УДК 004.934.1'1(043.2)

Ялковський А. Є.

ПРОБЛЕМИ РОЗПІЗНАВАННЯ МОВИ ЛЮДИНИ

Інститут комп'ютерних технологій
Національний авіаційний університет

Приведено аналіз існуючих методів розпізнавання мови людини. Були розглянуті основні типи задач, які не можуть існувати без систем розпізнавання. Теоретично обґрунтовано моделі і методи аналізу та розпізнавання сигналів багатьох змінних. Запропоновано методи, алгоритми і обчислювальні процедури аналізу сигналів на основі параметричних функцій систем, які створюють сигнал

Вступ

На сьогоднішній день суспільство вносить величезну кількість коштів на розвиток *know-how* і науково-дослідні розробки для вирішення проблем автоматичного розпізнавання і розуміння мови. Це стимулюється практичними потребами, пов'язанними із створенням систем воєнного і комерційного призначення. Можна зазначити, що тільки в Європі об'єм продаж систем цивільного призначення складає кілька мільярдів доларів. При цьому слід звернути увагу на те, що в практичному використанні відсутні системи, які вважаються вершиною розвитку автоматичного розпізнавання мови.

Актуальність

Проблема розпізнавання мови на сьогоднішній день вважається надзвичайно серйозною і відіграє надзвичайно велику роль у спілкуванні людини з машиною. Управління об'єктами за допомогою мови відкрило б широкі перспективи перед автоматизацією у багатьох галузях людської діяльності, відкрило б можливість спілкування з машинами, особливо користувачів персональних комп'ютерів, не знаючих мов програмування. Мовний контакт полегшує запис даних у машину, допомагає працювати людині і комп'ютеру в реальному масштабі часу: людина сказала – машина виконала.

Аналіз аналогів

На даний момент на ринку представлені наступні основні системи, які використовуються для автоматичного розпізнавання мови:

- *Dragon NaturallySpeaking*
- *IBM ViaVoice Gold*
- *L&H Voice Xpress Professional*
- *Philips FreeSpeech* 2000

Вони вважаються найкращими, але ні одна із них не є ідеальною, основні їхні недоліки:

– рівень безпомилковості розпізнавання мови не перевищує 85%;

– нерівномірна якість розпізнавання;

– низька якість розпізнавання власних назв і скорочених слів, повільна робота в середовищі деяких програм;

– затрата великого часу для настройки системи.

Наша мета розробити систему, яка буде максимально наближена до ідеальної, з високим рівнем безпомилковості, простою у використанні, швидкою настройкою.

Основні проблеми

Ідея створення системи керування електронним пристроєм, що базується не тільки на тактильній взаємодії людина-машина, але і на голосовому керуванні не нова.

Комп'ютерні системи розпізнавання мови поступово знаходять застосування не тільки в науковій сфері, але й у побутовий. Прикладом тому можуть служити офісні пакети й інше ПО з убудованим розпізнаванням мови дляголосового введення текстової інформації. Що ж стосується портативних пристроїв, то в них лише зараз починають упроваджува-

тися технології розпізнавання мови. На сьогоднішній день в умовах глобального розвитку інформатизації, конвергенція технологій голосового керування з мультимедійними функціями портативної і побутової техніки обумовлює науково-технічний прогрес у створенні нової функціональності передової техніки. Однією з проблем упровадження голосових технологій у портативній техніці був низький обчислювальний ресурс мікропроцесорів, і недостатній обсяг оперативної пам'яті. Крім цього алгоритми з достатньою надійністю розпізнавання мови в умовах складної шумової обстановки навколишнього середовища були занадто ресурсоємні для портативного застосування.

Існуючі сьогодні системи розпізнавання мови ґрунтуються на зборі всієї доступної (часом навіть надлишкової) інформації, необхідної для розпізнавання слів. Дослідники вважають, що в такий спосіб завдання розпізнавання зразка мови, засноване на якості сигналу, підданого змінам, буде достатнім для розпізнавання, але, проте, у цей час навіть при розпізнаванні невеликих повідомлень нормальної мови, поки неможливо після одержання різноманітних реальних сигналів здійснити пряму трансформацію в лінгвістичні символи, що є бажаним результатом.

Для того щоб машина навчилась розуміти людську мову, відповідати на питання потрібно затратити багато сил і часу, забиваючи її гігантською інформацією тільки для того, щоб розпізнати окремі звуки. У кожного звуку складна структура, яка включає в себе різні частоти і коливання, до того ж, те саме слово різні люди вимовляють по-різному: різний тембр голосу, різні інтонації, різна чистота вимови. Скільки людей, стільки й голосів. Голос – індивідуальна ознака особистості.

Щоб навчити машину впізнавати мову, її потрібно заставити прослуховувати слова, сказані як однією людиною, так і різними людьми. Задача машини – прослухавши всі дані, взяти середні значення особливостей вимови, повністю виключити індивідуальність, щоб потім, почувши слово, не зробити помилку.

Найбільші проблеми виникають в умовах: довільний користувач; спонтанна мова, яка супроводжується мовним «сміттям», наявність акустичних завад і скривлень; наявність мовних завад.

Для спрощення процесу розпізнавання мови доцільно було б використовувати шаблони окремих звуків єдині для всіх дикторів. На даний час таких шаблонів не існує через те, що не виявлено інформативних ознак звуків, які не залежать від характерних особливостей голосу.

Тому для реалізації ефективних дикторнезалежних систем автоматизованого розпізнавання мови необхідно виділити інформативні ознаки звуків мови, розробити математичні методи їх опрацювання з метою створення єдиних для всіх дикторів шаблонів.

У такому випадку система розпізнавання не буде потребувати навчання (створення набору шаблонів окремо для кожного диктора), її швидкодія збільшиться, оскільки відпаде необхідність створення набору шаблонів слів і з’явиться можливість розпізнавати мову незалежно від характерних особливостей голосу диктора.

Три типи задач систем розпізнавання мови

На сьогоднішній день розпізнавання мови зводиться до вирішення трьох типів задач:

1. Розпізнавання окремо вимовлених слів.
2. Розпізнавання злитої мови.
3. Ідентифікація по зразку мови.

Розпізнавання окремих слів по більшій степені використовується для мовного управління обчислювальною машиною.

Метою розпізнавання злитої мови є перетворення в текст звичайної мови людини.

Механізм розпізнавання для перших двох типів.

Для цих двох типів задач механізм розпізнавання мови буде виглядати так, як на рис. 1.

Рис. 1. Механізм розпізнавання для перших двох типів

Тут можна виділити 4 основних модуля:

- модуль збору даних;
- екстрактор;
- компаратор;
- інтерпретатор.

Модуль збору даних включає в себе отримання вхідного сигналу і його попередню обробку, яка може включити автоматичний регулятор посилення, приглушення еха, виявлення присутності або відсутності мови і виявлення інтонаційного кінця фрази.

Цей модуль також включає в себе виділення відрізка мови із вхідного сигналу. Існує декілька алгоритмів визначення початку і кінця мови. В одному із них визначається деякий граничний рівень сигналу. Початкова точка мови в цьому випадку відповідає моменту, коли вхідний сигнал починає перевищувати граничний рівень, а кінцева точка – моменту, де амплітуда вхідного сигналу менша граничної.

Другий метод використовує нормалізацію амплітуди вхідного сигналу у відповідності з мінімальною амплітудою. Отримані нормалізовані значення зрівнюються з граничним значенням.

Екстрактор виконує частотний аналіз сигналу. Акустично-фонетичний потік даних розбивається на короткі кадри, або вектори, тривалістю, як правило, біля 10 мс. Як правило, для кожного кадру визначається ряд параметрів, використовуючи швидке перетворення Фур'є. Крім того можна ще використовувати й інші характеристики, наприклад, спектральні.

Компаратор здійснює акустичні порівняння: кожен кадр, або вектор, порівнюється з акустично-фонетичними зразками, які зберігаються в спеціальній базі даних. При цьому можуть порівнюватись як окремі фонеми, так і слова, і навіть фрази. При невеликій кількості слів, використовуваних диктором, більш високу надійність і швидкість можна очікувати від розпізнавання цілих слів, але при збільшенні словника швидкість різко падає, і оптимальним стає розпізнавання окремих фонем.

В основному використовується три алгоритми для розпізнавання кадрів:

– алгоритм динамічної трансформації шкали часу;

– приховане Марківське моделювання;

– нейронна мережа з часовою затримкою.

Алгоритм динамічної трансформації шкали часу використовує оптимізаційний принцип для мінімізації кількості помилок, виникаючих при порівнянні розпізнаваного слова з еталонною моделлю.

Приховане Марківське моделювання використовує імовірнісні моделі слів. При використанні цієї технології для кожного можливого варіанта слова, яке розпізнається, вичислюється ймовірність, потім отримані ймовірності порівнюються і вибирається слово з найбільшою ймовірністю.

Нейронна мережа с часовою затримкою у випадку розпізнавання обмеженої кількості слів дає кращі результати ніж метод прихованого Марківського моделювання.

Один із методів, оснований на порівнянні фонем, використовує поняття «контекстна фонема». В даному методі фонема розглядається в поєднанні з попередньою і наступною фонемою. Далі в процесі розпізнавання визначається фонема, яка найбільше близько відповідає тій, яка розпізнається.

Інтерпретатор вирішує задачу динамічного програмування з метою знайти найкраще розбиття отриманого від ком-

паратора алфавітного потоку на слова і фрази. В залежності від об'єму використовуваного словника і діючих синтаксичних правил, застосовуються різні стратегії пошуку і відсіювання.

В даному блоці із розпізнаних фонем формуються слова, а із слів фрази. При цьому також часто використовується ймовірна система порівняння результатів.

Ідентифікація по зразку мови використовується для досягнення забезпечення безпеки. Вона складається із трьох стадій:

– регістрація;
– тестування;
– допуск.

Механізм розпізнавання для третього типу

Схема ідентифікації по зразку мови представлена на рис. 2.

Рис. 2. Механізм розпізнавання для третього типу

В процесі регістрації користувача запам'ятовуються особливості його голосу і формується так звана мовна модель. При тестуванні виконується порівняння запропонованого зразка мови із запам'ятованою мовною моделлю користувача, а також з моделлю «самозванця», складеною на базі голосів інших людей. Якщо результат порівняння виявиться позитивним для першого випадку і негативним для другого, можна вважати , що тестування пройшло успішно. Ідентифікацію по голосу можна використовувати і в поєднанні з іншими засобами забезпечення безпеки.

Системи розпізнавання мови можуть також поділятися на:

– диктороорієнтовані;
– дикторонезалежні.

Системи першого типу потребують наявності етапу «навчання», тобто налаштування системи на конкретного користувача, якому необхідно промовити визначений набір слів для того, щоб еталонні моделі його вимови були занесені у базу даних. Згодом при розпізнаванні мови цього користувача система опирається на еталонні моделі, які зберігаються у базі даних.

У випадку використання системи іншим користувачем необхідне повторне навчання.

Висновки

Приведено аналіз існуючих методів розпізнавання мови людини. Були розглянуті основні типи задач, які не можуть існувати без систем розпізнавання. Теоретично обґрунтовано моделі і методи аналізу та розпізнавання сигналів багатьох змінних. Запропоновано методи, алгоритми і обчислювальні процедури аналізу сигналів на основі параметричних функцій систем, які створюють сигнал.

На сьогоднішній день існує багато методів вирішення цих проблем, але ні один метод не є ідеальним, їхня точність не перевищує [85%]. Наша мета добитися результатів, які максимально будуть наближені до ідеальних.

Список літератури

1. *Оппенгейн А.В., Шафер Р.В.* Цифровая обработка сигналов – М.: Радио и связь, 1979. – 347 с.

2. *Кузнецов В., Отт А.* Автоматический синтез речи. – Таллинн: Валгус, 1989. – 135 с.