

МЕТОДИ ОБРОБКИ МОВНИХ КОМАНД СИСТЕМ ГОЛОСОВОГО КЕРУВАННЯ

В роботі сформовані основні задачі систем обробки мовних команд. Здійснена класифікація систем голосового керування. Проведено аналітичний огляд алгоритмів та етапів обробки мовних команд. Розроблено структурну та функціональну схеми електричної системи голосового керування комплексу "Smart home" за бездротовим протоколом передачі даних Z-Wave. Програмно-апаратно реалізована система голосового керування на базі модуля обробки, модуля Z-Wave, модуля візуалізації, модуля реєстрації, модуля живлення і модуля аудіо з відповідними колами комутації та регулювання. Розробка актуальна з огляду на популярність даних систем, а особливо для людей з обмеженими можливостями. Голосове керування в даному випадку розглядається як основний інструмент керування, який вкрай необхідний через відсутність можливості у людей з обмеженими можливостями впливати на звичайні органи керування.

Ключові слова: мовні команди, голосове керування, система, фільтрація, шум, мікроконтролер, мікрокомп'ютер, "розумний будинок".

V.I. STETSUK, V.V. KOVALENKO

Khmelnytskyi National University

METHODS FOR PROCESSING SPEECH COMMANDS OF VOICE CONTROL SYSTEMS

For a long time, the process of human interaction with software electronic hardware was purely tactile. The invention of the graphical interface has led to the proliferation of personal computers and other gadgets, but people have always sought a more universal and natural way of communication - voice control. Voice control is based on speech recognition technology, which is the processing of speech signals to filter and suppress noise, segmentation into informative areas, determine informative parameters that reflect the unique properties of speech and recognition. Language technologies are a key factor in the development of an automated human environment, ranging from the improvement of working and research processes and ending with the field of personal application of modern technologies. Despite the high speed of development of computers and information technologies, the main problems of voice control systems remain relevant. The main tasks of language command processing systems are formed in the work. Classification of voice control systems. An analytical review of algorithms and stages of processing language commands. Structural and functional diagrams of the electric control systems of the "Smart home" complex over the Z-Wave wireless data communication protocol have been developed. The hardware-based system of voice control is based on the processing module, the Z-Wave module, the visualization module, the registration module, the power module and the audio module with the respective switching and control circuits. The development is relevant in view of the popularity of these systems, especially for people with disabilities. Voice control in this case is considered as the main management tool, which is extremely necessary due to the lack of ability for people with disabilities to influence conventional controls.

Keywords: voice commands, voice control, system, filtering, noise, microcontroller, microcomputer, smart home.

Вступ

Мова являє собою найважливіший засіб людського спілкування. За свою здатність передавати інформацію голосом людина і виділяється як розумна, високорозвинута істота. Серед величезної кількості живих істот, що населяють наш світ, тільки людина виявилася здатною розвинути існуючий в рудиментарному стані голосовий апарат для кодування і передачі інформації. Довгий час процес взаємодії людини з програмними електронними технічними засобами був доступний тільки вузьким фахівцям, які спілкувалися з машиною за допомогою діалогового інтерфейсу. Винахід графічного інтерфейсу призвів до поширення персональних комп'ютерів та інших гаджетів, однак людина завжди прагнула до більш універсального і природного способу взаємодії – голосового керування. Голосове керування ґрунтується на технології розпізнавання мови, що представляє собою обробку мовних сигналів з метою фільтрації і пригнічення шуму, сегментації на інформативні ділянки, визначення інформативних параметрів, що відображають унікальні властивості мови і розпізнаванні. Коректне розпізнавання мови – одна із найцікавіших і складних завдань голосового керування. Мовні технології є ключовим фактором у розвитку автоматизованого оточення людини, починаючи від вдосконалення робочих і дослідних процесів і закінчуючи областю персонального застосування сучасних технологій. Робота переважної більшості мовних додатків неможлива без здійснення попередньої часової сегментації мови, тобто поділу мовного сигналу на квазістаціонарні за певними характеристиками часові фрагменти. Залежно від того, яка задача стоїть перед конкретним мовним додатком, застосованого методу рішення і умов роботи, необхідний рівень сегментації мовного сигналу буде відрізнятися. Це породжує велику різноманітність методів обробки мовних сигналів, в тому числі і сегментації. Незважаючи на високу швидкість розвитку обчислювальної техніки та інформаційних технологій, основні проблеми мовних додатків до сих пір залишаються актуальними. Основною причиною є складність структури мовного сигналу: величезна різноманітність фонетичних одиниць мови, інтонаційних барв, особистісних особливостей мовця. Слід врахувати також різноманітність зовнішніх факторів, що впливають на запис і передачу голосу. Крім того, ситуація ускладнюється великою кількістю мов світу (порядку 7000 мов). Тому продукти які виникають на ринку обмежені буквально кількома спорідненими мовами і взагалі не працюють з іншими типами мов. Внаслідок чого мовні сигнали досить складно детально дослідити і описувати за допомогою математичних моделей. Показовим є фактична

відсутність систем розпізнавання української мови з великим словником. Перераховані фактори визначають і основні недоліки існуючих алгоритмів часової сегментації мовних сигналів: недостатня точність визначення меж сегментів, висока ресурсомісткість, значне погіршення роботи за наявності шумів. Через низьку ефективність традиційних методів обробки нелінійних і нестационарних сигналів, використання існуючих алгоритмів і вузлів в спеціалізованих системах голосового керування стають малоєфективними. Виходячи з вищевикладеного, можна зробити висновок про актуальність вдосконалення наявних та створення нових підходів до вирішення завдання підвищення ефективності систем голосового керування шляхом розробки нових алгоритмів і вузлів, підвищення ефективності фільтрації, сегментації, виділення інформативних параметрів і розпізнавання мовних сигналів системи голосового керування.

Основна частина

Мовні технології дозволяють створювати інтуїтивно зрозумілі, легкодоступні і швидкі в застосуванні інтерфейси для «спілкування» людини з комп'ютеризованим технічним оточенням. Перелік актуальних додатків, які можуть бути реалізовані за рахунок мовних технологій досить великий. Це може бути голосовий помічник для електронних і аудіо книг, детектор телефонних номерів та електронних адрес, протоколювання нарад, в тому числі з ідентифікацією поточного оратора, голосовий пошук інформації, голосова навігація (синтез і розпізнавання), системи типу «розумний будинок» тощо [1]. Основні задачі систем обробки мовних команд наведені на рис. 1.

Дослідження [2] показали, що коливання голосових зв'язок характеризуються нерегулярністю, яка проявляється у вигляді значних змін тривалості періодів основного тону (на 10–30 %) і у вигляді невеликих флуктуацій сусідніх періодів тону. Нерегулярності виникають через неповне змикання голосових зв'язок на початку і в кінці тональних ділянок, при цьому періоди з неповним змиканням голосових зв'язок чергуються з періодами з повним змиканням. Розподіл частот основного фону є нормальним в логарифмічному масштабі і описується наступними формулами для чоловічого і жіночого голосів відповідно:

$$p_{ч} = 2,26T_0 \exp \left[-84,5 \left(\lg \frac{137}{T_0} \right)^2 \right] \quad (1)$$

$$p_{ж} = 3,16T_0 \exp \left[-166 \left(\lg \frac{247}{T_0} \right)^2 \right] \quad (2)$$

де T_0 – період основного тону, $p_{ч}$, $p_{ж}$ – розподіл частот основного тону для чоловічого і жіночого голосів відповідно.

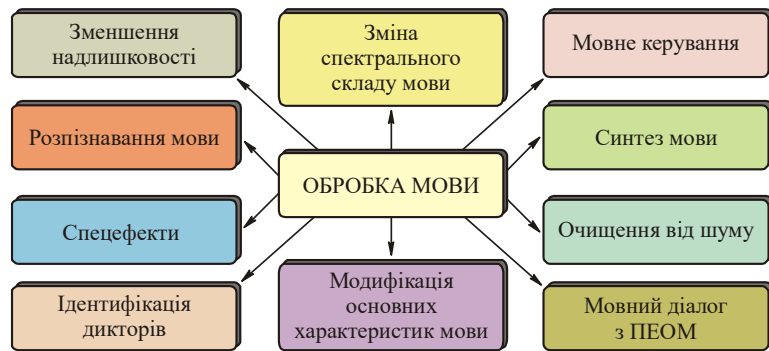


Рис. 1. Основні задачі систем обробки мовних команд

Голосове керування це спосіб взаємодії з керуючою системою за допомогою мовних команд. Воно зводиться до задачі розпізнавання, проте на відміну від систем розпізнавання мови, голосове керування призначене для розпізнавання окремих керівних команд. На рис. 2 представлена класична структурна схема системи голосового керування (СГК), яка складається з мікрофону, блоку обробки і блоку розпізнавання. Мова у вигляді звукових хвиль фіксується мікрофоном, який перетворює їх в аналоговий мовний сигнал. У блоці обробки аналоговий сигнал перетворюється в цифрову форму, проходить фільтрацію, сегментується на фрагменти, в яких відбувається виділення інформативних параметрів для подальшого аналізу. У блоці розпізнавання відбувається безпосереднє порівняння мовної команди, що надійшла в систему із шаблоном з бази даних, отриманим в ході навчання системи. Зазвичай база даних шаблонів формується на етапі розробки системи, може доповнюватися і коректуватися в подальшому при експлуатації під конкретного користувача. Процес створення бази даних шаблонів часто проходить в інтерактивному режимі для навчання системи.

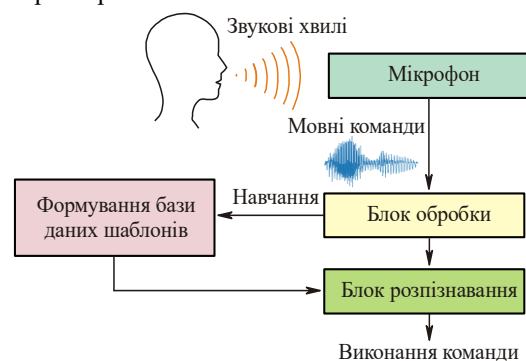


Рис. 2. Класична структурна схема СГК

Розпізнавання окремих мовних команд простіше, ніж розпізнавання злитого мовлення і не вимагає великих обчислювальних потужностей. Саме з цієї причини на сьогодні існує величезний вибір програмного і апаратного забезпечення із задовільною якістю розпізнавання. Однак дослідження показує, що системи досі не подолали рівень точності розпізнавання в 85 %, тоді як у людини цей показник становить 96–98 %.

Для повної оцінки сучасного стану систем голосового керування (СГК), на рис. 3 наведено класифікацію за такими основними параметрами: технічне виконання, призначення, персоналізація, тип мовлення, тип структурної одиниці і розмір словника. Таким чином, системи голосового керування можна класифікувати за параметрами, що описують споживчі властивості систем та за параметрами технічної реалізації систем. Системи з технічної реалізації включають в себе програмні продукти і програмно-апаратні засоби.

Найчастіше задачею зводиться до аналізу ізольованого слова або словосполучення з чітко вираженими межами (мовної команди), що включає в себе наступні завдання: реєстрація, попередня корекція, фільтрація, сегментація сигналу / пауза, визначення інформативних параметрів і розпізнавання.

На рис. 4 представлені етапи обробки мовних команд в системі голосового керування, що виконує перераховані вище завдання.

Реєстрація являє собою запис мовної команди в режимі реального часу, формування електричного аналогового сигналу та перетворення його у цифрову форму.

Попередня корекція використовується для усунення природних спотворень спектру звуку (мінус 6 дБ на октаву), що виникають в мовному апараті людини при розмові. Сигнал мовної команди пропускають через корегуючий фільтр з функцією передачі:

$$W(z) = \sum_{k=0}^m a_k z^{-k}, \quad (3)$$

де a_k – постійні коефіцієнти, m – ціле число ($m > 0$), k – номер коефіцієнта. Найчастіше $m=1$, а передавальна функція має вигляд:

$$W(z) = a_0 - a_1 z^{-1} \quad (4)$$

Фільтрація дозволяє підвищити розбірливість і зменшити частку шумів, викликаних як акустичними, так і технологічними причинами. Стосовно мовних сигналів, шум – це сукупність аперіодичних звуків різної інтенсивності і частоти, які змінюють інформативні ознаки сигналу. За взаємодією з корисним мовним сигналом шуми поділяються на адитивні та мультиплікативні. Адитивні шуми складаються з корисним сигналом і вносять незначну похибку. Мультиплікативні шуми перемножуються з корисним сигналом, вносять найбільшу похибку і можуть змінювати інформативні параметри мовних команд. У загальному вигляді, комбінація сигналу і шуму, виглядає наступним чином:

$$S(t) = [(k_s(t) + k_n(t)) \cdot e(t) + n(t)], \quad (5)$$

де $S(t)$ – мовний сигнал, $e(t)$ – корисний мовний сигнал; $k_s(t)$ – коефіцієнт, що характеризує корисний мовний сигнал; $k_n(t)$ – коефіцієнт, що характеризує мультиплікативний шум; $n(t)$ адитивний шум.

Сегментація в обробці мовних команд – лінійний розподіл мовного потоку на складові відрізки (фрагменти). Сегментація сигналу / пауза представляє собою завдання визначення моментів початку і закінчення фрази. При наявності шуму дана задача є однією з найважливіших в області обробки мовних команд, так при голосовому керуванні важливо точно визначити моменти початку і закінчення команди.

Визначення інформаційних параметрів – завдання визначення унікальних властивостей і характеристик мовних сигналів. Основні поняття, що характеризують інформативні параметри мовлення людини, пов'язані з формою, розмірами, динамікою зміни мовного апарату і описуються емоційним станом людини. Розпізнавання мовних команд є кінцевою метою обробки в системах голосового керування. Принцип роботи досить простий і полягає в порівнянні мовної команди, яка надійшла в систему, з шаблоном із бази даних, отриманою в ході навчання системи. Головна особливість, що впливає на якість розпізнавання полягає в застосовуваних методах порівняння.

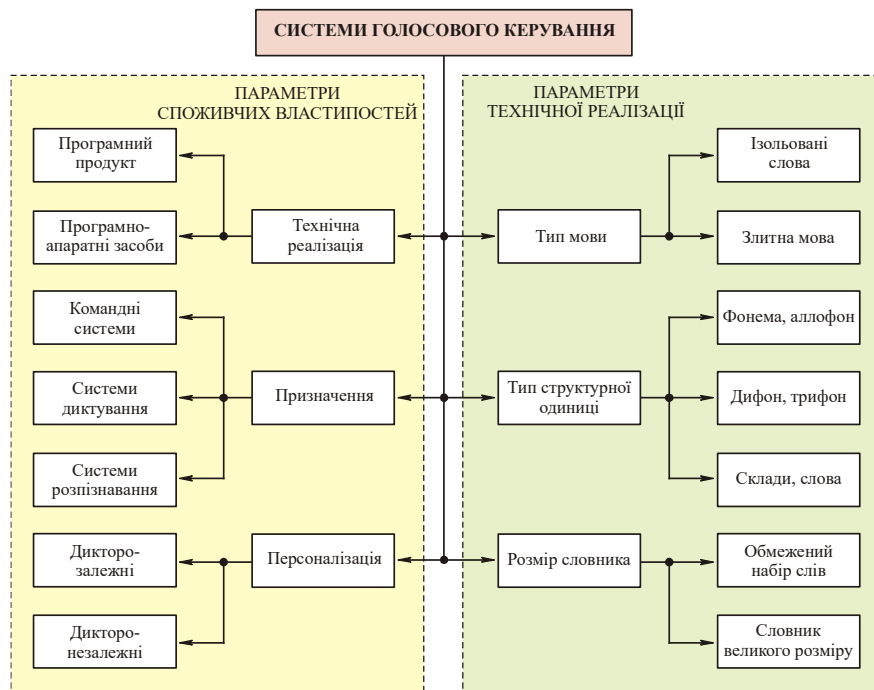


Рис. 3. Класифікація систем голосового керування

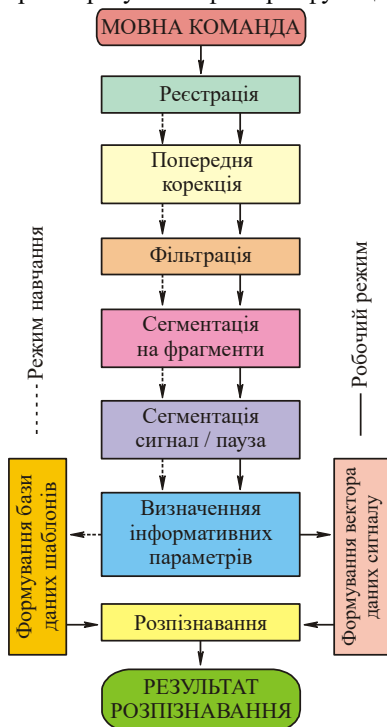


Рис. 4. Етапи обробки мовних команд в SGK

На рис. 5 наведено класифікацію методів обробки в системах голосового керування.

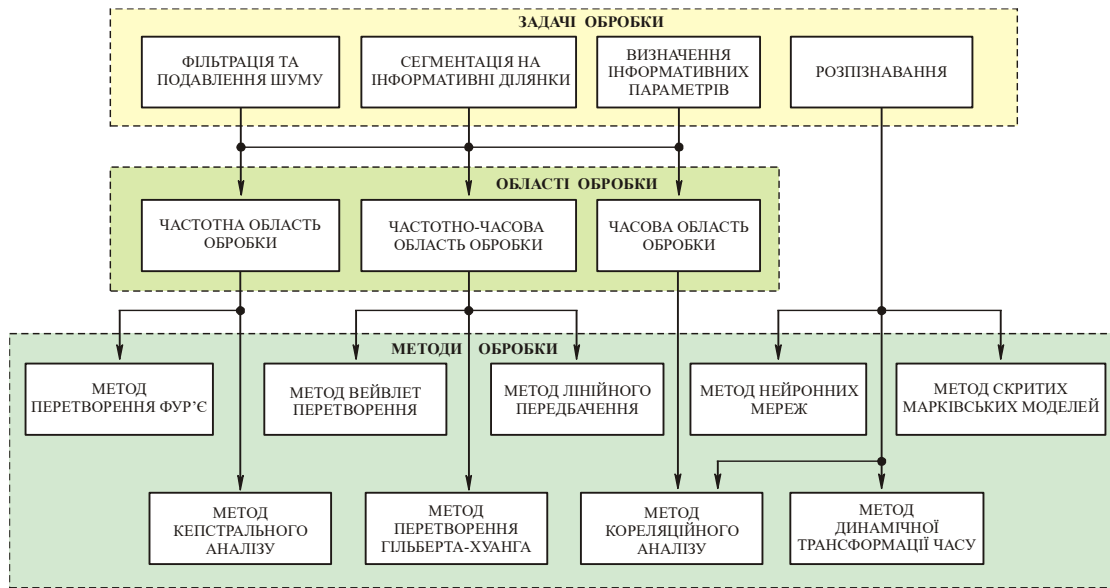


Рис. 5. Класифікація методів обробки мовних сигналів

Класифікація методів виконана з урахуванням основних завдань обробки в підсистемах голосового керування: фільтрація і пригнічення шуму, сегментація на інформативні ділянки, визначення інформативних параметрів і розпізнавання. Області обробки сигналів можна розділити на: частотну, часову і частотно-часову [3]. В залежності від обробки, методи слід розділити на групи, реалізовані різними видами аналізу:

- метод перетворення Фур'є (ПФ);
- метод вейвлет-перетворення (ВП);

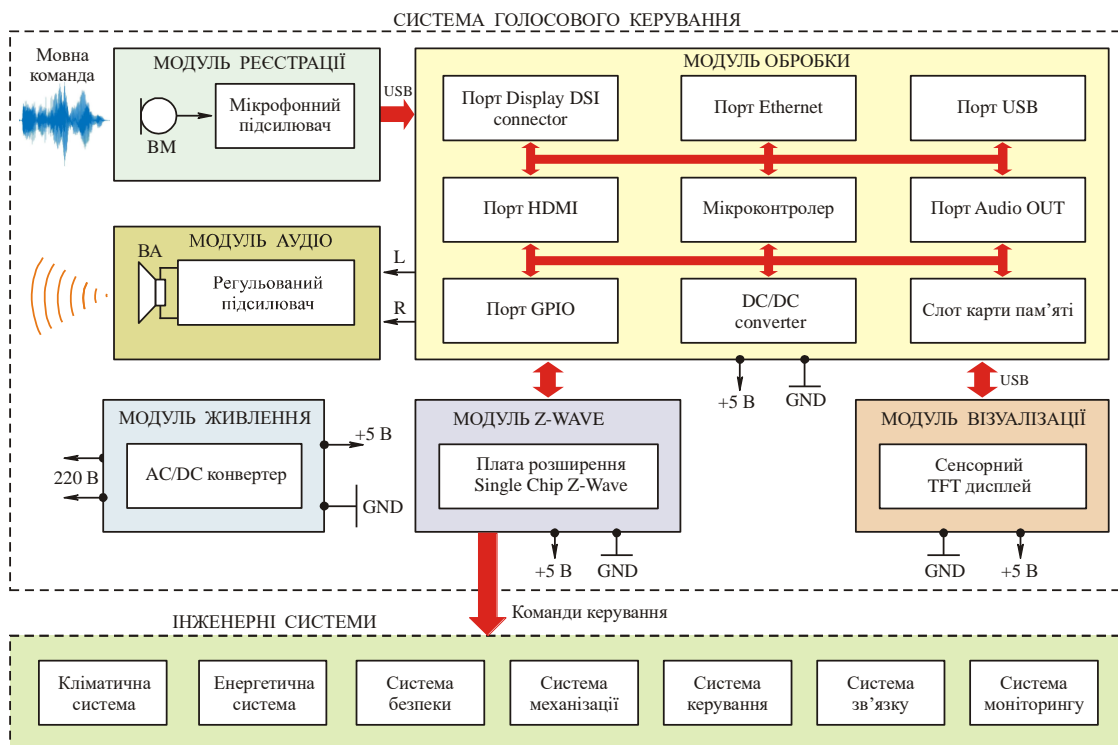


Рис. 6. Структурна схема системи голосового керування комплексу "Smart home"

- метод перетворення Гільберта-Хуанга (ПГС);
- метод кепстрального аналізу (КА);
- метод лінійного передбачення (ЛП);
- метод кореляційного аналізу (КРА);
- метод нейронних мереж (НМ) ;
- метод прихованих марківських моделей (ПММ);
- метод динамічного трансформування часу (ДТЧ).

На основі проведеного аналізу, була програмно-апаратно реалізована системи голосового керування за технологією “Smart home” (рис. 6, 7). Вона складається з наступних модулів: реєстрації, обробки, аудіо, живлення, Z-Wave та візуалізації.

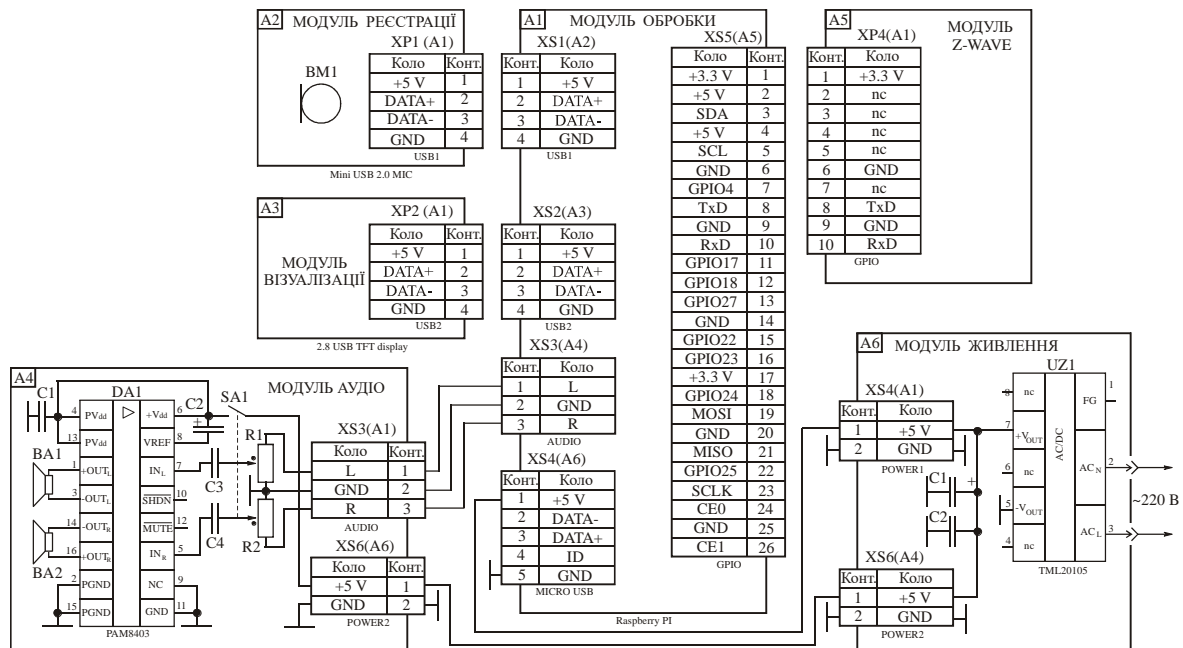


Рис. 7. Схема електрична функціональна системи голосового керування комплексу “Smart home”

Реєстрація сигналів мовних команд функціонально реалізована за допомогою модуля реєстрації, що складається з USB мікрофону. Шумозниження мікрофонного каналу реалізовано програмно.

В якості модуля обробки можна використати будь-яку сучасну платформу – STM32, Arduino, BeagleBone Black, Raspberry Pi, Orange Pi тощо. В даному випадку застосовано одноплатний мікрокомп'ютер Raspberry Pi на основі мікрочіпу Broadcom BCM2836 (рис. 8). Мікроконтролер спільно з додатковим периферійним обладнанням функціонально реалізований у вигляді модуля обробки на борту Raspberry Pi. Периферійне обладнання модуля відповідно до функціоналу забезпечує нормальну працездатність системах голосового керування.

Режими роботи системи голосового керування передбачають попереднє налаштування і навчання. Для цього передбачено модуль візуалізації, реалізований у вигляді 2.8" USB TFT display for Raspberry Pi (рис. 9). Попереднє налаштування здійснюється з використанням сенсорного TFT дисплею для введення і відображення інформації про налаштування. Також даний інтерфейс зручний в якості інформаційної системи. Як і модуль реєстрації, так і модуль візуалізації мають USB інтерфейс, що спрощує як самі відповідні тракти побудови, так і їх комутацію.



Рис. 8. Зовнішній вигляд платформи Raspberry Pi разом із встановленим модулем Z-Wave RaZberry



Рис. 9. Зовнішній вигляд 2.8" USB TFT display for Raspberry Pi

Контроль режимів роботи системи реалізується не тільки візуально, але і за допомогою модуля аудіо обробки, що здійснює звукову індикацію режимів роботи та поточні сповіщення. Ця опція особливо важлива для людей з обмеженими можливостями і дозволяє обрати будь-який зручний спосіб спілкування із системою “Smart home”.

Система голосового керування передбачає керування інженерними підсистемами – клімат-комфорту, енергетичними блоками, засобами механізації та автоматизації, тощо. Для забезпечення бездротового зв'язку використовується модуль Z-Wave у вигляді додаткової плати розширення RaZberry, яка встановлюється безпосередньо у верхню частину слоту GPIO Raspberry Pi (рис. 7, 8).

Висновки

В роботі сформовані основні задачі систем обробки мовних команд, здійснена класифікація систем голосового керування, проведено аналітичний огляд алгоритмів та етапів обробки мовних команд. Розроблено структурну та функціональну схеми електричні системи голосового керування комплексу “Smart home” по бездротовому протоколу передачі даних Z-Wave. Програмно-апаратно реалізована система голосового керування на базі модуля обробки, модуля Z-Wave, модуля візуалізації, модуля реєстрації, модуля живлення і модуля аудіо з відповідними колами комутації та регулювання. Розробка актуальна з огляду на популярність даних систем, а особливо для людей з обмеженими можливостями. Голосове керування в даному випадку розглядається як основний інструмент керування, який вкрай необхідний через відсутність можливості у людей з обмеженими можливостями впливати на звичайні органи керування.

Література

1. Smart Home. Your Trusted Advisor in Home Automation [Електронний ресурс]. – 2019. – Режим доступу : <https://www.smarthome.com/>.
2. Алимуратов А. К. Обработка речевых команд в системах голосового управления / А. К. Алимуратов // Измерение. Мониторинг. Управление. Контроль. – 2014. – № 1 (7). – С. 50–57.
3. Стецюк В. І. Системи автоматизованого управління телекомунікаційних мереж / В. І. Стецюк, В. В. Мішан // Вісник Хмельницького національного університету. Серія: технічні науки. – 2018. – № 6. Том 2. – С. 178–182.

References

1. Smart Home. Your Trusted Advisor in Home Automation [Elektronnyi resurs]. – 2019. – Rezhym dostupu : <https://www.smarthome.com/>.
2. Alimuradov A. K. Obrabotka rechevyh komand v sistemah golosovogo upravleniya / A. K. Alimuradov // Izmerenie. Monitoring. Upravlenie. Kontrol. – 2014. – № 1 (7). – S. 50–57.
3. Stetsiuk V. I. Systemy avtomatyzovanoho upravlinnia telekomunikatsiinykh merezh / V. I. Stetsiuk, V. V. Mishan // Visnyk Khmelnytskoho natsionalnoho universytetu. Serii: tekhnichni nauky. – 2018. – № 6. Tom 2. – S. 178–182.

Рецензія/Peer review : 27.10.2019 р.

Надрукована/Printed : 02.01.2020

Стаття рецензована редакційною колегією