УДК 519.852.67

Р.П. Нестерчук, студ., V курс, гр. ПІ-41м
Науковий керівник – Панішев А.В., д.т.н., професор
Житомирський державний технологічний університет

Розробка голосового інтерфейсу для керування програмними продуктами

Перед автором була поставлена задача – програмний модуль призначений для виконання різного роду інструкцій шляхом голосових команд. Він включає в себе компонент для зчитування голосу, парсингу отриманого сигналу в систему програмних інструкцій, а також синтезатор для відтворення голосу.

Важливим завданням розробки технічних систем є забезпечення інтуїтивного і природного інтерфейсу з користувачем, оскільки сучасні комп'ютерні програми орієнтовані на користувачів і розвиваються відповідно до їх зростаючих потреб. Однією з природних форм взаємодії для людини є мова. Голосовий інтерфейс користувача спроможний забезпечити зручний і гнучкий спосіб взаємодії людини з комп'ютером, оскільки для його використання не потрібно опановувати новими навичками.

Голосовий інтерфейс якісним чином змінює спосіб, а отже і ефективність взаємодії користувача з системою. Голосовий пошук від компанії Google і голосовий асистент Siri від компанії Apple є цьому яскравими прикладами, підтверджуючи нагальну необхідність впровадження мовних технологій, зокрема розпізнавання і синтез мови.

Складність розпізнавання мови полягає в тому, що сукупність таких характеристик голосу і мови як тембр, гучність, висота, темп, інтонація, якість дикції роблять мову кожної людини неповторною і унікальною як відбитки пальців. Завданням комп'ютерної техніки та програмного забезпечення є розпізнавання сказані людиною слова в будь-яких умовах без попередньої адаптації під конкретний голос.

Інтерфейс користувача (user interface) - різновид інтерфейсів взаємодії керованих людиною систем. Термін в основному застосовується по відношенню до комп'ютерних програм. Одним з найважливіших показників інтерфейсу користувача є usability - зручність програми або системи в користуванні, логічність і простота елементів управління, що скомпоновані та розташованими розумно і зрозуміло, і відповідають психофізіології людини. Збільшення в пристрої засобів введення-виведення дає спрощення побудови методів управління та спрощення правил користування, але призводить до складності сприйняття інформації користувачем - інтерфейс стає перевантаженим. І навпаки - зменшення засобів відображення і контролю призводить до ускладнення правил управління, оскільки кожен елемент несе на собі занадто багато функцій.

У зв'язку зі збільшенням інтенсивності обміну інформацією в системі «людина-машина» особливе значення має зниження навантаження на тактильно-зорові канали людини. Спроби навчити комп'ютери спілкуватися з людьми за допомогою природного голосового інтерфейсу робилися з перших років історії комп'ютерної техніки.

Реалізація мовного діалогу відбувається за допомогою діалогу, при якому запит і відповідь з боку користувача ведеться спрощеною природною мовою. Користувач вільно формулює завдання, але з набором встановлених програмним середовищем слів, фраз і синтаксисом.

Мова у фізичному сенсі - це акустичний сигнал, згенерований артикуляційними органами людини, що передається через фізичне середовище і сприймається вухом людини. При природній або штучній генерації мови в акустичному сигналі змінюються фізичні параметри. Ці зміни впливають на мембрану вуха, створюють траєкторії звукових образів, що розуміються людиною як відповідні звуки даної мови, чи інакше кажучи, при проголошенні слів людина генерує звуки (фонеми), які містять інформацію про ті символи, за допомогою яких ці слова можуть бути записані в вигляді тексту.

Математичну модель генерації звуку можна представити у вигляді збуджуючих генераторів тонового і білого шуму, групи резонаторів, модуляторів і ключів (рот, ніс, язик, губи), які забезпечують формування відчуття певного звуку.

Системи розпізнавання мови - це системи, що аналізують акустичний сигнал алгоритмами, заснованими на різноманітних теоріях, що припускають, які характеристики мовного сигналу створюють відчуття звуків даної мови, і математичних методах, які з певною точністю виділяють значущі параметри акустичного сигналу і перетворюють його в різній спосіб в необхідну форму.

Завчасно формується база фонем мови, що містить шаблони базового набору слів «усередненої» промови, тобто незалежної від диктора. Мова переводиться в фонемний опис і надходить у файл опису фонем, звідки цей опис надходить до блоку розпізнавання, який проводить порівняння інформації, що надійшла з тією, яка зберігається в базі. Формуються розпізнані слова, які перетворюються в текстові дані або команду.

Системи розпізнавання мови складаються з двох частин - акустичної та лінгвістичної.

1. Акустична частина - відповідає за подання мовного сигналу, за його перетворення до певної форми, в якій в явному вигляді присутня інформація про зміст мовного повідомлення.

2. Лінгвістична частина - інтерпретує інформацію, що отримується від акустичної моделі, і відповідає за подання результату розпізнавання до користувача. У загальному випадку можуть містити фонетичну, фонологічну, морфологічну, лексичну, синтаксичну та семантичну моделі мови.

При вирішенні задачі розпізнавання неперервної мови людина застосовує свої знання про природну мову, а також сенс сказаного для усунення неоднозначності при відновленні тексту речення. Завданням розпізнавання мови є автоматичне відновлення тексту вимовлених людиною слів, фраз або речень природною мовою, ідентифікація, очищення від шуму, оцінка психофізичного стану людини.

Відомі методи розпізнавання мови мають ряд основних загальних властивостей:

1. Для розпізнавання використовується метод порівняння з еталонами;
2. Сигнал може бути представлений або у вигляді безперервної функції, або у вигляді слова в деякому кінцевому алфавіті;
3. Для скорочення обсягу обчислень використовуються методи динамічного програмування.

Динамічне програмування - метод вирішення завдань шляхом складання послідовності з підзадач таким чином, що:

1. Перший елемент послідовності (кілька елементів) має тривіальне рішення;
2. Останній елемент цієї послідовності є вихідним завданням;
3. Кожну задачу цієї послідовності можна вирішити з використанням рішення підзадач з меншими номерами.

Методи розпізнавання мовлення можна розділити на дві великі групи: непараметричні і параметричні.

Непараметричні методи використовують міру близькості до еталонів на множині мовних сигналів (на основі формальних граматик чи метрик). Перевагами непараметричних методів є простота реалізації та навчання. До недоліків можна віднести складність обчислення міри близькості, яка пропорційна квадрату довжини сигналу і великий обсяг пам'яті, необхідний для зберігання еталонів команд - пропорційний довжині сигналу і кількості команд в словнику.

Параметричні методи застосовують теорію прихованих моделей Маркова - подвійні стохастичні процеси і ланцюги Маркова по переходах між станами і множині стаціонарних процесів в кожному стані ланцюга. Перевагами методу прихованих моделей Маркова є швидкий спосіб обчислення значень функції відстані (ймовірності) та істотно менший об'єм пам'яті. Основними недоліками є велика складність його реалізації та необхідність використання великих фонетично збалансованих мовних корпусів для навчання параметрів.

Класифікація систем розпізнавання мови:

1. Системи автоматичного розпізнавання ізольованих слів для розпізнавання вимовлених людиною команд послівно;
2. Системи автоматичного розпізнавання неперервного мовлення - з можливістю виділяти слова в природному частково неперервному потоці людської мови;
3. Системи розуміння мови - з елементами інтелекту, що дозволяє, по-перше, на основі змістовного аналізу більш правильно виділяти слова в потоці мови, а, по-друге, зберігати інформацію в базі знань, звідки її можна витягнути для вирішення певних інтелектуальних завдань.

Основні компоненти систем розпізнавання мови:

1. Графічне середовище для розробки, компіляції та оптимізації граматичних і лексичних блоків розпізнавання, перевірки і редагування лексиконів;
2. Система для протоколювання діалогів з працюючою програмою з метою оцінки якості розпізнавання і налаштування системи;
3. Інструмент оцінки якості роботи системи для перевірки відповідності слова, сказаного абонентом до використаної граматики;
4. Система для створення «тренованих» мовних моделей, що підвищують продуктивність і пришвидшують процес розпізнавання;
5. Система для розподілу багатьох паралельних запитів різних типів і прозорою інтеграцією різних мовних модулів в мережі.

Висновок. У даній роботі було описано суть і переваги використання голосового інтерфейсу. Було розглянуто, що собою являє і з яких елементів складається система розпізнавання мови. А також описано методики для реалізації даного інтерфейсу.

ВІДОМОСТІ ПРО АВТОРІВ:

НЕСТЕРЧУК Роман Петрович, студент групи ПІ-41м кафедри програмного забезпечення систем Житомирського державного технологічного університету. Наукові інтереси: робота з графікою, робота зі звукозаписом, інтернет-технології, розробки в сфері мобільного програмного забезпечення, розробка десктопних застосунків.