу. Код финалног производа се очекују додатне уштеде. Чињеница да је сензор у режиму мировања у значајном временском периоду утиче повољно на потрошњу енергије. Успостава везе се остварује периодично у присуству довољно звучних података за слање.

Разлог за коришћење додатног ADC је тај што интегрисани ADC на ESP8266 модулу није намењен одабирању звука. Коришћени ADC има својства која захтевају комбинацију високог динамичког распона, малог изобличења и мале потрошње енергије.

Развијен је и унапређен сензор. Код унапређеног сензора коришћен је MEMS микрофон. Веза са MEMS микрофоном остварује се помоћу I2S спреге док је сама конверзија аналогног сигнала интегрисана у MEMS. На слици 5.4.1 су приказана два прототипа бежичног сензора у различитим изведбама које су коришћене у току истраживања.



Слика 5.4.1 Прототипови јефтиног акустичног сензора

Предвиђање века трајања акустичног сензора на батеријском напајању је кратко дискутовано у раду. Показало се да је највећи потрошач ипак бежични пренос.

У току су истраживања употребе унапређеног решења које се заснива на модулу ESP32 који подржава и Блутут (енгл. *Bluetooth*) комуникацију. Прелиминарни резултати показују велике предности ове комуникације у погледу ефикасности пошто се пренос, управљање и координација сензорима одвија преко новог стандарда Bluetooth Low Energy - BLE. Уведен је иновативни принцип да се SNR информације појединачних сензора преносе путем BLE. Након одабира микрофона успоставља се класична веза где се звук преноси путем Advanced Audio Distribution Profile - A2DP комуникације за пренос мултимедијалног садржаја. Потенцијални проблем је свакако домет овакве везе и могућност губитака пакета. Ово је потребно детаљније истражити.

Овде су предложена нека решења за ниско буџетни бежични акустички сензор. Приказани су изазови током његове конструкције. Главни делови сензора су описани. Дата је анализа утицаја одабраног аудио формата и видова бежичне комуникације са акцентом на потрошњу енергије. Постигнути резултати показују

добар потенцијал. Посебно ако узмемо у обзир цену конкурентних бежичних акустичких сензора. Постоји простор за даљу оптимизацију, експерименте и побољшања. Ово ће свакако бити предмет нашег будућег рада.

5.5 Систем за испитивање гласовних команди у паметним кућама

У раду [Рајић] је приказано једно решење за испитивање успешности система за управљање гласом у паметним кућама. Примена гласовних команди у животном простору са собом и даље носи велике изазове. У раду је дат приказ система за испитивања успешности препознавања команди у различитим ситуацијама. Дат је пример где се аутоматско препознавање говора одвија у клауд сервису компаније Амазон. У поглављу два је дат опис овог сервиса. Важно својство које је овде искоришћено је проширивост система користећи подршку имплементације вештина (енгл. *skills*).

Систем за испитивање гласовних команди се састоји од корисничке апликације која се извршава на локалном рачунару и вештином која се извршава на Амазон серверима. Апликација се даље састоји из два дела. Први део је серверски модул, који служи за аутентификацију. При покретању апликације, захтева се пријављивање на Амазон налог, који обезбеђује токене за аутентификацију. Да би регистровани уређај био коришћен, потребно је направити само-потписни сертификат и користити га као верификацију уређаја при слању и примању захтева. Сертификат се може направити помоћу OpenSSL програма и при прављењу се уноси ID (енгл. *Identification*) уређаја који је наведен на Амазон сервису. Да би систем функционисао потребно је направити профил на Амазон сервису и регистровати кориснички уређај уз сигурносни профил. Приликом прављења сигурносног профила, потребно је дефинисати URL (енгл. *Uniform resource locator*) са којег долази захтев и на који URL се шаље одговор.

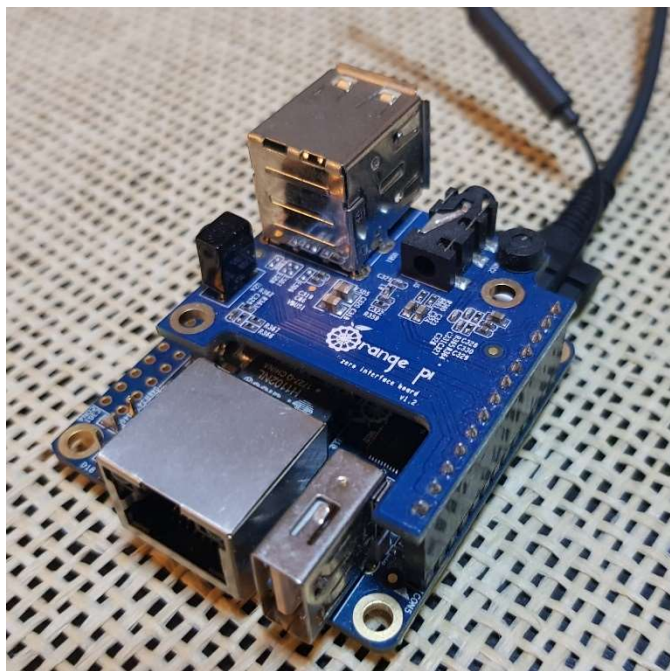
Други део апликације омогућава учитавање аудио датотеке, његово слање кроз AVS клијента до Амазон сервиса за препознавање говора. Сва комуникација се одвија шифрованим порукама користећи токене за аутентификацију. Речи из скупа W које изговара говорник се преносе као вибрације кроз ваздух. Микрофонским склопом се ова информација преводи у дигиталан облик. На

страни аутоматског препознавања заснованог на рачунару, систем покушава да реконструише изговорене речи и даје скуп препознатих речи W' .

У [Рајић] је приказан изглед корисничке спреге и дат је приказ архитектуре система. Развијене су две верзије овог система. Први је развијен у Јава програмском језику. Користи VLC, Node.js и Maven. Ова верзија система је инспирисана отвореним кодом компаније Амазон, доступном у пакетима за подршку развоја прототипа уређаја који користи Амазон Алекса сервисе. Велики део софтвера је заснован на јавно доступном коду.

Даљим радом на дисертацији уочено је неколико недостатака. Имплементација сервиса за комуникацију у Јава језику је и даље актуелна и удобна. Раније развијени прототипови који су приказани у овој дисертацији су сензори базирани на С и С++ језицима. Може се закључити да ово нису језици у којима се експерименти брзо реализују. Са друге стране Пајтон као језик постиже све већу популарност у научним круговима. Наредна верзија система је реализована у програмском језику Пајтон и ова верзија ће се одржавати. Решење је доступно на репозиторијуму отвореног кода [Cetic4].

Верификација система је урађена снимањем говорних команди уз помоћ бежичних акустичких сензора базираних на Orange Pi Zero развојном систему са додатном плочом која нуди проширења у виду додатних спрега. Једна од битних спрега за ово истраживање је постојање омнидирекционе микрофонске капсуле. Коришћени сензор је приказан на слици 5.5.1. Сваки од сензора комуницира путем интернет протокола и у раду користи веб технологије. Веб сервер је реализован уз помоћ Фласк Пајтон библиотеке.



Слика 5.5.1 Бежични сензор за испитивање гласовних команди

Грешка препознатих секвенци је релативно мала и постоји само у случају када је сензор максимално удаљен од говорника (енгл. *Far field*). Резултати су добијени на релативно малом узорку и у условима са добрим SNR. Овде је битно нагласити да у овом сегменту рада акценат није био на испитивању акустике простора и да су добијени резултати били довољни за потврду принципа и рада система за испитивање. Касније се приступило обимнијим испитивањима где је развијени систем послужио за проверу развијеног модела сензора, што је главни фокус ове дисертације. Резултати су приказани у поглављу 4 и прилогу В.

Допринос овог рада је реализација приступачног и елегантног система за испитивање система дистрибуираних акустичких сензора са акцентом на испитивање над великим бројем испитних случајева. Ово је добра основа и за даља испитивања оваквих система у различитим акустичким окружењима и сценаријима. Систем за испитивање гласовних команди је део рада на ширем истраживању примене говорних технологија и њихову примену у интелигентним системима.

ПОГЛАВЉЕ 6.

ЗАКЉУЧАК

Основни циљ истраживања обухваћеног дисертацијом била је примена дистрибуираних акустичких сензора за подршку командовања гласом у паметним кућама. Дат је преглед стања у области који покрива интернет технологије и њихове примене у паметним кућама. Истакнута је техника моделовања акустике затвореног простора помоћу пресликавања извора на основу које је реализован модел простора са реверберацијом. Затим је моделован акустички сензор са проценом односа сигнала и шума. На основу ова два модела, добијен је модел изворишног саобраћаја бежичних акустичких сензора у простору са реверберацијом. Повезивањем више микрофона добијен је систем микрофона за командовање гласом у виду физичког прототипа. Испитивањем рада система потврђено је да је могуће реализовати симулационо окружење које довољно добро описује реалан простор, и реализовати систем за подршку командовања гласом.

Примена дистрибуираних микрофона у односу на традиционални приступ употребом микрофонског низа има своје предности. Експериментални резултати показују да се постигло побољшање у погледу озвучености, односно способности да се изговорени глас боље препозна. Приказани модел може послужити као добра основа за испитивања сродних система у различитим акустичким условима. Посебна предност је истакнута код примене симулације којом је

могуће заменити обимне експерименте у физичком простору и тиме остварити уштеде у времену и новцу. Реализован систем за испитивање гласовних команди може послужити за даља истраживања препознавања говора и акустике затворених простора.

На основу представљених резултата, доприноса и спроведене анализе може се закључити да је постављена хипотеза потврђена.

Предложен систем има предности у односу на постојећа решења на бази микрофонских низова. Међутим, сама дистрибуираност микрофона отвара питања у погледу преноса података, поузданости, напајања електричном енергијом, синхронизације, исплативости, како енергетске тако и материјалне и многе друге. Предложени систем у коме се бира најповољнији микрофон је само један од могућих приступа решењу проблема. Велики потенцијал се види у техникама фузије сигнала који би довео до нових алгоритама за побољшања разумљивости говора. Нека од ових питања ће бити предмет даљих истраживања аутора и његових сарадника. Могући правци даљег истраживања су:

- решавање синхронизације дистрибуираних микрофона која би омогућила корелацију сигнала са више сензора,
- примена алгоритама за уклањање шума и ефекта реверберације применом дистрибуираних сензора,
- опширнија анализа узимајући у обзир: трошкове развоја, серијске производње, загађења у етру услед бежичне комуникације и многе друге аспекте,
- испитати могућност примене других познатих метода и техника моделовања акустике простора код истраживања дистрибуираних акустичких сензора.

ЛИТЕРАТУРА

- [AADSP1] V. Kovačević, M. Popović, M. Temerinac, N. Teslić, “Arhitekture i algoritmi digitalnih signal procesora I,” Fakultet Tehničkih Nauka, Novi Sad, 2005
- [ACE] J. Eaton, N. D. Gaubitch, A. H. Moore, P. A. Naylor "The ACE challenge - Corpus description and performance evaluation," 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) doi: 10.1109/waspaa.2015.7336912
- [Alaa] Mussab Alaa, A.A. Zaidan, B.B. Zaidan, Mohammed Talal, M.L.M. Kiah, "A review of smart home applications based on Internet of Things", Journal of Network and Computer Applications, vol. 97, pp. 48, 2017, ISSN 10848045 doi: 10.1016/j.jnca.2017.08.017
- [Allen] J. Allen and D. Berkley, “Image method for efficiently simulating small-room acoustic”, The Journal of the Acoustical Society of America, vol. 65, no. 4, pp. 943-950, 1979 doi: 10.1121/1.382599
- [Alumäe] Tanel Alumäe, Kaarel Kaljurand, “Open and extendable speech recognition application architecture for mobile environments,” in SLTU, 2012.
- [APSNArc] January 10, 1919: Death of Wallace Sabine, pioneer of architectural acoustics, American Physical Society, News Archives, Vol 20, No 1, Jan,2011
- [Atmoko] H. Atmoko, D. C. Tan, G. Y. Tian, B. Fazenda, “Accurate sound source localization in a reverberant environment using multiple acoustic sensors”, Measurement Science and Technology Vol. 19 no. 2, 2008 doi: 10.1088/0957-0233/19/2/024003

- [Barker1] Jon Barker, Ricard Marxer, Emmanuel Vincent, Shinji Watanabe, “The third 'CHiME' Speech Separation and Recognition Challenge: Dataset, task and baselines,” IEEE 2015 Automatic Speech Recognition and Understanding Workshop (ASRU), Scottsdale, ISBN: 978-1-4799-7291-3 pp. 504-511 doi: 10.1109/ASRU.2015.7404837
- [Barker2] Jon Barker, Ricard Marxer, Emmanuel Vincent, Shinji Watanabe, “The third 'CHiME' speech separation and recognition challenge: Analysis and outcomes. Computer Speech and Language,” Elsevier, 2017, 46, pp.605-626. doi:10.1016/j.csl.2016.10.005
- [Basicevic1] I. Basicevic, M. Popovic and D. Kukolj, “Comparison of SIP and H.323 Protocols,” 2008 The Third International Conference on Digital Telecommunications (icdt 2008), Bucharest, 2008, pp. 162-167. doi: 10.1109/ICDT.2008.14
- [Basicevic2] I. Basicevic, N. Cetic, M. Popovic and M. Krunic, “A light-weight SIP-based execution management service for bag-of-tasks grid system,” 2014 22nd Telecommunications Forum Telfor (TELFOR), Belgrade, 2014, pp. 919-922. doi: 10.1109/TELFOR.2014.7034555
- [Bertrand] Alexander Bertrand, “Applications and trends in wireless acoustic sensor networks: a signal processing perspective”, 18th IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT), 2011, pp. 1-6 doi: 10.1109/SCVT.2011.6101302
- [Biing-Hwang] Juang, Biing-Hwang, and Lawrence R. Rabiner. “Automatic speech recognition—a brief history of the technology development,” Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara 1 (2005): 67.
- [Bojovic] Z. Bojovic, Z. Peric, V. Delic, E. Secerov, M. Secujski, V. Senk, “Comparative Analysis of the Performance of Different Codecs in a Live VoIP Network using SIP Protocol,” Elektronika ir elektrotehnika Vol 117 No1 (2012) ISSN 1392 – 1215 doi: 10.5755/j01.eee.117.1.1050
- [Borish] Jeffrey Borish, “Extension of the image model to arbitrary polyhedral,” The Journal of the Acoustical Society of America, Acoustical Society of America, 1984, doi: 10.1121/1.390983
- [Bouzid] O. M. Bouzid, G. Y. Tian, J. Neasham, B. Sharif, “Investigation of sampling frequency requirements for acoustic source localisation using wireless sensor networks,” Applied Acoustics Vol. 74 no. 2, pp. 269-274, 2013. doi: 10.1016/j.apacoust.2010.12.013

- [BPopović1] Branislav Popović, Stevan Ostrogonac, Edvin Pakoci, Nikša Jakovljević, Vlado Delić, “Deep Neural Network Based Continuous Speech Recognition for Serbian Using the Kaldi Toolkit,” In: Ronzhin A., Potapova R., Fakotakis N. (eds) Speech and Computer. SPECOM 2015. Lecture Notes in Computer Science, vol 9319. Springer, Cham doi: 10.1007/978-3-319-23132-7_23
- [BPopović2] Popovic, B., Pakoci, E., Jakovljevic, N., Kocis, G., & Pekar, D. “Voice assistant application for the Serbian language,” 2015 23rd Telecommunications Forum Telfor (TELFOR). doi: 10.1109/telfor.2015.7377600
- [Brush] Brush, A. J., Hazas, M., & Albrecht, J. “Smart Homes: Undeniable Reality or Always Just around the Corner?” IEEE Pervasive Computing, 17(1), 82–86. doi: 10.1109/mprv.2018.011591065
- [Cetic1] Nenad Cetic, Miroslav Popovic, Miodrag Đukic, Jelena Kovacevic, “Source traffic modeling in WSN for acoustic sensing in reverberant environment”, Elektronika ir elektrotehnika, Vol.21, Issue 5, pp.64-68, ISSN:1392-1215, doi: 10.5755/j01.EEE.21.5.13335
- [Cetic2] Nenad Cetic, Miroslav Popovic, Miodrag Djukic, Momcilo Kronic, “A run-time library for parallel processing on a multi-core DSP,” ECBS-EERC 2013, Budapest, Hungary, pp.41-47, DOI: 10.1109/ECBS-EERC.2013.14
- [Cetic3] Nenad Četić, Srđan Furman, Miroslav Popović, Momčilo Krunic, “Cloud based voice interface for internet of things devices,” 1St Ieee Consumer Electronics Workshop Novi Sad, 2015, pp. 19 – 22 doi: 10.1109/CEWS.2015.7867144
- [Cetic4] Nenad Cetic, Nebojsa Rajic, Orange-pi-audio, доступно онлајн: <https://github.com/chedadsp/orange-pi-audio>
- [Christensen] Claus Lyng Christensen, George Koutsouris, Jens Holger Rinde, “111The ISO3382 parameters: Can we simulate them? Can we measure them?” ISRA2013 International Symposium on Room Acoustics, 2013
- [Davis] Don Davis, Eugene Patronis, Pat Brown, “Sound System Engineering,” 4th edition, Focal Press, 2013 ISBN: 978-0-240-81847-4
- [Delic1] Delic Vlado, Secujski Milan, Jakovlevic Niksa, Gnjatovic Milan, Stankovic Igor, “Challenges of Natural Language Communication with Machines,” Chapter 19 in DAAAM International Scientific Book 2013, pp. 371-388, B. Katalinic & Z. Tekic (Eds.), Published by DAAAM International, ISBN 978-3-901509-94-0, ISSN 1726-9687, Vienna, Austria doi: 10.2507/daaam.scibook.2013.19

- [Delic2] Vlado Delić, Zoran Perić, Milan Sečujski, Nikša Jakovljević, Jelena Nikolić, Dragiša Mišković, Nikola Simić, Siniša Suzić, Tijana Delić, “Speech Technology Progress Based on New Machine Learning Paradigm” Computational Intelligence and Neuroscience Volume 2019 doi: 10.1155/2019/4368036
- [Demirkol] I. Demirkol, F. Alagoz, H. Delic and C. Ersoy, “Wireless sensor networks for intrusion detection: packet traffic modeling,” in IEEE Communications Letters, vol. 10, no. 1, pp. 22-24, Jan 2006. doi: 10.1109/LCOMM.2006.1576557
- [Denes] P. Denes, M. V. Mathews, “Spoken Digit Recognition Using Time-Frequency Pattern Matching,” The Journal of the Acoustical Society of America 32, 1450 (1960); Published Online: 03 August 2005 doi: 10.1121/1.1907936
- [Doder1] D. Doder, N. Četić, M. Popović and J. Kovačević, “Realisation of server application for acoustic sensors based on RTSP, RTP protocols,” 2015 23rd Telecommunications Forum Telfor (TELFOR), Belgrade, 2015, pp. 516-519. doi: 10.1109/TELFOR.2015.7377520
- [Doder2] Doder Davor, Cetic Nenad, Kovacevic Jelena, Popovic Miroslav, “Implementation of low cost wireless acoustic sensor based on ESP8266 module,” 3rd International conference on electrical, electronic and computing engineering ICETAN 2016
- [Djukic] Miodrag Djukic, Miroslav Popovic Nenad Cetic, Ivan Povazan, “Embedded Processor Oriented Compiler Infrastructure,” August 2014 Advances in Electrical and Computer Engineering 14(3): 123-130 doi: 10.4316/AECE.2014.03016
- [ESP8266EX] ESP8266EX Datasheet Version 4.3, Espressif Systems IOT Team <http://bbs.espressif.com/> Copyright© 2015, Available:https://www.adafruit.com/images/product-files/2471/0A-ESP8266__Datasheet__EN_v4.3.pdf
- [Evans] Dave Evans, “The Internet of Things: How the Next Evolution of the Internet Is Changing Everything,” White Paper, Cisco IBSG, April 2011
- [Gartner] Alfonso Velosa, W. Roy Schulte, Benoit J. Lheureux, Nathan Nuttall, Hype Cycle for the Internet of Things, 2018, Gartner, Inc. 2018
- [Giannoulis] P. Giannoulis, A. Brutti, M. Matassoni, A. Abad, A. Katsamanis, M. Matos, G. Potamianos, P. Maragos, “Multi-room speech activity detection using a distributed microphone network in domestic environments,” European Signal Processing Conf. (EUSIPCO-2015), Nice, France, Sep. 2015

- [Gill] Khusvinder Gill, Shuang-Hua Yang, Fang Yao, Xin Lu, “ZigBee-Based Home Automation System,” *IEEE Transactions on Consumer Electronics*, Vol. 55, No. 2, May 2009, pp. 422 – 430. doi: 10.1109/TCE.2009.5174403
- [Gnjatovic] M.Gnjatovic, V. Delic, “Cognitively-inspired representational approach to meaning in machine dialogue,” *Knowl. Based Syst.* (2014), doi: 10.1016/j.knosys.2014.05.001
- [Goetze] S. Goetze, J. Schroder, S. Gerlach, D. Hollosi and J.-E. Appell, “Acoustic Monitoring and Localization for Social Care,” *Journal of Computing Science and Engineering*, vol. 6, no. 1, pp. 40-50, 2012. doi: 10.5626/JCSE.2012.6.1.40
- [Gomez] Carles Gomez, Josep Paradells, “Wireless Home Automation Networks: A Survey of Architectures and Technologies,” *IEEE Communications Magazine*, Vol. 48, No. 6, June 2010, pp. 92 –101. doi: 10.1109/MCOM.2010.5473869
- [Guamán] Steven Guamán, Adrián Calvopiña, Pamela Orta, Freddy Tapia, and Sang Guun Yoo. “Device Control System for a Smart Home using Voice Commands: A Practical Case,” 10th International Conference on Information Management and Engineering (ICIME 2018). ACM, New York, NY, USA, 86-89. doi: 10.1145/3285957.3285977
- [Hollosi1] D. Hollosi, G. Nagy, R. Rodigast, S. Goetze, P. Cousin “Enhancing Wireless Sensor Networks with Acoustic Sensing Technology: Use Cases, Applications & Experiments,” *IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, 2013, doi: 10.1109/GreenCom-iThings-CPSCom.2013.75
- [Hollosi2] B. Kelly, D. Hollosi, P. Cousin, S. Leal, B. Iglár, A. Cavallaro, “Application of Acoustic Sensing Technology for improving Building Energy Efficiency,” *Procedia Computer Science* Vol. 32, pp. 661-664, 2014 doi: 10.1016/j.procs.2014.05.474
- [Hwang] Juang, Bing Hwang, and Laurence R. Rabiner, “Hidden Markov models for speech recognition,” *Technometrics* 33.3 (1991): 251-272.
- [IEEE1588] IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems, in *IEEE Std 1588-2008 (Revision of IEEE Std 1588-2002)* , pp.1-300, July 24 2008 doi: 10.1109/IEEESTD.2008.4579760
- [Isikdag] Umit Isikdag, “Internet of Things: Single-Board Computers,” *SpringerBriefs in Computer Science*, pp. 43–53. 2015 doi: 10.1007/978-3-319-21825-0_4
- [ISO 3382-2] ISO 3382-2:2008 Acoustics -- Measurement of room acoustic parameters -- Part 2: Reverberation time in ordinary rooms, International Organization for Standardization, 2008

- [ISO 354:2003] ISO 354:2003 Acoustics -- Measurement of sound absorption in a reverberation room
- [Jain] M. Jain, A. Singh, and V. Chandan, "Portable+: A Ubiquitous and Smart Way Towards Comfortable Energy Savings," Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 1, no. 2, 2017; doi: 10.1145/3090079
- [Jie] Yin Jie, Ji Yong Pei, Li Jun, Guo Yun, Xu Wei, "Smart Home System Based on IOT Technologies," 2013 International Conference on Computational and Information Sciences, Shiyang, 2013, pp. 1789-1791 doi: 10.1109/ICCIS.2013.468
- [Jokic] Ivan Jokic; Stevan Jokic; Vlado Delic ; Zoran Peric, "Towards Enabling Measurement of Similarity of Acoustical Environments Using Mobile Devices," 2017 13th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS), ISBN: 978-1-5386-1800-4 doi: 10.1109/TELSIKS.2017.8246261
- [Kinoshita] Keisuke Kinoshita, Marc Delcroix, Sharon Gannot, Emanuël A. P. Habets, Reinhold Haeb-Umbach, Walter Kellermann, Volker Leutnant, Roland Maas, Tomohiro Nakatani, Bhiksha Raj, "A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research," EURASIP Journal on Advances in Signal Processing 2016, ISSN: 1687-6180 (electronic)
- [Kinsella2018] Bret Kinsella, Ava Mutchler, "Smart Speaker Consumer Adoption Report," Voicebot.ai, March 2018
- [Kinsella2019] Bret Kinsella, Ava Mutchler, "Smart Speaker Consumer Adoption Report," Voicebot.ai, March 2019 U.S.
- [Krunic] Momcilo Krunic, Miroslav Popovic, Vlado Krunic, Nenad Cetic, "Energy Consumption Estimation for Embedded Applications," ELEKTRONIKA IR ELEKTROTEHNIKA, vol.22,no.3, 2016, doi: 10.5755/j01.eie.22.3.15313
- [Kurtović] Husnija Kurtović, "Ozvučavanje", peto izdanje, Novinsko-izdavačka radna organizacija, Tehnička knjiga, Beograd, 1985
- [Lehmann1] E. A. Lehmann and A. M. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," The Journal of the Acoustical Society of America, vol. 124, no. 1, pp. 269-277, 2008. doi: 10.1121/1.2936367
- [Lehmann2] E. A. Lehmann, "Fast image-source method: Matlab code," Oct. 2013.<http://www.eric-lehmann.com>. [преступљено: Feb. 2014]

- [Levenshtein] V. Levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions and Reversals," *Soviet Physics-Doklady*, vol. 10, no. 8, pp. 707-710, 1966.
- [Luftman] Jerry Luftman, Hossein S Zadeh, Barry Derksen, Martin Santana, Eduardo Henrique Rigoni, Zhengwei (David) Huang, "Key information technology and management issues 2012–2013: an international study," *Journal of Information Technology* (2013) 28, 354–366. doi:10.1057/jit.2013.22
- [Lukic] N. Lukic, N. Teslic, T. Maruna, V. Mihic, "A java API interface for the search of DTV services in embedded multimedia devices," *Consumer Electronics, IEEE Transactions on*, vol.59, no.4, pp.875,882, November 2013 doi: 10.1109/TCE.2013.6689702
- [Markit] B. Morelli, J. Howell, J. Watson, S. Lucero, L. Ratliff, Y. Kanasin, C. Kim, P. Tomasi, "IoT trend watch 2018," IHS Markit, 2018.
- [May] David May, "The XMOS Architecture and XS1 Chips," *Micro, IEEE*, vol. 32, no. 6, pp. 28-37, 2012 .doi: 10.1109/MM.2012.87
- [Messier] G. G. Messier and I. G. Finvers, "Traffic models for medical wireless sensor networks," in *IEEE Communications Letters*, vol. 11, no. 1, pp. 13-15, Jan. 2007. doi: 10.1109/LCOMM.2007.061291
- [Mijić1] Miomir Mijić and Draško Mašović, "Reverberation Radius in Real Rooms," in *Telfor Journal*, vol. 2, no. 2, 2010, Telecommunications Society, Belgrade, Academic Mind, Belgrade, ISSN 2334-9905
- [Mijić2] Miomir Mijić and Draško Mašović, "Comments on Reverberation Radius in Real Rooms," Letter to the Editor Response to, in *Telfor Journal*, vol. 4, no. 2, 2012, Telecommunications Society, Belgrade, Academic Mind, Belgrade, ISSN 2334-9905
- [Mills] David L. Mills, "Internet time synchronization: the network time protocol," *IEEE Transactions on Communications*, vol. 39, no. 10, pp. 1482-1493, Oct 1991. doi: 10.1109/26.103043
- [Milošević] Miloš Milošević, Nenad Četić, Jelena Kovačević, Tihomir Anđelić, "Lighting Control Using Raspberry Pi and Oblo Living Home Automation System," *SERBIAN JOURNAL OF ELECTRICAL ENGINEERING* Vol. 16, No. 1, February 2019, 45-54 doi: <https://doi.org/10.2298/SJEE1901045M>
- [Mrazovac] Bojan Mrazovac, Branislav Todorovic, Dragan Kukulj, Miodrag Temerinac, "Device-Free Human Presence Detection Using Frequency Domain," Conference: IEEE 1st International Workshop on Consumer ElectronicsAt: Novi Sad, March 2015 doi: 10.1109/CEWS.2015.7867145

- [Nirjon] N. Nirjon, R. F. Dickerson, P. Asare, Q. Li, D. Hong, J. A. Stankovic, P. Hu, G. Shen and X. Jiang, “Auditeur: a mobile-cloud service platform for acoustic event detection on smartphones,” 11th Annual International Conference on Mobile Systems, Applications, and Services MobiSys’13, 2013. pp. 403–416. doi: 10.1145/2462456.2464446
- [OBLO] OBLO Living, OBLO Living LLC, December 2018, Available at: <http://www.oblolving.com/>
- [Öçal] O. Öçal, I. Dokmanic and M. Vetterli. “Source Localization and Tracking in Non-Convex Rooms,” 39th International Conference on Acoustics, Speech, and Signal Processing ICASSP, 2014, Florence, Italy, doi: 10.1109/ICASSP.2014.6853833
- [Pakoci] Edvin Pakoci, Branislav Popović, Darko Pekar, “Using Morphological Data in Language Modeling for Serbian Large Vocabulary Speech Recognition,” Computational Intelligence and Neuroscience, Vol. 2019, Article ID 5072918, DOI: 10.1155/2019/5072918, ISSN: 1687-5265, Hindawi, 03 March 2019.
- [Panayotov] Vassil Panayotov, Guoguo Chen, Daniel Povey, Sanjeev Khudanpur, “Librispeech: An ASR corpus based on public domain audio books,” 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, DOI: 10.1109/ICASSP.2015.7178964
- [Pejovic] Veljko Pejovic, Mirco Musolesi, “Anticipatory mobile computing: A survey of the state of the art and research challenges,” ACM Comput. Surv. 47, 3, Article 47, April 2015, doi: 10.1145/2693843
- [Pekar] Darko Pekar, Dragisa Miskovic, Dragan Knezevic, Natasa Vujnovic Sedlar, Milan Secujski and Vlado Delic, “Applications of Speech Technologies in Western Balkan Countries, Advances in Speech Recognition,” Noam Shabtai (Ed.), ISBN: 978-953-307-097-1,
- [Pham1] C. Pham and P. Cousin, “Streaming the Sound of Smart Cities: Experimentations on the SmartSantander test-bed,” Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing, 2013. doi: 10.1109/GreenCom-iThings-CPSCom.2013.114
- [Pham2] C. Pham, P. Cousin A. Carer, “Real-time On-Demand Multi-Hop Audio Streaming with Low-Resource Sensor Motes,” 39th Local Computer Networks Workshops (LCN Workshops), IEEE, 2014, doi: 10.1109/LCNW.2014.6927700

- [Popovic] Popovic M., Kovačević V. “An Approach to Internet-Based Virtual Call Center Implementation,” *Lecture Notes in Computer Science*, Springer Verlag, 2001, Vol. 2093, No. Jan 2001, str. 75- 84, ISSN 0302-9743
- [Rajić] Nebojša Rajić, Nenad Četić, Miroslav Popović, Sreten Tanacković, “System for testing voice commands in smart home,” *International Conference IcETRAN 2018*, pp. 477 - 480, ISBN 978-86-7466-752-1
- [Raju1] Anirudh Raju, Behnam Hedayatnia, Linda Liu, Ankur Gandhe, Chandra Khatri, Angeliki Metallinou, Anu Venkatesh, Ariya Rastrow, “Contextual Language Model Adaptation for Conversational Agents,” *Interspeech 2018*, Hyderabad, India, 2018 arXiv:1806.10215
- [Raju2] Anirudh Raju, Sankaran Panchapagesan, Xing Liu, Arindam Mandal, Nikko Strom, “Data Augmentation for Robust Keyword Spotting under Playback Interference,” *The Computing Research Repository*, 2018, arXiv:1808.00563
- [RFC4330] David L. Mills, *Simple Network Time Protocol (SNTP) Version 4 for IPv4, IPv6 and OSI*, RFC Editor, 2006, doi: 10.17487/RFC4330
- [RFC5905] Jim Martin, Jack Burbank, William Kasch, David L. Mills, *Network Time Protocol Version 4: Protocol and Algorithms Specification*, RFC Editor, 2010, doi: 10.17487/RFC5905
- [RFC7822] Tal Mizrahi and Danny Mayer, *Network Time Protocol Version 4 (NTPv4) Extension Fields*, RFC Editor, 2016, doi: 10.17487/RFC7822
- [Shalkouhi] P. Jafari Shalkouhi, “Comments on Reverberation Radius in Real Rooms,” *Letter to the Editor in Telfor Journal*, vol. 4, no. 2, 2012, Telecommunications Society, Belgrade, Academic Mind, Belgrade, ISSN 2334-9905
- [SBAS342B] *Low Power, 24-Bit, Single Channel Audio Analog-to-Digital Converter*, SBAS342B–DECEMBER 2004–REVISED APRIL 2006, Available: <http://www.ti.com/lit/ds/sbas342b/sbas342b.pdf>
- [Schalkwyk] J. Schalkwyk, D. Beeferman, F. Beaufays, B. Byrne, C. Chelba, M. Cohen, M. Kamvar and B. Strobe, “Your Word is my Command: Google Search by Voice: A Case Study,” in *Advances in Speech Recognition*, Springer, 2010, pp. 61-90. doi: https://doi.org/10.1007/978-1-4419-5951-5_4
- [Schulzrinne1] H. Schulzrinne, Columbia U, A. Rao and R. Lanthier, “RFC 2326 Real Time Streaming Protocol (RTSP),” April 1998. Available: <http://www.ietf.org/rfc/rfc2326.txt>

- [Schulzrinne2] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, "RFC 3550 RTP: A Transport Protocol for Real-Time Applications," July 2003. Available: <https://tools.ietf.org/html/rfc3550>
- [Škraba] Andrej Škraba, Andrej Koložvari, Davorin Kofjač, Radovan Stojanović, "Prototype of Speech Controlled Cloud Based Wheelchair Platform for Disabled Persons," in 3rd Mediterranean Conference on Embedded Computin, Budva, Montenegro, 2014 doi: 10.1109/MECO.2014.6862683
- [Soda] Shimpei Soda, Masahide Nakamura, Shinsuke Matsumoto, Shintaro Izumi, Hiroshi Kawaguchi, Masahiko Yoshimoto, "Handsfree voice interface for home network service using a microphone array network," in Third International Conference on Networking and Computing, 2012. doi:10.1109/ICNC.2012.37
- [Stanivuk] Stefan Stanivuk, Nenad Cetic, Jelena Kovacevic, Milenko Boric-Herget, "Twitter controlled real time audio synthesizer based on XMOS platform" Telecommunications Forum (TELFOR), 2013 21st pp.857 - 860, doi: 10.1109/Telfor.2013.6716365
- [Stefanovic] Miroslav Stefanovic, Nenad Cetic, Milan Kovacevic, Jelena Kovacevic, Miloš Jankovic, "Voice control system with advanced recognition," 20th Telecommunications Forum (TELFOR), pp. 1601-1604, 2012, doi: 10.1109/TELFOR.2012.6419529
- [Sullivan] Dan Sullivan, "The definitive guide to cloud computing," Realtime Nexus 1 (2010): 4-11.
- [Szöke] Igor Szöke, Miroslav Skácel, Ladislav Mošner, Jakub Paliesek, Jan Černocký, "Building and evaluation of a real room impulse response dataset," IEEE Journal of Selected Topics in Signal Processing, Volume: 13, 2019, pp. 863 – 876, doi: 10.1109/JSTSP.2019.2917582
- [Tang] Konglong Tang, Yong Wang, Hao Liu, Yanxiu Sheng, Xi Wang, Zhiqiang We, "Design and Implementation of Push Notification System Based on the MQTT Protocol," International Conference on Information Science and Computer Applications (ISCA 2013), Tel-Aviv, Israel, October 2013, pp. 116 – 119
- [TIMIT] John S. Garofolo, Lori F. Lamel, William M. Fisher, Jonathan G. Fiscus, David S. Pallett, Nancy L. Dahlgren, Victor Zue, "TIMIT Acoustic-Phonetic Continuous Speech Corpus," Linguistic Data Consortium, 1993

- [Tokuda] Keiichi Tokuda, Yoshihiko Nankaku, Tomoki Toda, Heiga Zen, Junichi Yamagishi. Keiichiro Oura, “Speech Synthesis Based on Hidden Markov Models,” Proceedings of the IEEE, vol. 101, no. 5, pp. 1234-1252, 2013. doi: 10.1109/JPROC.2013.2251852
- [Touloupis] E. Touloupis, A. Meliones, S. Apostolacos, “Speech codecs for high-quality voice over ZigBee applications: Evaluation and implementation challenges,” IEEE Communications Magazine, vol. 50, no. 4, pp. 122-128, 2012. doi: 10.1109/MCOM.2012.6178844
- [Tozlu1] Tozlu, S., “Feasibility of Wi-Fi enabled sensors for Internet of Things,” Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International, pp.291,296, 4-8 July 2011 doi: 10.1109/IWCMC.2011.5982548
- [Tozlu2] Tozlu, S.; Senel, M.; Wei Mao; Keshavarzian, A., “Wi-Fi enabled sensors for internet of things: A practical approach,” Communications Magazine, IEEE, vol.50, no.6, pp.134,143, June 2012 doi: 10.1109/MCOM.2012.6211498
- [Vrga] Nenad Vrga, Nenad Cetic, Jelena Kovacevic, Dejan Bardek, “HTTP server for control of home appliances based on XMOS platform,” Telecommunications forum (TELFOR), 2013 21st, pp.877 – 880 doi: 10.1109/TELFOR.2013.6716370
- [Wang] Q. Wang, T. Zhang “Source traffic modeling in wireless sensor networks for target tracking,” 5th ACM symposium on Performance evaluation of wireless ad hoc, sensor, and ubiquitous networks (PE-WASUN '08), pp. 96-100, doi: 10.1145/1454609.1454629
- [Webb] Willis L. Webb, Walter I. Christensen, Eldon P. Varner and Jack F. Spurling, Inter-Range Instrumentation Group Participation in the Meteorological Rocket Network, Bulletin of the American Meteorological Society Vol. 43, No. 12, pp. 640-649, 1962
- [Wei] H. Wei and S. Xincun, “VUI system of the portable maintenance aids based on cloud computing,” in Computational Problem-Solving (ICCP), 2012 International Conference on, Leshan, China, 2012. doi: 10.1109/ICCP.2012.6384258
- [Xiuhong] Li Xiuhong; Cheng Xiao; Yang Rongjin; Zhang Haijing; Zhanr Jialin, “Design and implementation of an audio-visual integrated wireless sensor remote monitoring network on Wetland,” Mechatronic Sciences, Electric Engineering and Computer (MEC), Proceedings 2013 International Conference on , vol., no., pp.609,614, 20-22 Dec. 2013 doi: 10.1109/MEC.2013.6885136

- [Ye] Jiaxing Ye, Takumi Kobayashi, Masahiro Murakawa, “Urban sound event classification based on local and global features aggregation,” *Applied Acoustics*, Vol 117, part B, 2017, pp. 246-256 doi: 10.1016/j.apacoust.2016.08.002
- [Zheng] Falk, C. Zheng, W.-Y. Chan, “A Non-Intrusive Quality and Intelligibility Measure of Reverberant and Dereverberated Speech,” *IEEE Trans Audio Speech Lang Process*, Vol.18, No.7, pp.1766-1774, 2010
- [Zordan] Davide Zordan, Borja Martinez, Ignasi Vilajosana, Michele Rossi, “To Compress or Not To Compress: Processing vs Transmission Tradeoffs for Energy Constrained Sensor Networking,” *CoRR*, vol. abs/1206.2129, 2012. arXiv: 1206.2129

Прилог А - ГРЕШКЕ ПРИ АУТОМАТСКОМ ПРЕПОЗНАВАЊУ ГОВОРА БЕЗ ИЗОБЛИЧЕЊА

За испитивање перформанси предложеног решења употребљен је јавно доступан, чист говорни сигнал снимљен у глувој соби [АСЕ]. Као референтни систем за препознавање говора коришћен је Амазонов сервис Транскрајб (енгл. *Amazon Transcribe*). Детаљни система за испитивање гласовних команди дати су у поглављу 5.5. Детаљни резултати препознавања су дати у табели А.1. Укупна грешка износи 17,31%.

Табела А.1 WER говора без изобличења

ИМЕ ДАТОТЕКЕ	C	S	D	I	WER	ИМЕ ДАТОТЕКЕ	C	S	D	I	WER
F1S1	1	0	0	0	0,00%	F3s4	162	22	11	0	16,92%
F1S2	5	0	0	0	0,00%	F3s5	9	1	0	0	10,00%
F1S3	21	1	0	0	4,55%	F4s1	1	0	0	0	0,00%
F1S4	125	28	5	12	28,48%	F4s2	4	1	0	0	20,00%
F1S5	3	4	3	0	70,00%	F4s3	24	1	0	1	8,00%
F2S1	0	1	0	0	100,00%	F4s4	105	6	6	0	10,26%
F2S2	4	0	0	0	0,00%	F4s5	5	3	2	0	50,00%
F2S3	32	4	3	1	20,51%	F5s1	1	0	0	0	0,00%
F2S4	187	32	13	2	20,26%	F5s2	4	0	0	0	0,00%
F2S5	10	0	0	0	0,00%	F5s3	15	1	0	0	6,25%
F3S1	1	0	0	0	0,00%	F5s4	207	21	3	4	12,12%
F3S2	4	0	0	0	0,00%	F5s5	2	3	5	0	80,00%
F3S3	19	0	0	0	0,00%						

Табела А.1 WER говора без изобличења - наставак

ИМЕ ДАТОТЕКЕ	C	S	D	I	WER	ИМЕ ДАТОТЕКЕ	C	S	D	I	WER
M1S1	1	0	0	0	0,00%	M4s3	17	1	1	0	10,53%
M1S2	5	0	0	0	0,00%	M4s4	120	25	5	2	21,33%
M1S3	14	0	1	0	6,67%	M4s5	4	1	5	0	60,00%
M1S4	88	17	6	2	22,52%	M5s1	1	0	0	0	0,00%
M1S5	7	2	1	0	30,00%	M5s2	4	0	0	0	0,00%
M2S1	1	0	0	0	0,00%	M5s3	28	1	3	0	12,50%
M2S2	0	5	0	0	100,00%	M5s4	250	39	8	3	16,84%
M2S3	27	1	1	0	6,90%	M5s5	8	2	0	0	20,00%
M2S4	140	29	12	4	24,86%	M6s3	17	1	0	0	5,56%
M2S5	6	3	1	0	40,00%	M6s4	147	13	5	1	11,52%
M3S1	0	1	0	0	100,00%	M7s3	19	4	0	0	17,39%
M3S2	4	0	0	0	0,00%	M7s4	164	18	12	1	15,98%
M3S3	21	3	1	1	20,00%	M8s3	12	1	0	1	15,38%
M3S4	193	18	6	2	11,98%	M8s4	135	6	3	2	7,64%
M3S5	10	0	0	1	10,00%	M9s3	26	6	1	0	21,21%
M4S1	1	0	0	0	0,00%	M9s4	25	10	3	2	39,47%
M4S2	4	0	0	0	0,00%						

ПРИЛОГ Б - МЕРЕЊЕ ВРЕМЕНА РЕВЕРБЕРАЦИЈЕ

У наставку је дат приказ мерења времена реверберације, RT60, у просторијама које су послужиле као окружење за испитивања у оквиру ове дисертације. Време реверберације зависи од величине просторије и коефицијената апсорпције материјала од којих се звук рефлектује. RT60 је зависна од фреквенције. Мерене су вредности на трећинама октаве од 125 Hz до 4 kHz. На основу мерења добијено је средње RT60 за сваку просторију. Средње вредности су узете као параметри симулација које су описане у поглављу 4 и прилогу В.

Табела Б.1 Време реверберације у просторији 1

МЕРЕЊЕ	RT60 ЦЕНТРАЛНА ФРЕКВЕНЦИЈА (HZ)						
	125	250	500	1000	2000	4000	
1	0,52	0,25	0,47	0,49	0,53	0,48	0,52
2	0,53	0,36	1,19	0,40	0,59	0,53	0,49
3	0,51	0,45	0,46	0,44	0,54	0,48	0,54
4	0,50	0,47	0,49	0,46	0,56	0,53	0,50
5	0,52	0,33	0,60	0,45	0,49	0,57	0,53
6	0,54	0,32	0,65	0,49	0,55	0,54	0,49
7	0,50	0,49	0,49	0,47	0,53	0,54	0,51
8	0,53	0,51	0,63	0,54	0,57	0,54	0,55
9	0,52	0,58	0,59	0,52	0,55	0,54	0,52
10	0,54	0,21	0,40	0,46	0,56	0,64	0,57

Табела Б.2 Време реверберације у просторији 2

МЕРЕЊЕ	ЦЕНТРАЛНА ФРЕКВЕНЦИЈА (HZ)						
	RT60	125	250	500	1000	2000	4000
1	0,39	0,24	0,38	0,44	0,44	0,35	0,38
2	0,44	0,14	0,45	0,43	0,41	0,39	0,38
3	0,43	0,44	0,23	0,43	0,51	0,49	0,44
4	0,43	0,14	0,27	0,40	0,51	0,48	0,44
5	0,47	0,27	0,25	0,44	0,59	0,55	0,48
6	0,46	0,21	0,18	0,43	0,56	0,47	0,45
7	0,42	0,33	0,36	0,45	0,40	0,42	0,41
8	0,43	0,24	0,39	0,33	0,52	0,43	0,44
9	0,44	0,18	0,29	0,35	0,55	0,47	0,49
10	0,46	0,22	0,26	0,31	0,57	0,52	0,50

Табела Б.3 Време реверберације у просторији 3

МЕРЕЊЕ	ЦЕНТРАЛНА ФРЕКВЕНЦИЈА (HZ)						
	RT60	125	250	500	1000	2000	4000
1	0,47	0,20	0,42	0,62	0,52	0,47	0,43
2	0,45	0,43	0,44	0,49	0,46	0,40	0,42
3	0,46	0,58	0,47	0,46	0,43	0,46	0,41
4	0,47	0,30	0,27	0,38	0,51	0,50	0,46
5	0,47	0,41	0,37	0,49	0,51	0,47	0,43
6	0,46	0,54	0,47	0,54	0,49	0,46	0,40
7	0,47	0,32	0,40	0,41	0,54	0,48	0,44
8	0,39	0,24	0,19	0,41	0,41	0,46	0,40
9	0,46	0,23	0,37	0,48	0,49	0,47	0,41
10	0,42	0,18	0,35	0,45	0,47	0,44	0,41

Табела Б.4 Време реверберације у просторији 4

МЕРЕЊЕ	ЦЕНТРАЛНА ФРЕКВЕНЦИЈА (HZ)						
	RT60	125	250	500	1000	2000	4000
1	0,88	0,28	0,67	0,96	0,87	0,86	0,79
2	0,87	0,79	1,27	0,79	0,86	0,82	0,75
3	0,88	0,55	1,02	0,98	1,07	0,88	0,79
4	0,85	0,92	1,06	0,93	0,85	0,79	0,75
5	0,86	0,40	1,16	0,76	0,94	0,77	0,71
6	0,85	0,78	1,26	0,76	0,83	0,83	0,71
7	0,85	0,76	1,21	0,88	0,87	0,79	0,71
8	0,86	0,29	1,23	0,92	0,86	0,78	0,75
9	0,82	0,64	1,25	0,65	0,80	0,86	0,76
10	0,83	0,20	1,17	0,91	0,88	0,75	0,78

Табела Б.5 Време реверберације у просторији 5

МЕРЕЊЕ	ЦЕНТРАЛНА ФРЕКВЕНЦИЈА (HZ)						
	RT60	125	250	500	1000	2000	4000
1	0,30	0,19	0,07	0,25	0,41	0,36	0,33
2	0,33	0,15	0,24	0,30	0,39	0,38	0,39
3	0,34	0,14	0,19	0,33	0,47	0,38	0,37
4	0,30	0,14	0,15	0,20	0,41	0,45	0,39
5	0,33	0,14	0,18	0,38	0,39	0,40	0,35
6	0,30	0,15	0,10	0,26	0,41	0,42	0,35
7	0,37	0,12	0,16	0,23	0,46	0,43	0,44
8	0,31	0,15	0,19	0,27	0,38	0,35	0,30
9	0,31	0,14	0,12	0,27	0,32	0,30	0,30
10	0,36	0,15	0,17	0,26	0,50	0,42	0,36

Прилог В - Испитни случајеви

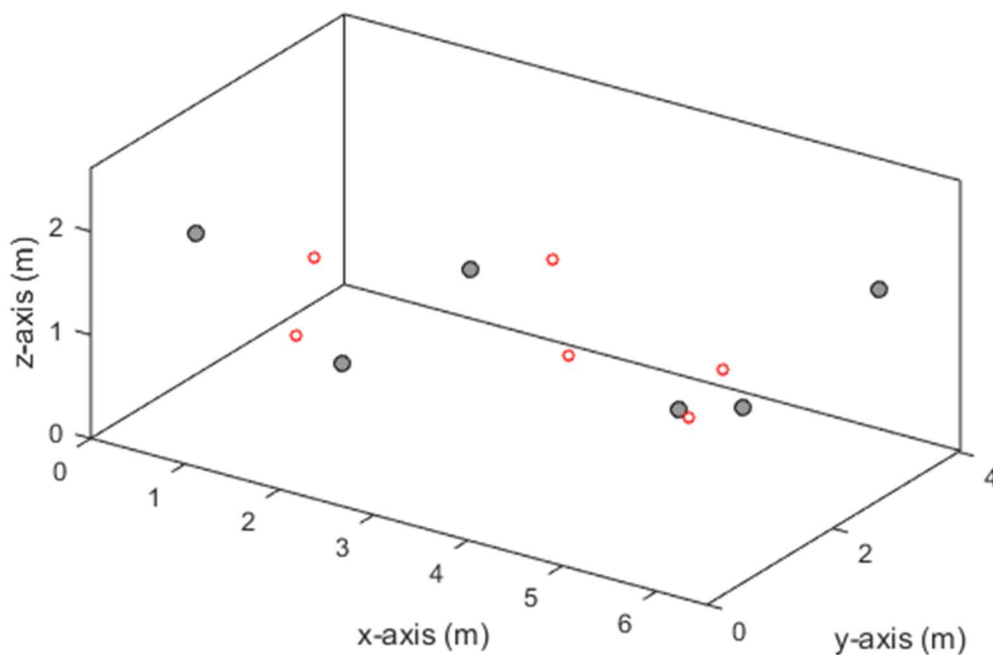
Овде су приказани сви параметри симулације код пет изведених експеримената за просторије описане у поглављу 4. Све димензије, раздаљине и позиције су одређене мерном траком и могућа су незнатна одступања. Раздаљине и позиције су изражене у метрима са два децимална места. У овом прилогу су дати и коефицијенти апсорпције. Да би се постигла верна симулација потребно је пажљиво одредити параметре модела. Приказани параметри симулације су уједно и параметри експеримената у физичком простору.

Позиције су одабране тако да њихово постављање не омета животни простор где је то било могуће. То су места где се лако може довести напајање зидног исправљача. Комуникација се одвија бежично те за њу нису потребне никакве додатне инсталације. Ово је велика предност коју истичу многи аутори и сматра се за битан услов за успешност прихватања нових IoT уређаја.

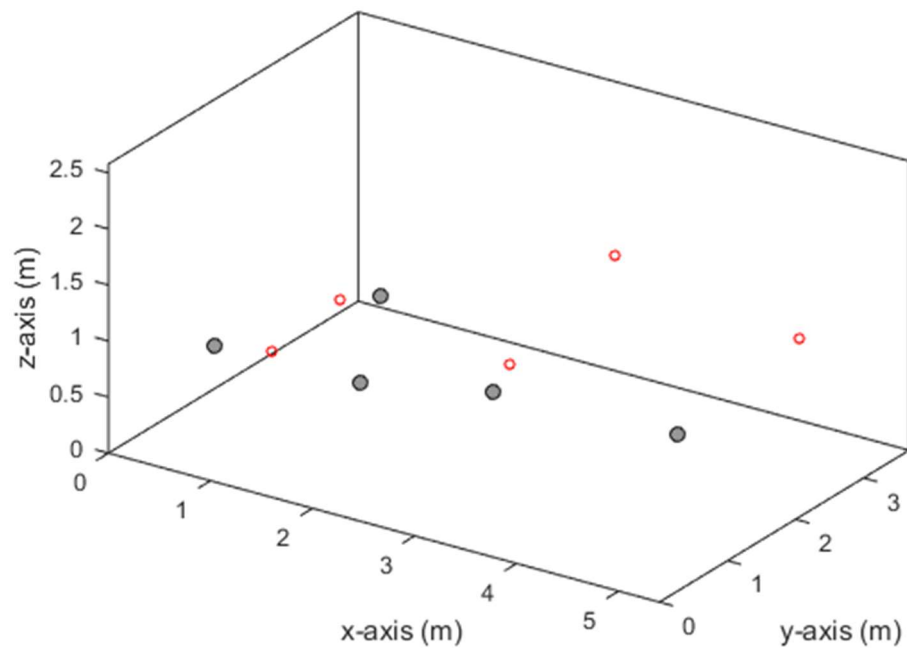
Позиција говорника у симулацији представља тачку где се налазе уста говорника. Извор звука у симулацији није усмерен. Код експеримената у реалном простору се на то место поставља звучник. Коришћен је Tannoy Mercury MX1 звучник и звучно појачало интегрисано у аудио-видео ресиверу (енгл. *audio video receiver - AVR*) Yamaha RX-V550. За везу са звучном картицом је коришћена је дигитална веза путем S/PDIF оптичког кабла. Подешавање гласноће система за репродукцију звука је постигнуто тако што се емитовани бели шум снимао

бежичним сензором у пробним циклусима. Добијени ниво звука се поредио са познатим нивоом звука добијеним у симулацији. Корекције нивоа су рађене док се није постигао исти ниво. Након тога је подешавање гласноће остало непромењено у току свих експеримената.

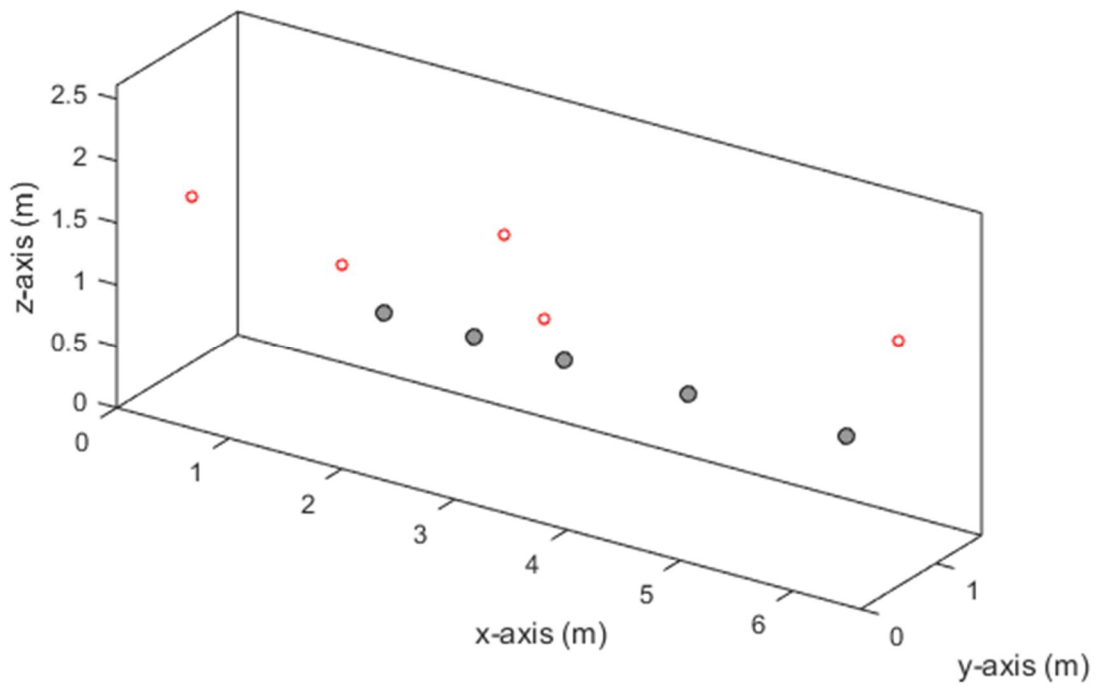
Симулиране просторије са реверберацијом су приказане на илустрацијама од В.1. до В.5. На илустрацијама су положаји сензора приказани испуњеним кружницом док су положаји говорника приказани неиспуњеном кружницом. Димензије просторија су изражене у метрима.



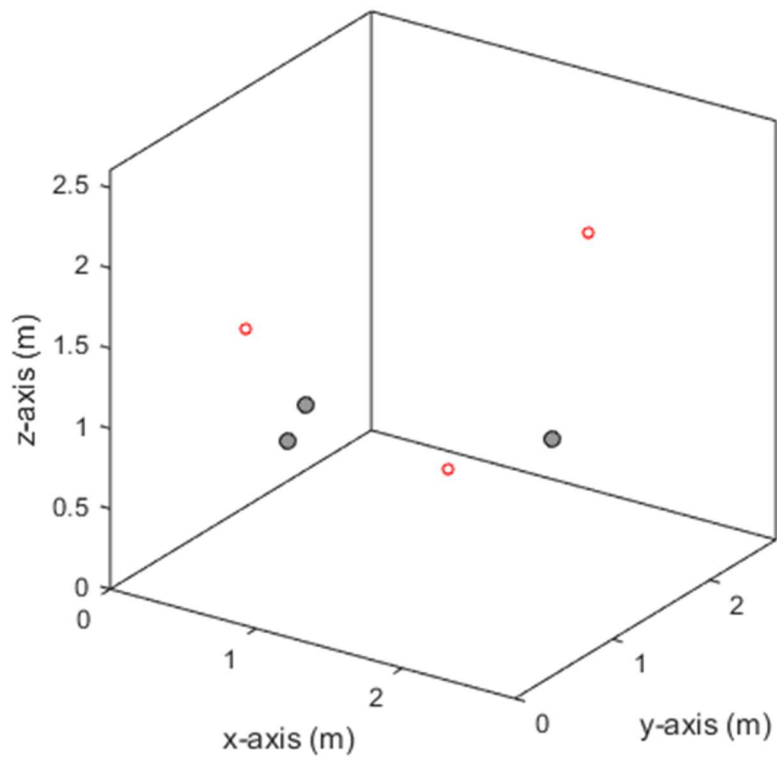
Илустрација В.1 Распоред говорника и сензора у просторији 1



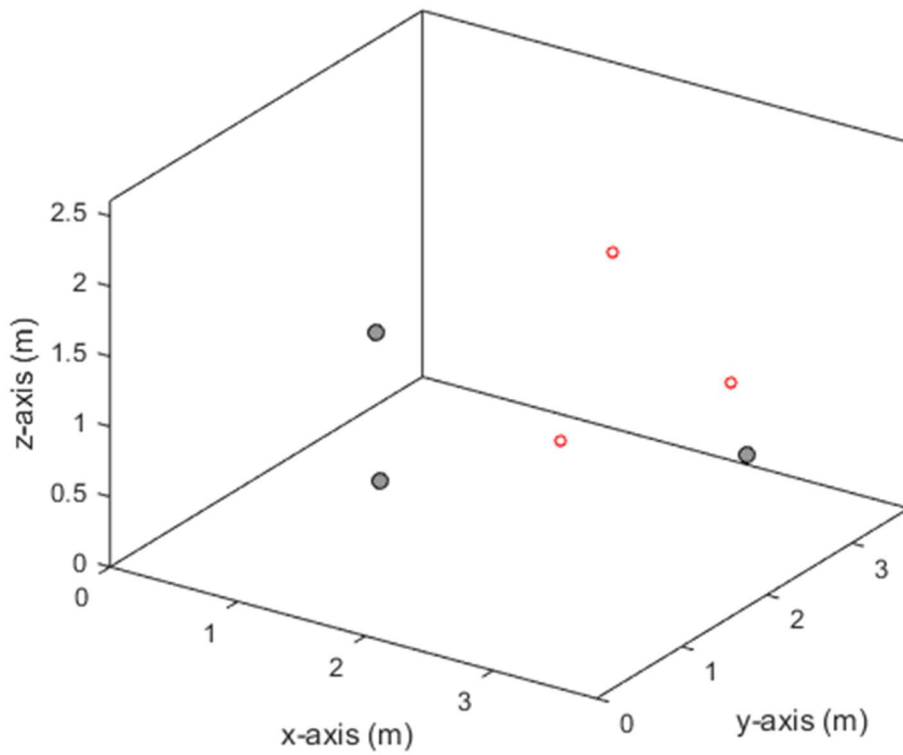
Илустрација В.2 Распоред говорника и сензора у просторији 2



Илустрација В.3 Распоред говорника и сензора у просторији 3



Илустрација В.4 Распоред говорника и сензора у просторији 4



Илустрација В.5 Распоред говорника и сензора у просторији 5

Параметри симулације дати ту у табелама од В.1. до В.5. Позиције су одређене у декартовом координатном систему какав је приказан на сликама изнад од В.1 до В.5.

Табела В.1 Параметри симулације у просторији 1

Параметар	Опис параметра	Вредност параметра
d_{room}	Димензије собе	[6,53 m 4,03 m 2,61 m] ^T
$p_{source,1}$	Позиције извора звука	[1,5 m 1 m 1 m] ^T
$p_{source,2}$	-	[0,7 m 2,5 m 1 m] ^T
$p_{source,3}$	-	[3,1 m 2,7 m 1,5 m] ^T
$p_{source,4}$	-	[4 m 1,6 m 1,2 m] ^T
$p_{source,5}$	-	[5,3 m 2,1 m 1,2 m] ^T
$p_{source,6}$	-	[6 m 0,5 m 1,5 m] ^T
$p_{sensor,1}$	Позиције сензора	[1,5 m 1,75 m 0,45 m] ^T
$p_{sensor,2}$	-	[1,5 m 3,78 m 0,6 m] ^T
$p_{sensor,3}$	-	[0,3 m 1,23 m 1,6 m] ^T
$p_{sensor,4}$	-	[5,85 m 3,75 m 1,6 m] ^T
$p_{sensor,5}$	-	[6,29 m 0,93 m 1,5 m] ^T
$p_{sensor,6}$	-	[4,93 m 1,94 m 0,77 m] ^T
A	Коефицијенти апсорпције	[0,2417 0,3022 0,3022 0,3022 0,6715 0,3022]
RT_{60}	Време реверберације	520 ms

Табела В.2 Параметри симулације у просторији 2

Параметар	Опис параметра	Вредност параметра
d_{room}	Димензије собе	[5,4 m 3,67 m 2,58 m] ^T
$p_{source,1}$	Позиције извора звука	[0,6 m 1,5 m 0,5 m] ^T
$p_{source,2}$	-	[0,6 m 2,5 m 0,6 m] ^T
$p_{source,3}$	-	[3,6 m 0,5 m 1,5 m] ^T
$p_{source,4}$	-	[3,1 m 2,8 m 1,5 m] ^T
$p_{source,5}$	-	[4,9 m 2,8 m 1,2 m] ^T
$p_{sensor,1}$	Позиције сензора	[0,7 m 0,5 m 0,95 m] ^T
$p_{sensor,2}$	-	[2,1 m 0,55 m 0,95 m] ^T
$p_{sensor,3}$	-	[2,5 m 1,91 m 0,45 m] ^T
$p_{sensor,4}$	-	[4,7 m 1,33 m 0,83 m] ^T
$p_{sensor,5}$	-	[0,55 m 3,17 m 0,36 m] ^T
α	Коефицијенти апсорпције	[0,2937 0,2937 0,2937 0,2546 0,2219 0,2937]
RT_{60}	Време реверберације	435 ms

Табела В.3 Параметри симулације у просторији 3

Параметар	Опис параметра	Вредност параметра
d_{room}	Димензије собе	[6,6 m 1,61 m 2,61 m] ^T
$p_{source,1}$	Позиције извора звука	[0,2 m 0,7 m 1,5 m] ^T
$p_{source,2}$	-	[1,8 m 0,3 m 1,5 m] ^T
$p_{source,3}$	-	[2,5 m 1,4 m 1,5 m] ^T
$p_{source,4}$	-	[3,6 m 0,3 m 1,5 m] ^T
$p_{source,5}$	-	[6 m 1,4 m 1,5 m] ^T
$p_{sensor,1}$	Позиције сензора	[1,8 m 0,85 m 0,91 m] ^T
$p_{sensor,2}$	-	[2,6 m 0,85 m 0,91 m] ^T
$p_{sensor,3}$	-	[3,4 m 0,85 m 0,91 m] ^T
$p_{sensor,4}$	-	[4,5 m 0,85 m 0,91 m] ^T
$p_{sensor,5}$	-	[5,9 m 0,85 m 0,91 m] ^T
α	Коефицијенти апсорпције	[0,1729 0,1729 0,1306 0,1306 0,0615 0,1729]
RT_{60}	Време реверберације	855 ms

Табела В.4 Параметри симулације у просторији 4

Параметар	Опис параметра	Вредност параметра
d_{room}	Димензије собе	[2,77 m 2,68 m 2,61 m] ^T
$p_{source,1}$	Позиције извора звука	[2,05 m 0,4 m 1,1 m] ^T
$p_{source,2}$	-	[0,5 m 0,65 m 1,5 m] ^T
$p_{source,3}$	-	[1,7 m 2,35 m 1,77 m] ^T
$p_{sensor,1}$	Позиције сензора	[1 m 0,32 m 1,05m] ^T
$p_{sensor,2}$	-	[2,05 m 1,47 m 0,9 m] ^T
$p_{sensor,3}$	-	[0,3 m 1,55 m 0,65 m] ^T
α	Коефицијенти апсорпције	[0,1435 0,1614 0,1704 0,1435 0,1793 0,4035]
RT_{60}	Време реверберације	460 ms

Табела В.5 Параметри симулације у просторији 5

Параметар	Опис параметра	Вредност параметра
d_{room}	Димензије собе	[3,81 m 3,67 m 2,61 m] ^T
$p_{source,1}$	Позиције извора звука	[3,1 m 2,65 m 1,1 m] ^T
$p_{source,2}$	-	[3,2 m 0,5 m 1,5 m] ^T
$p_{source,3}$	-	[1,8 m 3,2 m 1,5 m] ^T
$p_{sensor,1}$	Позиције сензора	[3,45 m 2,3 m 0,8 m] ^T
$p_{sensor,2}$	-	[1,95 m 0,25 m 1 m] ^T
$p_{sensor,3}$	-	[0,85 m 1,85 m 1,2 m] ^T
α	Коефицијенти апсорпције	[0,312 0,312 0,312 0,2689 0,2344 0,3102]
RT_{60}	Време реверберације	330 ms

У наставку су дате грешке препознавања, WER, за сваку локацију говорника на сваком сензору. У табелама од В.6 до В.10 су дати парови укупног WER за симулацију и одговарајући експеримент у реалној просторији.

Табела В.6 Грешке препознавања у просторији 1

Сензор:	1		2		3		4		5		6	
Локација:	сим.	реал.	сим.	реал.	сим.	реал.	сим.	реал.	сим.	реал.	сим.	реал.
1	19,66	23,03	35,04	29,09	19,66	26,06	38,46	36,97	45,40	43,64	31,62	36,97
2	19,66	26,06	23,08	22,42	18,80	31,52	27,35	44,85	27,35	38,79	31,62	29,70
3	23,08	35,76	19,66	33,94	34,19	32,73	33,33	36,97	27,35	34,55	22,22	37,58
4	32,48	36,36	25,64	41,21	37,61	32,73	28,21	35,15	26,50	24,85	20,51	24,24
5	27,35	38,79	25,64	34,85	35,04	38,18	29,06	23,64	23,08	29,09	15,37	21,82
6	25,64	48,48	23,08	36,36	31,62	39,39	22,22	29,70	14,54	18,18	29,06	26,06

Табела В.7 Грешке препознавања у просторији 2

Сензор:	1		2		3		4		5	
Локација	сим.	реал.	сим.	реал.	сим.	реал.	сим.	реал.	сим.	реал.
1	23,93	24,24	37,61	34,55	28,21	27,88	29,06	30,30	20,51	34,55
2	34,19	26,67	46,15	36,97	26,50	29,70	46,15	38,18	20,51	21,82
3	44,44	40,61	29,91	28,48	28,21	35,76	22,22	26,06	32,48	24,85
4	41,88	38,18	27,35	32,12	30,77	32,73	43,59	38,79	25,64	37,58
5	52,99	40,00	38,46	41,82	27,35	32,12	28,21	30,30	25,64	37,58

Табела В.8 Грешке препознавања у просторији 3

Сензор:	1		2		3		4		5	
Локација	сим.	реал.	сим.	реал.	сим.	реал.	сим.	реал.	сим.	реал.
1	67,52	53,94	53,50	53,94	62,39	53,94	64,96	48,48	63,25	55,76
2	55,56	46,06	46,15	46,06	60,68	46,06	58,12	58,18	62,39	52,73
3	42,74	40,61	36,75	40,61	46,15	40,61	53,85	57,58	55,56	60,00
4	56,41	52,73	43,59	52,73	40,17	52,73	52,99	44,85	63,25	61,21
5	58,12	42,42	59,83	42,42	47,01	42,42	41,03	39,39	32,48	39,39

Табела В.9 Грешке препознавања у просторији 4

Сензор:	1		2		3	
Локација	сим.	реал.	сим.	реал.	сим.	реал.
1	35,90	32,73	38,46	32,73	22,22	32,73
2	29,06	30,30	35,90	30,30	24,79	30,30
3	33,85	37,58	42,74	37,58	32,48	37,58

Табела В.10 Грешке препознавања у просторији 5

Сензор:	1		2		3	
Локација	сим.	реал.	сим.	реал.	сим.	реал.
1	17,09	20,00	35,04	28,48	35,04	26,67
2	17,09	29,70	22,22	26,06	28,21	26,67
3	24,79	29,09	38,46	29,70	30,77	31,52