

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

магістр

(назва освітнього ступеня)

на тему: «Обґрунтування методу розпізнавання голосових звуків
для голосового керування IoT дівайсами»

Виконав(ла): студент(ка) VI курсу, групи РРМ-61
спеціальності 172 Телекомунікації та радіотехніка

(шифр і назва спеціальності)

(підпис)

Дембічак А.А.

(прізвище та ініціали)

Керівник

(підпис)

Яськів В.І.

(прізвище та ініціали)

Нормоконтроль

(підпис)

Хвостівська Л. В.

(прізвище та ініціали)

Завідувач кафедри

(підпис)

Дунець В. Л.

(прізвище та ініціали)

Рецензент

(підпис)

(прізвище та ініціали)

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя
(повне найменування вищого навчального закладу)

Факультет Прикладних інформаційних технологій та електроінженерії

Кафедра Радіотехнічних систем

Освітній ступінь Магістр

Напрямок підготовки _____

(шифр і назва)

Спеціальність 172 «Телекомунікації та радіотехніка»

(шифр і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри _____

« _____ » _____ 201__ р.

ЗАВДАННЯ
НА ДИПЛОМНИЙ ПРОЕКТ (РОБОТУ) СТУДЕНТУ

Дембічак Анатолій Андрійович

(прізвище, ім'я, по батькові)

1. Тема проекту (роботи) «Обґрунтування методу розпізнавання голосових звуків для голосового керування IoT дівайсами»

Керівник проекту (роботи) к.т.н., доцент, декан ФІТ Яськів В.І.

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом по університету від « 24 » листопада _____ 2020 року № 4/7-870

2. Термін подання студентом проекту (роботи) 19.12.2020

3. Вихідні дані до проекту (роботи) Частотний спектр – 200 Hz – 20 kHz

Рівень тембру звуку ≥ 5 db

Гучність ≥ 3 db

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

1. Аналітична частина. Голосове управління технічними пристроями. Модель голосового управління. Схема управління.

2. Основна частина. Методи розпізнавання мережі, нейромережа.

3. Науково-дослідницька частина. Розробка методу розпізнавання голосових звуків.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

слайди 1-2 - актуальність теми, мети і задачі.

слайди 3-5 - архітектура, схемні рішення мережі

слайд 6 - схеми Хеммінга

слайди 7-9 - метод розпізнавання мови

слайд 10 - висновки

АНОТАЦІЯ

Тема кваліфікаційної роботи: «Обґрунтування методу розпізнавання голосових звуків для голосового керування IoT дівайсами» // кваліфікаційна робота // Дембічак Анатолій Андрійович // Тернопільський національний технічний університет імені Івана Пулюя, факультет прикладних інформаційних технологій та електроінженерії, група РРм-61 // Тернопіль, 2020 // с.- 59, рис.- 19, табл. – 2, додат. – 1, бібліогр. -32.

Ключові слова: ГОЛОСОВИЙ КОНТРОЛЬ, ПРИСТРОЇ ІОТ, НЕЙРОННІ МЕРЕЖІ, РОЗПІЗНАВАННЯ ГОЛОСУ, ПРИХОВАНІ МОДЕЛІ МАРКОВА, ГОЛОСОВІ sМОНОЛОГИ, МІКРОФОНИ, АУДІО ПОВІДОМЛЕННЯ.

Дипломну роботу присвячено дослідженню голосового управління технічними пристроями, модель голосового управління периферійними пристроями, схема управління електричними пристроями. Основні етапи розпізнавання мови - ключові моменти, можливість використання нейромереж для побудови системи розпізнавання мови, модель нейромережі та навчання нейромережі. Здійснюється аналіз структури розпізнавання мови та розрахунок mel-фільтрів.

Зв'язок між пристроями відбувається через відповідний інтерфейс (у нашому випадку голосовий). Прикладами таких комунікаційних інтерфейсів є голосовий пошук Google та Siri Assistant від Apple. Створюючи повністю функціональний та надійний інтерфейс такого типу, слід враховувати інформатику, лінгвістику та особисту поведінкову психологію. Для того, щоб використовувати штучний інтелект для підпорядкування голосових команд певного пристрою, вам потрібно вирішити наступні проблеми. Ці проблеми ґрунтуються на складності розпізнавання голосу залежно від емоційного стану, а саме тембру, гучності, висоти, швидкості, інтонації, інтонації та якості мови тощо. Навіть в екстремальних умовах ці завдання з розпізнавання мови повинні бути вирішені.

ANNOTATION

Theme of qualification work: "Substantiation of the method of voice sound recognition for voice control of IoT devices" // Thesis // Dembichak Anatoliy Andriyovych // Ternopil National Technical University named after Ivan Pulyuy, Faculty of Applied Information Technologies and Electrical Engineering, group PPM-61 // Ternopil, 2020 // p.- 59, fig.- 19, tab. - 2, Add - 1, Ref. -32.

Keywords: VOICE CONTROL, IOT DEVICES, NEURAL NETWORKS, VOICE RECOGNITION, HIDDEN MARKOV MODELS, VOICE MONOLOGUES, MICROPH.

Thesis is devoted to the study of voice control of technical devices, the model of voice control of peripherals, control circuit of electrical devices. The main stages of language recognition are key points, the possibility of using neural networks to build a language recognition system, neural network model and neural network training. The analysis of the structure of language recognition and calculation of mel-filters is carried out.

Communication between devices takes place through the appropriate interface (in our case, voice). Examples of such communication interfaces are Google voice search and Apple's Siri Assistant. Creating a fully functional and reliable interface of this type should take into account computer science, linguistics and personal behavioral psychology. In order to use artificial intelligence to obey the voice commands of a particular device, you need to solve the following problems. These problems are based on the difficulty of voice recognition depending on the emotional state, namely timbre, volume, pitch, speed, intonation, intonation and quality of speech, and so on. Even in extreme conditions, these language recognition tasks must be solved.

ЗМІСТ

ВСТУП.....	
РОЗДІЛ 1 АНАЛІТИЧНА ЧАСТИНА.....	
1.1 Голосове управління технічними пристроями	
1.2 Модель голосового управління периферійними пристроями.....	
1.3 Схема управління електричними пристроями.....	
1.4 Висновок до розділу 1.....	
РОЗДІЛ 2 ОСНОВНА ЧАСТИНА.....	
2.1 Розпізнавання мови - ключові моменти	
2.2 Опис системи розпізнавання мови яка самонавчається.....	
2.3 Модель нейромережі.....	
2.4. Навчання нейромережі.....	
2.5 Висновок до розділу 2.....	
РОЗДІЛ 3 НАУКОВО-ДОСЛІДНА ЧАСТИНА.....	
3.1 Основи розпізнавання мови.....	
3.2 Структура розпізнавання мови.....	
3.3 Розрахунок mel-фільтрів.....	
3.4 Приховані Марковські моделі.....	
3.5 Висновок до розділу 3.....	
РОЗДІЛ 4 ОХОРОНА ПРАЦІ БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ.....	
4.1 Забезпечення безпеки життєдіяльності підприємств електротехнічної галузі у воєнний час.....	
4.2 Класи виробничих та складських приміщень по вибуховій та пожежній небезпеці. Вогнестійкість будівельних конструкцій і матеріалів.	
4.3 Висновок до розділу 4.....	
ВИСНОВКИ.....	
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	
ДОДАТКИ.....	

ВСТУП

Актуальність. У зв'язку з тим, що на теперішній час система безпеки відіграє одну із пріоритетних основ функціонування IoT-дівайсів, тому цей напрям повинен достовірно і точно відображати всі команди керування пристроями. Повинна забезпечуватись інформативна достовірність голосових команд (голосових потоків керування). Враховуючи запровадження в стільниковій мережі технологій 3G-5G, швидкодія керування IoT-дівайсами має відображати технічні можливості сучасних і в найблищому майбутньому нового покоління пристроїв мобільного зв'язку. Крім цього данні функції повинні забезпечувати комфортне керування цифровими пристроями. Збільшення нових видів виконавчих пристроїв для побутового призначення призводить до розширення основних функцій та модернізації уже закладених у програмних платформах. Деякі системи зв'язку, що використовуються для передачі інформації, - це супутникові цифрові системи та радіорелейні цифрові системи зв'язку та передачі інформації. З метою оптимізації смуги частот та ефективного використання смуги частот, особливо в двох площинах поляризації, потік інформації можна подвоїти, а також використовувати модернізоване, інтегроване, сумісне з частотою мікрохвильове антенне обладнання;

Мета та задачі дослідження. Метою є аналіз, проектування та дослідження розпізнавання мови, нейронних мереж, які можуть використовуватись для голосового керування IoT-дівайсами на основі аналізу методів, які вже існують (наукові статті, конференції, патенти, результати досліджень) та створення математичної моделі голосового керування з використанням технічних вимог до звукових потоків;

Об'єкт дослідження. Аналіз та дослідження mell-фільтрів, прихованих Марковських моделей, алгоритму «прямого-зворотного» ходу. Створення алгоритму розпізнавання мови;

Предмет дослідження. Приховані Марковські моделі, методи створення нейромереж та алгоритм прямого-зворотнього ходу;

Методи дослідження. Виконувався аналіз, дослідження та систематизування типових алгоритмів розпізнавання мови при використанні розширених технічних характеристик;

Наукова новизна одержаних результатів. Основні результати, що становлять наукову новизну, отриману під час вирішення поставлених у дослідженні завдань, такі:

-Аналіз та створення нового алгоритму розпізнавання мови на основі фільтра *tell*;

-Аналізована можливість навчання та використання нейронних мереж для побудови систем розпізнавання мови;

-Створений алгоритм голосового введення, метою є покластись на штучний інтелект для надійного розпізнавання голосових команд конкретних пристроїв.

Практичне значення одержаних результатів. Вирішення проблем, на яких базується складність розпізнавання мови залежно від емоційного стану, а саме: тембру, гучності, висоти, темпу, інтонації, якості дикції та самої мови тощо. Крім того, потрібно «навчити» пристрої реагувати або виконувати команди строго відповідно до своїх завдань технічні характеристики, запрограмовані на пристрій.

Публікації. Апробація та оприлюднення результатів досліджень відбулось на VIII науково-технічній конференції ТНТУ ” Інформаційні моделі системи та технології ” 09-10 грудня 2020, збірник тез доповідей.

У тезах VIII Міжнародної науково-технічної конференції опубліковані тези доповідей.

Структура роботи. Робота складається із розрахунку та пояснювальної записки. Розрахунки та пояснення складаються із вступу, 4 частин, висновків, списку використаної літератури та додатків. Обсяг роботи: врегулювання та пояснення - 60 аркуш. Формат А4 та додатки.

РОЗДІЛ 1

АНАЛІТИЧНА ЧАСТИНА

1.1. Голосове управління технічними пристроями

Завдання машинного розпізнавання мови привертає увагу фахівців вже дуже давно. Проте, просунутися в цьому напрямку вдалося відносно недалеко. Процес розпізнавання мови - це процес перетворення звукового сигналу, що надходить від мікрофона, у послідовність слів. Потім використовуйте отриманий набір гіпотез ланцюжків слів для розуміння мови [2].

Це спричинило багато проблем. По-перше, людина, як правило, не робить пауз між словами, а у разі злиття висловлювань завдання вилучення слів із мовного потоку додається до завдання розпізнавання, що свідомо складніше. Слова, яким потрібно виділити один склад - він з'являється з ними і пов'язаний з максимальною кількістю помилок існуючої системи. Ви можете попросити людину вимовити одне слово по одному, зробити паузу на тривалий час або вимовити кожне наступне слово після звукового сигналу. Але цей метод незручний і може використовуватися лише для видачі простих команд [3].

Наступне питання - це голос, діалект, диктант, диктовка, різниця у віці, емоційні та фізичні відмінності. Акустика має великий вплив, а саме заміна мікрофона, положення мікрофона відносно рота та акустичні умови в кімнаті [4,5,6,7,8].

Саме через ці та багато інших проблем повного вирішення проблеми розпізнавання мови все ще недостатньо. Існує два різних режими роботи: голосові налаштування з певними динаміками та робота без таких налаштувань. Тепер при використанні налаштувань, що стосуються динаміка (динаміка), розмір словника може досягати тисяч (або навіть тисяч) слів у комбінованому оголошенні. Етапи налаштування динаміків такі: динаміки читають спеціально складений текст, а комп'ютер розпізнає слова та надає

опції розпізнавання. Диктор позначив помилку і знову прочитав текст. Після декількох таких ітерацій процес сходиться і комп'ютер може розпізнати мову.

Нарешті, останнє є найскладнішим у досягненні, але також і найбільш перспективним режимом роботи - його можна розпізнати без налаштування динаміка. Це гарантує, що система розпізнає будь-яке слово, що міститься у словнику, незалежно від того, хто його вимовляв. Зазвичай словник тут містить лише невелику кількість слів (як правило, не більше двох), і існує лише відносно невелика кількість мов (близько тридцяти). Хоча російська мова включена до цієї цифри, кількість розпізнаних російських слів дуже мала.

Кут зняття звуку.

Під кутом зняття звуку (рис. 1.1) розуміється зона можливого розташування джерела звукового сигналу, всередині якої не спостерігається значної втрати ефективності мікрофона.



Рис. 1.1. Схема зняття звуку.

Звук прямого поля.

Термін "пряме звукове поле" описує ситуацію, коли звук досягає мікрофона, не відбиваючись попередньо від стіни, стелі, підлоги чи іншої поверхні. (див. Рис. 1.2).

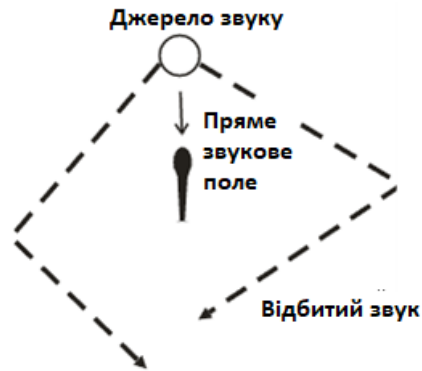


Рис. 1.2. Пряме звукове поле і відбитий звук.

Якщо мікрофон знаходиться в прямому звуковому полі, його можна спрямувати на нульову чутливість до шкідливих джерел шуму. Ця технологія допомагає значно зменшити вплив зворотного зв'язку та уникнути витоку звуку.

Відбите звукове поле.

Термін "відбивне (ревербераційне) звукове поле" відноситься до ситуації, коли звук відбивається стіною, стелею, підлогою або іншою поверхнею до того, як досягне мікрофона (див. Рис. 1.3).

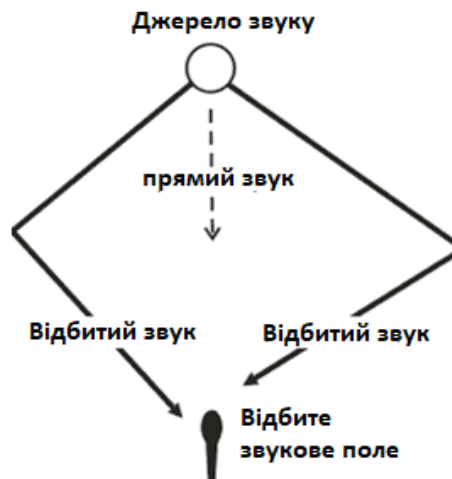


Рис. 1.3. Відбите звукове поле

У цьому випадку роботи зони нульової чутливості мікрофона недостатньо для повного контролю над небажаними шумами. Однак навіть у відбитому звуковому полі односпрямовані мікрофони забезпечують більш

надійний захист, ніж всеспрямовані мікрофони, запобігаючи зворотному зв'язку та зовнішнім сигналам. Чим більше кут нахилу мікрофона, тим краще він справляється з цим завданням. У зв'язку з цим за найефективнішим суперкардіоїдним мікрофоном слідує кардіоїдний мікрофон. Чисельним показником ефективності мікрофона у відбитому звуковому полі є показник спрямованості, який обернено пропорційний куту звуку. У таблиці 1.1 наведено значення коефіцієнтів мікрофонів з різними характеристиками випромінювання:

Таблиця 1.1.

Діаграми спрямованості

Направленість мікрофона	Коефіцієнта направленості	Відсічення шумів (дБ)
Всенаправлений	1	0 дБ
Кардіоїдний	1.7	4.8 дБ
Суперкардіоїдний	1.9	5.7 дБ

Ці таблиці показують, що спрямованість суперкардіоїдного мікрофона в 1,9 рази перевищує спрямованість мікрофона. У реальних умовах це виражається так: у відбитому звуковому полі суперкардіоїдний мікрофон записує на 5,7 дБ менше шуму реверберації

1.2. Модель голосового управління периферійними пристроями

В результаті цього об'єднання відбулася чудова річ - стандарт EIB забезпечує повну сумісність всіх пристроїв і, як пояснюють фахівці, обмін інформацією між ними.

«Інтелектуальна система» складається з сенсорів - пристроїв, що дають або беруть команду, і акторів - виконуючих її.

Сенсори і актори з'єднані інформаційної шиною через ПЕОМ та систему управління периферійними пристроями. На вигляд ця шина являє собою всього лише кілька мідних дротів.

Архітектурна схема розумної системи представлена на рис 1.4.

Джерелом сигналів є сама людина. Пристрої, що стежать за людиною (мікрофони, відеокамери, сенсори і ін.) Розташовуються по всій кімнаті (будівлі) і постійно сканують, які виходять від людини звуки і що проробляються їм жести, обробляючи їх у блоці обробки аудіо та відео сигналів (чим більше кількість цих пристроїв, тим краще і точніше він розпізнається) [22]. Після чого порівнює їх з базою даних, визначаючи при цьому яку команду їй потрібно виконати. Далі ця інформація надходить в систему управління периферійними пристроями у вигляді двійкового коду. В кінцевому підсумку ця система, в залежності від коду, включає або вимикає ту чи іншу електричний пристрій.

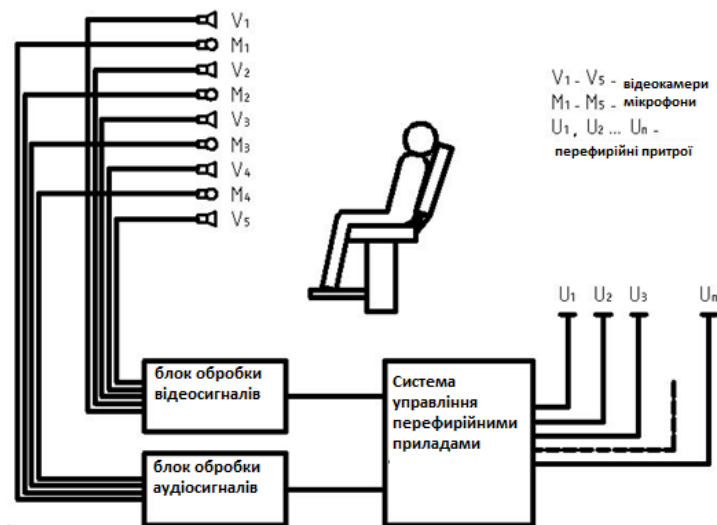


Рис. 1.4. Архітектурна схема «розумної системи».

1.3. Схема управління електричними пристроями

Схема голосового управління електричними пристроями представлена на рис. 1.5 (додаток 1). Вона складається з трьох основних частин: чотирьох поверхневих всепрямованих мікрофонів Crown PZM-11LLWRS1, блоку обробки аудіо сигналів і системи управління периферійними пристроями.

Принцип дії її простий: звук від людини за допомогою мікрофона передається в блок обробки аудіо сигналів, який представляє собою ПЕОМ з ОС Windows'2007 і вище з відповідним програмним забезпеченням, з вбудованою картою аудіо-захоплення.

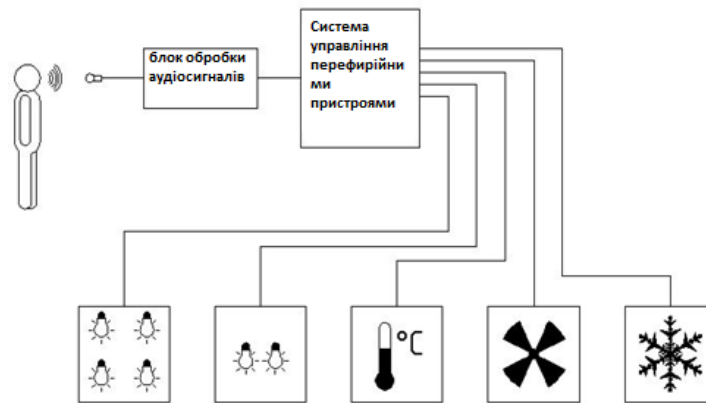


Рис. 1.5. Схема голосового управління електричними пристроями.

Прикладом може послужити аудіокарта Audiomedia III від фірми Digidesign, що має 8 вбудованих процесорів звукової обробки, електронних регуляторів гучності і балансу і еквайзерів, вбудований стереорегулятор майстер каналу і панорами і чотири електронних регулятора для входів. У цьому блоці скануються всі сигнали йдуть від мікрофона. При розпізнаванні з потоку слів голосової команди, наприклад «включити світло», програма шукає в базі даних голосових команд який код відповідає цій самій команді. Таким чином голосова команда перетвориться в найпростіший двухразрядний сигнал і по шині даних надходить в супу, де відбувається вже безпосереднє включення електромеханічних пристроїв.

Було поставлено завдання наступного плану: необхідно було розробити «розумну систему», яка за допомогою голосу могла включати і вимикала в приміщенні електромеханічні пристрої - це кондиціонер, обігрівач, вентилятор і три види освітлення: верхнє бічне і повне. А так же забезпечувала їх незалежну один від одного роботу.

Завдання розробки системи було вирішене (рис. 1.6) за допомогою програмованої мікросхеми КР1533ІР35 [23]. призначеної для фіксації даних, що приходять з ПЕОМ. Як ключі використані оптрони (V1 - V5) ТСО-10 ($U_{обр\max}$ до 1000 В і $I_{нр\phi} \leq 10$ А), замикають ланцюг при подачі на них напруги [10]. Для управління ними використані сигнали D0 - D4. Так як тиристори вимагають харчування не п'ять вольт, як мікросхема, а всього лише три, то було поставлено додатковий опір R_{ϕ} .

$$R_{\phi} = \frac{\Delta U_{KB}}{I_{\gamma}} = \frac{2}{0,08} = 250\text{Ом} \quad (1.1)$$

потужність опору

$$P_{R\phi} = I^2 R_{\phi} = 0,08^2 \cdot 25 = 0,16\text{Вт} \quad (1.2)$$

Остаточню вибираємо опір МЛТ-0,25 на 250м.

Мікросхема запрограмована таким чином, що при приході сигналу «1» на висновок D0, ланцюг замикається по Q0. У таблиці 1.2 наведені голосові команди і відповідні їм комбінації сигналів

Таблиця 1.2

Комбінації голосових команд і кодів.

Вихід регістра	Голосова команда				
	«включити освітлення»		«включити обігрів»	«включити вентилятор»	«включити кондиціонер»
	«включити бокове освітлення»	«включити верхнє освітлення»			
D0	1	0	0	0	0
D1	0	1	0	0	0
D2	0	0	1	0	0
D3	0	0	0	1	0
D4	0	0	0	0	1

При команді «вимкнути все» на все висновки D0 - D4 подаються «0», і все ланцюга розмикаються.

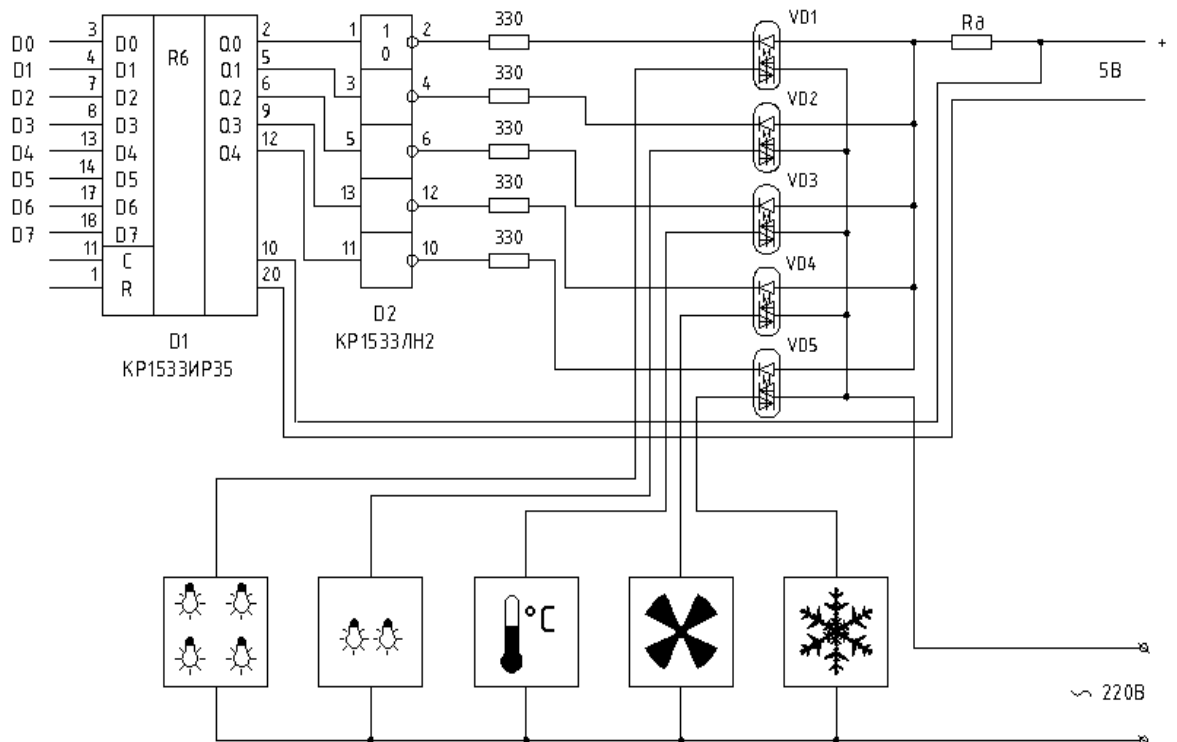


Рис. 1.6. Принципова схема включення електромеханічних пристроїв

Розглянемо принцип дії цієї схеми на одному з прикладів. Нехай на блок обробки аудіосигналів прийшла команда «включити світло». По шині даних від ПЕОМ до мікросхеми з частотою в декілька кілогерц черзі починають приходити сигнали «1,0,0,0,0» і «0,1,0,0,0» по входах D0, D1, D2, D3 і D4 відповідно. Регістр порівнює ці сигнали з запрограмованими і замикає ланцюг на висновки Q0 і Q1, через які починає текти струм. В результаті цього в роботу включаються оптрони V1 і V2, включаючи ланцюг бічного і верхнього освітлення. В результаті у нас в кімнаті спалахує весь світ!

Якщо за цим послідує команда «вимкнути верхнє світло», блок обробки аудосигналов перестає подавати сигнал з кодом «0,1,0,0,0», але не припиняє подачу імпульсів бокового світла. В кінцевому підсумку ланцюг по виходу Q1 розмикається, на оптрон V1 перестає приходити струм і він розмикає ланцюг верхнього освітлення., В результаті чого залишається горіти тільки бічне світло.

Для роботи оптронів і блоку управління периферійними пристроями в цілому було розроблене джерело вторинного живлення напругою п'ять вольт, який живиться від загальної мережі змінного струму.

1.4. Висновок до розділу 1

Для досягнення цих цілей необхідно використовувати інтерактивні засоби, щоб повною мірою використовувати навички міжособистісного спілкування. Людям надається багато можливостей сприймати та передавати інформацію: зір, слух (включаючи голос), жести та рухи, міміку, дотики тощо. У взаємодії між людиною та комп'ютерами існує два інформаційні потоки: команди управління та дані, що передаються на комп'ютер для обробки; результати розрахунків та інша інформація, що надається комп'ютером користувачеві.

РОЗДІЛ 2

ОСНОВНА ЧАСТИНА

2.1. Розпізнавання мови - ключові моменти

Що означає розпізнавання мови? Це може бути перетворення мови в текст, розпізнавання та виконання певних команд, вибір будь-яких характеристик з мови (наприклад, ідентифікація мовця, визначення його емоційного стану, статі, віку тощо) - все це може відбуватися в різних джерелах. Відповідно до цього визначення. У моїй роботі розпізнавання мови стосується присвоєння мовних або фонетичних послідовностей (фонем, букв, слів) будь-якому класу. Тоді в цьому класі можна порівняти символи алфавіту - ми отримуємо систему для перетворення мови в текст або певні дії - отримуємо систему для виконання мовних команд. Як правило, цей метод обробки голосової інформації може бути використаний на першому етапі будь-якої системи з набагато більш складною структурою. Ефективність класифікатора визначатиме ефективність всієї системи.

Що станеться при побудові системи розпізнавання мови? Головною характеристикою мовного сигналу є те, що він сильно варіюється за багатьма параметрами: тривалістю, ритмом, висотою тону, спотвореннями, спричиненими величезною мінливістю голосового каналу людини, різними емоційними станами мовця та сильними відмінностями в голосі різних людей. Навіть якщо це одна і та ж людина, два часові подання голосу, записані одночасно, не перекриватимуться. Необхідно шукати ці параметри мовного сигналу, які повністю описують мовні параметри (тобто можна відрізнити один мовний звук від іншого), але певною мірою не мають нічого спільного із вищезазначеними змінами мови. Потім отримані таким чином параметри слід порівняти із зразком. Це не має бути випадковістю, а простим порівнянням для пошуку найбільшого збігу. Це змушує вас знайти форму необхідної відстані у знайденому просторі параметрів.

Крім того, обсяг інформації, яку може зберігати система, не необмежений. Як запам'ятати майже необмежену кількість змін голосового сигналу? Очевидно, що середньої статистичної форми не існує.

Іншою проблемою є швидкість пошуку в базі даних. Чим більший його розмір, тим менша швидкість пошуку - це речення правильне, але стосується лише звичайних послідовних комп'ютерів. Які ще машини можуть вирішити всі вищезазначені проблеми? - Ви запитаете. Абсолютно правильно, це нейронна мережа.

2.2. Опис системи розпізнавання мови яка самонавчається

Для вивчення характеристик системи самонавчання модель розпізнавання та синтезу мовлення об'єднується в систему, щоб вона мала деякі характеристики системи самонавчання. Ця комбінація є одним із ключових атрибутів створеної моделі.

По-перше, система має можливість виконувати дії (синтез) та аналізувати дії (розпізнавання), а саме атрибути (2). По-друге, існує атрибут (1), оскільки ніяка інформація не буде розміщуватися в процесі розробки системи, а здатність розпізнавати та синтезувати мовні звуки - це результат навчання.

Перевага отриманої моделі полягає в тому, що її можна навчити автоматично. Механізм навчання такий.

Ще однією дуже важливою функцією є можливість перетворення зображень, що зберігаються в новому просторі параметрів, у менший розмір. Ця функція ще не впроваджена в систему, що розробляється, і не перевірена на практиці, але, незважаючи на це, я все одно спробую узагальнити її суть на прикладі розпізнавання мови.

Припустимо, що вхідний сигнал задається основним вектором ознак у N -мірному просторі. Для зберігання такого сигналу потрібно N елементів. На стадії розробки ми не знаємо деталей сигналу, або сигнал занадто складний,

щоб розглядати. Це призводить до великого представлення сигналу, який ми використовуємо. Далі ми припускаємо, що ми маємо здатність синтезувати один і той же сигнал (тобто синтезовану мову), але синтезований сигнал є функцією вектора параметрів у M -мірному просторі, і $M \ll N$ (насправді кількість параметрів моделі синтезу мовлення більше, ніж основна Набагато менше) функція моделі розпізнавання мови). Але тоді ми можемо пам'ятати, що вхідний сигнал запам'ятовується не своїми основними характеристиками в N -мірному просторі, а параметрами інтегрованої моделі в M -мірному просторі. Виникає питання: як передати сигнал з одного простору параметрів в інший простір параметрів? Є всі підстави вважати, що для завершення цього перетворення можна використовувати досить просту нейронну мережу. Більше того, я думаю, що цей механізм пам'яті працює у реальних біологічних системах, особливо у людей.

Далі описується модель автоматичного розпізнавання та синтезу мовлення. Описує механізм введення звуку в нейронну мережу, модель синтезу мовлення, модель нейронної мережі та проблеми, що виникають у процесі побудови моделі.

Вхід аудіо здійснюється в режимі реального часу за допомогою звукової карти або файлів Microsoft Wave з кодуванням PCM (16 біт, частота дискретизації 22050 Гц). Краще використовувати файли, оскільки це дозволяє повторювати їх обробку неодноразово через нейронну мережу, що особливо важливо під час навчання.

Для того, щоб застосувати звук до входу нейронної мережі, на ній необхідно виконати певне перетворення. Очевидно, тимчасова форма звукового представлення неефективна. Він не може відображати характеристики звукового сигналу. Частотний спектр мови має більш довідкове значення. Для отримання спектра використовується набір смугових фільтрів або дискретні перетворення Фур'є, налаштовані на різні частоти. Потім виконайте різні перетворення отриманого спектру, наприклад, логарифмічне масштабування (в амплітудному та частотному просторі). Це

дозволяє розглянути певні функції мовного сигналу, що зменшують кількість інформації у високочастотній частині спектра, логарифмічну чутливість людського вуха тощо.

Як правило, неможливо описати мовний сигнал повністю на основі його частотного спектру. Окрім спектральної інформації, потрібна також інформація про динаміку голосу. Щоб отримати його, використовуйте додаткові параметри, які походять від основних параметрів.

Отримані таким чином параметри мовного сигналу вважаються його основною ознакою і представляють сигнал на наступному рівні обробки.

Процес введення звукової інформації показаний на рисунку 2.1:

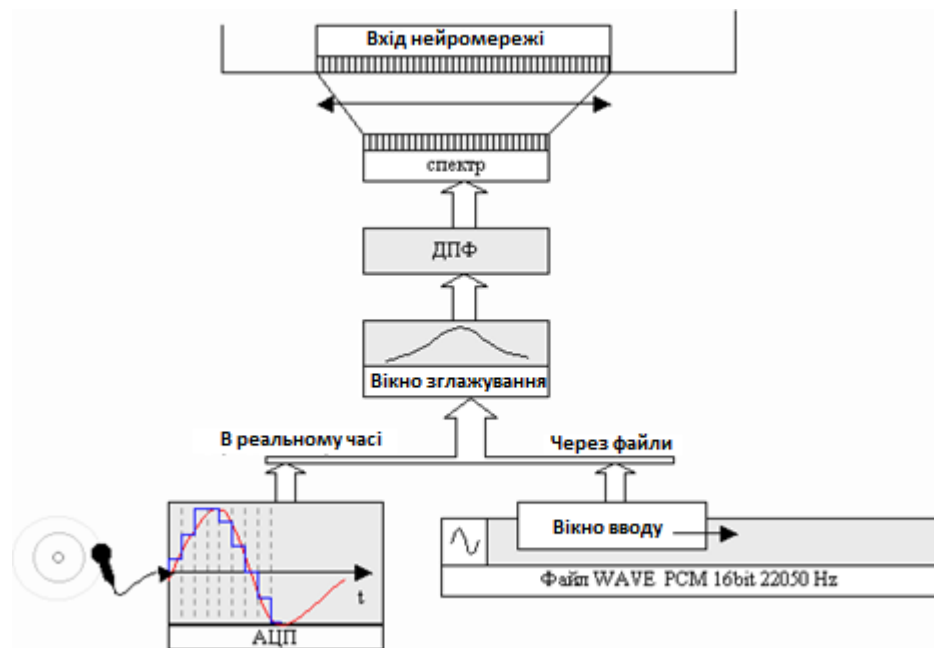


Рис.2.1. Введення звуку

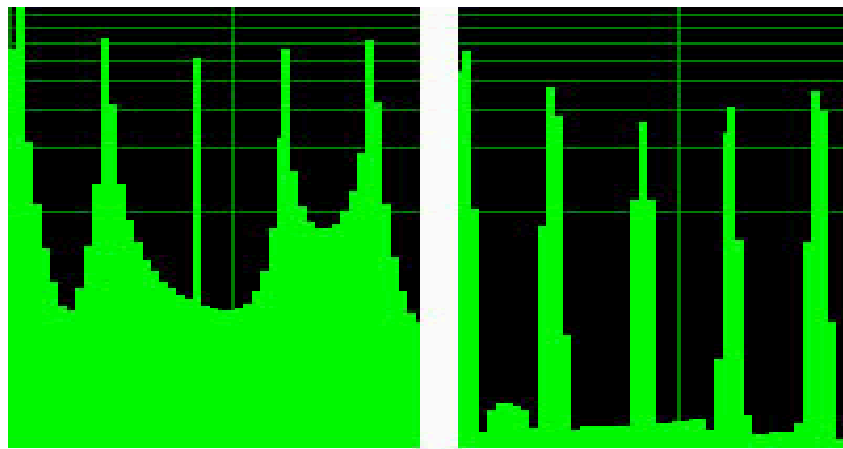
Під час обробки файлу вікно введення рухатиметься вздовж файлу, і його розмір дорівнює розміру вікна дискретного перетворення Фур'є (DFT). Зміщення вікна щодо попереднього положення можна регулювати. Кожне положення вікна заповнене даними (система працює лише зі звуком, де кожен відлік кодується 16 бітами). Коли ви вводите звук у режимі реального часу, він буде записаний у блоки однакового розміру.

Після введення даних у вікно та обчислення DFT накладіть на нього вікно згладжування Хеммінга:

$$newData[i] = Data[i] \left(0.54 - 0.46 \cos \frac{2\pi i}{N-1} \right) \quad (2.1)$$

де N - розмір вікна ДПФ

Накладення вікна Хеммінга трохи зменшує контрастність спектру, але дозволяє видалити бічні пелюстки різких частот (рис. 2.2), і це поєднання мов є особливо хорошим.



а) без вікна згладжування б) з вікном згладжування Хеммінга

Рис. 2.2. Дія вікна згладжування Хеммінга (логарифмічний масштаб)

Після цього дискретне перетворення Фур'є обчислюється відповідно до алгоритму швидкого перетворення Фур'є ([XX]). В результаті отримують амплітудний спектр та інформацію про фазу з дійсними та уявними коефіцієнтами. Інформація про фазу відкидається, а енергетичний спектр розраховується:

$$E[i] = \sqrt{\operatorname{Re}C[i] \cdot \operatorname{Im}C[i]}, \quad i = 0 \dots NS-1, NS = N/2 \quad (2.2)$$

Оскільки оброблені дані не містять уявної частини, атрибут DFT призводить до симетрії, тобто $E[i] = E[N-i]$. Отже, розмір інформаційної частини спектру NS дорівнює $N/2$.

Всі розрахунки в нейронній мережі виконуються з числами з плаваючою точкою, і більшість сигналів обмежуються діапазоном $[0,0, 1,0]$, тому отриманий спектр нормується до 1,00. Для цього розділіть кожен компонент вектора на його довжину:

$$e[i] = \frac{E[i]}{|E|} \quad (2.3)$$

$$|E| = \sum_{i=0}^{NS-1} E[i] \quad (2.4)$$

Обсяг інформації в різних частинах спектра неоднаковий: область низьких частот містить більше інформації, ніж область високих частот. Тому, щоб запобігти надмірному споживанню входів нейронної мережі, необхідно зменшити кількість елементів, які отримують інформацію з високочастотної області, або, аналогічним чином, стиснути високочастотну область спектра в частотному просторі. Найпоширеніший метод (завдяки своїй простоті) - логарифмічне стиснення.

$$m = 1125 \cdot \log(0.0016 f + 1) \quad (2.5)$$

де f - частота в спектрі, Гц,

m - частота в новому стиснутому частотному просторі

Після нормалізації та стиснення частотний спектр накладається на вхід нейронної мережі. Вхід нейронної мережі не виконує жодних ключових функцій, а лише надалі передає сигнал до нейронної мережі. Вибір кількості входів є складним завданням, оскільки розмір вхідного вектора може втратити важливу інформацію для розпізнавання, а кількість вхідних векторів значно збільшить складність розрахунку (лише тоді, коли моделювання виконується на ПК, у власне нейронній мережі У мережі це неправильно, оскільки всі елементи працюють паралельно).

Завдяки високій роздільній здатності (великий вхід) можна виділити гармонійну структуру голосу та визначити висоту голосу. При низькій

роздільній здатності (невелике введення) можна визначити лише структуру формантів.

Як показало подальше вивчення проблеми, для ідентифікації достатньо лише інформації про структуру форманту. Насправді людина також може розпізнавати нормальну мову і шепіт, хоча останній не має джерела мови. Джерело голосу надає іншу інформацію у формі інтонації (зміни висоти тону під час вимови), і ця інформація дуже важлива при обробці мовлення вищого рівня. Але в першому наближенні ми можемо обмежитися отриманням формантової структури, з цієї причини, враховуючи стиснення неінформаційної частини спектра, достатня кількість входів вибирається в діапазоні 50-100.

2.3. Модель нейромережі

Структура нейронної мережі досить проста і складається з трьох шарів: вхідного шару, шару символу та ефектору (рис. 2.4). Кожен нейрон наступного шару з'єднаний з усіма нейронами попереднього шару. Функції передачі всіх шарів є лінійними, а конкуренція моделюється на вхідному рівні.

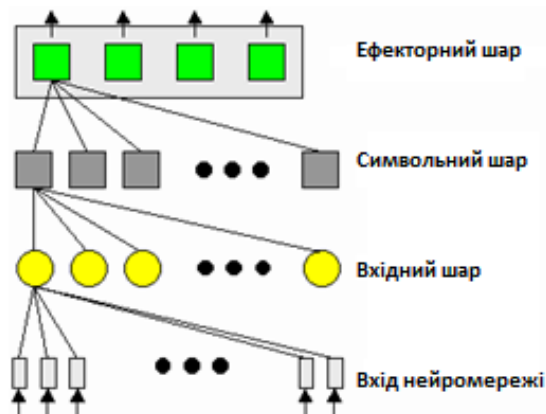


Рис. 2.4 Архітектура нейромережі

1. Вхідний рівень - цей рівень приймає сигнал безпосередньо від входу нейронної мережі (вхід не обробляє сигнал, а лише надалі розподіляє його до нейронної мережі). Це один із варіантів самоорганізованої картки Кохонена,

яку можна вивчити без вчителя. Основним завданням вхідного рівня є формування нейронного набору для кожного типу вхідного вектора, і ці нейронні набори представляють відповідні сигнали наступного рівня обробки. Насправді саме цей рівень визначає ефективність подальшої обробки сигналів, і моделювання цього шару є найскладнішим.

Нейрони в цьому шарі функціонують за принципом конкуренції, тобто завдяки певній кількості ітерацій нейрон або колекція нейронів залишаються активними (група нейронів, що працюють одночасно). Цей механізм викликаний бічними зв'язками і називається бічним гальмуванням. Багато джерел це детально обговорювали ([1], [6]). Оскільки для розробки цього механізму потрібно багато обчислювальних ресурсів, він штучно змодельований у моїй моделі, тобто є нейрон з найбільшою активністю, його активність встановлена в 1, а інші - в 0.

Провести мережеве навчання відповідно до правила (7):

$$w_n = w_c + v(x - w_c)a, \quad (2.6)$$

де w_n - нове значення ваги,

w_c - старе значення,

v - швидкість навчання, $v < 1$

x - нормований вхідний вектор,

a - активність нейрона.

Геометрично це правило ілюструє малюнок 2.5.

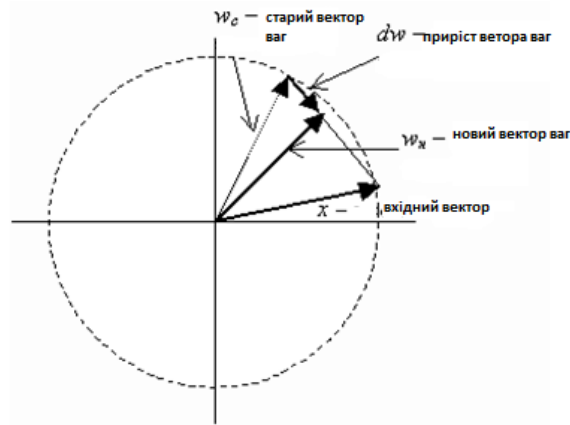


Рис. 2.5. Корекція ваг нейрона Кохонена

Перед подачею вхідного вектора x на вхід нейронної мережі він нормалізується, тобто знаходиться в гіперсфері одиничного радіуса в масштабному просторі. Коли вага коригується згідно з правилом (6), вектор ваги обертається до вхідного вектора. Поступове зменшення скорочення обертання може статистично усереднювати вхідний вектор, на який реагує нейрон.

Проблеми, що виникають при навчанні прошарку Кохонена.

2.Символьний шар-нейрони в цьому шарі пов'язані із символами букв (це не обов'язково повинна бути проста буква, але це може бути буква будь-якої фонемі). Цей шар генерує символи під час розпізнавання та вхідні символи під час синтезу. Він є частиною Гросберга і навчається у свого вчителя. Нейрони цього шару функціонують звичайним чином: він обчислює загальний зважений сигнал на вході і передає його на вихід за допомогою лінійної функції. Під час тренування ваги з'єднання змінюються наступним чином:

$$w_{ij}^n = w_{ij}^c + v (u_j - w_{ij}^c)x_i, \quad (2.7)$$

де w_{ij}^n , w_{ij}^c - ваги зв'язків до і після модифікації

v - швидкість навчання, $v < 1$

u_j - вихід нейрона

x_i - вхід нейрона

Відповідно до цього правила, вектор, що з'єднує ваги, прагне до вихідного вектора, але лише в тому випадку, якщо активний вхід (тобто зв'язок) модифікується лише відповідно до поточних активних нейронів шару Кохонена. Вихідний рівень символного шару є двійковим, тобто нейрони можуть бути активними ($u_j = 1$) або неактивними ($u_j = 0$), що відповідає включенню певного символу. Вхідний шар і шар символів разом дозволяють кожному типу вхідного сигналу порівнювати певний символ алфавіту.

3. Ефекторний шар - Цей шар приймає сигнал від шару символу, а також є шаром Гроссберга. Вихід цього шару є вектором ефектора - елемент ефектора контролює зазначені параметри в моделі синтезу. Зв'язок ефекторів та параметрів синтетичної моделі здійснюється за допомогою графіків ефекторів. Цей шар дозволяє кожному нейрону символного шару (i , отже, кожному символу літери) порівнювати деякі ефекторні вектори (i , отже, із специфічними синтезованими звуками). Вивчення цього рівня схоже на шар символів.

2.4. Навчання нейромережі

Навчання нейромережі включає три етапи. Спочатку в систему подаються лише звукові зразки, а у вхідному шарі формується нейронний набір, а нервові ядра забезпечуються зразками. Потім подаються звук та відповідні символи алфавіту. У той же час нейрони вхідного рівня пов'язані з нейронами символного шару. На завершальному етапі інтегрується систематичне навчання. У цій системі не відображаються зразки, а використовується інформація, накопичена на попередніх етапах. Використання механізму випадкового навчання: нейрони ефекту шару піддаються випадковим змінам, а потім генерують звуки, розпізнають їх і порівнюють результати із символами генерованих звуків. Водночас зміни виправлені. Повторюйте цей процес, поки всі звуки не будуть правильно створені.

Вибір швидкості навчання.

Чому в правилах дослідження шару Кохонена існує коефіцієнт (ν) ? Якщо він дорівнює 1, то для кожного вхідного вектора ключовий вектор активного нейрона буде дорівнює йому. Зазвичай для кожного нейрона існує багато вхідних сигналів його можна активувати, і його вектор підключення буде постійно змінюватися. Якщо $\nu < 1$, вектор підключення буде слабо реагувати на кожен вхідний сигнал. Зменшуйте ν під час навчального процесу, і ми отримуємо середнє статистичне значення подібних вхідних сигналів в кінці навчання з цією ж метою швидкість навчання вводиться в усі інші правила навчання.

Від чого залежить швидкість навчання? Тут головна роль визначається порядком відображення зразків. Припустимо, існує велика навчальна вибірка, порядок її елементів представлений нейронною мережею. Якщо рівень навчання високий, в середині цього прикладу нейронна мережа «забуде» попередній елемент. Більше того, якщо кожен зразок з'являється кілька разів поспіль, у наступному зразку нейронна мережа забуде попередній зразок. Тому головним критерієм вибору швидкості навчання є незначна зміна відносин у всій навчальній вибірці. Однак не слід забувати, що час навчання обернено пропорційний швидкості навчання. Тому тут потрібно знайти компроміс.

Запам'ятовування рідко зустрічаються елементів.

Зазначений алгоритм навчання підходить для часто повторюваних сигналів. Якщо сигнал рідкісний на тлі всієї навчальної вибірки, він взагалі не буде збережений. У цьому випадку необхідно задіяти механізм уваги [5]. Коли з'явиться невідома нейронні мережа вибірки, швидкість навчання буде збільшена в рази, і рідкісний елемент буде запам'ятовуватися в нейронній мережі. У цій системі навчальні зразки штучно побудовані, тому таких проблем не виникатиме, а механізм уваги не реалізований. При навчанні в природних умовах, якщо зразки навчання не передбачаються заздалегідь, потрібен механізм уваги.

Проблеми, що виникають при навчанні про шарку Кохонена

З метою вивчення динаміки навчання та характеристик шару Кохонена було створено інструмент «модель нейронної мережі», який імітує шар Кохонена в двовимірному сигнальному просторі (рис. 2.6).

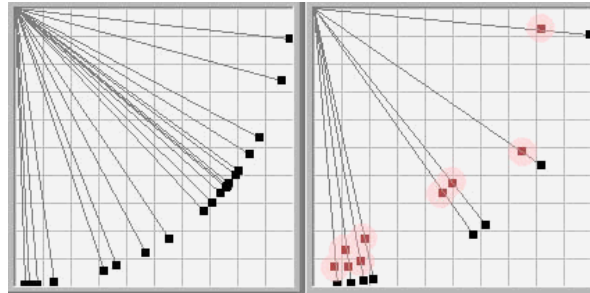


Рис.2.6. Моделювання шару Кохонена

Модель створює нейронну мережу з двома входами, тому вона може класифікувати вхідні вектори в двовимірному сигнальному просторі. Хоча функція цієї нейронної мережі відрізняється від функції нейронної мережі в просторі сигналів, але розмірність більша, основні характеристики та ключові моменти цього алгоритму нейронної мережі можуть бути вивчені в цій простій моделі. Головна перевага полягає в тому, що він може візуалізувати динаміку навчання нейронної мережі за допомогою двох входів. В процесі експерименту з цією моделлю під час вивчення нейронних мереж були виявлені наступні проблеми.

1. Вибір початкового значення ваги

Оскільки в кінці навчання вектори шкали будуть знаходитися в одному колі, найкраще їх нормалізувати до 1,00 на початку. У моїй моделі вектор ваги випадковим чином вибирається на колі з одиничним радіусом (рис. 2.6).

2. Використання всіх нейронів.

Якщо вектор ваги знаходиться далеко від вхідного сигналу, він ніколи не зможе задовольнити умову в найбільшій мірі, завжди матиме нульовий вихід, а тому не буде коригуватися і буде марним. Решти нейронів може бути недостатньо, щоб розділити простір вхідного сигналу на кілька категорій. Для вирішення цієї проблеми було запропоновано багато алгоритмів ([1], [8]). У

моїй роботі застосовується правило "прагнення до роботи": якщо якийсь нейрон тривалий час неактивний, він збільшить вагу з'єднання, поки не стане активним і не почне проходити тренінг. Цей метод також вирішує проблему тонкої класифікації: якщо формується група вхідних сигналів, близьких один до одного, група сигналів асоціюється з великою кількістю нейронів Кохонена, і вони поділяються на кілька категорій (рис. 2.6). Правило "охочий працювати" написано у такій формі:

$$w_n = w_c + w_c v_1 (1 - a), \quad (2.8.)$$

де w_n - нове значення ваги,

w_c - старе значення,

v_1 - швидкість модифікації,

a - активність нейрона.

Чим менше активність нейронів, тим більша вага зв'язок.

Вибір коефіцієнта v_1 залежить від наступних міркувань: постійне збільшення ваги нейронів компенсується правилом (2.8) (компенсується правилом 2.6) (активні нейрони, як правило, повертаються в гіперсферу одного радіуса), і за одну ітерацію нейронної мережі Клітини або нервові агрегати. У зв'язку з цим, враховуючи кількість нейронів у шарі, коефіцієнт v_1 у (2.8) повинен бути обраний набагато меншим, ніж коефіцієнт v у (2.6).

3. Нерівномірний розподіл вхідних векторів у сигнальному просторі та відсутність нейронів.

Зазвичай основна частина вхідного вектора не рівномірно розподілена по всій поверхні гіперсфери, а зосереджена на деяких невеликих ділянках. У цьому випадку лише невелика кількість весових векторів зможе виділити вхідний вектор, і в цих областях буде бракувати нейронів, а в районах зі значно меншою щільністю сигналу кількість нейронів буде надмірною.

Для того, щоб вирішити цю проблему, ви можете використовувати правило "знайти центроїд", тобто на початковій стадії навчання існує невеликий попит на всі весові вектори вхідного вектора. Як результат, буде

багато векторів ваги, де є вхідні сигнали високої щільності. Правила написані таким чином:

$$w_n = w_c + v_2 (x - w_c) \quad (2.9)$$

де w_n - нове значення ваги,

w_c - старе значення,

v_2 - швидкість модифікації,

x - вхідний вектор

Якщо нейрони зосереджені в одному місці, це правило добре працює. Якщо існує кілька груп нейронів, це правило не дасть бажаного результату.

Іншим рішенням є використання "відпалу" вагового вектора. У нашому випадку цього можна досягти, додавши трохи шуму при зміні масштабу, що дозволить їм рухатися по гіперсфері. Під час тренувального процесу рівень шуму поступово зменшується, а вектор ваги концентрується в положенні найбільшої щільності сигналу.

Недоліком цього правила є те, що швидкість навчання дуже низька. Якщо нейрон у двовимірному просторі більш-менш успішно «виявив» вхідний вектор, то в багатовимірному просторі ймовірність цієї події значно зменшиться.

Найефективнішим рішенням є більш точне моделювання механізму бічного гальмування. Як і раніше, є один нейрон з найбільшою активністю. Потім за допомогою правил (2.10) штучно встановлюють бічну комунікаційну діяльність периферичних нейронів:

$$a_j = e^{\frac{-(i-j)^2}{\theta}} \quad (2.10)$$

де a_j - активність нейрона

i - виграв нейрон

j - індекс нейрона

θ - визначає радіус дії латеральних зв'язків, зменшується в процесі навчання

Припустимо, що всі нейрони мають конкретне положення щодо інших нейронів. Цей топологічний зв'язок є одновимірним і лінійним, і положення кожного нейрона визначається його індексом. Правило (2.10) стверджує, що порушується не один нейрон, а група тісно пов'язаних нейронів. В результаті навчання формується впорядкована одновимірною карта об'єктів. Порядок означає, що наступні два нейрони відповідають двом найближчим векторам в просторі сигналів, і навпаки (оскільки неможливо постійно відображати багатовимірний простір до одного виміру). Спочатку радіус бічного з'єднання був великим, і майже всі нейрони брали участь у тренуванні. У той же час вони виявили "центр ваги" всієї навчальної вибірки. У процесі зменшення коефіцієнта навчання нейрони поділяються на групи, що відповідають місцевим центроїдам. Згодом радіус бічної зв'язки зменшується настільки, що нейрони функціонують незалежно один від одного і можуть ділити дуже близькі вектори.

2.5. Висновок до розділу 2

У цьому розділі представлена модель автоматичного розпізнавання та синтезу мовлення. Описано механізм голосового введення до нейронної мережі, модель синтезу мовлення, модель нейронної мережі та проблеми, що виникають у процесі побудови моделі.

РОЗДІЛ 3

НАУКОВО-ДОСЛІДНА ЧАСТИНА

3.1. Основи розпізнавання мови

В основі застосування ПММ лежать рекурсивні процедури, що володіють обчислювальною складністю. Щодо кількості станів моделі N і довжини спостерігається послідовності T . Працюючи з великим словником і використанні Трифонов в якості моделей фонем число станів досягає сотень, а довжина спостерігається послідовності при розпізнаванні зливої промови може бути, в принципі, необмеженої. При цьому від систем автоматичного розпізнавання мови (АРМ) часто потрібно, щоб вони працювали в режимі реального часу, тому підвищення швидкодії для таких систем є актуальною проблемою. В основі застосування прихованих марковських моделей лежать рекурсивні процедури, що володіють обчислювальною складністю. При цьому від систем автоматичного розпізнавання мови часто потрібно, щоб вони працювали в режимі реального часу, тому підвищення швидкодії для таких систем є актуальним завданням. Матеріали та методи. Одним із шляхів вирішення даної задачі є реалізація апаратної підтримки обчислень в асоціативної осциляторних середовищі. Вона володіє малими апаратними витратами через простоту базових клітинних ансамблів і виконуваних ними функцій і високою швидкістю, не залежних від довжини спостерігається послідовності і кількості станів прихованих марковських моделей, завдяки масовому паралелізму і конвеєрного характеру обчислень.

Не дивлячись на те, що назвою даної роботи є пошук ключових слів, тема розпізнавання мови не могла не з'явитися в даній роботі. Більш того, пошук цілком і повністю базується на розпізнаванні, адже щоб шукати щось у промові необхідно щоб це була саме мова, а не набір незрозумілих шумів.

Почнемо з того, що наша мова - це серія звуків. Звук - це суперпозиція (перекриття) звукових коливань (хвиль) різних частот. З фізики ми знаємо, що

ця хвиля має дві властивості - амплітуду і частоту.

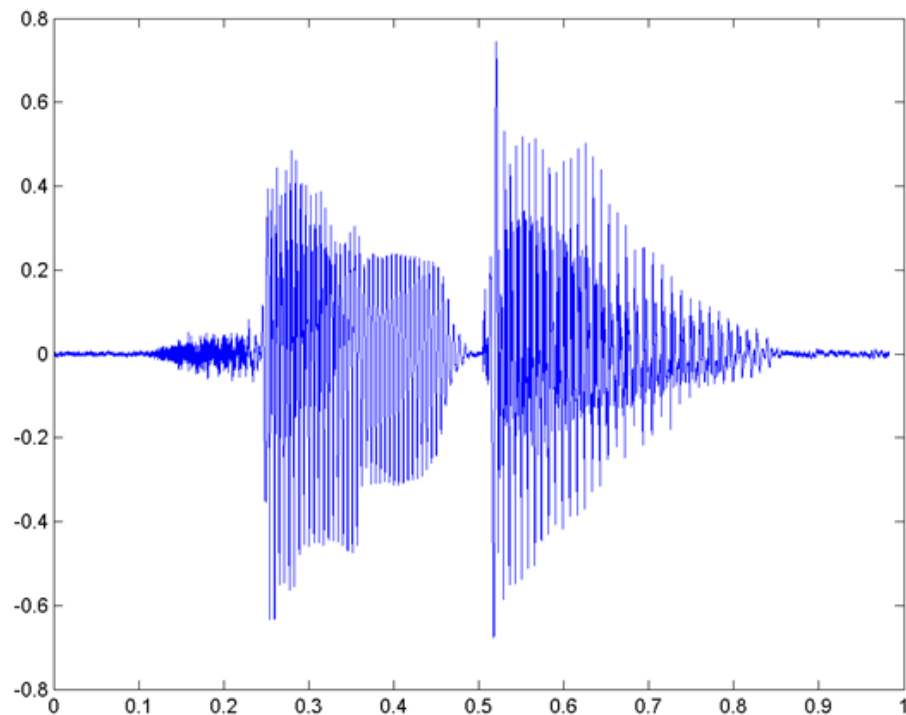


Рис. 3.1. Вид сигналу

Для того щоб зберегти звуковий сигнал на цифровому носії, його потрібно розділити на кілька інтервалів, і в кожному інтервалі береться деяке «середнє значення».

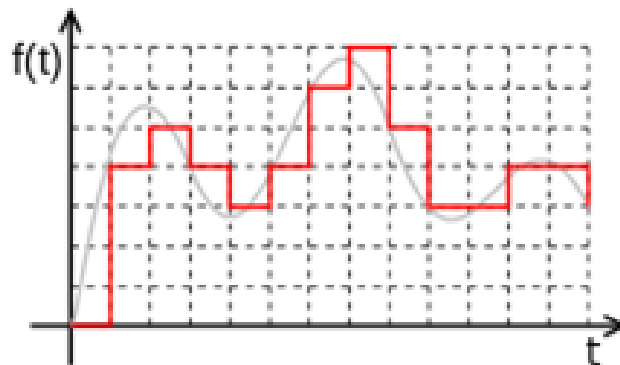


Рис. 3.2. Середнє значення

Таким чином, механічні коливання перетворюються в набір чисел, придатних для обробки на сучасних комп'ютерах. Тому завдання розпізнавання мови спрощується, щоб "порівняти" багато значень (цифрові сигнали) та слова (наприклад, українську мову) у словнику.

Вхідні дані:

Припустимо, є кілька файлів / потоків із аудіоданими. По-перше, потрібно зрозуміти його розташування та спосіб читання. Розглянемо найпростіший варіант - WAV-файли.

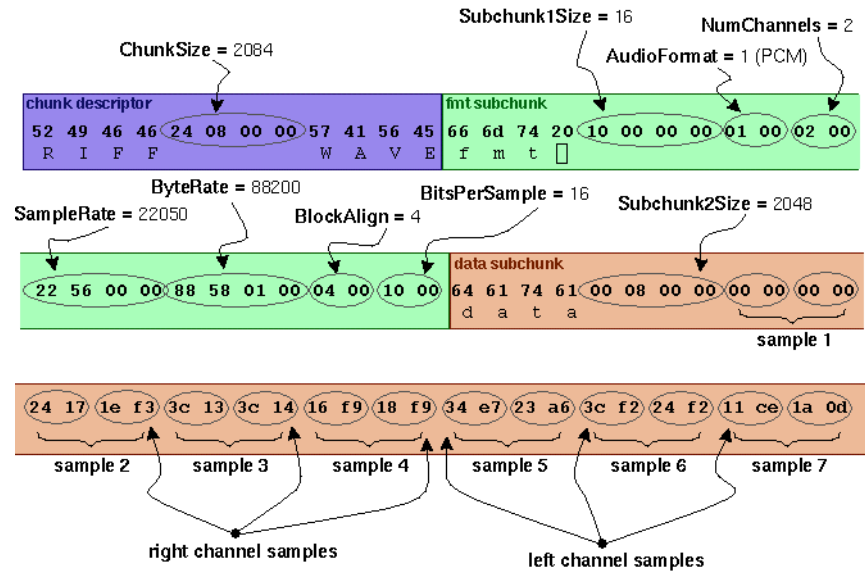


Рис. 3.3. Побудова вихідного файла.

Цей формат передбачає, що у файлі є два блоки. Перший блок - це заголовок, який містить інформацію про аудіопотік: бітрейт, частоту, кількість каналів, довжину файлу тощо. Другий блок містить "необроблені" дані - той самий цифровий сигнал, набір значень амплітуди.

У цьому випадку логіка зчитування даних дуже проста. Прочитайте заголовок, перевірте деякі обмеження (наприклад, відсутність стиснення) та збережіть дані у спеціально вибраному масиві.

3.2. Структура розпізнавання мови

Чисто теоретично, тепер ми можемо порівнювати існуючі приклади з будь-яким іншим відомим текстом (елемент за елементом). Іншими словами, спробуйте «розпізнати» мову, але цей метод для нас не підходить.

Метод повинен бути стабільним (принаймні трохи менше), щоб змінити інтонацію (людина, що говорить слово), гучність та швидкість вимови. Звичайно, поелементного порівняння двох аудіосигналів досягти неможливо.

Спочатку розберемо дані на менші інтервали часу-кадри. І кадри не повинні строго «перекриватися» один за одним. Тобто кінець одного кадру повинен перетинати початок іншого кадру.

Фрейми є більш придатною одиницею аналізу даних, ніж конкретне значення сигналу, оскільки набагато зручніше аналізувати хвилі через певні інтервали, ніж у конкретних точках. "Накладання" положення кадру дозволяє згладити результати аналізу кадру, перетворюючи ідею кадру у "вікно", яке рухається вздовж вихідної функції (значення сигналу).

Експерименти виявили, що оптимальна довжина кадру повинна відповідати інтервалу 10 мс, тобто "перекриття" -50%. Якщо припустити, що середня довжина слова (принаймні в моєму експерименті) становить 500 мс - цей крок забезпечить нам приблизно $500 / (10 * 0,5) = 100$ кадрів на слово.

Розбиття слів.

Перше завдання, яке необхідно вирішити при розпізнаванні мови, - розбити мову на окремі слова. Для простоти припустимо, що в нашому випадку мова містить деякі паузи (інтервали мовчання), які можна розглядати як "роздільники" слів.

У цьому випадку вам потрібно знайти якість значення, поріг - вище цього значення - це слово, нижче - мовчати. Варіантів може бути декілька:

-Встановити константу (якщо вихідний сигнал завжди генерується однаково за однакових умов, константа буде працювати);

-Згрупуємо значення сигналу та чітко виберемо багато значень, що відповідають тиші (працює лише тоді, коли тиша займає значну частину вихідного сигналу);

-Аналіз ентропії;

Як ви вже здогадалися, зараз ми обговоримо останній пункт. Почнемо з того, що ентропія є мірою безладу, "мірою невизначеності будь-якого досвіду"

(с). У нашому випадку ентропія означає, наскільки наш сигнал "коливається" в даному кадрі.

Для того, щоб обчислити ентропію конкретного кадру, слід виконати наступні кроки:

Припустимо, наш сигнал нормалізований і всі його значення знаходяться в $[-1; 1]$;

Побудуємо гістограму значень сигналу кадру (розподіл щільності): обчислимо ентропію як:

$$E = \sum_{i=0}^{N-1} P[i] * \log_2(P[i]) \quad (3.1)$$

Так, ми отримуємо значення ентропії. Але це лише ще одна особливість фреймворку, щоб відокремити звук від тиші, нам все одно потрібно порівняти його з чимось. Деякі статті пропонують встановити поріг ентропії рівним середньому значенню між його максимальним та мінімальним значенням (у всіх кадрах). Однак, що стосується мене, цей метод не дав хороших результатів.

На щастя, ентропія (на відміну від того самого середнього квадратного значення) є відносно незалежною величиною. Це дозволяє мені вибрати поріг у вигляді константи (0,1).

Однак проблема не обмежується лише цим: "Ентропія може провиснути в середині слова (на голосну) і може раптово потрапити через невеликий шум. Друга проблема полягає у використанні" мінімальної довжини слова "та Видалити всіх кандидатів, які не пройшли відбір (не використовуються в першому абзаці).

Якщо мова в принципі не "чітко виражена", ви можете спробувати розкласти оригінальний набір кадрів на підпорядковані певним чином послідовності, і кожна послідовність пройде процес розпізнавання. Але це вже зовсім інша історія.

Так, у нас є набір кадрів, які відповідають певним словам. Ми можемо піти шляхом найменшого опору і використати середній квадрат як числову характеристику кадру. Однак ця метрика містить лише дуже мало інформації і підходить для подальшого аналізу.

Тут набуває значення коефіцієнт цепструма частоти Мела. Згідно з Вікіпедією, MFCC є поданням спектральної енергії сигналу. Переваги його використання такі:

Використовуючи частотний спектр сигналу (тобто, заснований на декомпозиції ортогональної функції синуса), це дозволяє враховувати "властивість" сигналу в подальшому аналізі.

Спектр проектується за спеціальною шкалою Мела, і найважливіші частоти можна вибрати для сприйняття людей.

Кількість розрахункових коефіцієнтів може бути обмежена будь-яким значенням (наприклад, 12), що дозволяє "стискати" кадр і, отже, обсяг обробленої інформації.

Розглянемо процес обчислення коефіцієнтів MFCC для конкретного кадру. Уявіть, наш кадр є вектором:

$$x[k], 0 \leq k < N \quad (3.2)$$

де N - розмір кадру.

Розкладання в ряд Фур'є

Насамперед розраховуємо спектр сигналу за допомогою дискретного перетворення Фур'є (бажано його "швидкої" FFT реалізацією).

$$X[k] = \sum_{n=0}^{N-1} x[n] * e^{-2*\pi*i*k*n/N}, 0 \leq k < N \quad (3.3)$$

Так само до отриманими значеннями рекомендується застосувати віконну функцію Хеммінга, що б "згладити" значення на кордонах фреймів.

$$H[k] = 0.54 - 0.46 * \cos(2 * \pi * k / (N - 1)) \quad (3.4)$$

Тобто результатом буде вектор такого вигляду:

$$X[k] = X[k] * H[k], 0 \leq k < N \quad (3.5)$$

Важливо розуміти, що після цього перетворення вздовж осі X ми отримуємо частоту (гц) сигналу та амплітуду сигналу вздовж осі Y (амплітуду (як спосіб позбавлення від складних значень)).

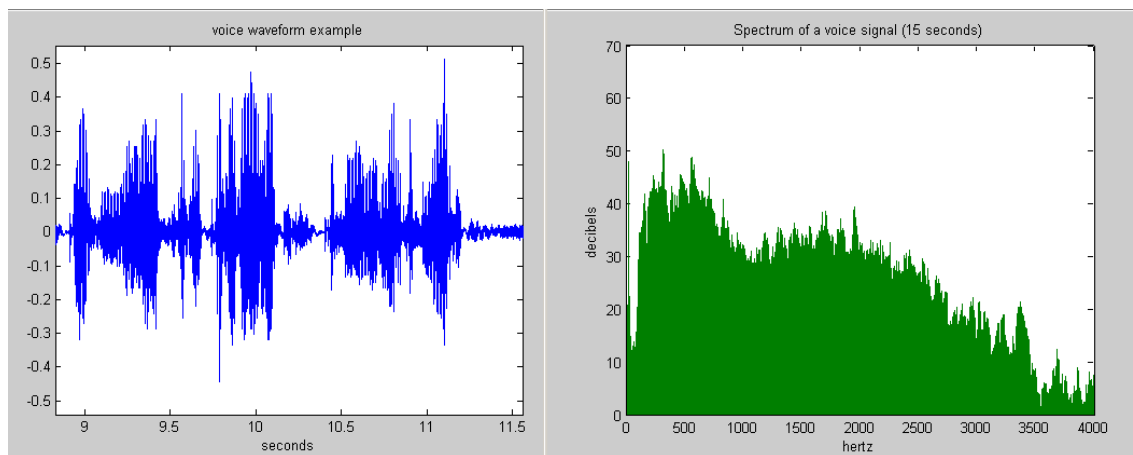


Рис. 3.4. Перетворення

3.3. Розрахунок mel-фільтрів

Почнемо з того, що таке мел - "психофізична одиниця висоти", заснована на суб'єктивному сприйнятті простих людей. В основному це залежить від частоти (та гучності та тону) звуку. Іншими словами, значення показує, наскільки звук певної частоти для нас «важливий». Ви можете перетворити частоту в крейду за такою формулою (не забудьте це висловити як "Формула 1"):

$$M = 1127 * \log(1 + F/700) \quad (3.6)$$

Зворотне перетворення виглядає так (запам'ятаємо її як «формула-2»):

$$F = 700 * (e^{M/1127} - 1) \quad (3.7)$$

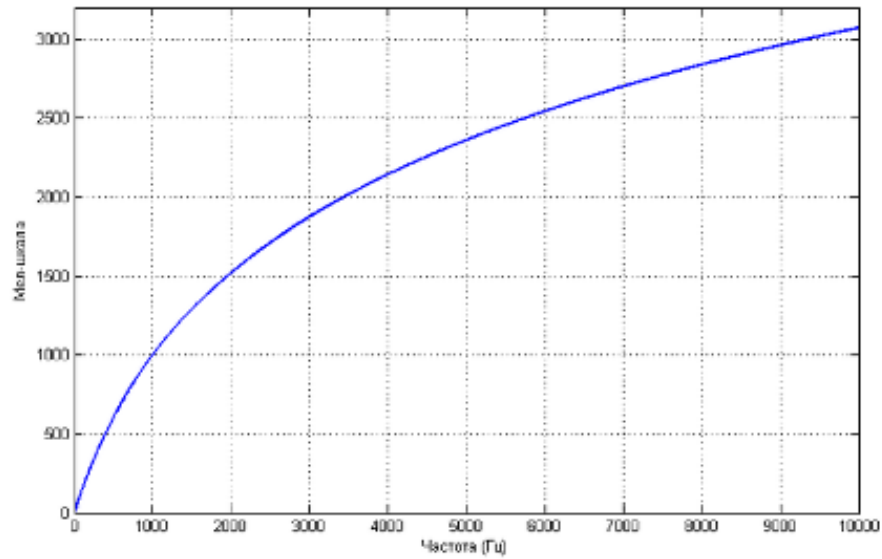


Рис. 3.5. Графік залежності мел / частота

Але повернемося до нашого завдання. Припустимо, у нас є кадр з 256 елементів. З даних аудіоформату ми знаємо, що частота звуку в цьому кадрі становить 16000 Гц. Припустимо, що діапазон людської мови становить [300; 8000] герц. Кількість необхідних коефіцієнтів крейди встановлюється на $M = 10$ (рекомендоване значення).

Для того, щоб розкласти отриманий вище спектр на рівні Мела, нам потрібно буде створити фільтр «гребінець». По суті, кожен фільтр розплаву є трикутною віконною функцією, яка дозволяє підсумовувати енергію в певному діапазоні частот для отримання коефіцієнта розплаву. Знаючи кількість коефіцієнтів крейди та проаналізований діапазон частот, ми можемо побудувати набір таких фільтрів.

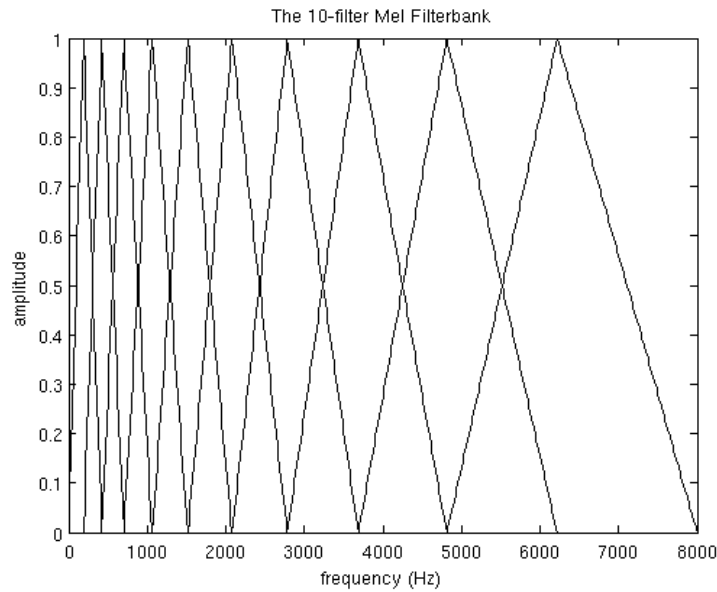


Рис. 3.6. Фільтри.

Зверніть увагу, що чим більший порядковий коефіцієнт крейди, тим ширша основа фільтра. Це пов'язано з тим, що поділ частотного діапазону, що цікавить, на частотну смугу, оброблену фільтром, відбувається за шкалою Мілове.

Але ми знову відволікаємось. Тому в нашому випадку діапазон частот, який нас цікавить, дорівнює [300, 8000]. Відповідно до Формули 1 у шкалі крейди, цей діапазон перетворюється на [401,25; 2834,99].

Далі, щоб побудувати 10 трикутникових фільтрів, нам знадобиться 12 опорних точок: $[i] = [401.25, 622.50, 843.75, 1065.00, 1286.25, 1507.50, 1728.74, 1949.99, 217.24, 2392.49, 2113.74, 2834.99]$

Зверніть увагу, що бали на крейдовій шкалі розподілені рівномірно. Використовуйте рівняння 2 для перетворення відношення назад у герц: $[i] = [300, 517.33, 781.90, 1103.97, 1496.04, 1973.32, 2554.33, 3261.62, 4122.63, 5170.76, 6446.70, 8000]$

Як бачите, шкала поступово подовжувалась, тим самим згладжуючи імпульс зростання "важливості" на низьких і високих частотах.

Тепер нам потрібно накласти отриману шкалу на спектр кадру. Ми пам'ятаємо, що вздовж осі X є частоти. Довжина спектру становить 256

елементів, і він підтримується на рівні 16000 Гц. Вирішуючи просте співвідношення, ви можете отримати таку формулу:

$f(i) = \text{floor}((\text{frameSize} + 1) * h(i) / \text{sampleRate})$ (3,8) що в нашому випадку еквівалентно $f(i) = 4, 8, 12, 17, 23, 31, 40, 52, 66, 82, 103, 128$.

От і все! Знаючи опорні точки на осі X нашого спектра, легко побудувати необхідні нам фільтри за такою формулою:

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \leq k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \quad (3.8)$$

Застосування фільтрів, логарифмування енергії спектра

Застосування фільтра полягає у множенні його значення та значення спектру в парах. Результатом цієї операції є коефіцієнт Мела. Оскільки ми маємо M фільтрів, коефіцієнти будуть однаковими.

$$S[m] = \log\left(\sum_{k=0}^{N-1} |X[k]|^2 * H_m[k]\right), 0 \leq m < M \quad (3.9)$$

Однак нам потрібно застосовувати фільтр Мела не до значення спектру, а до його енергії. Потім прологарифмувати результати. Вважається, що це знижує чутливість коефіцієнта до шуму.

Косинусне перетворення

Для того, щоб отримати однакові коефіцієнти "Keerpral", використовується дискретне косинусне перетворення (DCT). Сенс полягає в тому, що він буде «стискати» результати, отримані за рахунок підвищення ефективності першого коефіцієнта та зниження ефективності останнього.

У цьому випадку не множите (коефіцієнт масштабу) при використанні DCTII.

$$C[l] = \sum_{m=0}^{M-1} S[m] * \cos(\pi * l * (m + \frac{1}{2})/M), 0 \leq l < M \quad (3.10)$$

Тепер для кожного фрейма ми маємо набір з M mfcc-коефіцієнтів, які можуть бути використані для подальшого аналізу.

3.4. Приховані Марковські моделі

Модель прихованого Маркова - статистична модель, що використовується для моделювання роботи процесу, подібного до процесу Маркова з невідомими параметрами. Завдання полягає у вирішенні невідомих параметрів на основі спостережень. Отримані параметри можна використовувати для подальшого аналізу, наприклад, для розпізнавання зразків. Паливно-мастильні матеріали можна вважати найпростішою мережею обходу довіри.

Процес Маркова - випадковий процес, еволюція після будь-якого заданого значення його часового параметра t не залежить від еволюції, що відбулася до t , за умови, що значення процесу фіксовано в цей час ("майбутнє" процесу не залежить від відомого Інше пояснення (Венцель): "Майбутнє" процесу залежить лише від "минулого" і лише від "минулого".

Процес Маркова - модель авторегресії першого порядку:

$$X_t = c + \alpha X_{t-1} + \epsilon_t \quad (3.11)$$

Коли простір станів дискретний, ланцюжок Маркова є окремим випадком марковського процесу.

Розглянемо простий приклад випадкового процесу Маркова. Точка рухається хаотично вздовж абсциси. У нульовий момент часу точка знаходиться у початку координат і залишається там одну секунду. Через одну секунду киньте монету - якщо значок відкладено, то точка X рухається вліво на одну одиницю довжини (якщо це число) - вліво. Через секунду знову переверніть монету і виконайте ті самі випадкові рухи тощо. Процес зміни положення точки ("блукання") є випадковим процесом з дискретним часом ($t = 0, 1, 2, \dots$) і набором злічуваних станів. Цей випадковий процес називається

марковським, оскільки наступний стан точки залежить лише від поточного (поточного) стану, а не від минулого стану (незалежно від того, як і коли точка зіткнеться з поточними координатами).

Структура прихованої марковської моделі.

У звичайній моделі Маркова спостерігачем може бути держава, тому єдиним параметром є можливість переходу. У основній моделі Маркова ми можемо спостерігати лише змінні, на які впливає цей стан. Кожен стан має розподіл ймовірностей між усіма можливими початковими значеннями. Отже, послідовність символів, що генеруються паливом, забезпечує інформацію про послідовність станів.

На малюнку нижче показано загальну структуру палива. Еліпси - це змінні з випадковими значеннями. Випадкова величина $x(t)$ - це значення прихованої змінної в момент часу t . Випадкова величина $y(t)$ - це значення змінної, що спостерігається в момент часу t . Стрілки на графіку позначають умовну залежність.

З рисунка видно, що значення прихованої змінної $x(t)$ (у момент часу t) залежить лише від значення прихованої змінної $x(t-1)$ (у момент часу $t-1$). Це називається марківською властивістю. Хоча спостерігається одночасно значення змінної $y(t)$ залежить лише від значення прихованої змінної $x(t)$ (все в момент часу t).

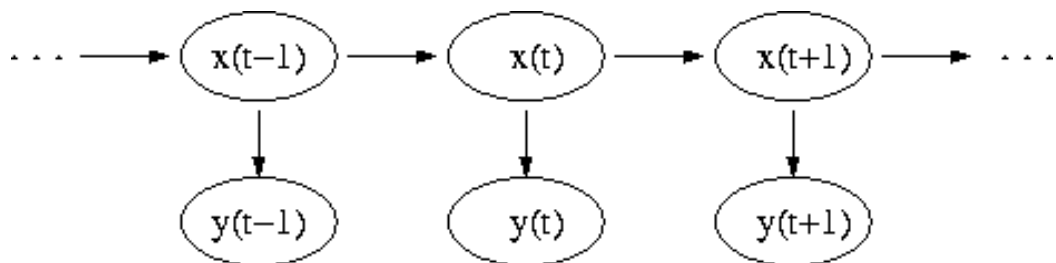


Рис. 3.7. Схема переходів Марковського процесу.

Імовірність побачити послідовність

$$Y = y(0), y(1), \dots, y(L-1) \quad (3.12)$$

довжини L дорівнює

$$P(Y) = \sum_x P(Y | X)P(X), \quad (3.13)$$

тут сума пробігає по всіх можливих послідовностей прихованих вузлів

$$X = x(0), x(1), \dots, x(L - 1). \quad (3.14)$$

Через те, що кількість можливих послідовностей прихованих вузлів дуже велика, для багатьох реальних завдань метод підрахунку шляхом повного пошуку значення $P(Y)$ дуже трудомісткий. Але використання процесу прямого та зворотного руху може значно збільшити швидкість обчислення.

3.5. Висновок до розділу 3

Цей розділ представляє основи розпізнавання мови, алгоритми розпізнавання та структуру прихованих моделей Маркова. Найшвидша та найефективніша взаємодія між людьми - через усне мовлення. За допомогою мови можна передавати різні емоційні дії та емоції, а головне - корисну інформацію. Немає сумнівів, що для аудіо-вводу-виводу потрібно створювати комп'ютерні інтерфейси, оскільки їх ефективність базується на майже безмежних можливостях вираження в різних сферах людської діяльності.

РОЗДІЛ 4

ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1. Забезпечення безпеки життєдіяльності підприємств електротехнічної галузі у воєнний час

Забезпечення стійкості роботи об'єкта радіотехніки в умовах НС - одне з основних завдань ЦЗ.

Під стійкістю функціонування об'єкта розуміють здатність його в умовах НС випускати продукцію в запланованому обсязі й номенклатурі, виконувати всі свої функції, а у випадку аварії, катастрофи, пошкодження - відновлювати виробництво в мінімально короткий термін.

На стійкість функціонування об'єкта господарювання (ОГ) у НС впливають такі чинники:

- надійність захисту робітників і службовців від наслідків НС - аварій, катастроф, від первинних і вторинних факторів зброї масового ураження (ЗМУ);
- здатність інженерно-технічного комплексу об'єкта протистояти цим впливам;
- надійність системи постачання об'єкта всім необхідним для виробництва продукції (сировиною, паливом, енергією, газом, водою і т.д.);
- стійкість і безперервність керування виробництвом і ЦЗ;
- підготовленість об'єкта до проведення рятувальних та інших невідкладних робіт (РіНР) і відбудовних робіт;

З переліченого вище впливають вимоги до стійкості функціонування ОР в умовах мирного часу, щоб виключати ситуацію типу Чорнобильської. Ці вимоги закладені в нормах проектування інженерно-технічних заходів (ІТЗ) ЦО, а також у розроблених на їхній основі нормативних документах.

Захист робітників і службовців досягається трьома способами:

- укриття людей у захисних спорудах.
- застосування засобів індивідуального захисту.

Проведення евакуаційних заходів для робітників і службовців та членів їх сімей.

Засоби індивідуального захисту забезпечують захист людей при перебуванні на виробничих місцях і на місцевості, що забруднена РР, ОР, НХР, БЗ.

Укриття в захисних спорудах - найбільш ефективний спосіб захисту виробничого персоналу працюючої зміни. Захисні споруди повинні будуватися на кожному об'єкті своєчасно і забезпечувати укриття найбільшої працюючої зміни.

Евакуаційні заходи забезпечують захист членів сімей робітників, службовців і виробничого персоналу непрацюючих змін.

Надійність захисту виробничого персоналу досягається завдяки застосуванню всіх трьох способів захисту з урахуванням конкретної обстановки.

Захист засобів виробництва полягає у підвищенні опірності будівель, споруд і конструкцій об'єкта впливу можливих вражаючих чинників і захисту виробничого процесу.

Забезпечення стійкого постачання досягається проведенням заходів щодо захисту комунально-енергетичних мереж, транспортних комунікацій і джерел постачання, а також створенням необхідних запасів палива, сировини, напівфабрикатів і комплектуючих виробів.

Підготовка до відновлення порушеного виробництва здійснюється своєчасно. Вона передбачує планування відповідних робіт за різними варіантами, підготовку ремонтних бригад, створення необхідного запасу матеріалів, обладнання і направлена на поновлення випуску необхідної продукції в мінімальні строки.

Підвищення стійкості й оперативності управління виробництвом досягається створенням на об'єкті системи зв'язку, високою професійною

підготовкою керівного складу до виконання функціональних обов'язків і заходами ЦЗ в повсякденній діяльності і в умовах НС, а також своєчасним прийняттям правильних рішень і постановкою завдань підлеглим щодо обстановки. Таким чином, підвищення стійкості роботи об'єктів промисловості в умовах НС досягається своєчасним проведенням комплексу інженерно-технічних, технологічних і організаційних заходів, направлених на максимальне зниження впливу вражаючих чинників і створення умов для ліквідації наслідків НС. Інженерно-технічні заходи включають комплекс робіт, направлених на підвищення стійкості виробничих будівель, споруд, технологічного обладнання, комунально-енергетичних систем.

Технологічні засоби забезпечують підвищення стійкості роботи об'єкта шляхом зміни технологічних процесів, сприяють спрощенню виробництва продукції і виключають можливість виникнення аварій і катастроф. Організаційні заходи передбачають розробку і планування дій керівного складу, штабу, служб і формувань ЦЗ із захисту робітників і службовців, проведення рятувальних та інших невідкладних робіт, відновлення виробництва, а також випуск продукції на обладнанні, що збереглося.

Заходи щодо підвищення стійкості роботи підприємства в умовах НС

Підвищення стійкості роботи об'єкта в умовах НС досягаються:

- моніторингом зовнішніх і внутрішніх небезпечних факторів для підприємства;
- підвищенням надійності роботи й створенням дублюючих джерел енерго, газо- та водопостачання, а також створенням запасів сировини, палива, комплектуючих деталей, обладнання і матеріалів;
- вдосконаленням технологічних процесів виробництва, забезпеченням автоматичного відключення при виході з ладу установок;
- забезпеченням співробітників підприємства засобами індивідуального захисту;

- підготовкою в заміській зоні баз для розміщення науково-дослідних, конструкторських відділів та інших невиробничих підрозділів об'єкта;
- постійною готовністю аварійно-рятувальних формувань до проведення рятувальних та невідкладних аварійних робіт;
- будівництвом і обладнанням сховищ на підприємствах для робітників і службовців (для цього можуть бути використані шахти та інші виробітки);
- створенням на об'єктах захисних споруд для пунктів керування;
- проведенням організаційних та інженерно-технічних заходів щодо підготовки об'єкта до особливого режиму роботи.

Підвищення надійності електропостачання підприємства може здійснюватися підключенням до резервних мереж (ліній електропередач), розподілом схеми мереж на частини, що працюють незалежно, встановленням дизельних електростанцій. Особливу увагу необхідно приділяти надійності електро-, водо-, теплопостачання на комунальних підприємствах, медичних, дитячих установах, підприємствах з неперервним циклом виробництва.

З метою надійного забезпечення водою підприємства підключають до дублюючих джерел, створюють резервуари, влаштовують артезіанські свердловини.

Для підвищення стійкості систем газо-, тепло- й паливопостачання закріплюють газопроводи, газорозподільні станції, якомога більшу частину газопроводів влаштовують під землею, встановлюють автоматичні пристрої відключення в разі аварії, максимально збільшують запаси вугілля, мазуту, бензину, влаштовують власні системи обігріву.

Стійкість підприємства підвищується при збільшенні запасів сировини, інструментів та матеріалів, але при цьому зменшується ефективність використання коштів, тому в практиці підприємницької діяльності запаси збільшують при підвищенні ризику виникнення НС. При створенні запасів необхідно враховувати не тільки можливість розвитку НС в районі

функціонування підприємства, але і в регіонах, де працюють постачальники й через які пролягають транспортні магістралі. Зрив поставок тільки однієї комплектуючої призводить до зупинки всього підприємства.

Розміщуючи небезпечні виробництва, враховують ризик виникнення НС техногенного походження.

Будівництво широких магістралей і створення необхідної транспортної мережі дає можливість в НС мирного і воєнного часу при зруйнуванні будівель запобігти суцільним завалам, які ускладнюють дії формувань ЦО і евакуацію потерпілих з району ураження в заміську зону.

Ширина магістралі в метрах $L = H_{\text{макс}} + 15\text{м}$

де H - висота найвищого будинку на магістралі, крім висотних громадських будівель каркасної конструкції.

Міська транспортна мережа повинна забезпечувати надійність сполучення між житловими і промисловими районами, вільний вихід до магістралі, що веде за межі міста, а також найкоротший і найзручніший зв'язок з центром міста, житлових і промислових районів із залізничними і автобусними вокзалами, вантажними станціями, річковими і морськими портами і аеропортами. Міжміські автомобільні шляхи повинні прокладатись в обхід міста. Навколо великих міст краще прокладати кільцеві дороги і з'єднувальні обхідні шляхи. Це зменшить забруднення повітряного басейну в межах міста від автомобільного транспорту і не порушить транспортних зв'язків в НС мирного і воєнного часу.

Лісопарковий пояс при застосуванні противником сучасної зброї може служити для розміщення робітників, службовців підприємств і населення.

Розміщення об'єктів повинно здійснюватись з урахуванням зон можливих руйнувань. Нові важливі промислові підприємства, основні склади і бази мають розміщуватись за межами зони можливих руйнувань.

За зонами можливих сильних руйнувань повинні розміщуватись:

- бази, склади з продовольчими і промисловими товарами першої необхідності;

- базові склади легкозаймистих і горючих матеріалів, головні споруди водозабезпечення;
- насосні й компресорні станції магістральних трубопроводів;
- міжміські кабельні магістральні мережі та інші важливі об'єкти.

У зоні можливих сильних руйнувань дозволяється розміщувати комунальні гаражі, тролейбусні депо, склади поточного забезпечення, підземні магістральні трубопроводи, одну з груп головних споруд системи водопостачання та інші підприємства обслуговування населення міста.

Нові промислові підприємства (об'єкти) треба будувати з урахуванням вимог, виконання яких сприяє підвищенню стійкості інженерно-технічного комплексу об'єкта.

Будівлі й споруди на об'єкті необхідно розміщувати розосереджено. Відстань між будівлями має забезпечувати протипожежні розриви. Ширину протипожежного розриву визначають за формулою

$$L = H_1 + H_2 + 15 \text{ м}, \quad (4.1)$$

де H_1 і H_2 - висота сусідніх будинків.

4.2 Класи виробничих та складських приміщень по вибуховій та пожежній небезпеці. Вогнестійкість будівельних конструкцій і матеріалів.

Відповідно до чинних будівельних норм і правил (БНіП) будівлі й споруди за вогнестійкістю поділяються на п'ять ступенів.

Ступінь вогнестійкості будівель і споруд визначається границями вогнестійкості основних будівельних конструкцій, границями поширення вогню по цих конструкціях, а також відповідними їм групами займання.

Границя вогнестійкості елементів і будівельних конструкцій - це проміжок часу (в годинах або хвилинах) від початку вогневого стандартного випробування зразків до виникнення одного з граничних станів елементів і конструкцій. Стандартне випробування виконують відповідно до стандарту.

Для того щоб перевірити, чи відповідає вогнестійкість будівлі або споруди" що проектується, протипожежним вимогам, необхідно зробити розрахунок вогнестійкості конкретних будівельних конструкцій, який зводиться до визначення розрахункової тривалості горіння і потрібної границі вогнестійкості. Розрахунок вогнестійкості завершується порівнянням фактичної і потрібної границі вогнестійкості.

Границя поширення вогню - це проміжок часу (в годинах або хвилинах) від початку вогневого стандартного випробування зразків до настання будь-якої з ознак, які характеризують поширення вогню по конструкціях.

Ступінь вогнестійкості будівель і споруд характеризується групою займання будівельних матеріалів, з яких виготовлена споруда і вогнестійкості несучих будівельних конструкцій та їхніх частин. Відповідно до цього всі об'єкти поділяються на п'ять ступенів займистості. Нормований ступінь вогнестійкості будівель за вимогами СНІП залежить від категорії пожежної безпеки виробництва, висотності будівель тощо. Із зростанням номера категорії приміщення збільшується ступінь займання конструкції і зменшується границя їхньої вогнестійкості.

За ступенем вогнестійкості вибирають матеріал для стін, перекриттів і колон та визначають необхідні границі між будівлями або спорудами. Необхідну кількість води для зовнішнього гасіння пожеж можна визначити, знаючи категорію пожежної безпеки виробничої будівлі, ступінь її вогнестійкості і об'єм. Вибір необхідних систем вентиляції, опалення, водопостачання, освітлення і електропроводки, електрообладнання і засобів для гасіння пожеж здійснюють залежно від категорії пожежної безпеки будівель або споруд.

Категорія А (вибухопожежонебезпечна). Горючі гази, легкозаймисті рідини з температурою спалаху $t < 28^{\circ}\text{C}$ в такій кількості, що можуть утворювати вибухонебезпечні парогазоповітряні суміші, під час займання яких розвивається розрахунковий надлишковий тиск $\Delta P_{\text{ф}} > 5$ кПа у фронті ударної хвилі вибуху в приміщенні. Речовини і матеріали, здатні вибухати і

горіти при взаємодії з водою, киснем повітря або один з одним в такій кількості, що розрахунковий тиск $\Delta P_{\text{ф}} > 5$ кПа у фронті ударної хвилі у приміщенні. До них відносяться водневі станції, склади балонів для зріджених газів, приміщення стаціонарних кислотних і лужних акумуляторних установок, насосні станції з перекачування рідин з температурою спалаху до $t = 28^{\circ}\text{C}$ та склади цих рідин, приміщення малярних цехів, де застосовують нітрофарби, лаки і нітроемалі, ацетиленові станції і приміщення ацетиленових генераторів та ін.

Категорія Б (вибухопожежонебезпечна). Горючі порохи або волокна, легкозаймисті рідини з температурою спалаху $t > 28^{\circ}\text{C}$, горючі рідини в такій кількості, що можуть утворювати вибухонебезпечні порохоповітряні або пароповітряні суміші, під час займання яких розвивається розрахунковий надлишковий тиск вибуху у приміщенні $\Delta P_{\text{ф}} = 5$ кПа. До них відносяться, наприклад, виробництва із застосуванням аміаку, пального для ТГД і ТРДД, а також складські приміщення, в яких зберігаються дизельне пальне, мінеральні мастила, закриті склади вугілля, пакгаузи змішаних вантажів, гаражі тощо.

Категорія В (пожежонебезпечна). Легкозаймисті, горючі й важкозаймисті рідини, тверді горючі й важкозаймисті речовини і матеріали, здатні при взаємодії з водою, киснем повітря або один з одним горіти тільки за умови, що приміщення, в якому вони знаходяться (обертаються), не відносяться до категорії А або Б.

Категорія Г. Негорючі речовини і матеріали в горючому, розжареному або розплавленому стані, процес обробки яких супроводжується виділенням променевого тепла, Іскор та полум'я; горючі гази, рідини й тверді речовини, що спалюються або утилізуються як паливо (ливарні цехи, станції випробування двигунів, котельні і т. ін.).

Категорія Д Негорючі речовини і матеріали в холодному стані. Це - цехи холодної обробки металів (крім магнієвих сплавів), повітродувки, інструментальні цехи, насосні станції для перекачування негорючих рідин тощо.

Категорію виробництва вибирають за нормами технологічного проектування або за спеціальним переліком виробництв.

Відповідно до ПУЕ існує також класифікація виробничих приміщень і зовнішніх установок з вибухової і пожежної безпеки. Віднесення приміщень і зовнішніх установок до вибухо- або пожежонебезпечних залежить від умов створення можливих вибухо- і пожежонебезпечних середовищ.

Вибухонебезпечними називаються приміщення і зовнішні установки, в яких згідно з умовами технологічного процесу можуть створюватися вибухонебезпечні суміші: горючих газів або пари з повітрям чи киснем та іншими газами-окислювачами (наприклад, з хлором), горючих пилу або волокна з повітрям у разі переходу їх у завислий стан. Усі вибухонебезпечні виробничі приміщення і установки відповідно до ПУЕ поділяються на такі класи: В-1, В-1а, В-1б - приміщення з безпекою вибуху пари і газів, В-1 г - зовнішні установки з безпекою вибуху горючого пилу або волокна. Наприклад, в приміщеннях класу В-1 виділяються горючі гази або пара в об'ємі, що створює з повітрям або іншими окислювачами вибухонебезпечні суміші не тільки при аварійних, але й при нормальних режимах роботи, невеликих за тривалістю (при зберіганні або зливанні, наливанні, переливанні легкозаймистих і горючих речовин, які знаходяться у відкритих посудинах, тощо).

У приміщеннях класу В-1а при нормальній експлуатації вибухонебезпечні суміші горючих газів і пари з повітрям та іншими окислювачами не виникають, а створюються тільки в результаті аварій або несправностей. У приміщеннях класу В-1б на відміну від класу В-1а є такі особливості: горючі гази мають високу нижню границю вибуховості (15 % і більшу) і різкий запах при граничнодопустимих за санітарними нормами концентраціях (наприклад, машинні зали аміачних і холодильних абсорбційних установок); створення вибухонебезпечної концентрації для всього приміщення за умов технологічного процесу виключається, можливе створення місцевої вибухонебезпечної концентрації; легкозаймисті і горючі

речовини є в приміщеннях в невеликих кількостях і роботу з ними проводять у витяжних шафах або під витяжними зонтами (приміщення для зберігання проб легкозаймистих і горючих речовин, лабораторії і т. ін.).

До класу В-1г відносяться зовнішні установки, які містять у собі вибухонебезпечні гази, пару, легкозаймисті речовини (газгольдери, зливно-наливні естакади для легкозаймистих речовин та ін.).

У приміщеннях класу В-II виділяються горючі пил або волокна, які переходять у завислий стан і здатні створювати з повітрям та іншими окислювачами вибухонебезпечні суміші не тільки при випадкових, але й при нормальних, недовгочасних режимах роботи (наприклад, при завантажуванні технологічних апаратів).

У приміщеннях класу В-IIа, на відміну від В-II, небезпечні стани не виникають під час нормальної експлуатації, вони можливі тільки в результаті аварій або несправностей.

Найбільш вибухонебезпечними є приміщення класів В-I і В-II, найменш вибухонебезпечними — приміщення класів В-1б, установки класу В-1г.

Пожежонебезпечними називаються приміщення і установки, в яких використовують або зберігають горючі речовини. Відповідно до ПУЕ вони поділяються на класи II-I, II-II, II-IIa, II-III.

До класу II-I відносяться приміщення, в яких використовують або зберігають горючі рідини з температурою спалаху пари вище 45°C (склади мінеральних мастил, регенераційні установки для них тощо).

До класу II-II відносяться приміщення, в яких небезпека виникає внаслідок виділення горючих пилу або волокна, які переходять у завислий стан. Виникаюча при цьому небезпека призводить до пожежі (але не до вибуху).

До класу II-IIa відносяться виробничі й складські приміщення, в яких знаходяться тверді або волокнисті горючі речовини (дерево, тканини тощо) і в яких горючих пилу або волокна, що переходять у завислий стан, немає.

До класу II-III відносяться зовнішні установки, в яких використовують або зберігають горючі рідини з температурою спалаху пари вищою 45°C, наприклад, відкриті склади мінеральних масел, а також тверді горючі речовини [96].

4.3. Висновок до розділу 4

У четвертому розділі охорона праці розглянуті такі питання як класи виробничих та складських приміщень по вибуховій та пожежній небезпеці, вогнестійкість будівельних конструкцій і матеріалів. У підрозділі розглянуто питання з підвищення стійкості роботи радіотехніки. Визначені основні заходи, які використовуються на надійність роботи об'єктів. Визначено основні способи захисту робітників і службовців та заходи щодо підвищення роботи підприємств радіотехнічної галузі.

ВИСНОВКИ

У кваліфікаційній роботі здійснено обґрунтування методу розпізнавання голосових звуків для голосового керування IoT дівайсами.

У першому розділі було проведено дослідження голосове управління технічними пристроями, модель голосового управління периферійними пристроями, схема управління електричними пристроями. Розглянуті можливості голосового управління технічними пристроями, модель голосового управління периферійними пристроями, схема управління електричними пристроями. Основні етапи розпізнавання мови - ключові моменти, можливість використання нейромереж для побудови системи розпізнавання мови, модель нейромережі та навчання нейромережі. Здійснюється аналіз структури розпізнавання мови та розрахунок mel -фільтрів.

У другому розділі показано дослідження розпізнавання мови - ключові моменти, проаналізовано можливість використання нейромереж для побудови системи розпізнавання мови, описана модель нейромережі та введення звуку.

У третьому розділі проведений аналіз основ розпізнавання мови та структури розпізнавання мови, проведено розрахунок mel -фільтрів. Виконано дослідження характеристик прихованих Марковських моделей також виконано алгоритм ПММ (прихованої марковської моделі) в розпізнаванні мови.

Відповідно у магістерській роботі проводився аналіз та систематизування типових технічних характеристик голосового управління.

У четвертому розділі описано забезпечення безпеки життєдіяльності підприємств електротехнічної галузі у воєнний час. Класи виробничих та складських приміщень по вибуховій та пожежній небезпеці. Вогнестійкість будівельних конструкцій і матеріалів

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. <https://hi-tech.mail.ru/review/iot-devices/>
2. <https://coderlessons.com/articles/mobilnaia-razrabotka-articles/kak-ispolzovat-voice-dlia-upravleniia-ustroistvami-iot-s-pomoshchiu-android>
3. <https://habr.com/ru/post/409289/>
4. <https://iot.ru/gadzhety/potrebiteley-interesuet-integratsiya-golosovogo-upravleniya-v-umnye-ustroystva-issledovanie>
5. <https://gadgetpage.ru/smart-house/3930-golosovye-pomoschniki-dlja-umnogo-doma.html>
6. <http://abbgroup.ru/news/golosovoe-upravlenie-vozglavilo-trend-iot-2017>
7. <https://dou.ua/lenta/articles/voice-controlled-smart-home/>
8. https://pikabu.ru/story/umnyiy_dom__golosovoe_upravlenie_v_tri_shaga_raspberry__homekit_7183707
9. <https://moluch.ru/archive/153/43302/>
10. <https://www.bestreferat.ru/referat-168577.html>
11. https://studentlib.com/diplom-205110-razrabotka_sistemy_golosovogo_upravleniya_elektromehaniceskimi_ustroystvami.html
12. <https://nauchkor.ru/uploads/documents/5b887daf7966e1073081b4e8.pdf>
13. <https://habr.com/ru/post/237589/>
14. <https://dou.ua/lenta/articles/voice-controlled-smart-home/>
15. <https://www.bestreferat.ru/referat-140657.html>
16. <https://referatbank.ru/referat/preview/45906/referat-sistemy-raspoznavaniya-rechi.html>
17. <https://www.bibliofond.ru/detail.aspx?id=871454>
18. <https://bot.konveier.com/2018/12/24/osobennosti-primeneniya-tehnologij-raspoznavaniya-golosa/>
19. <https://moluch.ru/archive/163/45163/>

20. <https://www.primavista.ru/blog/2017/08/03/programmyi-raspoznavaniya-rechi>
21. https://studentlib.com/diplom-205110-razrabotka_sistemy_golosovogo_upravleniya_elektromehanicheskimi_ustroystvami.html
22. <https://www.bibliofond.ru/view.aspx?id=528193>
23. <https://www.bestreferat.ru/referat-168577.html>
24. <https://works.doklad.ru/view/JfQion7ow-4/all.html>
25. <https://blog.dti.team/voice-assistants-3/>
26. <https://vc.ru/flood/27565-preimushchestva-i-nedostatki-golosovyh-interfeysov>
27. http://cinref.ru/razdel/05600voenoe_delo/03/179870.htm
28. Балахов Н.Н. Развитие форм и способов ведения военных действий в начале XXI века // Зарубежное военное обозрение. – 2003. – № 4. – С. 33—42.
10 4/2014
29. Попов А.О., Ремесло В.Я. Загальні тенденції розвитку радіоелектронної боротьби за досвідом локальних війн і збройних конфліктів сучасності // Труди Академії. – 2003. – № 43. – С. 314—316.
30. Перунов Ю.М., Любин М.Д. Радиоэлектронная борьба: исторический аспект // Военная мысль. – 2012. – № 12. – С. 58—74.
31. Ямпольский Л.С. Обобщенный анализ применения средств воздушного нападения ОВС НАТО при проведении военной операции в Югославии «Решительная сила» и в других локальных войнах в 90-х годах. – Ульяновск: УлГТУ, 2000. – 80 с.
32. Балыбин В.А., Батулин Ю.О., Гулидов А.А. О совершенствовании системы вооружения радиоэлектронной борьбы // Военная мысль. – 2013. – № 11. – С. 14—20.

ДОДАТКИ

УДК [004.383.8](#)

Дембічак А.А., студент - магістр кафедри радіотехнічних систем,

Химич Г.П., науковий керівник, ст. викл.

Тернопільський національний технічний університет імені Івана Пулюя

ОБҐРУНТУВАННЯ МЕТОДУ РОЗПІЗНАВАННЯ ЗВУКІВ ДЛЯ ГОЛОСОВОГО КЕРУВАННЯ ІОТ ДЕВАЙСАМИ.

Dembichak A., student - master of the department of radio engineering systems,

Khymych G., scientific supervisor, senior lecturer

Ternopil Ivan Puluj National Technical University

JUSTIFICATION OF SOUND RECOGNITION METHOD FOR VOICE CONTROL OF IOT DEVICES.

Впровадження технологій мобільних стільникових наземних систем 4G, 5G стало поштовхом для розвитку технології ІоТ «інтернет речей». Одна із задач, яку необхідно вирішити при створенні діалогу між людиною та девайсом – це голосове спілкування, яке є практично однією із складових системи захисту, а також можливістю керувати даним пристроєм на основі голосового монологу (діалогу), передачею відповідних команд. Орієнтовна структура взаємодії між ІоТ на рівні обмеженого простору (квартира, будинок) показано на рис.1.



Рисунок 1. Структура взаємодії між девайсами ІоТ за допомогою комунікацій.

Зв'язок між девайсами відбувається за допомогою відповідних інтерфейсів, у нашому варіанті – голосових. Прикладами таких комунікаційних інтерфейсів є голосовий пошук від компанії Google та асистент Siri від компанії Apple. Вони стають складовими частинами технологій «smart home». При створенні повноцінних та достовірних таких інтерфейсів слід враховувати знання комп'ютерних наук, лінгвістичних, психологію поведінки індивідууму. Для того, щоб за допомогою штучного інтелекту підпорядкувати голосові команди для конкретного девайса, потрібно вирішувати наступні задачі на яких ґрунтується складність розпізнавання мови в залежності від емоційного стану, а саме: тембр, гучність, висота, темп, інтонація, якість дикції та самої мови і ін. Ці задачі з розпізнавання мови повинні вирішуватись навіть при екстремальних умовах. Крім цього потрібно «навчити» девайси реагувати або виконувати команди суґубо у відповідності до свої технічних задач, які запрограмовані у пристрої.

Даний створений метод, де основним протоколом обміну даних на невеликих відстанях є Bluetooth та враховані основні технології, які безпосередньо пов'язані з ІОТ:

- БСС (бездротова сенсорна мережа, де сенсори зв'язуються один з одним за допомогою одного радіочастотного каналу),

- RFID (радіочастотна ідентифікація, при якому дані зчитуються за допомогою радіочастотних сигналів), використовує систему базових складових характеристики голосу та

смугових, обмежувальних, адаптивних фільтрів у вузьких спектральних частотних смугах, за рівнем амплітуд.

Такий метод дає можливість досить точно ідентифікувати голос, який використовується індивідумом для створення керуючих команд девайсами.