УДК 004.522

АНАЛІТИЧНИЙ ОГЛЯД ПІДХОДІВ ДО ВИДІЛЕННЯ ОЗНАК В ЗАДАЧІ РОЗПІЗНАВАННЯ ФОНЕМ

Романенко А.Ю.[1)], Олійник В.В.[2)]
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
[1)] студент, [2)] к.т.н., старший викладач

Системи розпізнавання мови набирають широку популярність сьогодні у зв'язку із зручністю використання та збільшенням точності і швидкості їх роботи. Однією із задач розпізнавання мови є задача розпізнавання голосових команд. Вона є найпростішою та досить корисною з практичної точки зору.

Для спрощення процесу розробки систем розпізнавання голосових команд була запропонована модель у статті [1] та описано принцип функціонування системи та функції, які повинні виконувати блоки у системі. Розглянемо детальніше популярні алгоритми, що можуть використовуватись для реалізації блоку виділення ознак у підсистемі розпізнавання фонем. Так як точність та якість виділення ознак впливає на точність розпізнавання фонем та відповідно точність роботи системи в цілому.

Алгоритми виділення ознак звуку

Як відомо, мова представляє собою нестаціонарний сигнал. Тобто під час мовлення голосовий тракт людини постійно міняє свої параметри. Внаслідок чого виникає нестаціонарний сигнал, який сприймається як мова. Алгоритм виділення ознак має вказати проміжки часу, коли сигнал стаціонарний та визначити значення характеристик сигналу на цих проміжках.

Віконне перетворення Фур'є

Як відомо перетворення Фур'є працює для стаціонарних сигналів. Для нестаціонарних сигналів перетворення Фур'є дає результати, тобто неможливо визначити, коли сигнал змінив свої характеристики.

Щоб застосовувати ефективніше перетворення Фур'є до нестаціонарних сигналів, припускається, що на малому проміжку функція стаціонарна. Виконавши перетворення Фур'є таким чином для кожного проміжку можна отримати значення характеристик сигналу та час їх появи.

Переваги перетворення Фур'є: розроблені ефективні алгоритми для його обчислення. Спектральна картина однаково точна як для голосних, так і для приголосних. Здійснюється значне стискання інформації про сигнал, що спрощує його подальше розпізнавання.

Недоліки: спектральна картина залежить від ширини вікна. Чим вужче вікно, тим менше компонент зберігає спектр. Якщо сигнал має складові, що знаходяться у проміжках між спектральними компонентами, то вони розподіляються між компонентами результуючого спектру, що робить картину змазаною.

Область застосування**:** перетворення Фур'є варто застосовувати у тому випадку, коли разом із ним застосовується алгоритм розпізнавання, що може компенсувати "розмазаність" спектру. Крім цього, варто проходитись по сигналу вікном із незначним зсувом, щоб створювати більш точну картину. З урахуванням вище наведених міркувань, даний спосіб може використовуватись як для словників малої так і великої розмірності.

Мел-частотні кепстральні коефіцієнти

Мел-частотні кепстральні коефіцієнти (MFCC) враховують особливості сприйняття звуків людиною та ще більше стискають інформацію про спектр сигналу [2]. Це перетвореннння надає більшої ваги частотам, до яких людина більш чутлива та згортає всі гармоніки частоти, наявні у сигналі, до одного відліку.

Переваги: Зберігаються переваги перетворення Фур'є. Частоти, які мають гармоніки у спектрі, стискаються до 1 кепстрального коефіцієнта, що особливо ефективно у випадку голосних літер. Даний спосіб отримання характеристик враховує спосіб сприйняття звуків людиною, внаслідок чого незначущі компоненти отримують меншу вагу.

Недоліки: Зберігаються недоліки перетворення Фур'є. Інформація про приголосні звуки ще більше стискається та розпорошується, внаслідок чого погіршується якість їх розпізнавання.

Область застосування: як і у випадку перетворення Фур'є варто застосовувати алгоритм розпізнавання, що може компенсувати недоліки втрати інформації. Даний алгоритм можна використовувати для невеликого словника в парі із простим алгоритмом розпізнавання, особливо у

тому випадку, коли слова мають багато голосних. З іншої сторони використовуючи потужний алгоритм розпізнавання можна отримати високу якість розпізнавання для середніх та великих словників.

Вейвлет-перетворення

Вейвлет перетворення є інтегральним частотно-часовим представленням сигналу. Його особливістю є те, що воно дає різне розширення для різних частот [3]. Вейвлет перетворення дозволяє отримати хороше розширення по частоті і погане по часу для низьких частот, а для високих – навпаки: хороше розширення по часу і погане по частоті.

Переваги: цей спосіб отримання ознак має різну роздільну здатність для різних частот, що дозволяє отримати більш точне частотно-часове представлення сигналу, порівняно із перетворенням Фур'є. Крім того, при використанні вейвлет-перетворення блок виділення звуків опускається у підсистемі розпізнавання звуку, бо це перетворення автоматично робить часову локалізацію.

Недоліки:вейвлет-перетворення складніше та менш популярне, ніж перетворення Фур'є, тому знайти бібліотеки для його застосування буде складніше. Для ефективного використання цього перетворення з метою визначення ознак сигналу, необхідно мати спеціальні знання про типи вейвлет-перетворення, типи вейвлетів та відмінності між ними.

Область застосування: у зв'язку з тим, що дане перетворення забезпечує більш гнучке та детальне частотно-часове представленням сигналу може застосовуватись для словників середньої та великої розмірності.

Банк фільтрів

Ідея полягає у тому, що вся спектральна шкала розділяється на відрізки певної довжини. Кожному відрізку призначається свій фільтр, який пропускає ті спектральні складові сигналу, що знаходяться в межах діапазону фільтрації. У результаті пропускання сигналу через банк фільтрів ми отримуємо N сигналів, спектральні характеристики кожного з яких лежать у діапазоні спектральних характеристик відповідного фільтра [4]. Таким чином, для кожного каналу можна визначити енергію сигналу для кожного діапазону частот та час, коли змінюється розподіл енергії у спектрі.

Переваги: простота реалізації та інтерпретації результатів. На виході можна отримати декомпозицію сигналів за складовими, що лежать у певних частотних інтервалах. Відповідно можна оцінити енергію кожної складової та час її існування.

Недоліки: чим точніше цифровий фільтр обмежує смугу пропускання, тим більшу затримку він вносить у вихідний сигнал. Відповідно можуть розмиватись короткострокові зміни сигналу, і таким чином може втрачатись істотна інформація стосовно приголосних.

Область застосування: зважаючи на легкість реалізації та високу гнучкість методу, він може застосовуватись для словників від малої до великої розмірності. При використанні цього методу блок виділення звуків може бути опущеним, а у якості ознаки будуть використовуватись енергії сигналів на виході фільтрів у певному вікні.

Коефіцієнти лінійного передбачення

Метод базується на припущенні, що голосовий сигнал є результатом проходження сигналу через фільтр, який пропускає всі частоти, але з різним рівнем підсилення для кожної частоти.

Припускається, що голосові звуки утворюються при пропусканні через фільтр послідовності одиничних імпульсів. Приголосні у свою чергу утворюються при пропусканні білого шуму через фільтр. У такому разі можна характеристиками фонеми можуть виступати тип вхідного сигналу та набір коефіцієнтів фільтра, який перетворює вхідний сигнал у вихідний звук [5].

Переваги: простота реалізації – для отримання коефіцієнтів треба розв'язати систему лінійних рівнянь. Хороше стискання інформації про сигнал із високою якістю для голосних, що спрощує подальше розпізнавання. Даний спосіб аналізу дозволяє отримати моменти, появи та закінчення фонеми (за помилкою передбачення), що дозволяє досить точно відокремити їх один від одного.

Недоліки: низька точність формування ознак для приголосних.

Область застосування: на практиці часто використовують кількість коефіцієнтів у проміжку від 8 до 16. Зважаючи на те, що даний метод дає кращу якість для голосних, його варто застосовувати для словників, у яких велика кількість голосних і мало приголосних. Це зазвичай словники малого або середнього розміру.

Чистий сигнал

У якості ознак може виступати чистий сигнал, адже він у собі має всю необхідну інформацію для розпізнавання звуку. Складність у тому, що звуковий сигнал несе у собі надмірну кількість інформації не тільки про фонему, а й про людину, що говорить та середовище, у якому відбувається

мовлення. Відповідно сам алгоритм розпізнавання повинен відокремлювати інформацію, що стосується фонеми з усього об'єму даних.

Переваги: простота реалізації – не потрібно проводити ніяких додаткових операцій для виділення ознак.

Недоліки: велика складність розпізнавання, у зв'язку із надмірною кількістю параметрів.

Область застосування: можна застосовувати для словників малого розміру (наприклад з 2 словами: "так", "ні"). Для використання із словниками середнього та великого розміру потрібно застосовувати алгоритми розпізнавання, що самі внутрішньо можуть виділити значущі ознаки із сигналу зазвичай методом навчання.

Алгоритми розпізнавання фонем

Для подальшого розпізнавання фонем на основі отриманих ознак можуть використовуватись різні класичні та некласичні алгоритми розпізнавання образів. Для словників малого розміру (кілька слів 1-5) може використовуватись алгоритм мінімуму відстані. Для малих та середніх (до 100 слів) можуть використовуватись алгоритми динамічного програмування на зразок DTW. Для словників великого розміру (десятки тисяч слів) довгий час найбільш потужним засобом розпізнавання були приховані моделі маркова, зараз глибокі нейромережі показують кращі результати, зокрема рекурсивні та згорткові мережі.

Висновки

Представлений огляд методів виділення ознак в задачі розпізнавання фонем дозволяє структурувати підходи до розв'язання цієї задачі за рахунок визначення сильних та слабких сторін кожного алгоритму та вказанням умов ефективного застосування. В результаті цього зменшується час необхідний для пошуку та первинного дослідження методів та розробки першого прототипу системи. Маючи цю інформацію, інженери отримують можливість реалізувати декілька алгоритмів та обрати той, що найкраще справляється із поставленими задачами. Варто враховувати, що велике значення має і алгоритм розпізнавання фонем, який класифікує набір ознак отриманих від алгоритму виділення ознак. Чим якісніше і точніше будуть розпізнані фонеми, тим точнішими будуть результати розпізнавання команд та тим більшим буде можливий розмір словника.

Список використаних джерел

1. Романенко А.Ю.Узагальнена модель розпізнавання голосових команд / А.Ю. Романенко, В.В. Олійник // Адаптивні Системи Автоматичного Управління. Міжвідомчий науково-технічний збірник–2017. –№1(30).
2. Мел-кепстральные коэффициенты (MFCC) и распознавание речи / Хабрахабр: [Електронний ресурс] – Електронні дані. – https://habrahabr.ru/post/140828/ – Назва з екрана.
3. "Введение в вейвлет-преобразование" RobiPolicar, IowaStateUniversity, Автор перекладу: Грібунін В.Г.
4. Lawrence R. B. Fundamentals of speech recognition / Lawrence Rabiner Biing, Hwang Juang – Upper Saddle River, NJ, USA : Prentice-Hall, Inc., 1993 – 507с.
5. Linear Predictive Coding is All-Pole Resonance Mod…: [Електроннийресурс] – Електронні дані. – https://ccrma.stanford.edu/~hskim08/lpc/ – Назва з екрана.
6. Распознавание речи от Яндекса. Под капотом у Yandex.SpeechKit: [Електронний ресурс] – Електронні дані. – Режим доступу: https://habrahabr.ru/company/yandex/blog/198556/ – Назва з екрана.