

**МЕТОД РОЗПІЗНАВАННЯ МОВЛЕННЯ, ПО КОРОТКОМУ СЛОВНИКУ,
З ВИКОРИСТАННЯМ MEL-КЕПСТРАЛЬНИХ КОЕФІЦІЄНТІВ**

В статті розглядається розпізнавання мовлення, а саме використання цієї технології для андроїд пристроїв. Так як вони є дуже розповсюдженими, постає задача створити голосовий інтерфейс, для більшої збільшення зручності у їх використанні. Головною ідеєю є - створення додатку, який буде виконувати розпізнавання не на сервісі, як зазвичай це відбувається, а на самому пристрої, за допомогою відкритої технології розпізнавання мовлення Pocketsphinx. Ця технологія мультиплатформена, що дозволяє скомпілювати її в бібліотеку під ARM процесор, за допомогою Android NDK.

В ній використовується обмежений словник вимови та граматики запитів, що дозволяє підвищити точність розпізнавання саме для конкретного набору слів і зробити це прямо на пристрої, без підключення до мережі.

Ключові слова: розпізнавання, короткий словник, Pocketsphinx, андроїд.

S. OZHAROVSKIY, A. MYASISHEV
Khmelnitsky National University

SPEECH RECOGNITION METHOD, IN BRIEF DICTIONARIES, WITH MEL-CEPSTRAL COEFFICIENTS

In the paper the speech recognition, namely the use of this technology for android devices. Since they are very common, there is the problem of a voice interface for greater ease of increasing their use. The main idea is - to create applications that will not perform recognition on service, as usually happens, and on the device itself, using speech recognition technology open Pocketsphinx. multiplatform This technology, which allows it to compile the library under the ARM processor, using the Android NDK.

It uses a limited vocabulary and grammar pronunciation queries that can improve the accuracy of recognition is for a particular set of words and do it right on your device without connecting to the network.

Key words: recognition, brief dictionary, Pocketsphinx, android.

Постановка задачі

Створення інтерфейсів, які підтримують та надають більш природні форми діалогу між користувачем та андроїд пристроєм, рухається та прискорюється впровадженням інформаційних технологій в слід за зростаючими потребами професійної та повсякденної діяльності людини. В області інформаційних технологій засоби взаємодії користувача з технічною системою називають інтерфейсом. Інтерфейси бувають різні і реалізуються різними засобами та методами. Одною з найважливіших задач розробки сучасних технічних систем є забезпечення найбільш інтуїтивного і природного інтерфейсу з користувачем, тобто сучасні андроїд додатки орієнтовані на користувача.

Одною з природних форм взаємодії для людини є мовлення. Голосовий інтерфейс може покращити існуючий користувацький інтерфейс – він забезпечує більш зручний та менш обмежений спосіб взаємодії людини з пристроєм. Якісний голосовий інтерфейс допомагає подолати несприйняття технології користувачами, так як для його використання не потрібно опанувати нові навички. Голосовий інтерфейс якісним чином змінює спосіб, а отже і ефективність взаємодії користувача з системою. Голосовий пошук від компанії Google та голосовий асистент Siri від компанії Apple є яскравими прикладами цьому, які підтверджують існуючу необхідність впровадження мовленнєвих технологій, зокрема розпізнавання мовлення та голосових інтерфейсів.

Важливий та практичний аспект даних проблем зв'язані з тим, що голосовий інтерфейс є необхідною складовою, коли мова йде про створення комфортних умов життя для людей з порушенням опорно-рухового апарату, а також спеціалістів які втратили можливість використання стандартних засобів в результаті професійного захворювання або травми. Такі системи з часом увійдуть в повсякденний побут в процесі реалізації концепції «розумних будинків».

В зв'язку з вищесказаним стає актуальною проблема можливості створення простого голосового інтерфейсу для користувачів андроїд пристроїв, враховуючи неможливість миттєвого переходу і необхідність адаптації до нових інструментів та засобів, очевидною стає проблема в інтеграції з вже існуючими системами.

Далеко не всі задачі розробки голосового інтерфейсу на даний час можна вважати вирішеними. Проблема розробки голосового інтерфейсу є достатньо складною та комплексною, що потребує від розробника знань в різних предметних областях, таких як комп'ютерні науки, лінгвістика і психологія поведінки людини. Навіть при наявності сучасних засобів проектування, розробка ефективного користувацького інтерфейсу потребує від його творців детального розуміння як задач, виконуваних системою, так і психологію користувачів системою.

Проведений бібліографічний пошук та аналіз інформації в Internet підтвердив актуальність теми тим, що в даний час багато провідних компаній посилюють роботу в напрямку розвитку голосових інтерфейсів і технологій розпізнавання мовлення. Впровадження голосового інтерфейсу залишить очі та руки користувача вільними від перевантаження, що підвищить надійність та якість управління.

Використання мовного діалогу в системах масового обслуговування також актуальне[1]. Крім виключної зручності для населення, такі системи підвищують комерційну вигоду за рахунок залучення додаткових клієнтів.

Виклад основного матеріалу досліджень

Переваги голосового інтерфейсу:

- Оперативність та природність;
- Мінімум спеціальної підготовки користувача;
- Можливість користування пристроєм в темряві, за межами його візуальної видимості;
- Можливість використання одночасно ручного та голосового введення інформації;

До основних класів задач голосового інтерфейсу слід віднести:

- Синтез мовлення – ця задача включає в себе комплекс підзадач і полягає в перетворенні друкарського тексту у мовний сигнал;
- Аналіз та розпізнавання мовлення – комплекс задач, які включають в себе запис, оцифрування і аналіз мовлення для розпізнавання отриманого мовного сигналу системою;
- Розуміння (інтерпретація) мовлення – це комплекс задач, зв'язаних з аналізом змісту мовленнєвих повідомлень і формування реакції (відповіді) системи.

Загальна структура голосового інтерфейсу включає два основних компоненти:

- Синтез мовлення;
- Розпізнавання мовлення.

Кожна з задач голосового інтерфейсу є достатньо складною, відповідно вказаним компонентам виділяють два класи систем:

- Системи синтезу мовлення;
- Системи розпізнавання мовлення.

Реалізація мовленнєвого діалогу відбувається за допомогою діалогу, в якому запит і відповідь зі сторони користувача ведеться на мові, близькій до природної. Користувач вільно формує задачу, але з набором встановлених програмним середовищем слів. Різноманітністю інтерактивного природного діалогу є мовленнєве спілкування з андроїд системою. В цьому випадку людський голос може перетворюватись, наприклад, в текст, або використовуватись для інтерактивного управління системою, або для ідентифікації особи. В основі даних процесів лежить технологія і вирішення задачі розпізнавання мовлення.

Мовлення в фізичному сенсі – це акустичний сигнал, генерований артикуляційними органами людини, який передається через фізичне середовище, та сприймається вухом людини. При природній або штучній генерації мовлення в акустичному сигналі змінюються фізичні параметри. Ці зміни діють на мембрану вуха, створюють траєкторії звукових образів, зрозумілих для людини як відповідні звуки даної мови, або інакше кажучи, при вимові слів людина генерує звуки (фонем), які несуть інформацію про ті символи, за допомогою яких ці слова можуть бути записані у 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 відповідного звуку.

Системи розпізнавання мовлення – це системи, які аналізують акустичний сигнал алгоритмами, заснованих на різних теоріях, передбачаючих, які характеристики мовленнєвого сигналу створюють відчуття звуків даної мови, і математичних методах, які з тою чи іншою точністю виділяють значущі параметри акустичного сигналу і перетворюють його в різній повноті в необхідну форму.

Завчасно формується база фонем мови, яка містить шаблони базового набору слів за «усередненої» мови, тобто незалежній від диктора. Мовлення переводиться в фонемний опис і надходить у файл опису фонем, звідки цей опис надходить в блок розпізнавання, де проводиться порівняння інформації яка надійшла з тією, яка зберігається в базі. Формується розпізнавання слова, яке перетворюється в текстові дані або команду.

Системи розпізнавання мовлення складаються з двох частин – акустичної та лінгвістичної. В загальному випадку можуть включати в себе фонотечну, фонологічну, морфологічну, лексичну, синтаксичну і семантичну моделі мови.

Акустична – відповідає за представлення мовленнєвого сигналу, за його перетворення в деяку форму, в якій в більш явному вигляді присутня інформація в змісті мовленнєвого повідомлення.

Лінгвістична – інтерпретує інформацію, отриману від акустичної моделі, і відповідає за представлення результатів розпізнавання користувачу.

Задачі розпізнавання мовлення – автоматичне відновлення тексту вимовлених людиною слів, фраз або пропозицій на природній мові і проблеми ідентифікації, шумоочистки, розпізнавання мов, оцінка психофізичного стану людини. При вирішенні задачі розпізнавання злитого мовлення людина застосовує свої знання про природне мовлення, а також зміст сказаного усунення неоднозначності при встановленні тексту речення.

Методи розпізнавання мовлення можна розділити на дві великі групи: непараметричні – з використанням непараметричних мір близькості до еталону, і параметричні (імовірнісні) – на основі методу прихованих моделей Маркова, нейромережні[2], [3].

Параметричні методи застосовуються для розпізнавання мовлення в наш час, були вперше

запропоновані американськими вченими (Бейкер[4] і Джелінек[5]) в 1970 роки минулого століття. В них використовується теорія прихованих моделей Маркова – двічі стохастичні процеси і ланцюги Маркова[6] по переходах між станами і множинами стаціонарних процесів в кожному стані ланцюга.

Перевагами методу прихованих моделей Маркова є:

- Швидкий спосіб розрахунку значень функції відстані (імовірності);
- Значно менший обсяг пам'яті, в порівнянні з методом «динамічної деформації часу», необхідний для зберігання еталонів команд.

Основний недолік - це важкість його реалізації.

В нашому випадку розпізнаватись буде визначений набір команд, для цієї задачі найкраще підходить розпізнавання на основі обмеженого словника та граматики запитів. Це дозволить підвищити точність розпізнавання саме для конкретного набору слів та здійснювати це прямо на пристрої, без підключення до мережі. Було вирішено використовувати відкриту технологію розпізнавання мовлення Pocketsphinx.

Pocketsphinx - система розпізнавання зливої промови - розробляється інженерами з Університету Карнегі-Мелон і компілюється в бібліотеку під ARM за допомогою Android NDK. При цьому можна згенерувати JNI-обгортки для використання нативних методів з коду на Java.

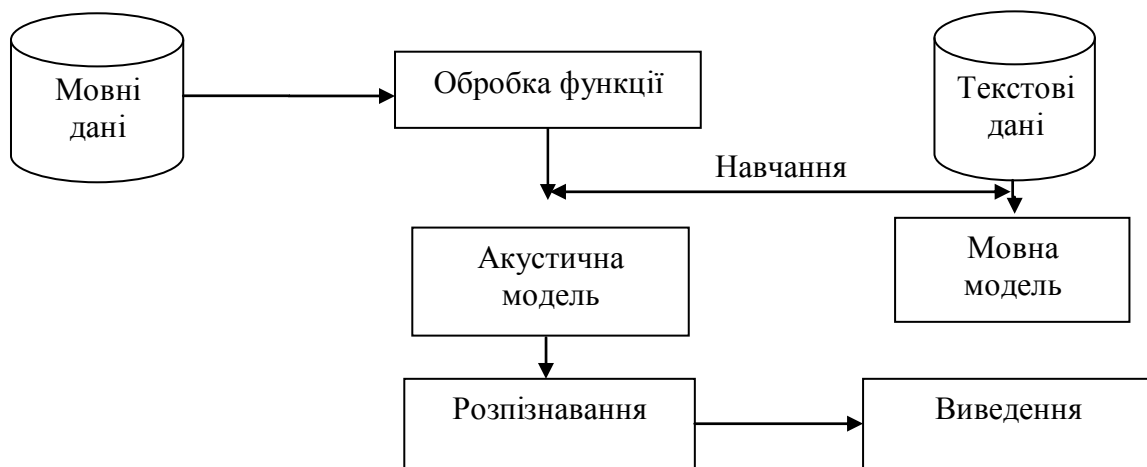


Рис. 1. Схема роботи Pocketsphinx

Поклавши скомпільовану бібліотеку в `libs / armeabi` проекту, можна потім завантажити її з коду таким чином:

```
static {
    System.loadLibrary("pocketsphinx_jni");
}
```

Основними класами для ініціалізації та роботи з розпізнаванням мови є `Config` і `Decoder`. З допомогою першого вказуються параметри розпізнавання, а також повні шляхи до мовної моделі, словника вимов і файлу граматики запитів у форматі JSGF (Java Speech Grammar Format). `Decoder` надає конструктор, що приймає в якості аргументу цю конфігурацію, і ініціалізує власне інтерфейс для розпізнавання мови.

Для цього можна скористатися методами `startUp`, `process Raw`, `getHub` і `endUtt`. Перший і останній відповідно починають і закінчують сесію, а `processRaw` і `getHub` власне розпізнають мова і повертають гіпотези.

`process Raw` приймає на вході байтові фрейми - дані з мікрофону, а `getHub` повертає гіпотезу `Hypothesis`, що містить припущення про те, що вимовив користувач. Таким чином, організувавши в одному потоці збір даних з мікрофона, і передаючи їх у чергу на обробку декодером, можна отримати результати розпізнавання.

Для отримання даних з мікрофону пристрою використовується стандартний клас Android SDK `AudioRecord`. У легу конструкторі потрібно вказати тип джерела звуку (стандартний мікрофон, розпізнавання мови, камкодер тощо), частота дискретизації (залежить від моделі, найчастіше це 8 або 16 kHz), формат аудіо-даних і розмірність буфера (також залежить від моделі, чим менше тим краще). Варто відзначити, що тип джерела звуку безпосередньо впливає на якість розпізнавання, тому операційною системою задіюються деякі можливості по оптимізації вихідних даних для кожного конкретного типу (докладніше про це написано в самій документації по класу `AudioRecord`).

Висновки

Pocketsphinx надає можливість використовувати мовлення у проектах для мобільних платформ, забезпечуючи прийнятну якість розпізнавання і відносну простоту впровадження. Звичайно, багато проблем

залишаються невирішеними. Різноманіття пристроїв на платформі Android породжує найчастіше різну поведінку на різних девайсах - від швидкості розпізнавання до якості звуку з мікрофону і впливу сторонніх шумів. Проте, продемонстрований підхід дозволяє реалізувати реальний продукт, для управління андроїд пристроєм за допомогою голосового інтерфейсу.

Література

1. Жожикашвили В. А. Применение распознавания речи в автоматизированных системах массового обслуживания//Автоматизация и современные технологии, 2003, No 11, с. 23-29 Лобанов Б. М. 2002 В стране и мире.
2. Винцюк Т. К. Анализ, распознавание и интерпретация речевых сигналов. Киев: Наукова думка, 1987.
3. Винцюк Т. К. Распознавание слов устной речи методами динамического программирования / Кибернетика. 1968. с. 81-88.
4. Baker J. K. Stochastic modeling for automatic speech understanding / Speech Recognition / ed.: D. R. Reddy. New York: Academic Press, 1975. P. 521-542.
5. Дженилек Ф. Распознавание непрерывной речи с помощью статических методов/ ТИИЭР. 1976 с. 131-160.
6. Марков А. А. Пример статистического исследования над текстом «Евгения Онегина», иллюстрирующий связь испытаний в цепь / Известия академии наук. СПб. 4. Т. 7. 1993. С. 153-162.
7. CMU Sphinx (Pocketsphinx on android) [Electronic resource]. - <http://cmusphinx.sourceforge.net/wiki/tutorialandroid>.

Рецензія/Peer review : 27.11.2015 р. Надрукована/Printed :6.12.2015 р.
Рецензент: д.т.н., проф., Сорокатий Р.В.

УДК 519.21

І.І. ЧЕСАНОВСЬКИЙ, В.А. СОБЧЕНКО

Національна академія Державної прикордонної служби України імені Богдана Хмельницького

РОЗРОБКА УЗАГАЛЬНЕНОЇ МОДЕЛІ НАДІЙНОСТІ ПІДСИСТЕМИ МОБІЛЬНОГО ТЕПЛОВІЗІЙНОГО КОМПЛЕКСУ

У статті на основі загальної методики використання напівмарківських процесів розроблено узагальнену модель надійності мобільного тепловізійного комплексу. На основі розробленої моделі проаналізовано характер залежності комплексного показника надійності коефіцієнта технічного використання за різних умов.

Ключові слова: надійність, технічна система, мобільний тепловізійний комплекс.

I.I. CHESANOVSKYY, V. A. SOBCHENKO

The National Academy of State Border Service of Ukraine named after B. Khmelnytsky

DEVELOPMENT OF THE GENERALIZED MODEL OF RELIABILITY OF SUBSYSTEM MOBILE THERMOVISION COMPLEXIS

Abstract – In the article on the basis of general methodology of the use of markov processes the generalized model of reliability of mobile thermo vision complex is worked out. On the basis of the worked out model character of dependence of complex reliability of coefficient of the technical use index is analysed at different terms.

Keywords: Reliability, technical system, mobile thermo vision complex.

Постановка задачі

З точки зору надійності мобільний тепловізійний комплекс (МТК) може розглядатись як складна технічна система, яка складається з окремих підсистем. Враховуючи велику кількість підсистем з різноманітними показниками і стратегіями експлуатації, поєднання їх в єдину модель з можливістю поточної оцінки надійності є практично недосяжною метою. Проте, для досягнення цієї мети, можна вирішити задачу іншим шляхом – розробивши узагальнену модель підсистеми, що включає всі можливі стани окремих підсистем і трансформується під кожен, в ході експлуатації, за рахунок формування вектору інтенсивностей і ймовірностей переходів. Такий підхід, дасть змогу виробити єдиний підхід в експлуатації окремих підсистем, що поєднані в комплекс, застосовуючи гнучку систему профілактичних та відновлювальних заходів яка забезпечить екстремальні значення основних показників надійності.

Аналіз досліджень та публікацій

В теорії надійності відома велика кількість методів, інженерних методик, які доведені до алгоритмів і програм. Їх авторами є загальновідомі вчені Барзилович Е. Ю. [1], Барлоу Р. Е., Острейковський В. А. [2], Дружинін Г. В., Коваленко І. Н., Ушаков К. З. [3] та ін. Дані роботи є фундаментальними в теорії, проте методи описані в даних роботах часто не вдається застосовувати на практиці.

Виклад основного матеріалу