

## АНАЛІЗ ПІДХОДІВ ДО РОЗПІЗНАВАННЯ МОВЛЕННЯ ПРИ ПОБУДОВІ ФОНЕМНИХ ВОКОДЕРІВ

Дзісь О. В.

Науковий керівник – доц., к.т.н. Ткаченко О. М.

Ущільнення мовленнєвих сигналів під час їх передавання дозволяє підвищити ефективність використання частотної смуги каналів зв'язку. Відомо, що найвищий ступінь ущільнення досягається при застосуванні фонемних вокодерів, здатних забезпечити передавання мовленнєвого потоку зі швидкістю 300 біт/с. При цьому необхідно розв'язати дві достатньо складні задачі:

- 1) розпізнати у мовленнєвому потоці деякі акустичні одиниці (фонеми);
- 2) синтезувати на їх основі якісний мовленнєвий сигнал.

У роботі розглянуто шляхи підвищення фонетичної коректності (ФК) розпізнавання. Для дослідження було обрано три найбільш поширені параметричні моделі представлення мовленнєвих даних:

- коефіцієнти лінійного прогнозування (LPC);
- лінійні спектральні частоти (LSF);
- кепстральні коефіцієнти (CC).

Для аналізу методів обчислення відстаней було обрано такі метрики:

- евклідова метрика (EM);
- зважена евклідова метрика (ЗЕМ);
- дві спеціальні метрики для LPC (СМ1 та СМ2 відповідно);

Для проведення дослідження було використано матеріал англійської бази даних ТІМІТ, яка включає достатньо великий набір дикторів та фраз, розмічених на фонемі.

Дослідження параметричних моделей показало, що використання кепстральних коефіцієнтів дає кращі результати ніж LPC та LSF в тих же умовах. При аналізі метрик найвищий показник (58%) було досягнуто при використанні ЗЕМ для кепстральних коефіцієнтів. Найкращий результат для евклідової метрики склав лише 40%. Спеціальні метрики для моделі LPC виявились не достатньо ефективними: 33% та 47% для СМ1 та СМ2 відповідно.

Складність розрахунків для обчислення ЗЕМ значно підвищилась порівняно з EM: кількість додавань зросла у 6 разів, кількість множень - в 11 разів. Таким чином, підвищення ФК при використанні ЗЕМ досягнуто за рахунок суттєвого зростання обчислювальних витрат.

Для підвищення показника ФК запропоновано використовувати методи кластеризації. Це дозволило підвищити ФК на 11-27% для різних метрик та параметричних моделей, при цьому найвище значення - 78% досягнуто при використанні ЗЕМ для кепстральних коефіцієнтів.

Отримані результати було застосовано для розробки програми автоматизованої сегментації акустичних мовленнєвих баз даних.