РОЗРОБКА СИСТЕМ РОЗПІЗНАВАННЯ ГОЛОСОВИХ КОМАНД

Зіміна А.Р.
Науковий керівник – к.т.н., доц.каф.СТ Губаренко Є.В.
Харківський національний університет радіоелектроніки
(61166, Харків, просп. Науки,14, каф. Системотехніки,
тел. (057) 702-13-06)
e-mail: anna.zimina@nure.ua

The given work is devoted to the development of modern system for speech recognition and voice commands. Speech recognition technologies are used for voice commands, in situations where speaking is much easier than typing. Speech recognition is used in interactive language self-service systems, for example, when a telephone is used by a robot that can deal with standard support questions. Another application of voice recognition technologies is dictation of texts, a kind of automatic secretary, which can be useful for people with disabilities. Finally, systems with voice control, with any kind of technique, such as "smart home", or a car, are increasingly appearing. The developed software can be used to recognize the language and dictate text, for example, for people with disabilities

При розпізнаванні звукових сигналів найгостріше стоїть проблема наявності шуму. Системи ефективно працюють в ідеальних умовах записи, але при цьому не справляються з фоновими шумами. Штучні шуми виділити цілком можливо, але як відрізнити голос людини, який потрібно розпізнати, від голосу людини, що розмовляє по сусідству, майже неможливо. Хоча проблема завадостійкості до сих пір не вирішена, є кілька способів, у ряді випадків дозволяють її подолати.

По-перше, мовний корпус, на якому навчається акустична модель, штучно розширюють, додаючи до мовних сигналів з нього різноманітні шуми. Проблема у тому, що інший тип шумів, не врахований при навчанні, буде викликати перешкоди у роботі.

По-друге, можна використовувати апаратні можливості підвищення завадостійкості, що застосовувалися, наприклад, ще в перших моделях iPhone, які реалізовували функцію розпізнавання. У них використовувалися два мікрофони: перший, основний, ловив мову користувача разом з фоновими звуками, а другий мікрофон на іншій стороні пристрою фіксував тільки фонові шуми. В результаті з сигналу, записаного першим мікрофоном, вичитали шуми, які були записані на другий мікрофон.

Друга проблема пов'язана з тим, що деякі акустичні характеристики користувача можуть не збігатися з типовими голосами, що використовуються в мовних корпусах. Стандартні алгоритми дикторонезалежного розпізнавання мови не дозволяють працювати з нестандартній промовою. Проблема особливо актуальна при розпізнаванні мови диктора, якщо він розмовляє на нерідній мові, тобто наявність

акценту. Іншими прикладами є порушення мовного апарату, розпізнавання дитячого мовлення.

Один з варіантів подолання зазначеної проблеми полягає в накопиченні великої кількості нестандартних фрагментів мови, з різними акцентами. Також можливо налаштувати акустичну модель на голос конкретного користувача. Користувач або в процесі експлуатації, або на попередньому етапі налаштування системи, читає певний текст, а параметри акустичної моделі адаптуються, щоб врахувати варіації голосу користувача. Такої адаптації досить, щоб система стала надійно розпізнавати голос користувача, навіть при наявності в мові дефектів. В ряді випадків доводиться формувати акустичну модель тільки на основі фрагментів мови користувача з використанням кластеризації мінімальних звукових одиниць і, наприклад, математичного апарату теорії інформації.

Основні поняття:

1. чіткість голосу – відносна кількість правильно прийнятих елементів мови (звуків, складів, слів, фраз), виражене у відсотках від загального числа переданих елементів;
2. якість мови – параметр, що характеризує суб'єктивну оцінку звучання мови в випробуваної системі передачі мови;
3. нормальний темп мови – проголошення промови зі швидкістю, при якій середня тривалість контрольної фрази дорівнює 2,4 с;
4. прискорений темп мови – проголошення промови зі швидкістю, при якій середня тривалість контрольної фрази дорівнює 1,5-1,6 с.;
5. впізнаваність голосу мовця – можливість слухачів ототожнювати звучання голосу, з конкретною особою, відомим слухачеві раніше;
6. смислова розбірливість – показник ступеня правильного відтворення інформаційного змісту промови;
7. інтегральне якість – показник, що характеризує загальне враження слухача від прийнятої мови.

Системи розпізнавання речі класифікуються: по розміру словника; за залежності від диктора (дикторозалежні і дикторонезалежні системи); по типу речі (злита або роздільна мова); призначення (системи диктовки, командні системи); по використовуваному алгоритму (нейронні мережі, скриті Марковські моделі, динамічне програмування); по типу структурної одиниці (слова, фонеми, дифони, аллофони); за принципом вибору структурних одиниць (розпізнавання по шаблону, вилучення лексичних елементів).