

**КЛАСТЕРНИЙ АНАЛІЗ ФОНЕМ МОВНИХ ПОВІДОМЛЕНЬ**

**Актуальність теми.** Головним завданням системи фонетичного розпізнавання мови є розбивка безперервного мовного сигналу на окремі фонемі і їх подальше розпізнавання. Точність сегментації - один з чинників надійності і вірогідності автоматичного розпізнавання мови. Ручна сегментація вимагає значних витрат сил і часу, і вона піддана помилкам. Крім того, внаслідок мінливості людського зорового й слухового сприйняття практично неможливо відтворити її результати. Автоматична сегментація не безпомилкова, однак вона несуперечлива по своїй суті, і результати такої сегментації можна відтворити. Ідеальна система розпізнавання повинна сегментувати безперервне мовлення будь-якого диктору і будь-якою мовою.

**Аналіз останніх досліджень.** Існує два основних типи алгоритмів сегментації мови [1,2]. До першого типу належать алгоритми, які роблять сегментацію мовного сигналу за умови, що відомо послідовність фонем даної фрази. Інший тип алгоритмів не використовує апріорної інформації про фразу, і при цьому границі сегментів визначаються по ступені зміни акустичних характеристик сигналу. Існують і інший тип алгоритмів, які приймають рішення як на основі апріорної інформації, так і на основі зміни акустичних характеристик.

**Постановка задачі.** Метою даної роботи є проведення експериментального дослідження значень формант спектра ударних голосних української мови за допомогою кластерного аналізу, порівняльний аналіз отриманих значень із формантами голосних російської мови.

**Теоретичне обґрунтування.** Обробка мовного сигналу. Голосовий тракт людини - це неоднорідна акустична трубка. Площа поперечного переріза трубки неоднакова й залежить від положення артикуляторів. Голосовий тракт має деякі стійкі види резонансних коливань, а саме форманти, які істотно залежать від положення артикуляторів. При аналізі й синтезі мови авторегресійна модель мовного сигналу

зарекомендувала себе як модель, що добре описує поведження мовного тракту при його описі у вигляді динамічної труби [3]. Вихідний мовний сигнал  $x(n)$  моделюється як лінійна комбінація попередніх відліків сигналу.

Рівняння, що описує авторегресійний процес, визначається наступною формулою:

$$x(n) = \sum_{i=1}^P a(i)x(n-i) + \xi(n), \quad (1)$$

де  $x(n)$  – масив даних розміром  $n$ ;  $a(i)$  – авторегресійні коефіцієнти процесу;  $P$  – порядок процесу;  $\xi(n)$  – процес, що породжує, типу гаусовського білого шуму. Параметри авторегресійної моделі мовного сигналу, а саме її вектор коефіцієнтів  $a(i)$ , обчислюються на основі мінімізації помилки лінійного проорокування по її дисперсії. Далі аналізується спектр сигналу, отриманого на основі коефіцієнтів лінійного проорокування.

Голосні звуки відіграють важливу роль у мовленні. Вони мають найбільшу енергію, що відповідає кращому відношенню сигнал/шум. Голосні характеризуються розташуванням та енергією формант. У спектральній області форманти звичайно виглядають як локальні максимуми енергії. Типові значення частот перших трьох формант для голосних російської мови наведені в таблиці 1.

Таблиця 1

Значення формантних частот голосних звуків російської мови, Герц

звук	$F_1$	$F_2$	$F_3$
А	600	1200	2400
Э	450	1700	2500
О	450	800	2500
У	350	750	2200
Ы	300	1800	2450
И	300	2200	3000
Е	400	1850	2600
Я	450	1500	2450

При оцінюванні формант можливою комбінацією піків будемо вважати набір піків спектра із частотами, розташованих строго по зростанню, тобто  $F_i < F_{i+1}$ , і які лежать в інтервалах відповідної форманти. Крім того, частота першої форманти повинна бути строго

більше основного тону  $F_0$ . Таким чином, обмежувальні умови можна записати так:

$$\begin{cases} F_1 \in [150,900]; \\ F_2 \in [700,2700]; \\ F_3 \in [1800,4000]. \end{cases} \quad (2)$$

Тут частоти зазначені в герцах. Частотні інтервали обрані з урахуванням всіх можливих положень формант для чоловічого й жіночого голосів.

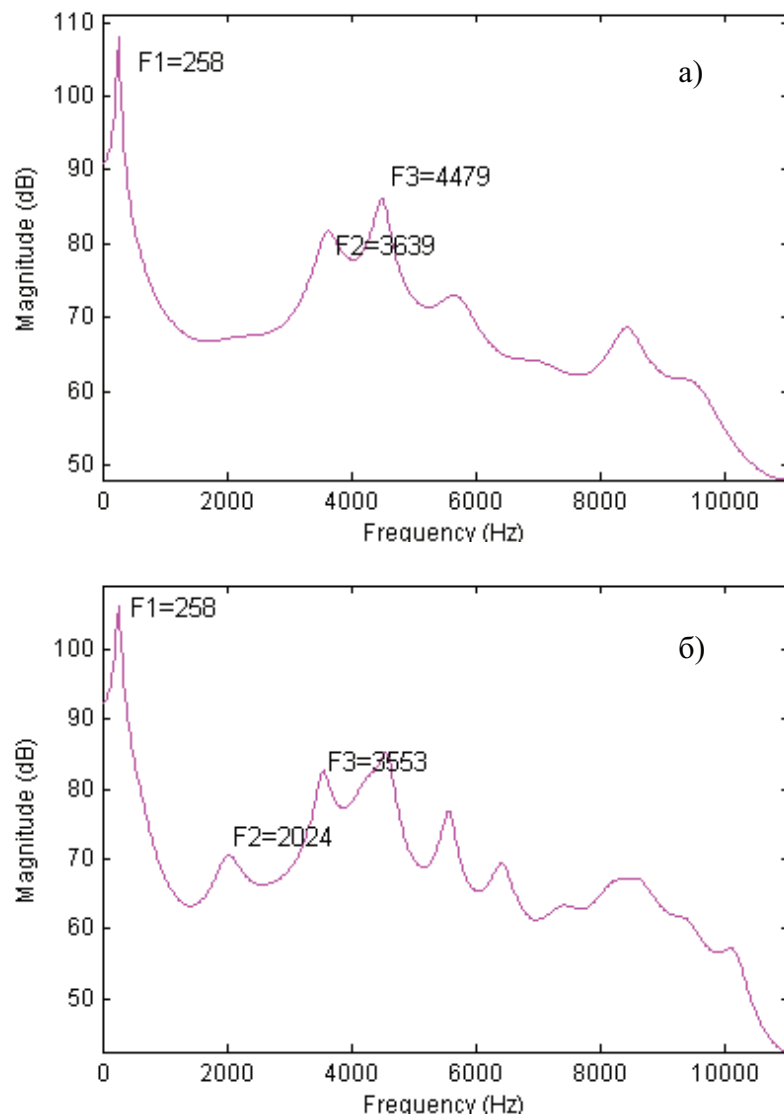


Рис. 1. Спектр сигналу звуку «І» на основі:  
а) 18 коефіцієнтів; б) 28 коефіцієнтів.

**Практична реалізація.** Для оцінювання голосних була створена мовна база даних українських слів, які вимовлялись 10 дикторами. Запис виконувався з використанням мікрофона в тихій кімнаті.

Частота дискретизації - 22050 Гц, кількість біт на відлік - 16. Розмітка мовних сигналів була виконана вручну. З метою визначення коефіцієнтів лінійного пророкування подальша обробка часового сигналу проводилася на ділянках довжиною 10 мс з використанням вагової функції Хемінга. Порядок авторегресійної моделі впливає на точність оцінювання формант. Звичайно для мовних повідомлень використовується порядок авторегресійної моделі - 18 [4], але його збільшення до 28 необхідно для більш точного визначення другої і третьої формант. Спектри сигналів, отриманих по 18 і 28 коефіцієнтам лінійного пророкування, наведені на рис. 1. Якщо використовується тільки 18 коефіцієнтів, другу форманту на частоті  $F_2=2024$  Герц неможливо визначити.

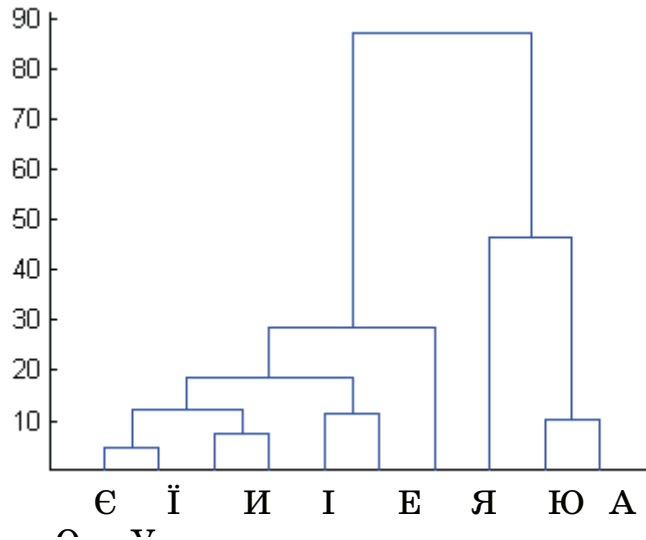
У кожній реалізації голосного звуку форманти можуть погано визначатися, або з'являться додаткові піки в спектрі. Причинами провалля піків можуть бути або втрати енергії в мовному тракті, або особливості амплітудно-частотної характеристики каналу мовного зв'язку, або високий рівень шумів. По цих же причинах можуть з'являтися й додаткові, "помилкові", піки. У даній роботі була почата спроба визначення «об'єктивно» існуючих класів у просторі формантних частот голосних звуків української мови. Із цією метою формантні частоти всіх ударних голосних були об'єднані. Потім за допомогою алгоритму К-середніх було проведено кластерний аналіз, тобто знайдені центри згущень у тривимірному розподілі. Кількість кластерів - це кількість голосних, що беруть участь у кластеризації. Результати оцінювання формант ударних голосних і ударних дифтонгів української мови наведені в таблиці 2.

Таблиця 2

Центри кластерів, отримані методом К-середніх

Звук	$F_1$	$F_2$	$F_3$
А	630	1233	2281
О	466	797	2169
У	286	711	2089
Е	517	1938	2246
Є	358	2013	2799
И	301	2304	2972
І	301	2138	3097
Ї	315	2161	2836
Ю	266	1637	2617
Я	463	1917	2491

Як видно з цієї таблиці, формантні частоти приблизно відповідають середньостатистичним значенням ударних голосних російської мови, зазначеним у таблиці 1. Можливо деяке підвищення значень частот, оскільки аналізувались в тому числі, і жіночі голоси. У жінок більш короткий мовний тракт, ніж у чоловіків. Для перевірки якості розбивки даних на кластери використовувався багатомірний дисперсійний аналіз ознак.



Процедуру дисперсійного аналізу можна застосовувати до рішень, отриманим будь-яким методом кластеризації, аби тільки він породжував розбивки. Результатом проведеного аналізу є кількість ознак, які є значимими. Дисперсійним аналізом було підтверджено, що саме три форманти дозволяють розбити спектри голосних звуків на відповідну кількість кластерів. Дендрограма зв'язків між центрами кластерів голосних звуків показана на рис.2. Центри кластерів звуків «О» і «У», «И» і «І», «Є» і «Ї» між собою розташовані досить близько. У результаті чого виникають помилки при розпізнаванні голосних звуків.

**Висновки.** Проведений кластерний аналіз дозволив визначити відповідність формантних частот голосних звуків української мови загальноприйнятим значенням. Знайдені центри кластерів можуть використовуватися для визначення, чи є даний мовний сегмент голосним звуком.

### ЛИТЕРАТУРА

1. В.Н.Сорокин, А.И. Цыплихин Сегментация и распознавание гласных // Информационные процессы, Том 4, № 2, 2004, стр. 202-220.
2. В.Н.Сорокин, А.И. Цыплихин Сегментация речи на кардинальные элементы // Информационные процессы, Том 6 , № 3, 2006, стр. 177-207.
3. Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов М.:«Мир», 1978, 848 с.
4. Loizou Ph. COLEA: A MATLAB software tool for speech analysis <http://www.utdallas.edu/~loizou/speech/>

Получено 30.11.06