

Жуковський В.Р.

Національний університет «Одеська політехніка»

АНАЛІЗ КОРИСНОСТІ РОЗРОБКИ КОМП'ЮТЕРНО-ІНТЕГРОВАНОЇ СИСТЕМИ ГОЛОСОВОГО УПРАВЛІННЯ БАЗОВИМИ ФУНКЦІЯМИ КОМП'ЮТЕРА ТА ПРОБЛЕМИ ВИКОРИСТАННЯ ГОЛОСОВИХ ПОМІЧНИКІВ

Аналіз публікацій наукових досліджень і розробок технологій персональних голосових асистентів дає змогу виявити тенденцію до появи широкого спектра голосових помічників із різноманітними функціями. Уперше це реалізували Apple у далекому 2011 році, представивши свого голосового помічника Siri, пізніше, у 2015 році, Microsoft випускає голосового помічника Cortana. У 2018 році Яндекс випустив розумну колонку Яндекс, яка працює з голосовою помічницею Алісою, а Amazon представила свою голосову помічницю Alexa. В результаті вивчення історії створення та розвитку технології голосового помічника було зроблено припущення про появу в найближчому часі значного числа різноманітних голосових асистентів. Дійсно, подібна тенденція існує, і вона підтверджена в концептуальних дослідженнях проблеми [3, с. 165].

Концепція, викладена В.І. Городецьким, про те, що кожна річ рано чи пізно повинна стати «розумною», а користувач – отримати свого асистента, який допомагає вирішувати проблеми реального життя в реальному часі, лежить в основі обґрунтування практичного значення цього дослідження. Будь-яка річ, підкреслює В.І. Городецький, рано чи пізно повинна буде володіти не тільки датчиками та пристроями впливу на об'єкти зовнішнього світу й комунікації, а й приймати рішення, причому по можливості узгоджено з іншими речами [4, с. 16].

У такого підходу є важливий у теоретичному і практичному плані наслідок – реалізація, на наш погляд, передбачає розробку нових, більш ефективних поколінь голосових асистентів. Водночас необхідно буде враховувати економічні, психологічні, фізіологічні та соціальні аспекти поведінки людини для створення технологій голосових асистентів дійсно персоналізованих, максимально безпечних, зручних у використанні [1, с. 102].

У статті розглядається тема використання голосових асистентів. Описано основні проблеми, пов'язані з використанням таких помічників. Досліджено способи їх вирішення. Описано план розробки програмного забезпечення «голосовий помічник»

Ключові слова: голосовий помічник, розпізнавання мови, інтернет, програмне забезпечення, управління, автоматизація.

Постановка проблеми. Технологія voice assistant розглядається в науковій літературі як перспективний напрямок організації роботи користувачів з системою сучасних пристроїв. Персональні голосові асистенти створюють для користувача можливість комфортного взаємодії з цифровими пристроями, наближаючи його до комунікації зі співрозмовником. Виділяючи серед основних тенденцій розвитку технологій голосових асистентів появу на ринку значної кількості розробок, зазначу також тенденцію до дедалі більшої персоналізації та створення все більш простих алгоритмів роботи з асистентами. Налаштування помічника відповідно до смаків і побажань користувача дає змогу складати індивідуальний список «гарячих» команд, кодових слів для активації, персоналізувати голос асистента, приємний

для конкретного споживача. Високий потенціал розвитку технологій персональних голосових помічників зумовлений і тим, що середня швидкість вимови слів людиною – 150 слів за хвилину, значно більше, ніж середня швидкість набору слів на пристрої – близько 40 слів [1, с. 49].

За прогнозами Adweek, до кінця 2022 року ринок платформ розпізнавання голосу досягне 40 млрд доларів. Голосові помічники допомагають у міжособистісній комунікації, роблячи спілкування більш комфортним, враховуючи той факт, що набирати текст на пристрої значно складніше й довше, ніж говорити. Не можна не помічати і психологічну складову в цьому питанні. Усе більше людей починають свій ранок з діалогу з голосовим помічником, виникає нова звичка, формуються нові навички. Оскільки ця тема активно

досліджується і розвивається, актуальність дослідження постійно зростає.

Постановка завдання. Мета статті – проаналізувати проблеми, пов’язані з використанням голосових помічників та корисність створення голосових асистентів.

Виклад основного матеріалу. Інтернет речей – це концепція, що з’явилася на початку 2000-х років і з того часу надійно закріпилася серед найбільш популярних напрямків в області інформаційних технологій. У її основі лежить ідея про обчислювальні мережі пристроїв, здатних взаємодіяти один з одним і з зовнішнім світом [2, с. 22].

Рішення інтернету речей пропонують різні способи взаємодії користувача і системи пристроїв. Успішні розробки в області розпізнавання й синтезу мови сприяють появі нового виду людиномашинних інтерфейсів: голосових асистентів, здатних вести з користувачем діалог на природній мові і виконувати різні голосові команди.

За даними літературних джерел, візуальний канал дає 80–90% всієї інформації [5, с. 62], а другий найбільш важливий, слуховий, практично не задіяний у виробництві. Наприклад, якщо колега стоїть у майстерні поряд із бажаним дисплеєм, набагато простіше запитати його, ніж іти й шукати потрібний параметр самостійно. Крім того, оператори часто глибоко залучені в робочий процес, і голосове управління залишається єдиним можливим способом для взаємодії з інформаційною системою.

Наразі можливості голосових помічників значно розширені – від простого пошуку інформації до повного голосового управління різними побутовими пристроями. Звичайно, використання голосових помічників у виробничому середовищі має свої нюанси, такі як шум у виробничих приміщеннях, одночасне взаємодія з декількома операторами і проблеми безпеки.

Голосові помічники використовують у багатьох областях людської діяльності. Однак більшість досліджень застосовні тільки до побутового використання, наприклад, установка нагадування про важливу дату або час чи установка кінцевого пункту маршруту. Навпаки, практично немає досліджень із використання цієї технології у виробничій сфері, незважаючи на те, що ідея отримання інформації таким способом може бути успішним рішенням і для автоматизованого виробництва, де, наприклад, замість пошуку потрібного графічного дисплея візуально швидше використовувати голос. [6, с. 69].

Принцип роботи голосових помічників такий само простий, як і їх використання:

1. Пасивне зчитування звуку за допомогою активації функції із вбудованим кодовим словом;

2. Фільтрація сигналу – етап усунення шуму і перешкод, що виникають під час запису голосового запиту;

3. Оцифровка звуку – відбувається перетворення звукового сигналу в цифровий вигляд, зрозумілий комп’ютеру;

4. Аналіз сигналу – виділяються ділянки з мовленням, відбувається оцінка параметрів, таких, як частина мови, форма слова, зв’язок в один запит;

5. Пошук шаблонних даних – штучний інтелект збирає різні вимови слова, порівнює з шаблонами і видає результат.

Голосовий помічник можна написати практично за допомогою будь-якої мови програмування, але найбільш популярним у 2021 році є Python. Для написання програмного коду знадобиться знання мови, вміння працювати з підключенням різних бібліотек і чітке формулювання вхідних і вихідних даних.

Е.А. Артозей у своїй статті «Розробка алгоритмів розпізнавання голосових команд для управління бортовою системою автомобіля» пише, що «для візуального представлення процесів введення і подальшого розпізнавання голосових команд системою голосового управління використовується діаграма діяльності підсистеми ідентифікації. Діяльність підсистеми бере початок із введення голосової команди. Отриманий сигнал набуває цифрового вигляду та проходить процедуру фільтрації зовнішніх шумів. Далі перетворений сигнал потрапляє в підсистему ідентифікації. У названій підсистемі для розпізнавання команди спершу подається запит до бази даних сигналів. За умови збігу вхідного і збереженого сигналів розпізнавання проходить успішно, і команда передається на виконавчий пристрій, яке виконує певну дію. Якщо голосова команда не розпізнана, то система повертається до початку – введення голосової команди, і алгоритм дій повторюється знову, поки не буде досягнутий позитивний результат, тобто доти, поки голосова команда не розпізнається» [12, с. 99].

Однак існує ряд проблем, пов’язаних із використанням голосових помічників.

Присутність шуму на тлі мови диктора є одним із найскладніших випадків для систем автоматичного розпізнавання мови (АРР) через надмірну нестационарність перешкоди. Повною мірою впоратися з завданням якісного придушення шуму музики нездатні як класичні одноканальні методи (наприклад, фільтр Вінера), так і адаптивні багатоканальні алгоритми націлювання, такі як алго-

ритм мінімуму дисперсії шуму MVDR (Minimum Variance Distortionless Response). Але все-таки існують способи вирішення цієї проблеми. Підвищення якості АРР в умовах шуму музики можливо досягти, застосовуючи акустичні моделі, реалізовані на штучних нейронних мережах із застосуванням методів глибокого навчання. Для поліпшення мови на багатоканальних записах за останні кілька років було запропоновано низку підходів підвищення робастності алгоритмів націлювання за допомогою оцінювання їхніх параметрів акустичними моделями, зокрема оцінки бінарної шумової маски [6, с. 54], реалізації детектора мови [7, с. 88], оцінки коваріантної матриці [8, с. 19] і коефіцієнтів алгоритму націлювання безпосередньо [9, с. 55; 10, с. 87], а також метод оцінки реальної шумової маски і підрахунку коваріантної матриці для алгоритму MVDR. [11, с. 3]

Як було зазначено вище, існують і інші проблеми голосових помічників. Натепер існує велика різноманітність комерційних систем автоматичного розпізнавання мови: Google Cloud Speech-to-Text1, Amazon Transcribe, IBM Speech to Text, Yandex SpeechKit. Вказані системи вирішують велику кількість завдань і звільняють розробника від необхідності створювати таку комплексну і складну систему, як система автоматичного розпізнавання мови. Наприклад, розпізнавання мови за допомогою перевірених часом алгоритмів, що демонструють високі результати, можна реалізувати з використанням високопродуктивних обчислювальних ресурсів хмарних систем. Хмарні платформи надають стійкий програмний інтерфейс програми (API, Application Programming Interface) для розпізнавання мови, а також мають екосистеми з великою кількістю користувачів, куди легко вбудовувати власні додатки з голосовим інтерфейсом. Крім безперечних переваг, згадані вище системи мають і недоліки, а саме:

1. Перераховані системи є хмарними системами, тому робота з ними залежить від їхньої доступності, пропускнуої спроможності та інших факторів, що впливають на швидкість розпізнавання.

2. Немає можливості налаштувати системи розпізнавання мови для вирішення конкретного завдання. Якість розпізнавання залежить від використовуваної мовної моделі. У різних прикладних областях різна ймовірність певних слів. Стандартні системи використовують усереднену модель мови, або модель, розроблену для вирішення проблем, поставлених перед творцями платформи, що далеко не завжди збігається із завданнями користувачів системи.

3. Зазвичай робота з хмарними системами вимагає додаткових фінансових витрат, заданих правилами використання таких систем.

4. Ресурсовитратність голосової активації. За постійного запису звуку додатково витрачається інтернет-трафік. Але ці проблеми вирішуються роботою в напрямку створення автономних голосових помічників, що можуть використовуватися не тільки для автоматизованого виробництва.

Після розбору основ пристрою та принципу роботи слід зробити свого примітивного але перспективного помічника. Почнемо з вибору мови програмування. Як було сказано вище, найбільш популярною мовою для розробки голосових асистентів на 2021 г. є Python, тому його і візьму. Далі потрібно встановити всі необхідні бібліотеки.

Для перетворення тексту на мову була встановлена бібліотека pyttsx3. Далі стоїть завдання розпізнавання мови, для цього скористаюся описаною вище системою Google Cloud Speech-to-Text, тобто бібліотекою SpeechRecognition. Також потрібно зчитувати мову з мікрофону, у цьому допоможе бібліотека PyAudio. Тепер є можливість перетворювати текст на мову і навпаки, мову на текст, і, по суті, це все, що потрібно, щоб написати свого голосового асистента.

Таким чином був створений примітивний голосовий асистент, який уміє показувати час, повторювати за нами різні фрази, вмикати радіо і розповідати прості анекдоти. Лістинг коду представлений у репозиторії GitHub (URL: <https://pastebin.com/tPD0g802>).

Як можна зрозуміти зі статті, створити такого асистента нескладно і досить актуально сьогодні, однак мій асистент далекий від ідеалу і вимагає допрацювань функціоналу. Далі планується впровадити нейронну мережу, щоб можна було не тільки давати команди, а й вести бесіду з асистентом, а також розширити функціонал згаданого голосового помічника.

Висновки. На описаному етапі роботи були досліджені і проаналізовані проблеми використання голосових помічників на виробництві. Рішення цих проблем може бути досягнуто різними способами, але часто ці способи потребують додаткових досліджень у конкретному виробництві, що вимагає деяких засобів. Також був створений примітивний голосовий асистент із базовими функціями. Отже, в подальшій роботі необхідно досліджувати рішення, які спростять інтеграцію голосових помічників. Надалі планується розробка навичок для голосового асистента і його оптимізація для найбільш комфортного використання, а також експерименти з ним.

Список літератури:

1. Шыпота Н.А., Маколкіна М.А. Анализ использования технологии voice assistant в сетях связи пятого поколения. *Информационные технологии и телекоммуникации*. 2020. Т. 8. № 3. С. 86–93. DOI 10.31854/2307-1303-2020-8-3-86-93.
2. Gershenfeld N.A. *When Things Start to Think*. New York : Henry Holt and Company, 2000. 224 p.
3. Шыпота Н.А. Развитие коммуникационной инфраструктуры с использованием облачных сервисов программно-конфигурируемых сетей // Актуальні проблеми гуманітарних та природничих наук : V Міжнародна науково-практична конференція. Харків, 2018. С. 165–166.
4. Gorodetsky V. Internet of Agents: From Set of Autonomous Agents to Network Object. *Second International Workshop on Internet of Agents*. 2017. pp. 1–17.
5. Линдгрэн Н. Органы чувств и их электронные аналоги. *Электроника*. 1962. Вып. 35. № 7. С. 22–27.
6. Heymann J., Drude L., Haeb-Umbach R. Neural network based spectral mask estimation for acoustic beamforming. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. Shanghai, China, 2016. P. 196–200. doi: 10.1109/icassp.2016.7471664
7. Higuchi T., Ito N., Yoshioka T., Nakatani T. Robust MVDR beamforming using time-frequency masks for online/offline ASR in noise. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. Shanghai, China, 2016. P. 5210–5214. doi: 10.1109/icassp.2016.7472671
8. Li B., Sainath T.N., Weiss R.J., Wilson K.W., Bacchiani M. Neural network adaptive beamforming for robust multichannel speech recognition. *Proc. INTERSPEECH*. 2016. P. 1976–1980. doi: 10.21437/interspeech.2016-173
9. Yoshioka T. et al. The NTT CHiME-3 system: advances in speech enhancement and recognition for mobile multimicrophone devices. *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding*. Scottsdale, USA, 2015. P. 436–443. doi: 10.1109/asru.2015.7404828
10. Du J. et al. The USTC-iFlyteck system for the CHiME4 challenge. *Proc. 4th Int. Workshop on Processing in Everyday Environments*. 2016.
11. Автоматическое распознавание речи в условиях шума музыки на многоканальных записях с удаленного микрофона / С.С. Астапов, Е.В. Шуранов, А.В. Лаврентьев, В.И. Кабаров. *Научно-технический вестник информационных технологий, механики и оптики*. 2019. Т. 19. № 3. С. 557–559. doi: 10.17586/2226-1494-2019-19-3-557-559
12. Артозей Е.А. Разработка алгоритмов идентификации голосовых команд для управления бортовой системой автомобиля. *Современная техника и технологии*. 2015. № 9. URL: <http://technology.snauka.ru/2015/09/7903>.

Zhukovskiy V.R. ANALYSIS OF THE USEFULNESS OF DEVELOPING A COMPUTER-INTEGRATED VOICE CONTROL SYSTEM FOR BASIC COMPUTER FUNCTIONS AND THE PROBLEM OF USING VOICE ASSISTANTS

The analysis of publications of scientific researches and developments of technologies of personal voice assistants allows to mark a tendency to emergence of a wide range of voice assistants with various functions. This was first implemented by Apple back in 2011 with the introduction of its voice assistant Siri, after in 2015 Microsoft released a voice assistant Cortana. In 2018, Yandex released a smart column Yandex, which works with voice assistant Alice, and Amazon presents its voice assistant Alex. As a result of studying the history of creation and development of voice assistant technology, it was assumed that a significant number of different voice assistants will appear in the near future. Indeed, a similar trend exists, and it is confirmed in conceptual studies of the problem [3, c. 165].

The concept set out by VI Gorodetsky, that every thing sooner or later must become “smart”, and the user – to get his assistant who helps to solve real-life problems in real time, is the basis for substantiating the practical significance of this research. Anything, emphasizes VI Gorodetsky, sooner or later will have not only sensors and devices to influence the objects of the outside world and communication, but also to make decisions, and if possible consistent with other things [4, c. 16].

This approach has an important consequence in theoretical and practical terms - the implementation, in our opinion, involves the development of new, more efficient generations of voice assistants. It will be necessary to simultaneously take into account the economic, psychological, physiological and social aspects of human behavior to create technologies for voice assistants really personalized, safe, easy to use [1, c. 102].

The article considers the use of voice assistants. The main problems associated with the use of such assistants are described. The ways of their solution are investigated. Describes the software development plan “voice assistant”.

Key words: Voice Assistant, Speech Recognition, Internet, Software, Microphone, Control, Automation.