

МЕТОД ВИЗНАЧЕННЯ ДІЛЯНОК ЗМІНИ ДИКТОРА В АУДІО ЗАПИСАХ ІНТЕРВ'Ю

Навроцький Є. В.

Науковий керівник – доц., к.т.н. Гришук Т. В.

Мовні технології – технології ХХІ століття, завдяки яким з'явилась можливість управляти комп'ютером за допомогою голосу, диктувати текст, слухати книги, спілкуватися з комп'ютером на інтелектуальному рівні.

В доповіді розглядається метод розв'язання задачі визначення ділянок зміни диктора в аудіо записах інтерв'ю. Розв'язання цієї задачі дозволить значно збільшити не тільки зручність обробки аудіо матеріалів, а також надасть можливість зменшувати розміри записів за рахунок видалення неінформативних ділянок тиші.

Авторами пропонується метод, що умовно складається з двох етапів.

На першому етапі пропонується виконувати попереднє маркування мовного сигналу з метою знаходження ділянок умовної тиші. Для виконання цього етапу можна використовувати класичний метод визначення границь мови. Даний алгоритм ґрунтується на вимірюванні двох простих характеристик: енергії сигналу та числа переходів через нуль.

На другому етапі пропонується аналізувати ділянки, що на першому кроці позначені як закінченні тиші, щоб визначити чи відбулась зміна диктора. Для цього використовуються такі характеристики мовного сигналу, як сила, основний тон та темп сигналу. Тут, як і на першому етапі, можна використовувати алгоритм пошуку з поверненням. Опишемо коротко основні кроки цього алгоритму: 1) формуємо перший вектор ознак $X = \{x_1, x_2, \dots, x_n\}$ для визначення ділянок зміни диктора при прямому проходженні та другий вектор $Y = \{y_1, y_2, \dots, y_m\}$ для визначення ділянок зміни диктора при зворотному проходженні; 2) для кожного маркера, координата якого визначена на першому етапі такою, що позначає можливий початок зміни диктора, рухаючись вправо, знаходимо перший семпл, на якому спостерігається різка зміна ознак першого вектора; 3) рухаючись в оберненому напрямку, визначаємо точку, в якій спостерігається різка зміна ознак другого вектора – точка зміни диктора.

В розробленому програмному комплексі розглянуто перший етап, а саме визначення початку і кінця фрази. Для моделювання використано середовище MatLab. За результатами дослідження розробленого ПЗ можна зробити висновок, що алгоритм дає 95% точності визначення ділянок, на яких може відбуватись зміна диктора. В подальшому авторами планується виконання додаткових досліджень характеристик мовного сигналу, що входять у перший та другий вектори ознак, які дають найбільш точний результат знаходження ділянок зміни диктора.