

АВТОМАТИЗОВАНА СИСТЕМА КОНТРОЛЮ ТЕМПУ МОВИ В КОНТАКТ-ЦЕНТРАХ

Вінницький національний технічний університет

Анотація

В даній роботі досліджено можливості автоматичного визначення і контролю темпу мови на основі інформації про тривалість складів і пауз в мовному потоці, розроблено метод і алгоритм надійного виділення складових сегментів, визначення їх кількості і тривалості.

Ключові слова: Темп мови, розпізнавання мови, мовний сигнал.

Abstract

The ways of possibility of automatic detection and control rate of speech based on the duration of syllables, pauses in the speech flow developed a reliable method, and algorithm for allocation of component segments to determine their number and length investigated in this work.

Keywords: Rate of speech, recognition of language, speech signal.

Вступ

Актуальність роботи полягає у вирішенні проблем визначення параметрів голосового зв'язку в комунікаційних каналах, що стоїть на порядку денному в багатьох дослідженнях, пов'язаних з мовними технологіями. Особливо актуальною є задача визначення такого показника, як темп мови. Контроль цього параметру в системах обробки телефонних викликів (контакт-центрах) дозволяє оптимізувати темп діалогу оператора з клієнтом і підвищити ефективність трафіку та комфортність обслуговування клієнта.

Під темпом мови розуміють швидкість вимовлення елементів мови – звуків, складів чи слів. Середній темп мови складає 120 слів на хвилину, або 240-260 складів/хвилину [1]. Найбільш відповідною мірою темпу мови вважають середню тривалість мовного складу, або кількість складів в одиницю часу (наприклад, в секунду чи хвилину).

Таким чином, визначення темпу мови вимагає сегментації безперервного потоку мови на склади і вимірювання їх кількості в одиницю часу. Сегментація мови на склади на тлі шумів, характерних для телефонної лінії, вимірювання їх тривалості - одна з основних задач розпізнавання мови. В роботі [4] досліджено, що використання, наприклад, тільки складової інформації дозволяє уже на верхньому рівні розпізнавання мови скоротити кількість кандидатів на класифікацію в 2 – 4 рази. Відомий цілий ряд алгоритмів, що використовують традиційну обробку мовного сигналу в частотній або часовій області, які виділяють формантні характеристики [5], [6]. Як альтернативні методи використовуються приховані марковські моделі (ПММ).

Одним з основних параметрів, які використовуються для розмежування складів у мовному сигналі, є його енергія [2,5]. При цьому ядро складу визначається в місці локалізації максимумів енергії, обмежених істотними (на 40 або 50 дБ) спадами енергії. Однак у ряді зазначених робіт, наприклад [6], відзначається часте виділення за цією ознакою помилкових складів, сформованих високоенергетичними фрикативними або сонорними звуками. Це підтверджують і експериментально зняті записи. У роботі [7] як параметр для виділення ознаки складу використовується функція "гучності", одержувана як зважена сума амплітуд сигналів 22-х частотних каналів, розміщених у критичних смугах. Очевидний недолік такого методу формування ознаки складу – великі апаратурні або обчислювальні витрати та недостатня надійність.

Метою роботи є розробка автоматизованої системи контролю темпу мови на основі ефективного методу надійного виділення складових сегментів, призначеної для управління швидкістю мовлення в контакт-центрах.

Результати досліджень

Для досягнення поставленої мети в даній роботі автори використали комбінацію алгоритмів, заснованих на обробці мови в часовій області.

В ході роботи було розроблено новий метод виділення ознак складових сегментів, для формування яких за первинні параметри використовуються огинаючі сигналів у частотних діапазонах $\Delta_1 = 800 - 2500$ Гц і $\Delta_2 = 250 - 540$ Гц. Результуючий параметр, який в подальшому використовується для виділення ознак складів, отримується кореляційним методом і записується у вигляді:

$$U_c(t) = U_{\Delta_1}(t) \cdot U_{\Delta_2}(t), \quad (1)$$

де $U_{\Delta_1}(t)$ – огинаюча енергії в смузі частот Δ_1 , а $U_{\Delta_2}(t)$ – огинаюча енергії в смузі Δ_2 .

Діапазон частот першого смугового фільтра, рівний 250 – 540 Гц, вибраний з огляду на те, що в ньому відсутня енергія високоенергетичних фрикативних звуків типу /ш/ і /ч/, які створюють помилкові складові ядра, а так само зосереджена значна частина енергії всіх дзвінких звуків, у тому числі і голосних. Проте в цьому діапазоні енергія сонорних звуків типу /л/, /м/, /н/ співвідносна з енергією голосних, тому визначення складових сегментів тільки за огинаючою мовного сигналу в цьому діапазоні супроводжуватиметься багатьма помилками. Тому діапазон частот другого смугового фільтра, вибраний в межах 800 – 2500 Гц, в якому енергія голосних звуків мінімум в два рази перевищує енергію сонорних звуків.

При виконанні операції множення огинаючих в результирующей часовій функції відбувається посилення ділянок кривої у області голосних звуків через кореляцію їх енергій в обох діапазонах, а помилкові максимуми енергії, зумовлені наявністю в діапазоні 800 – 2500 Гц значної частини енергії фрикативних звуків, усуваються їх множенням на практично нульове значення амплітуди фрикативних звуків в діапазоні 250 – 540 Гц.

Схема алгоритму визначення темпу мови, який підраховує кількість складів і пауз в мовному сигналі за одиницю часу, показана на рисунку 1.

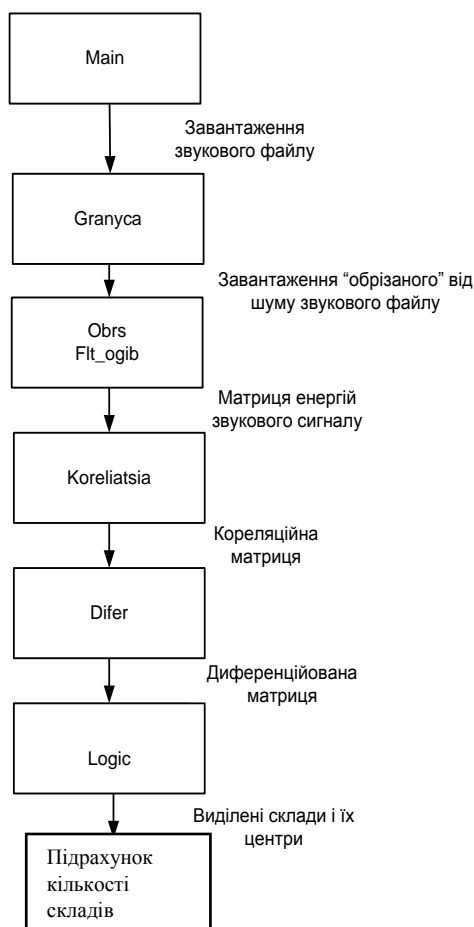


Рисунок 1 – Схема алгоритму визначення темпу мови

Обґрунтуємо дану схему. Звуковий файл завантажується за допомогою програми Granusa і в ході роботи цієї підпрограми знаходяться межі початку і кінця слова. Створюється файл, який в собі містить відділене від шуму слово.

Створений файл завантажується в програмі Obrs. За допомогою програми Flt_ogib він двічі фільтрується. Отримуємо відфільтрований двома фільтрами сигнал, який є матрицею енергій звукового сигналу від часових відліків з періодом дискретизації 12 мс. Далі в програмі Koreliatsia енергії перемножуються і в результаті отримуємо кореляційну матрицю по всім енергіях та відліках часу. Потім методом різницевого диференціювання кореляційна матриця в програмі Difer диференціюється, і обчислюється поріг. Після чого в програмі Logic відбуваються логічні перетворення диференційованої матриці кореляції. На останньому етапі підраховується кількість складів в слові за одиницю часу, і визначається темп мови.

Для тестування запропонованого методу було проведено експеримент, який полягав у сегментації 650 складів з їх використанням для обчислення темпу мови. Статистична обробка експериментальних даних дозволила розрахувати надійність даного методу, яка склала 96,4%. Надійність інших методів при еквівалентній тестуючій вибірці не перевищувала 76%.

Висновки

В результаті виконання роботи були створені алгоритми і методи автоматичного контролю темпу мови, що базуються на уточненій моделі слухової системи людини, дозволяють підвищити надійність сегментації мовного сигналу на складові сегменти. Одержані алгоритми та методи дозволяють вирішити низку задач:

- розробка методу, алгоритму і пристрою надійного виділення складів в сигналі мовлення;
- розробка алгоритму для автоматичного обчислення темпу мови;
- розробка програмного забезпечення для реалізації розроблених алгоритмів автоматичного контролю темпу мови.

Подальший вектор досліджень спрямований на створення автоматизованої системи корегування темпу мови клієнтів контакт-центрів .

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Лингвистический энциклопедический словарь // Ред. В.Н. Ярцева. – М.: Советская энциклопедия, 1990. – 688 с.
2. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. М.: Радио и связь, 1981. С. 113–119.
3. Сапожков М.А., Михайлов В.Г. Вокодерная связь. М.: Радио и связь, 1983. С. 156–158.
4. N.M. Bykov, I.V. Kuzmin, A.I. Yakovenko. Development of effective strategy of pattern recognition. – Proceedings of SPIE, 2001, Vol. 4225, pp.76 – 83.
5. Биков М.М., Гришук Т.В. Методи підвищення дикторонезалежності опису і розпізнавання мовної інформації в мережі INTERNET // “Інтернет – Освіта – Наука – 2002”, третя міжнародна конференція ІОН – 2002, 8 – 12 жовтня 2002 р. Збірник матеріалів конференції. – Вінниця: УНІВЕРСУМ – Вінниця, 2002. – Том 2.– С. 329 – 332.
6. Методы автоматического распознавания речи / Под ред. У. Ли.– М.: Мир, 1983. – Т.1. – 200 с.
7. Ruske C., Schotola F. An approach to speech recognition using syllabic decision units. – Proc. 1978, IEEE ICASSP, Tulsa, 1978. – N.Y., 1978, pp. 772 – 725.

Науковий керівник: **Биков Микола Максимович** – к.т.н., доцент, професор кафедри комп'ютерних систем управління, Вінницький національний технічний університет

Кудлаєнко Марія Ігорівна – студентка групи 1АВ-12б, факультет комп'ютерних систем та автоматики, Вінницький національний технічний університет, м. Вінниця, mary1995kudlaenko@gmail.com

Supervisor: **Mykola Bykov** – Ph.D., Associate Professor at the Department of Computer Control Systems, Vinnitsa National Technical University

Mariia Kudlaienko – group 1AV-12, Faculty of Computer Systems and Automatics.