УДК 004.056.53

д.т.н., проф. **Мясіщев О.А.** (ХмНУ)
к.т.н., доц. **Муляр І.В.** (ХмНУ)
к.т.н. **Лоза В.М.** (ВІКНУ)
Козак С.В. (ХмНУ)

ГОЛОСОВЕ КЕРУВАННЯ ВІДДАЛЕНИМИ ПРИСТРОЯМИ ЧЕРЕЗ МЕРЕЖУ ІНТЕРНЕТ

При роботі з технікою, для користувача в основному важливу роль відіграють, такі фактори, як легкість та швидкість. Для того щоб забезпечити користувача такими можливостями, важливою є розробка системи, яка буде працювати від голосу користувача.

Керування системою за допомогою голосу, може значно суттєво підвищити її надійність, тому що, при якісній розробці такої системи, її можна налаштувати тільки на той голос, від якого вона буде працювати. Це унеможливить використання системи кимось іншим, тому що, голос буде ідентифікуватись.

В статті описано принцип роботи такої системи, яка дозволяє користувачу лише за допомогою голосу, керувати електронними пристроями . Виокремленні завдання, які потрібно виконати для того щоб створити стабільну систему. Використання перетворень Фур'є для виділення набору основних частот і амплітуди, щоб проаналізувати спектральний склад каналу та представити сигнал у цифровій формі. Використання методів статистичного аналізу: метод прихованого Марківського моделювання, метод динамічного програмування та метод нейронних мереж.

Таким чином, описані усі важливі моменти і етапи для побудови системи керування віддаленими електронними пристроями через мережу Інтернет за допомогою голосу.

Ключові слова: прихована модель Маркова, перетворення Фур'є, нейронні мережі, динамічне програмування.

Вступ. В сучасному світі з кожним днем, з'являється велика кількість різноманітних електронних пристроїв. Для їх освоєння людині, потрібно затратити значний час та зусилля. Є певна категорія людей з обмеженими можливостями. Для того, щоб спростити керування пристроями та полегшити роботу з ними, є необхідними створити систему, яка могла б керувати пристроями за допомогою голосу користувача.

Крім цього, така система, може бути надзвичайно надійною, в тому випадку, коли система буде розрізняти голоси різних людей. Це особливо стосується, наприклад задіяння військового озброєння за допомогою конкретної особи.

Існують області застосування систем автоматичного розпізнавання мови, де описані проблеми проявляються особливо гостро через жорстко обмежені обчислювальні ресурси, наприклад, на мобільних пристроях. Виробники мобільних телефонів і планшетів знайшли вихід в перенесенні ресурсномістких обчислень з пристроїв користувачів на сервери хмарних технологій, де, власне, і проводиться розпізнавання. Користувацький додаток тільки відправляє туди мовні запити і приймає відповіді, використовуючи підключення до Інтернету.

Однак, для такої реалізації необхідні певні умови, наприклад, безперервний доступ до Інтернету, і створення компактного і надійного самостійного пристрою, що експлуатує тільки доступні «на місці» обчислювальні потужності. Описані труднощі виникають при створенні інтелектуальних пристроїв як у військовій сфері, так і в цивільній.

Постановка задачі. Беручи до уваги вище викладений матеріал стає зрозуміла уся актуальність такої теми, голосове керування. Комп'ютерні технології дедалі більше проникають у всі сфери нашого життя. Кількість пристроїв, під керуванням різних ОС, стрімко збільшується, а темпи їх поширення зростають.

Не дивлячись на те, що на сьогоднішній день розроблено різні методи керування, наявність недоліків у цих методах та поява все новіших приладів, гостро ставлять питання, щодо створення голосового керування цими приладами. Використання адаптивного метода є суттєвим кроком вперед, він дозволить зробити систему гнучкою та адаптованою до змін.

Враховуючи особливості мови, застосування такого методу можливе в трьох випадках:

1. Запис голосової команди від користувача на пристрій.
2. Посилання запису на сервер для оброки і перетворення сигналу.
3. Отримання результатів перетворення звукового сигналу.

У першому випадку такий метод буде корисний при зберіганні записаних команд на пристрій,.

У другому випадку їх аналізу та оновленню бази даних.

Це дає можливість створити велику базу записів, для того щоб, в подальшому використанні, можна було суттєво пришвидшити роботу відбору з бази вже перетворених записів.

Головним завданням є дослідження та створення методів голосового керування різними електронними пристроями. На основі проведеного аналізу, врахувати слабкі сторони голосового керування для вдосконалення і формулювання покращеного методу. Основною особливістю якого буде вміння до адаптації, зміна поведінки в залежності від умов. Це дасть можливість підвищити якість систем голосового керування.

Виклад основного матеріалу. Існує кілька способів вибору моделі опису характеристик проходження сигналів. Виділяють два основних типи моделей: детерміністичні та стохастичні.

Детерміністичні моделі зазвичай використовують деякі відомі властивості сигналу, наприклад, подання сигналу синусоїдальною хвилею або сумою експонент.

У цьому випадку специфікація моделі досить проста: необхідно лише оцінити параметри сигналу – амплітуду, частоту, фазу і т.д.

Стохастичні моделі намагаються описати лише статистичні властивості сигналу. Прикладами подібних моделей можуть служити гаусові процеси, Пуассонівські процеси, марківські процеси (у тому числі і приховані)[4]. В основі стохастичних моделей лежить

припущення про те, що сигнал може бути добре описаний як параметричний випадковий процес і що його параметри можуть бути оцінені досить точно.

Прихована марківська модель (ПММ) визначається як подвійний випадковий процес. Основа випадкових процесів являє собою однорідний марківський ланцюжок з кінцевим числом станів. Послідовність станів не спостерігається, і тому така модель називається прихованою. Цей ланцюжок станів впливає на інший випадковий процес, який і виробляє послідовність спостережень. Приховані марківські моделі є важливим класом моделей, які успішно використовуються в багатьох галузях знань, у тому числі і при моделюванні мовлення.

Можна виділити наступні переваги використання прихованих марківських моделей при використанні та розв'язуванні задачі розпізнавання мови:

- ПММ володіють простою математичною структурою;
- Структура ПММ дозволяє моделювати складні ланцюжки спостережень.

Традиційно в системах розпізнавання мовлення для визначення меж мовлення використовуються методи, засновані на обчисленні короткочасної енергії сигналу або спектральної енергії. Таким чином розв'язування задачі розбивається на наступні етапи:

1. Виділення із загальної звукової відцифрованої осцилограми тривалістю 2 секунди осцилограми конкретного слова – команди.
2. Розбиття осцилограми на окремі ділянки довжиною ~ 15-23 мілісекунд (довжина команди – слова ~ 0.65-0.90 секунди).
3. Застосування дискретного перетворення Фур'є (ДПФ) [3] до кожної ділянки слова (отримання спектра сигналу на ділянці).
4. Виділення на кожній ділянці *n* - точок локальних максимумів амплітуд з їх значеннями частот (виділення форманти мовного сигналу).
5. Дослідний пошук такого значення *n*, при якому після відновлення звуку за допомогою синусного перетворення Фур'є буде однозначне суб'єктивне "впізнавання" перетвореного слова з набору вибраних 7-ми команд.
6. Формування масиву даних для кожного слова з ділянок слова, які будуть характеризувати конкретне слово.
7. У зв'язку з тим, що одне і те ж слово, повторене навіть одним і тим же диктором, має осцилограми, що відрізняються одна від одної, створюється набір з декількох десятків масивів, характерних для одного і того ж слова.
8. Отримавши окремі набори масивів характерних для конкретних слів, використовується математичний апарат нейронних мереж для розпізнавання конкретного введеного через мікрофон слова. Тут для створення і роботи з нейронною мережею використовується широко розповсюджена бібліотека FANN [2].

На рис. 1 представлена осцилограма слова "зелений", відображена в аудіо редакторі audacity. Тут по осі абсцис представлено час в секундах, по осі ординат - нормоване значення амплітуди.

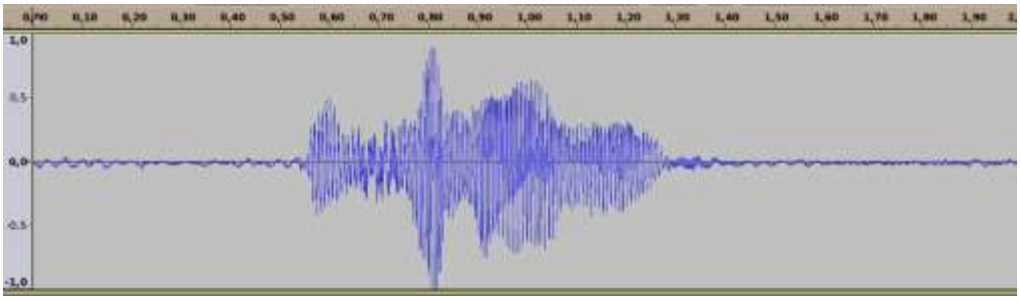

Рис. 1. Осцилограма слова "зелений" в 2-х секундному аудіо файлі a.wav

Для розпізнавання слова системою, необхідно знайти масив чисел, який би однозначно представляв саме це слово. В роботі це зроблено таким чином. Ділимо одне з заданих семи слів на 40 інтервалів. У цьому випадку довжина інтервалу знаходиться в межах 15...23мс. До кожного інтервалу застосовуємо дискретне перетворення Фур'є:

$$X_k = \frac{1}{N}\sum_{n=0}^{N-1} x_n e^{-\frac{2\pi k n}{N}} \qquad , \qquad k = 0,..., \frac{N-1}{2}.$$

Тут:

N - Кількість значень сигналу, виміряних в одному з 40-ка інтервалів;

X_π - виміряні значення сигналу в π-й точці інтервалу;

X_k - комплексна амплітуда;

k-й синусоїдальний сигнал (k-й індекс частоти на кривій спектру). Індекс k вимірюється від 0 до $(N/2)/2$, так як, друга половина з N комплексних амплітуд, фактично являється дзеркальним відображенням першої та не несе додаткової інформації.

Розкладемо експоненту по формулі Ейлера та отримаємо:

$$X_k = \frac{1}{N}\left[\sum_{n-0}^{N-1} x_n \cos\left(\frac{2\pi k n}{N}\right) - i\sum_{n-0}^{N-1} x_n \sin\left(\frac{2\pi k n}{N}\right)\right];$$

чи:

$$X_n = \frac{1}{N}\left(\operatorname{Re} X_k + \operatorname{Im} X_k\right).$$

Визначення дійсної амплітуди виконується за формулою:

$$A_k = \left(1/N\right)\sqrt{\operatorname{Re} X_k^2 + \operatorname{Im} X_k^2}$$

частоти:

$$Fr_k = \left(1/T\right)k.$$

Тут T - період часу, протягом якого бралися вхідні дані (тривалість одного із 40-ка інтервалів).

Геометрична інтерпретація дискретного перетворення Фур'є на рис. 2.

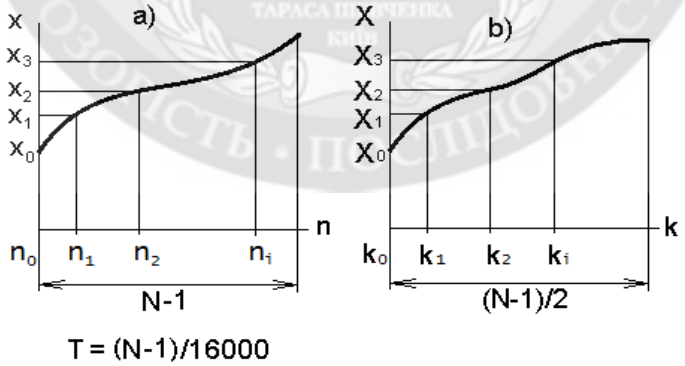

Рис. 2. Геометрична інтерпретація дискретного перетворення Фур'є
а) - осцилограма інтервалу (n - індекс часу), б) - спектрограма (k - індекс частоти)

Крім математичних моделей , для створення голосового інтерфейсу використовуються бібліотеки, створені колективами Google та Yandex. Робота полягає в тому, щоб використовуючи програмну модель звертатися до серверів Google та Yandex, які після

розпізнавання мови повертають набір розпізнаних символів. Саме такі бібліотеки були обрані для виконання даного завдання та створення програми для голосового керування віддаленими пристроями.

Одним з модулів, призначених для розпізнавання голосових команд є Voice recognition module V3.1 (рис. 3). Модуль являє собою невелику плату, на якій розташовані головний мікроконтролер, роз'єм jack 3,5 мм для підключення мікрофона, мікросхема flash пам'яті, контакти GPIO, UART і живлення, пара світлодіодів, резистори, конденсатори, кварц (рисунок 4). Для збільшення дальності роботи голосових команд необхідно використовувати мікрофон з підсилювачем. Дотримуючись прийнятної дальності, модуль можна використовувати в системах розумного будинку. Не збільшуючи дальність роботи модуля, його можна використовувати для настільних систем керування, а також в системах охорони (контролю та обмеження доступу). Модуль може працювати без зовнішнього мікроконтролера, завдяки тому, що має функціональну самостійність. Для цього слід записати голосові команди та задати налаштування для самостійної роботи з допомогою зовнішнього пристрою.

Рис. 3. Набір компонентів для роботи Voice recognition module V3.1

Характеристики модуля:

- Напруга джерела – 5 вольт
- Сила струму, що споживається – до 40 мА
- Інтерфейси UART, GPIO
- Точність розпізнавання – 99% (в ідеальних умовах)
- Дальність дії – залежить від мікрофона. Для штатного мікрофона з комплекту дальність становить 0,5 – 1 метр максимальної відстані при достатньо гучному голосі.

Рис. 4. Плата Voice recognition module

Програма була створена в середовищі розробки MIT App Inventor 2. App Inventor є середовищем візуальної розробки android-додатків. Для програмування в App Inventor використовується графічний інтерфейс користувача, візуальна мова програмування.

Трансформація візуальної блокової мови середовища App Inventor у байт-код Android здійснюється компілятором, заснованим на фреймворку GNU. Фреймворк - інфраструктура програмних рішень, що полегшує розробку складних систем.

У вікні перегляду безпосередньо відбувається написання програми. Для цього ми «перетягуємо» з секції блоків потрібні нам блоки, з'єднуємо та прописуємо необхідні команди. У лівому нижньому куті вікна показана інформація про помилки. Вона має дві позначки: жовтий та червоний трикутники зі знаками оклику всередині. Зручним є те, що при виникненні помилки, відповідний значок з'являється на блоці, де вона виникла.

У правому нижньому кутку розміщений «смітник». Він слугує для видалення з тексту програми непотрібних блоків. У верхньому правому кутку знаходиться «портфель». Він дає можливість копіювати блоки з одного проекту в інший, тобто є своєрідним буфером обміну.

Метою роботи є створення голосового інтерфейсу та впровадження його для пристроїв під керуванням ОС Android, тому завдання полягає в тому, щоб – прописати низку голосових команд для керування віддаленим пристроєм.

У якості віддаленого пристрою, використовується контролер Arduino, зв'язок з яким встановлюється через Bluetooth. До контролера приєднані датчик температури та три світлодіоди (як альтернатива трьом механізмам, наприклад світло, телевізор, кондиціонер) (рис. 5).

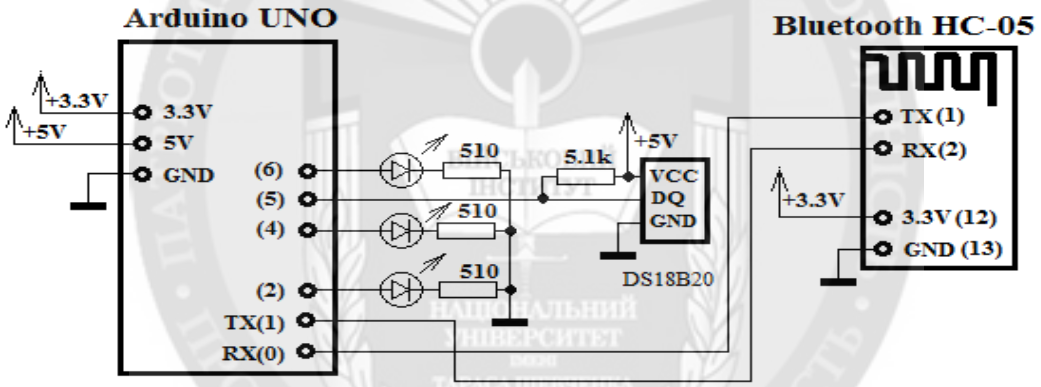

Рис. 5. Схема підключення пристроїв до контролера

Далі розглянемо програму для пристрою під керуванням ОС Android для формування голосового інтерфейсу і взаємодії з віддаленим пристроєм.

Спочатку прописуємо програму підключення до віддаленого пристрою через Bluetooth (рис. 6).

Рис. 6. Фрагмент коду підключення до віддаленого пристрою через Bluetooth

Після звернення користувача до програми, розпізнаний текст має з'явитись у першому полі для тексту. Якщо він співпадає з однією з прописаних команд, виконується певна дія. У протилежному випадку, перевіряються співпадання тексту з іншими командами. Якщо ж співпадання не відбулося, жодна з дій не виконується.

Розглянемо частину програми, яка відображає послідовність дій після голосової команди «включить синий». Якщо розпізнавач тексту отримує команду «включить синий» або «Включить синий» (програма може записати розпізнаний текст як з малої, так і з великої літери, слід розглядати два варіанти), клієнт Bluetooth передає на віддалений пристрій відповідний сигнал для ввімкнення синього світлодіоду. У четвертому полі для тексту (Label 5) з'являється текст «синий включен» і функція перетворення тексту в звук повторює цей текст. Таким чином ми чуємо те, що синій світлодіод увімкнено (рис. 7).

Рис. 7. Фрагмент коду голосового повідомлення “Синий включен”

Цього ж результату (ввімкнення синього світлодіоду) можна досягнути натиснувши на кнопку (рис. 8).

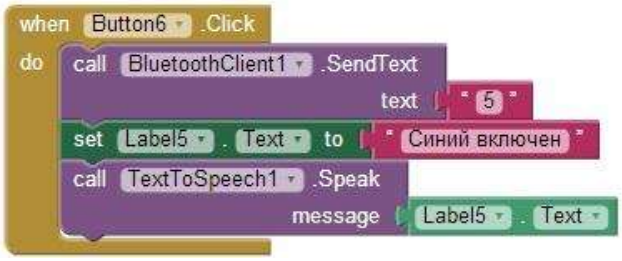

Рис. 8. Фрагмент коду натискання кнопки щоб включити синій

Після проведення всіх вище описаних дій, отримаємо готове до використання ПЗ.

Головне вікно програми (рис. 9).

Рис. 9. Головне вікно програми

Розглянемо функції кнопок та полів для тексту. При натисканні на клавішу «Разговор», користувач має змогу дати програмі голосову команду. Після цього розпізнані слова з'являються у першому полі для тексту (0). Після застосування кнопки «Bluetooth» перед користувачем відкривається список дивайсів, до яких можна приєднатися (рис. 10). Після успішного підключення, надпис «Bluetooth» змінюється на «подключен». Якщо ж підключитися не вдалося, висвічується текст «переподключение», що означає необхідність повторити спробу. Кнопки «включить», «выключить» і «температура» створені для неголосового управління віддаленими пристроями. У другому, третьому та четвертому текстових полях показана інформація про стан червоного, білого та синього світлодіодів відповідно. П'яте поле для тексту показує значення температури у градусах С після натискання кнопки «температура» або голосової команди «температура».

Рис. 10. Спискок можливих підключень до пристроїв

Висновки. Отже, процес побудови системи починається з вирішення певних питань, які ускладнюють створення. Для нормальної роботи системи віддаленого керування електронними пристроями через Інтернет, потрібно вирішити наступні питання: Необхідно

перетворити коливання повітря в електричні сигнали за допомогою мікрофона, відфільтрувавши при цьому перешкоди і шуми; Представлення в цифровій формі, доступній для обробки за допомогою комп'ютера, тобто відцифрувати; Виділити з відцифрованого звуку лінгвістичні конструкції (фонеми), застосувавши різні математичні методи. Таким чином, завдання, пов'язані з розпізнаванням голосу - це складні наукомісткі завдання. Але, якщо професійно вирішити виниклі труднощі, то система буде досить надійною та простою в роботі.

ЛІТЕРАТУРА:

1. Исследование надежности распознавания речи системой GOOGLE VOICE SEARCH / Бобкин Д.В., Жигалов К.Ю.//Gloud of science. – 2015. – № 3. – С. 465-472.

2. Мясищев А.А. Программы распознавания команд с помощью ДПФ и библиотеки FANN. [Electronic resource]. - Mode of access: http://webstm32.sytes.net/obrazec/prog.htm, 2015.

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

4. Степанкова Г.А., Баклан І.В. Побудова гібридних моделей на основі прихованих марківських моделей та нейронних мереж. / Степанкова Г.А., Баклан І.В. // Автоматика. Автоматизация. Электротехнические комплексы и системы. – 2008. – №1.

REFERENCES:

1. Yssledovanye nadezhnosty raspoznavanyya rechy systemoy GOOGLE VOICE SEARCH / Bobkyn D.V., Zhyhalov K.Yu.//Gloud of science. – 2015. – # 3. – S. 465-472.

2. Myasyshchev A.A. Prohrammы raspoznavanyya komand s pomoshch'yu DPF y byblyoteky FANN. [Electronic resource]. - Mode of access: http://webstm32.sytes.net/obrazec/prog.htm, 2015.

3. Dyskretnoe preobrazovanye Fur'e. Vykypedyya. [Electronic resource]. - Mode of access: https://ru.wikipedia.org/wiki/Dyskretnoe_preobrazovanye_Fur'e.

4. Stepankova H.A., Baklan I.V. Pobudova hibrydnykh modeley na osnovi prykhovanykh markivs'kykh modeley ta neyronnykh merezh. / Stepankova H.A., Baklan I.V. // Avtomatyka. Avtomatyzatsyya. Ɔlektrotekhnycheskye kompleksы y systemы. – 2008. – #1

Без рецензії.

д.т.н., проф. Мясищев А.А., к.т.н., доц. Муляр И.В., к.т.н. Лоза В.Н., Козак С.В.

ГОЛОСОВОЕ УПРАВЛЕНИЕ УДАЛЕННЫМИ УСТРОЙСТВАМИ ЧЕРЕЗ СЕТЬ ИНТЕРНЕТ

При работе с техникой, для пользователя в основном важную роль играют такие факторы, как легкость и скорость. Для того чтобы обеспечить пользователя такими возможностями, важной является разработка системы, которая будет работать от голоса пользователя.

Управление системой с помощью голоса, может значительно повысить ее надежность, потому что, при качественной разработке такой системы, ее можно настроить только на тот голос, от которого она будет работать. Это сделает невозможным использование системы кем-то другим, потому что, голос будет идентифицироваться.

В статье описан принцип работы такой системы, которая позволяет пользователю лишь при помощи голоса, управлять электронными устройствами . Выделены задачи, которые нужно выполнить для того чтобы создать стабильную систему. Использование преобразований Фурье для выделения набора основных частот и амплитуды, чтобы проанализировать спектральный состав канала и представить сигнал в цифровой форме. Использование методов статистического анализа: метод скрытого Марковского моделирования, метод динамического программирования и метод нейронных сетей.

Таким образом, описаны все важные моменты и этапы для построения системы управления удаленными электронными устройствами через сеть Интернет с помощью голоса.

Ключевые слова: скрытая модель Маркова, преобразование Фурье, нейронные сети, динамическое программирование.

Prof. Myasischev O.A., Ph.D. Mulyar I.V., Ph.D. Loza V.M., Kozak S.V.

VOICE CONTROL REMOTE DEVICES VIA THE INTERNET

When users working with technique the most significant role play factors such as speed and simplicity. To provide every user the opportunity such as written above it is very important to develope the system which is going to work using the voice of the user.

Control system by voice, can significantly improve its reliability because, while the qualitative development of such a system, it can only be set by the voice it will operate. It will prevent using the system of someone

The article describes the principle of operation of such a system only using your voice to control electronic devices. Selected tasks that must be performed in order to create a stable system. The use of Fourier transforms to highlight a set of core frequency and amplitude to analyze the spectral composition of channel and introduce the signal in digital form. The use of statistical analysis techniques: the method of hidden Markov modeling, dynamic programming method and neural network method.

As follows, described all the important points and stages for building a system of remote electronic devices over the Internet using voice.

Keywords: hidden Markov model, Fourier transformation, neural networks, dynamic programming.