

Державна служба України з надзвичайних ситуацій
Львівський державний університет безпеки життєдіяльності
Навчально-науковий інститут цивільного захисту
Кафедра інформаційних технологій та систем електронних комунікацій

«Допущено до захисту»
Начальник кафедри ІТ та СЕК

“ _____ ” _____ 20__ року

ДИПЛОМНА РОБОТА МАГІСТРА

на тему «Розроблення методу підвищення роздільної здатності голосових повідомлень для керування мікрокліматом приміщення»

Виконав:

здобувач VI курсу, групи КН-61м
спеціальності (освітньої програми)

122 «Комп'ютерні науки» (Комп'ютерні науки)
(шифр і назва спеціальності (освітньої програми))

Остап-Святослав Малець

(ім'я та прізвище)

Керівник _____ Юрій Борзов

(ім'я та прізвище)

Рецензент _____ Тарас Рак

(ім'я та прізвище)

Львів – 2023 року

Державна служба України з надзвичайних ситуацій
Львівський державний університет безпеки життєдіяльності
Навчально-науковий інститут цивільного захисту

Кафедра інформаційних технологій та систем електронних комунікацій

Освітній ступінь магістр

Спеціальність 122 “Комп’ютерні науки”

Освітня програма Комп’ютерні науки

ЗАТВЕРДЖУЮ

Начальник кафедри ІТ та СЕК

“ _____ ” _____ 20__ року

ЗАВДАННЯ

на дипломну роботу

Здобувач Остап-Святослав Малець

(ім’я, прізвище)

1. Тема Розроблення методу підвищення роздільної здатності голосових повідомлень для керування мікрокліматом приміщення

керівник роботи Юрій Борзов, к.т.н.

(ім’я, прізвище, науковий ступінь, вчене звання)

затверджені наказом ЛДУ БЖД від “ _____ ” _____ 20__ року № _____

2. Термін подання здобувачем роботи _____

3. Початкові дані до роботи:

1. Методичні вказівки до виконання дипломної роботи магістра для здобувачів другого рівня вищої освіти спеціальності 122 «Комп’ютерні науки». Укл. Ольга Смотр, Олександр Придатко, Назарій Бурак. – Львів: Вид-во ЛДУ БЖД, 2019. – 29 с.

4. Зміст дипломної роботи/проекту (перелік питань, які потрібно розробити)

Вступ

Розділ 1. Сучасне управління голосом

Розділ 2. Розроблення методу розпізнавання голосових команд з допомогою

штучних нейронних мереж

Розділ 3. Програмно-апаратна реалізація

Розділ 4. Конструкторська частина

Висновки

Список використаних джерел

5. Консультанти розділів роботи

Розділ	Ім'я, Прізвище та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів виконання дипломної роботи/	Термін виконання етапів роботи	Примітка
1	Аналітично-інформаційний огляд систем управління голосом		
2	Розроблення методу розпізнавання голосових команд з допомогою штучних нейронних мереж		
3	Програмно-апаратна реалізація		
4	Конструкторська частина		

Здобувач

Остап-Святослав МАЛЕЦЬ

(підпис) (ім'я та прізвище)

Керівник роботи

Юрій БОРЗОВ

(підпис) (ім'я та прізвище)

АНОТАЦІЯ

Останнім часом спостерігається зростання інтересу до технологій, що пов'язані з розпізнаванням мови. Завдання управління пристроями за допомогою голосових команд, інтерактивні платформи які надають інформацію після запиту в більш природній формі - за допомогою голосу, все це знаходить широке застосування в сучасному інформаційному світі. Багато задач виникає при бажанні взаємодіяти за допомогою голосу з мобільними пристроями. Наприклад, введення голосових команд для отримання інформації з Інтернету, прокладання маршруту руху, запуск програм користувача, диктування тексту. Останнім часом з'явилася можливість управління домашньою, офісною технікою за допомогою електронних пристроїв (смартфонів) голосовими командами.

Передумовою розвитку голосових технологій є значне збільшення обчислювальних можливостей, обсягу пам'яті при значному зменшенні габаритів комп'ютерних систем. Слід також відзначити розвиток математичних методів, що дозволяють виконати необхідну обробку аудіо сигналу шляхом виділення з нього інформативних ознак.

Для прикладу, широко використовується дискретне перетворення Фур'є, яке відоме з теорії цифрової обробки сигналів. Подальша обробка виконується з використанням акустичної моделі, яка ставить у відповідність виділених параметрах конкретні звуки (фонемі). Слід зазначити, що в роботі розглядається можливість використання методу динамічного програмування і метод нейронних рекурентних мереж.

Розглянуто взаємодію людини і домашніх приладів яка реалізовується на базі Arduino, за допомогою голосових команд. Тут передбачається, що обробка аудіо сигналу, побудова текстового рядка на основі виголошеної фрази виконується спеціальними бібліотеками (класами, методами), що покращує розпізнавання голосу в порівнянні з існуючими методами.

SUMMARY

Recently, there has been a growing interest in technologies related to speech recognition. The task of controlling devices using voice commands, interactive platforms that provide information after a request in a more natural form - with the help of voice, all this is widely used in the modern information world. Many problems arise when you want to interact with mobile devices using voice. For example, inputting voice commands to receive information from the Internet, planning a route, launching user programs, dictating text. Recently, it has become possible to control home and office equipment using electronic devices (smartphones) with voice commands.

A prerequisite for the development of voice technologies is a significant increase in computing capabilities, the amount of memory with a significant reduction in the dimensions of computer systems. It should also be noted the development of mathematical methods that allow performing the necessary processing of an audio signal by extracting informative features from it.

For example, the discrete Fourier transform, which is known from the theory of digital signal processing, is widely used. Further processing is performed using an acoustic model that matches specific sounds (phonemes) to the selected parameters. It should be noted that the paper considers the possibility of using the method of dynamic programming and the method of neural recurrent networks.

The interaction of a person and home appliances, which is implemented on the basis of Arduino, with the help of voice commands, is considered. Here it is assumed that the processing of the audio signal, the construction of a text line based on the spoken phrase is performed by special libraries (classes, methods), which improves voice recognition in comparison with existing methods.

ЗМІСТ

АНОТАЦІЯ	4
SUMMARY	5
ВСТУП	7
Розділ I. СУЧАСНЕ УПРАВЛІННЯ ГОЛОСОМ	9
1.1 Голосове та інтелектуальне управління.	9
1.2 Поняття, призначення і види голосового управління	13
1.2.1 Поняття голосового управління	13
1.2.2 Призначення пристроїв розпізнавання мови	15
1.2.3 Види голосового управління.	16
1.3 Схема пристроїв розпізнавання голосу	17
1.4 Якість мови та її синтез	22
1.4 Оцифрування звуку	27
1.4.1 Оцифрування сигналу	27
1.4.2 Шуми	28
1.6 Аналіз ринку систем голосового управління	30
1.7 Задача керування кліматом	33
1.8 Аналіз останніх досліджень і результатів.	34
Висновки за розділом 1	40
Розділ II. РОЗРОБЛЕННЯ МЕТОДУ РОЗПІЗНАВАННЯ ГОЛОСОВИХ КОМАНД З ДОПОМОГОЮ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ	41
2.1 Запропонований спосіб розпізнавання голосових команд	41
2.1 Розпізнавання	44
2.2.2. Мережі довгої короткотривалої пам'яті (LTSM).	50
2.2.3 Згорткові нейронні мережі (CNN)	52
2.2.4 Наскрізний підхід на основі CNN	58
2.3 Алгоритм методу розпізнавання голосових команд	60
Розділ III ПРОГРАМНО-АПАРАТНА РЕАЛІЗАЦІЯ	62
3.1 Структурна схема системи розпізнавання голосових даних	62
3.3 Експериментальне дослідження методу розпізнавання голосових команд	80
Розділ IV. КОНСТРУКТОРСЬКА ЧАСТИНА	83
4.1. Огляд проектування	83
4.2 Вибір мікроконтролера	86
4.3 Середовище для проведення дослідів, принципова схема взаємодії з приладами	93
Висновки за розділом 4	97
ВИСНОВКИ	98
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ	100

ВИСНОВКИ

У магістерській кваліфікаційній роботі вирішено актуальне наукове завдання – розробка пристрою розпізнавання голосу з високими показниками точності в розпізнаванні голосу за допомогою використання рекурентних нейронних мереж.

Основні наукові теоретичні та практичні результати полягають в наступному:

Проведений огляд процесу розпізнавання голосу на основі існуючих методів, що дозволило виявити переваги та недоліки існуючих методів та виділити особливості процесу розпізнавання мови, на які необхідно звертати увагу під час проектування пристрою.

Найбільш перспективним вбачається варіант побудови пристрою голосових команд на основі розпізнавання голосу за допомогою рекурентних нейронних мереж. Це дозволить досягти високої точності та завадостійкості в розпізнаванні голосових команд.

Проведений огляд відомих принципів побудови нейронних мереж та способів їх реалізації що дозволило виявити переваги та недоліки існуючих рішень та виділити рішення, які можуть розглядатись як прототипи.

Побудова пристрою розпізнавання мови на основі РНМ забезпечує можливість збільшення голосових команд шляхом простого додавання даних в базу НМ, що є дуже істотним, оскільки підключення багатьох пристроїв є суттєвою перевагою.

Спільне використання рекурентних НМ та перетворення Фур'є дозволяє поєднати переваги обох способів: забезпечити високу точність відтворення заданого сигналу в разі необхідності.

Запропоновано методику отримання масивів даних з отриманого сигналу.

У результаті дослідження було отримано результати розпізнання команд при визначеній базі даних, які дозволяють стверджувати, що при записі одного слова 42 рази в різних умовах, в подальшому сприяють на точність розпізнавання.

Було спроектовано схема при якій вимовляння голосових команд приводило в дію прилади.

Експериментальні дослідження показали, що математичні розрахунки та імітаційне моделювання, проведені у попередніх розділах були виконані коректно.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Ардуїно [Електронний ресурс] // Arduino.ua. – 2015. – Режим доступу до ресурсу: <http://arduino.ua/>.
2. Перспективи на ринку систем голосового управління [Електронний ресурс] // Хабрахарбр. – 2016. – Режим доступу до ресурсу: <https://habrahabr.ru/post/232613/>.
3. Розпізнавання мови. [Електронний ресурс] // Вікіпедія. – 2007. – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/Распознавание_речи.
4. Голосове управління [Електронний ресурс] // Вікіпедія. – 2007. – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/Голосовое_управление.
5. Уллі С. Программирование микроконтроллерных плат Arduino/Freduino / Соммер Уллі. – Петербург, 2012.
6. Ревич Ю. Цікава електроніка / Юрій Ревич. – Петербург, 2015.
7. Карвінен Т. Робимо сенсори. Проекти сенсорних пристроїв на базі Arduino і Raspberry Pi / Т. Карвінен, К. Карвінен, В. Валтокарі., 2015.
8. Петрін В. О. Проекти з використанням контролера Arduino. 2 изд. / Віктор Олександрович Петрін..
9. Голосове управління Arduino засобами Processing і Google Speech API [Електронний ресурс]. – 13. – Режим доступу до ресурсу: <https://habrahabr.ru/post/236673/>.
10. Голосове управління вимикачами на Arduino [Електронний ресурс]. – 30. – Режим доступу до ресурсу: <http://compcar.ru/forum/showthread.php?t=8016>.
11. Перетворення Лапласа [Електронний ресурс] // Вікіпедія – Режим доступу до ресурсу: https://ru.wikipedia.org/wiki/%D0%9F%D1%80%D0%B5%D0%BE%D0%B1%D1%80%D0%B0%D0%B7%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_%D0%9B%D0%B0%D0%BF%D0%BB%D0%B0%D1%81%D0%B0

12. Частота дискретизації [Електронний ресурс] // Вікіпедія – Режим доступу до ресурсу
[:https://uk.wikipedia.org/wiki/%D0%A7%D0%B0%D1%81%D1%82%D0%BE%D1%82%D0%B0_%D0%B4%D0%B8%D1%81%D0%BA%D1%80%D0%B5%D1%82%D0%B8%D0%B7%D0%B0%D1%86%D1%96%D1%97](https://uk.wikipedia.org/wiki/%D0%A7%D0%B0%D1%81%D1%82%D0%BE%D1%82%D0%B0_%D0%B4%D0%B8%D1%81%D0%BA%D1%80%D0%B5%D1%82%D0%B8%D0%B7%D0%B0%D1%86%D1%96%D1%97).
13. Навіщо потрібні Powerline адаптери [Електронний ресурс] // Lantorg. – 2017. – Режим доступу до ресурсу: <https://lantorg.com/article/zachem-nuzhny-powerline-adaptery>.
14. Інтернет з розетки: загальні принципи роботи технології і огляд Powerline-адаптера TP-LINK TL-PA6010 [Електронний ресурс] // 3dNews. – 2014. – Режим доступу до ресурсу: <https://3dnews.ru/821880>.
15. Домашній міні-клімат-контроль своїми руками [Електронний ресурс]. – 2013. – Режим доступу до ресурсу: <https://geektimes.ru/post/258012/>.
16. Система "Розумний будинок" для заміського будинку на Arduino Mega2560, HC-05, SIM900, DHT11,3-х DS18B20, RTC-DS1302 [Електронний ресурс] // Arduino.ru. – 2015. – Режим доступу до ресурсу: <http://arduino.ru/forum/proekty/sistema-umnyi-dom-dlya-zagorodnogo-doma-na-arduino-mega2560-hc-05-sim900dht113-kh-ds18>.
17. Фролов А. В. Синтез и распознавание речи. Современные решения. [Електронний ресурс] / А. В. Фролов, Г. В. Фролов. – 2003. – Режим доступу до ресурсу: <http://www.frolov-lib.ru/books/hi/index.html>.
18. Квитко М.В. Распознавание речи с помощью глубоких рекуррентных нейронных сетей [Електронний ресурс]/ Квитко М.В. // IASA –2016 р. –223 стр. –Режим доступу: http://sait.kpi.ua/media/filer_public/73/32/7332a68e-e93b-4c57-a3c8-66f11ee074cd/sait2016ebook.pdf
19. Голосове управління Arduino засобами Processing і Google Speech API [Електронний ресурс]. – 13. – Режим доступу до ресурсу: <https://habrahabr.ru/post/236673/>.
20. Мясищев А. А. Управление голосом с помощью Android и Arduino

[Электронный ресурс] / А. А. Мясищев. – 2015. – Режим доступа до ресурсу:
http://khnu.km.ua/root/kaf/ksm/my_syte_g/.

21. Mohri M. Speech recognition with weighted finite-state transducers. In Springer Handbook of Speech Processing / M. Mohri, M. Pereira, F. Riley. // Springer Berlin Heidelberg. – 2008. – С. 559–584.

22. Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N. et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Processing Magazine, IEEE, 29(6), 82-97.

23. Jurafsky D., Martin J.H. (2008) Speech and language processing, 2nd edition. Prentice Hall.

24. Голосове керування [Электронный ресурс] // Wikipedia. – 2007. – Режим доступа до ресурсу:
https://ru.wikipedia.org/wiki/Голосовое_управление.

25. Дискретное преобразование Фурье. Википедия. [Electronic resource]. - Mode of access:
https://ru.wikipedia.org/wiki/Дискретное_преобразование_Фурье.

26. Fast Artificial Neural Network Library(FANN). [Electronic resource].
 - Mode of access: <http://leenissen.dk/fann/wp/>.

27. Audacity. Википедия. [Electronic resource]. - Mode of access:
<https://ru.wikipedia.org/wiki/Audacity>.

28. Arduino. Официальный сайт. [Electronic resource]. - Mode of access: <http://arduino.cc> , 2015.

29. Круглов В. Искусственные нейронные сети / В. Круглов, В. Борисов. – Горячая Линия – Телеком, 2001.

30. Холоденко А.Б., “О построении статистических языковых моделей для систем распознавания русской речи” // Интеллектуальные системы, 2002. Т.6. Вип. 1-4. С. 381-394.

31. MIT Lectures 2003. <http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-345-automatic-speech-recognition-spring->

2003/downloadcourse-materials/

32. Фант. Г. Акустическая теория речеобразования. «Наука». Москва 1964.
4. Picone J. Fundamentals of speech recognition: a short course.1996.
http://speech.tifr.res.in/tutorials/fundamentalOfASR_picone96.pdf
33. Application to automatic speech recognition. In Proc. INTERSPEECH-2006, Pittsburgh, USA, 2006, pp. 1021-1024.
34. Schlippe T., Ochs S., Schultz T., Grapheme-to-Phoneme Model Generation for Indo-European Languages. In Proc. ICASSP-2012, Kyoto, Japan, 2012.
35. Huang C., Chang E., Zhou J., Lee K. Accent modeling based on pronunciation dictionary adaptation for large vocabulary Mandarin speech recognition. In Proc. INTERSPEECH-2000, Beijing, China, 2000, pp. 818-821
36. Hannemann M., “Combinations of Confidence Measures for the Detection of Out-of-Vocabulary Segments in Large Vocabulary Continuous Speech Using Differently Constrained Recognizers”, Otto-von-Guericke-Universitat Magdeburg, 21. April 2008.
37. Bourlard H., Wellekens C.J., “Links between Markov models and multilayer perceptrons”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12 , No. 12, 1990, pp. 1167-1178.
38. Bourlard H., Hermansky H., Morgan N., “Towards increasing speech recognition error rates”, Speech Communication, Vol. 18, 1996, p.p. 205–231.
39. Hornik K., Stinchcombe M., White H., “Multilayer feedforward networks are universal approximators”, Neural Netw. Vol. 2(5), 1989, pp. 359–366.
40. Hinton G., Deng L., Yu D., Dahl G., Mohamed A., Jaitly N., Senior A., Vanhoucke V., Nguyen P., Sainath T., Kingsbury B., “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups”, IEEE Signal Process. Mag., Vol. 29, No. 6, Nov. 2012, pp. 82–97.
55. Dong Yu, Li Deng, “Automatic Speech Recognition. A Deep Learning Approach”, Springer-Verlag, London. 2015, 321 p.
41. Чистович Л.А. и др., «Руководство по физиологии. Физиология

речи. Восприятие речи человеком», «Наука», Ленинград, 1976.

42. Hermansky H., Ellis D., Sharma S., “Tandem connectionist feature extraction for conventional HMM systems”, Proc. ICASSP-2000, Istanbul. 2000. V. 3. pp. 1635–1638.

43. Deng, L., Chen, J., “Sequence classification using high-level features extracted from deep neural networks.” In: Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pp. 6894-6898.

44. Hochreiter S., Schmidhuber J., “Long short-term memory.” Neural Computation, V. 9(8), 1997, pp. 1735–1780.

45. Pascanu R., Mikolov T., Bengio Y., “On the difficulty of training recurrent neural networks”, Cornell University Library, arXiv:1211.5063 [cs.LG], 2013.