

УДК 004.934

Ст. викладач Мальчиков В.В., студент Ільченко Б.І.

**Національний технічний університет України
«Київський політехнічний інститут»**

ПЕРЕТВОРЕННЯ МОВНИХ СИГНАЛІВ У ТЕКСТ

Abstract

V.V. Malchikov; Bogdan Ilchenko, student

Development system of automation converting linguistic signals into text

This work deals with researching of methods for recognition structured audio - signal and development the system of automatic converting linguistic signals into text.

Вступ

Усна мова і сьогодні залишається одним з найоперативніших і поширених способів передачі інформації в будь-якій сфері людської діяльності. Це продуктивний, природний і зручний спосіб передачі інформації. У сучасних комп'ютерних системах все більше уваги приділяється побудові інтерфейсу мовного вводу-виводу, ефективність якого основана на практично необмежених можливостях формулювання на природній мові всіляких завдань в самих різних областях людської діяльності. Системи мовного вводу є дуже перспективними на сьогоднішній день.

Існуючі системи розуміння мови поки що значно поступаються мовним здібностям людини, що свідчить про їх недостатню адекватність і обмежує вживання мовних технологій в промисловості і побуті [1].

Постановка задачі

Основна мета роботи полягає в дослідженні методів і алгоритмів перетворення мови в текст, а також у виборі методу для подальшої розробки програмного продукту.

Термінологія

Фонема — вся сукупність звуків, які мають однакові функції і не створюють змістових відмінностей в даній мові.

Частота́ дискретизації (англ. *sample rate*) — кількість сигналів за секунду (або за іншу одиницю) при перетворенні безперервного сигналу в дискретний сигнал (тобто його дискретизації). Як правило, частота дискретизації вимірюється в герцах (Гц). Поняття частоти дискретизації може бути застосованим лише в тому випадку, коли квантування здійснюється зі сталою періодичністю.

Розрядність перетворення — кількість одиниць інформації, яка кодує один відлік [2].

Загальна ідея

Розробка ефективних алгоритмів розпізнавання мови є ключовим моментом в рішенні задачі. Процес розпізнавання мови включає дві основні складові: лінгвістичну і акустичну. Яким чином мова (акустичний сигнал) перетвориться в текст (лінгвістична складова) залежить від використовуваного алгоритму.

Лінгвістична складова

Мовний сигнал

Акустична складова

Текст

Акустична складова виконує:

1. Перетворення мовного сигналу в цифрову форму, виділення фонем із слів;
2. Попередню обробку мовного сигналу (нормалізація, шумо-заглушення);
3. Обчислення ознак мовних одиниць.

Лінгвістична складова системи перетворення мови в текст включає етапи на яких залучається апріорна інформація, характерна для використовуваної мови:

1. Створення бази знань вірогідності поєднань фонем, що накладають обмеження на структуру слова, підкоряючись особливостям розпізнаваної мови;
2. Розробка бази словоформ мови;
3. Формування граматичної форми слова [3].

В якості алгоритму перетворення результату роботи акустичної складової системи, для лінгвістичної складової може виступати нейронна мережа. На даному етапі визначається приналежність обчислених ознак мовних одиниць до певного фонетичного класу.

Акустична складова системи

Для аналізу мови її необхідно перетворити у форму, зрозумілу обчислювальній системі. Це може бути аналогова форма, цифрова форма, спектральна вистава, вистава у вигляді оптичного випромінювання і т.д. Оскільки в роботі зачіпається дослідження мови на персональному комп'ютері, то розглядається лише один вигляд представлення мовного сигналу - в цифровій формі. Як відомо, звук є подовжніми хвилями розрідження-стискування, що поширюються в акустично-провідному середовищі. За допомогою звукозаписних пристроїв (наприклад, мікрофону) він перетвориться в електричний сигнал, коливання якого повторюють звукові коливання. Потім цей сигнал фільтрується з метою відсікання частот, що перевищують деяку частоту. Після цього він подається на аналого-цифровий перетворювач, який з деякою частотою (частотою дискретизації), записує поточний рівень сигналу в цифровій формі, тобто квантує сигнал за часом і по амплітуді.

Таким чином, параметрами, що визначають якість оцифрування сигналу, є *частота дискретизації* і *розрядність перетворення* (скільки одиниць інформації кодується один відлік). Частота дискретизації визначає максимальну частоту сигналу, яку можна записати. Типові значення - 11025, 22050, 44100 Гц. Від розрядності залежить точність кодування інформації при аналого-цифровому перетворенні. Стандартні значення - 4 біт, 8 біт, 16 біт на відлік. Природно, чим більше розрядність і частота дискретизації, тим точніше записується звук, але і тим більше потік інформації і тим складніше його обробити [4].

Попередня обробка сигналу

Вхідний мовний сигнал необхідно підготувати для подальшої обробки. На даному етапі здійснюється очищення сигналу від шуму, видалення пауз, нормалізація сигналу. Також проводиться розбиття сигналу на фіксовані інтервали (сегменти) в тимчасовій області, на яких розраховуватимуться описові ознаки. Тривалість цих сегментів вибирається, виходячи з таких міркувань: по-перше, вони мають бути значно менше тимчасових інтервалів, що витрачаються на нормальне вимовлення голосних і приголосних звуків; по-друге, вони повинні давати можливість виділяти характерні ознаки. Надалі ці сегменти можуть бути віднесені до однієї з фонетичних категорій.

Таким чином, маємо приблизний алгоритм для перетворення мовного сигналу у текст:

1. Лінгвістична складова

- а) Перетворення мовного сигналу в цифрову форму;
 - б) Попередня обробка мовного сигналу;
 - в) Обчислення ознак мовних одиниць.
- 2. Мовний сигнал
 - 3. Акустична складова
 - а) Створення бази знань вірогідності поєднань фонем;
 - б) Розробка бази словоформ мови;
 - в) Формування граматичної форми слова.
 - 4. Текст

Висновки

В данній статті були розглянуті методи і алгоритми системи перетворення мови в текст. Програмне забезпечення, що буде розроблено на основі даних алгоритмів, дозволить автоматизувати процес введення текстової інформації в ЕОМ, проводити експериментальні дослідження в області розпізнавання мови. База даних фонем мови може бути використана для розробки і експериментальної оцінки якості роботи мовних застосувань.

Надалі можливе визначення нових шляхів перетворення мовних сигналів, котрі будуть використовувати методи, що даватимуть точніші результати.

Література

- 1. *Сорокин В.Н., Цыплихин А.И.* Сегментация и распознавание гласных. // Информационные процессы. – 2004. - №2. – с.202-220.
- 2. <http://ru.wikipedia.org/>
- 3. *Медведев Максим Сергеевич.* Преобразование речи в текст. Фонемный подход : диссертация... канд. техн. наук : 05.13.17 Красноярск, 2007 – 111с.
- 4. *Назаров М.В., Прохоров Ю.Н.* Методы цифровой обработки и передачи речевых сигналов.- М.: Радио и связь, 2005 – 280с.