

УДК 004.021

КЕРУВАННЯ ПЕРСОНАЛЬНИМ КОМП'ЮТЕРОМ ЗА ДОПОМОГОЮ ГОЛОСУ

Лисков Сергій

Науковий керівник: канд.ф.-м. наук, доцент Гуртовий Ю.В.

Центральноукраїнський державний педагогічний університет імені

Володимира Винниченка, м. Кропивницький, Україна

Розглядається питання застосування систем розпізнавання мовлення для створення програмного забезпечення, що дозволяє керувати персональним комп'ютером за допомогою голосових команд для людей з обмеженими фізичними можливостями. Проаналізовано основні теоретичні положення про системи розпізнавання мовлення, висвітлено проблеми, які виникають під час використання даних систем в програмному забезпеченні і шляхи їх вирішення та зроблено огляд їх застосування на практиці.

Ключові слова: розпізнавання мовлення, глибинне навчання, голосове керування.

PC managing via voice recognition

S. Lyskov

Scientific supervisor: Candidate of Physics and Mathematics Sciences, Docent Hurtovyi Y.V.

The Volodymyr Vynnychenko Central Ukrainian State Pedagogical University,

Kropyvnytsky, Ukraine

The article examines the use of speech recognition systems to create software that allows you to manage a personal computer using voice commands for people with disabilities. The basic theoretical positions of speech recognition systems are analyzed, problems are encountered during the use of these systems in software and ways of their solution and an overview of their application in practice is made.

Keywords: speech recognition, deep learning, voice control.

Постановка проблеми:

В наш час ритм життя настільки збільшився, що інколи навіть бракує часу на виконання повсякденних справ на комп'ютері в звичний спосіб — за допомогою клавіатури та мишки. На зміну прийшли нові технології, що дозволяють керувати персональним комп'ютером за допомогою голосу. Програмне забезпечення для розпізнавання мовлення було розроблено, щоб забезпечити швидкий спосіб написання на комп'ютері та може допомогти людям з різними обмеженнями. Це корисно для людей з обмеженими

фізичними можливостями, які часто вважають, що друкувати текст важко, болісно або неможливо. Програмне забезпечення для розпізнавання голосу також може допомогти тим, хто має орфографічні труднощі, включаючи користувачів, у яких присутня дислексія, оскільки розпізнані слова завжди правильно написані.

Програми розпізнавання голосу працюють, аналізуючи звуки та перетворюючи їх у текст. Після правильного налаштування, система повинна розпізнавати приблизно 95% інформації, при умові, що користувач чітко вимовляє слова.

Аналіз досліджень і публікацій:

Системи розпізнавання голосу — це обчислювальні системи, які можуть визначати мову людини із загального потоку. Ця технологія пов'язана з технологією розпізнавання мови, яка перетворює вимовлені слова в цифрові текстові сигнали шляхом проведення процесу розпізнавання мови машинами.

Перший пристрій для розпізнавання усної мови з'явився в 1952 році, він міг розпізнавати вимовлені людиною цифри.

Комерційні програми з розпізнавання мовлення з'явилися на початку дев'яностих років. Зазвичай їх використовують люди, які через травми рук не в змозі набирати велику кількість тексту. Ці програми (наприклад, Dragon Naturally Speaking, VoiceNavigator) переводять голос користувача в текст, таким чином, розвантажуючи його руки. Надійність перекладу у таких програм не дуже висока, але з роками вона поступово покращується.

На початку 2000-х років в процесі розпізнавання мови все ще переважали традиційні підходи, такі як «Приховані моделі Маркова» у поєднанні з послідовними штучними нейронними мережами. Сьогодні, однак, багато аспектів розпізнавання мови були перейняті методом глибинного навчання, що називається «Довга короткочасна пам'ять» (LSTM) — періодична нейронна мережа, опублікована Sepp Hochreiter & Jürgen Schmidhuber в 1997 році. У 2007 році LSTM, що навчалася за часовою класифікацією Connectionist Temporal Classification (CTC), почала перевершувати традиційне розпізнавання мови в

певних програмах. У 2015 році розпізнавання мовлення Google, як повідомляється, зазнало суттєвого стрибка продуктивності на 49% за допомогою LSTM, тренованого за допомогою CTC, який зараз доступний через Google Voice для всіх користувачів смартфонів.

Деякі з основних завдань, що стоять перед світовим ринком розпізнавання голосу:

- Неможливість придушення зовнішніх шумів
- Висока вартість додатків розпізнавання голосу
- Проблеми з точністю розпізнавання
- Низький рівень безпеки в верифікації диктора [1]

На 2016 рік розробка з розпізнавання та синтезу мови від компанії Microsoft забезпечує якість розпізнавання, близьку до людини (5.9% помилок проти 5.1%) та здатна визначати контекст (спорт, комп'ютери тощо). В 2017 році компанія IBM досягла рівня помилок у 5.5%.

З точки зору технології, розпізнавання мовлення має довгу історію з кількома хвилями великих нововведень. Зовсім недавно ця область отримала вигоду від досягнень глибинного навчання та великих даних. Досягнення підтверджені не тільки надзвичайно високою кількістю наукових праць, опублікованих у цій галузі, але й, що більш важливо, завдяки впровадженню в усьому світі різноманітних методів навчання при розробці систем розпізнавання мовлення. Найбільш вагомими гравцями в галузі мовної індустрії є Google, Microsoft, IBM, Baidu, Apple, Amazon, Nuance, SoundHound, iFLYTEK, які оприлюднили основні технології в своїх системах розпізнавання мовлення, що базуються на глибинному навчанні [2].

Метою статті є висвітлення теоретичних положень щодо систем розпізнавання мовлення та наведення практичного застосування програмного забезпечення для керування комп'ютером за допомогою голосових команд.

Виклад основного матеріалу (результатів) дослідження:

Системи розпізнавання мовлення

Розпізнавання мовлення є міждисциплінарним підрозділом обчислювальної лінгвістики, що розробляє методики та технології, які дозволяють розпізнавати та перекладати розмовну мову на текст комп'ютерами. Він включає в себе знання та дослідження в галузі лінгвістики, інформатики та електротехніки.

Деякі системи розпізнавання мов вимагають "тренування" (також називається "реєстрацією"), де окремий оратор читає текст або ізольований словник. Система аналізує голос людини та використовує його для точного визначення мовлення цієї людини, що призводить до більшої точності. Системи, які не використовують тренування, називаються системами, що не залежать від спікера. Системи, що використовують навчання, називаються залежними від спікера [3].

Програми розпізнавання мов включають голосові користувацькі інтерфейси, такі як голосовий набір, маршрутизацію викликів, керування приладами, пошук, введення даних, підготовка структурованих документів та обробка мовлення до тексту.

Термін «розпізнавання голосу» або «ідентифікація диктора» стосується виявлення диктора, а не того, що вони кажуть. Визнання диктора може спростити завдання перекладу мови в системах, які були навчені за конкретним голосом людини, або вони можуть бути використані для аутентифікації або перевірки особистості диктора в рамках процесу безпеки.

Як Ви можете керувати комп'ютером своїм голосом?

Програмне забезпечення розпізнавання голосу забезпечує альтернативу друку на клавіатурі. Найпростіше застосування — це забезпечення швидкого способу запису тексту на комп'ютер — Ви говорите в мікрофон, а ваші слова з'являються в текстовому файлі на екрані [4].

Тим не менше, багато програм розпізнавання голосу є набагато складнішими, але в той самий час є все ще простими у використанні. Ви можете використовувати їх для запуску та управління комп'ютером, використовуючи прості розмовні команди, наприклад:

- відкривати та запускати різні програми
- створювати та формувати документ
- написати та відправити електронну пошту
- переглядати певні веб-сайти.

Кому може допомогти програмне забезпечення для розпізнавання мовлення?

Розпізнавання голосу надає значні переваги широкому колу потенційних користувачів. Найбільш очевидно, що це надзвичайно корисно для будь-кого з фізичною непрацездатністю, у кого наявні проблеми з набором тексту або управлінням комп'ютером.

Програми розпізнавання голосу також можуть сильно допомогти людям з дислексією — в даному випадку дане програмне забезпечення сприятиме правильному написанню та структуруванню речень.

У загальному випадку, розпізнавання голосу може допомогти полегшити роботу з будь-яким пристроєм, а також допомогти тим, хто може не володіти навичками набору тексту. Фактично, більшість людей можуть говорити набагато швидше, ніж вони можуть вводити з клавіатури. Окрім того, ПО для розпізнавання мовлення дає додаткові можливості для багатозадачності.

Як працює програма розпізнавання голосу?

Програми розпізнавання голосу аналізують звуки та перетворюють їх у текст. Програмне забезпечення використовує величезний словниковий запас, щоб визначити, що, найімовірніше, говорить оратор. Деякі програми дозволяють персоналізувати словник, додавши додаткові слова, наприклад, технічні терміни та прізвища людей.

Програмне забезпечення розпізнавання голосу захоплює та перетворює вашу мову через мікрофон. Деякі комп'ютери включають в себе вбудовані мікрофони, але краще та простіше використовувати мікрофонну гарнітуру. Ви підключаєте мікрофон до свого комп'ютера через гніздо звукової карти або USB (або подібне) з'єднання.

Всі голоси та фрази звучать трохи інакше, тому найефективнішим способом первинної ініціалізації програми є одноразовий процес, що називається "реєстрацією" для програмного забезпечення, щоб визначити, як Ви говорите. Це займає всього кілька хвилин і включає в себе читання екранної статті або певних слів та словосполучень. Така ініціалізація голосом є необхідною, щоб отримати найкращі результати.

Що Ви можете зробити з програмою розпізнавання голосу?

Диктування та коригування

Коли люди говорять, вони часто вагаються, бормочуть або вимовляють невиразно свої слова. Одним із ключових навичок використання програмного забезпечення для розпізнавання голосу є навчитися чітко говорити так, щоб комп'ютер міг розпізнати те, що Ви говорите. Це означає планувати те, що сказати, а потім говорити в повних фразах або реченнях. Програмне забезпечення для розпізнавання голосу буде неправильно розуміти деякі слова, що вимовляються, тому необхідно коректувати та виправляти будь-які помилки.

Поправки до ваших документів (листи, звіти, електронні листи тощо) можна легко зробити за допомогою миші та клавіатури або голосу.

Редагування та форматування

Для редагування тексту Ви можете виділити текст, який потрібно змінити, за допомогою команд, таких як "Select line" або "Select paragraph", а потім сказати зміни, які Ви хочете внести у вибраний текст.

Форматувати текст також дуже просто. Наприклад, якщо в документі міститься фраза "вступні думки", Ви можете підкреслити цю фразу, сказавши "Select «вступні думки»", а потім сказати "Underline".

Керування комп'ютером

Для керування вікнами на своєму комп'ютері необхідно використовувати голосові команди. Наприклад, щоб відкрити Word необхідно сказати «Open Word».

Для того, щоб викликати меню потрібно ненадовго зупинитись, а потім вимовити пункт меню. Наприклад, якщо сказати «File» відкриється меню «File», «Save» — меню збереження.

В Інтернет-браузері Ви можете диктувати веб-адреси та переглядати веб-сайти, просто вказавши текст у посиланнях.

Для завдань, що вимагають миші, існують розмовні команди, які дозволяють переміщати, перетягувати та натиснути мишу.

Робота «без рук»

Люди, які практично не користуються руками, повинні повністю користуватися комп'ютером за допомогою голосу. В такому випадку користувач має можливість робити наступне:

1. Внесення виправлень голосом.
2. Вибір меню та запуск програм голосом.
3. Переміщення та натискання миші голосом.
4. Навігація по Інтернету за допомогою голосових команд.
5. Натискання клавіш та комбінацій клавіш за допомогою голосових команд (наприклад Alt, Ctrl+C тощо).

Вирішення проблем

Програмне забезпечення для розпізнавання голосу може викликати проблеми, коли у користувачів виникають певні труднощі.

- Програмне забезпечення працює надто повільно: системи розпізнавання голосу потребують відносно нових комп'ютерів із великою кількістю пам'яті. Чим більше пам'яті Ви матимете у вашій системі, тим швидше система буде працювати.

- Погане розпізнавання: це може бути проблемою, якщо комп'ютер не отримує чіткий сигнал від мікрофона або якщо користувач не говорить чітко.

- Труднощі виявлення помилок під час роботи: це може бути проблемою, особливо для користувачів з дислексією. В даному випадку допоможе додаткове програмне забезпечення, що дозволяє прослухати текст та уникнути читання - тоді Ви можете прослухати помилки.

- Труднощі під час використання мікрофона для навушників: існує цілий ряд альтернативних мікрофонів, які можуть бути використані, в тому числі: стендові мікрофони, комірні мікрофони, магнітні мікрофони, ручні мікрофони - один з них повинен працювати. Всі мікрофони повинні бути сумісними з USB.

Висновки:

В ході дослідження були проаналізовані системи розпізнавання мовлення. Розглянуто основні теоретичні положення про системи розпізнавання мовлення, висвітлено проблеми, які виникають під час використання даних систем в програмному забезпеченні і шляхи їх вирішення та зроблено огляд їх застосування на практиці.

Список літератури

1. Перспективы развития систем распознавания речи [Електронний ресурс] – Режим доступу до ресурсу: <https://habr.com/post/232613/>.
2. Speech recognition [Електронний ресурс] – Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Speech_recognition.
3. У. А. Ли. Методы автоматического распознавания речи / У. А. Ли, Э. П. Нейбург, Т. Б. Мартин. – М.: Мир, 1983. – 328 с.
4. Voice recognition [Електронний ресурс] – Режим доступу до ресурсу: <https://www.scope.org.uk/Support/Disabled-people/Technology/Voice-recognition>.