

Світловський Є.В.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Трапезон К.О.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

АНАЛІЗ МОВНИХ АКУСТИЧНИХ СИГНАЛІВ В СИСТЕМАХ ЗВ'ЯЗКУ З ЧАСТКОВИМ ЗАШУМЛЕННЯМ

Метод лінійного кодування з прогнозуванням при практичному застосуванні в системах зв'язку є складовою частиною при створенні алгоритмів розпізнавання зашумленого мовного сигналу. На основі аналізу літературних джерел отримані підтвердження того, що метод лінійного кодування з прогнозуванням є зручним при створенні узагальненої дискретної моделі мовного сигналу в умовах зашумленого середовища. Розглянуто випадок, коли частина мовного сигналу знаходиться під окремою дією трьох різних видів шуму з однаковою амплітудою. Отримано, за допомогою інструментів програмного забезпечення PRAAT, енергетичні характеристики сигналу без шуму та з шумом, і на основі методу лінійного кодування з прогнозуванням знайдено коефіцієнти формантного фільтру на основі використання автокореляційної функції. За отриманими результатами виявлено, що присутність шуму призводить до зменшення усереднених значень енергетичних параметрів мовного сигналу. Причому, більше це помітно для рожевого шуму і менше для білого шуму. Крім цього виявлено, що при додаванні шуму до сигналу зменшується і середня інтенсивність сигналу (в середньому на 2-3 дБ), що може бути перешкодою при отриманні необхідної якості за розбірливістю мовного сигналу на приймальному кінці лінії зв'язку.

Додатково отримано підтвердження того, що присутність шумових компонент у мовному сигналі є напевно причиною зміни значень формантних частот моделі мовного сигналу. І ця перебудова не залежить від природи шуму, і його окремих характеристик. Тобто, можна зробити припущення, що на основі отриманих результатів, для обраного тестового мовного сигналу тривалістю близько 1 сек. метод лінійного кодування з прогнозуванням є цілком адаптивним до появи шумових складових.

Ключові слова: вокодер, акустика, сигнал, оброблення, метод, шум, прогнозування, відновлення.

Постановка проблеми. Технології з використанням мовної інформації стають все більш популярними у різних сферах життєдіяльності людини: охоронні системи, роботизовані комплекси, системи управління обладнанням та системи за типом «Розумний будинок», засоби телекомунікацій та зв'язку, тощо. У всіх цих сферах можуть бути наявними технічні реалізації аутентифікації об'єктів за голосом. Крім цього, дуже часто в системах мобільного і стільникового зв'язку мовні сигнали проходять через процедуру кодування. Кодування мовної інформації дозволяє отримати ущільнення каналу зв'язку задля забезпечення більшої пропускної здатності. Тобто, з'являється можливість збільшити об'єм мовної інформації, яка передається через канал зв'язку. Додатково, кодування мовних сигналів може бути використано і як засіб захисту інформації. Причому при організації радіоліній систем зв'язку, особливо коли йдеться про передавання та приймання акустичних мовних сигналів, дуже часто використовується саме низь-

кокошвидкісне кодування сигналів. Разом з тим, через помилки передавання і шуми, такий підхід може призвести водночас і до зменшення якості мовних сигналів при відтворенні на приймальному кінці лінії зв'язку. Тобто, вказані чинники можуть бути визначальними при досягненні потрібної якості мовних сигналів. Їх врахування особливо є важливим для систем зв'язку з використанням ідентифікації абонента за голосом, адже від цього залежить точність розпізнавання.

При створенні систем передавання мовних сигналів, особливо коли розглядаються технології передавання мови на основі Інтернет протоколу, використовуються різні пристрої синтезу мови, серед яких можна виділити вокодери. Реалізація таких пристроїв передбачає, що в якості однієї з ланок в схемі присутні кодери звукових та мовних сигналів, де при обробленні мовних сигналів дуже часто використовуються різні методи та алгоритми низькошвидкісного кодування. Розглядаючи ці методи необхідно враховувати, що при обробленні мовних

сигналів слід зберігати їх індивідуальні особливості. Насамперед це стосується таких параметрів, як частота основного тону, значення формантних частот, спектральний баланс, тощо. Іншими словами, необхідно передбачити точну ідентичність фонетичних особливостей мовного сигналу.

Одним з перспективних методів низькошвидкісного кодування можна вважати метод лінійного кодування з прогнозуванням (LPC, linear predictive coding). Метод LPC дуже просто адаптується до створення узагальненої дискретної моделі мовного сигналу. Реалізація цього методу передбачає використання в структурній схемі вокодера формантного фільтру. При цьому модель формантного фільтру має відповідати спрощеній моделі голосового тракту. Модель голосового тракту є інтерполяцією реальної акустичної системи. Саме такий підхід дозволяє врахувати індивідуальні особливості мовного акустичного сигналу при конструюванні вокодерів.

Метод лінійного кодування з прогнозуванням при практичному застосуванні використовується і в алгоритмах розпізнавання зашумленого мовного сигналу [1, с. 165]. На рисунку 1 наведена структурна схема, яка пояснює реалізацію методу. Сигнал з шумом надходить на кодер, де відбувається дискретизація сигналу та розбиття його на часові інтервали (сегменти або кадри) тривалістю до 20 мс з 50% запасом за перекриттям. І вказане розбиття є обмеженням методу, оскільки більш

короткі сегменти, тривалістю, наприклад 10 мс, будуть взяті з сусідніх сегментів, що може бути причиною помилок.

Додатково проводиться оброблення сигналу шляхом підвищення амплітуд окремих складових сигналів. Ця процедура покликана підвищити загальне відношення сигнал-шум. Далі, на основі відповідного аналізу, за схемою визначаються кепстральні коефіцієнти та автокореляційна функція. Ці коефіцієнти ще називаються лінійними коефіцієнтами прогнозування і якщо їх позначити як k_l , а сам сигнал на виході системи кодування визначити як $s(n)$, де n - номер відліку, то можна записати

$$s(n) = \sum_{l=1}^p k_l s(n-l).$$

Ідея розрахунків полягає у тому, щоб отримати в результаті мінімальну похибку прогнозування. Похибку прогнозування $e(n)$ можна визначити як

$$e(n) = s(n) - \sum_{l=1}^p k_l s(n-l).$$

Найбільш поширеними є три базових алгоритмів розрахунку коефіцієнтів лінійного прогнозування: коваріаційний, автокореляційний і сходовий [2, с. 426]. В даному дослідженні обрано алгоритм на основі розрахунку функції автокореляції. Ці дані необхідні щоб сконструювати модель формантного фільтру. При цьому автокореляційна функція дозволяє отримати дані про висоту тону мовного сигналу (рис. 1).

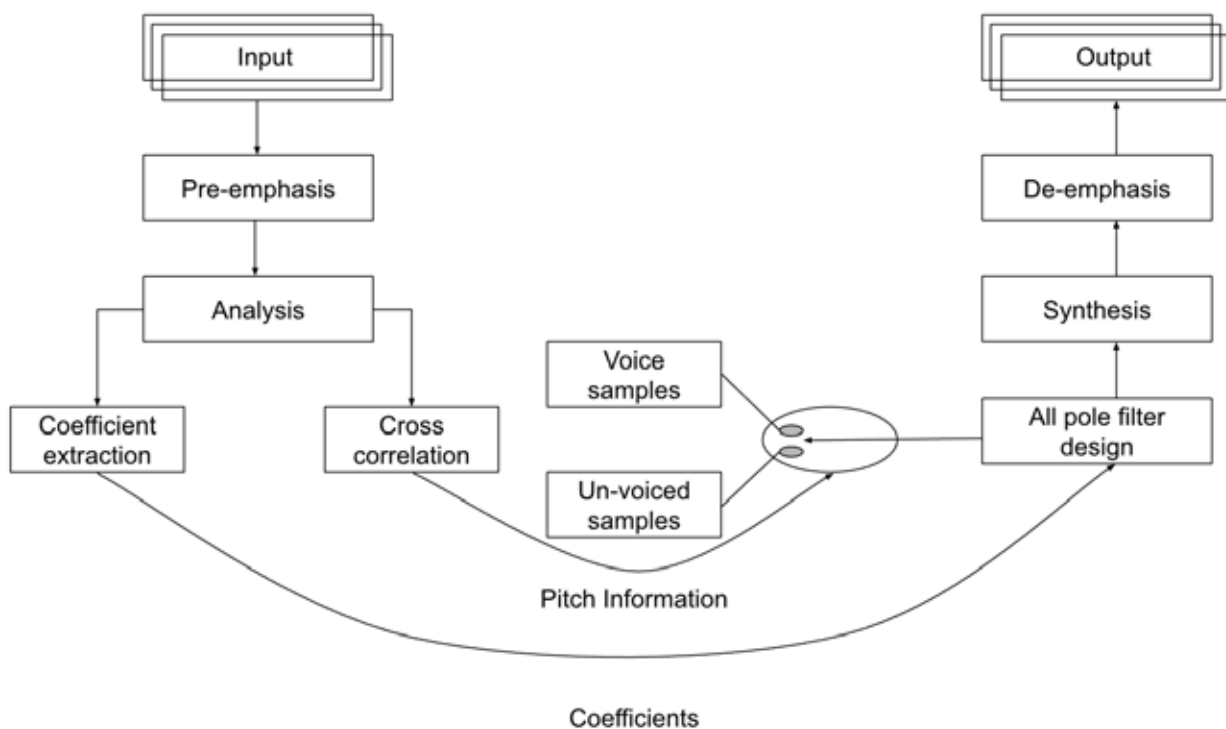


Рис. 1. Спрощена модель [1, с. 168]

Тобто сигнал розбивається на часові вікна фіксованого розміру, де як передбачається, параметри сигналу не змінюються. Далі до кожного вікна застосовується алгоритм вилучення ознак у формі коефіцієнтів кепстру, тобто коефіцієнтів лінійного прогнозування. На основі значень цих коефіцієнтів можна оцінити форманти мовного сигналу і до того ж коефіцієнти можуть бути використані при розрахунку параметрів мовного тракту. Тобто вони описують параметри мовного тракту [3, с. 151].

Відповідно до положень методу лінійного кодування з прогнозуванням, аналіз мовних сигналів проводиться за умови розгляду та прорахунку значень функції формантного фільтру, яка складається з відповідних коефіцієнтів, за умови що сигнал джерела звучання проходить крізь процедуру фільтрації за аналогією передавання сигналу через одиночну акустичну трубку зі змінним перерізом. Така реалізація дозволяє адекватно змодельовувати мовний тракт. Тобто, тракт являє собою резонатор з певними піками в амплітудно-частотній характеристиці. Ці резонанси призводять до виникнення формант (розширених смуг частот). Модель мовного тракту, як приклад лінійної стаціонарної системи, включає в себе з'єднання N секцій акустичних трубок, в яких A_i – площі поперечного перерізу у напрямку, перпендикулярному поздовжній осі трубки (рис. 2).

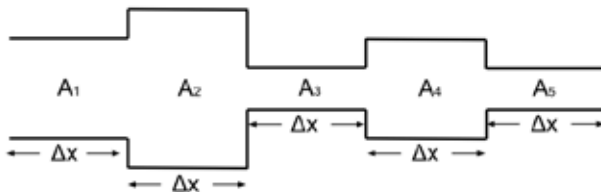


Рис. 2. Спрощена модель мовного тракту

Передаточна функція моделі має лише полюса і не має нулів. Ці полюса визначають формантні частоти F_i і утворюють F -картину сигналу. Тобто, модель є формантним смуговим фільтром. Кількість ланок моделі визначає порядок фільтру, а центральна частота контуру фільтру є частотою відповідної форманти. Тобто, підсумовуючи сказане, метод LPC означає, що даний відлік мовного сигналу апроксимується лінійною комбінацією попередніх відкликів. Коефіцієнти лінійного прогнозування є параметрами формантного фільтру моделі мовного тракту і саме вони містять інформацію про характерні властивості мовного сигналу. Набір числових характеристик дозволяє отримати вираз для виходу фільтру у часовій області.

Метою статті є аналіз розподілу формант записаного сигналу з частковим зашумленням, і подальше виділення його характерних особливостей на основі методу лінійного кодування з прогнозуванням. При цьому, в рамках дослідження передбачається, що сам сигнал у виділеному часовому діапазоні перебуває під дією різних видів шуму з однаковою амплітудою. Тобто, необхідно перевірити ефективність методу за умови, що частина мовного сигналу з шумом.

Для досягнення поставленої мети необхідно виконати наступні завдання:

- в спеціалізованому програмному пакеті для аналізу даних завантажити мовний сигнал без шуму і проаналізувати його основні параметри;
- використовуючи модель лінійного прогнозування знайти кепстральні коефіцієнти сигналу та трьох його копій з шумом;
- визначити вплив шумів різної природи на сонограму та енергетичні параметри мовного сигналу;
- розробити рекомендації щодо оцінки впливу шумів на критерії відновлення мовних сигналів з точки зору розбірливості при конструюванні системи ідентифікації за голосом.

Виклад основного матеріалу дослідження. З інженерної точки зору, шум – це спонтанний або псевдовипадковий сигнал, який можна визначити в діапазоні сигналу розподілом енергії. «Білий» шум має рівномірний розподіл енергії. Кольорові шуми передбачають нерівномірний розподіл енергії [4, с. 742]. Рожевий шум має величину незмінної потужності на октаву, а не постійну потужність на Герц, що робить його більш придатним для слухових досліджень. У той час, білий шум є фізично ідентичним розподілом енергії, а рожевий шум звучить так, ніби має еквівалентний розподіл енергії, і це пов'язано зі структурою внутрішнього вуха. З метою зменшення шуму зазвичай використовують спектр стиснення голосу за допомогою обчислювальних процедур методу лінійного кодування з прогнозуванням. Згідно з експериментальними даними, голосовий сигнал зазнає значного стиснення після використання цього методу [5, с. 4].

На основі інструментів програмного забезпечення PRAAT визначимо характеристики сигналу без шуму та з шумом, а потім на основі методу лінійного кодування з прогнозуванням знайдемо коефіцієнти та проведемо оцінку можливості очищення сигналу від шумового забруднення [6, с. 4].

В якості вихідних даних для проведення експерименту виберемо акустичний мовний сигнал відомої фрази "I'll be back" тривалістю 0,988 сек

(рис.2). Параметри сигналу наступні: верхня частота сигналу 11,025 кГц, частота дискретизації 16 кбіт/сек, рівень квантування 32 біт, формат сигналу – моно. За умовами експерименту частина оригінального мовного сигналу в діапазоні від 0 сек до 0,3 сек (рис. 3) знаходиться під впливом послідовно трьох видів шуму: білий (шум № 1), рожевий (шум №2) та броунівський (шум № 3) (рис.4-6). Рівень амплітуди шуму для трьох видів однаковий і дорівнює у нормованій формі 0,2. На рисунках 4-6 червоним штрихом показані уздовж горизонтальних осей форманти сигналу з шумом. Визначимо, як шум впливає на якість мовного сигналу і чи змінюються при цьому енергетичні характеристики сигналу та відповідні коефіцієнти за методом LPC.

В таблиці 1 наведено енергетичні параметри сигналу без шуму, та за умови присутності на фрагменті сигналу трьох різних видів шуму.



Рис. 3. Сонограма оригінального сигналу

З аналізу даних таблиці 1 можна відмітити наступні особливості. Додавання шуму до сигналу навіть на відокремленому фрагменті оригінального сигналу призводить до зменшення

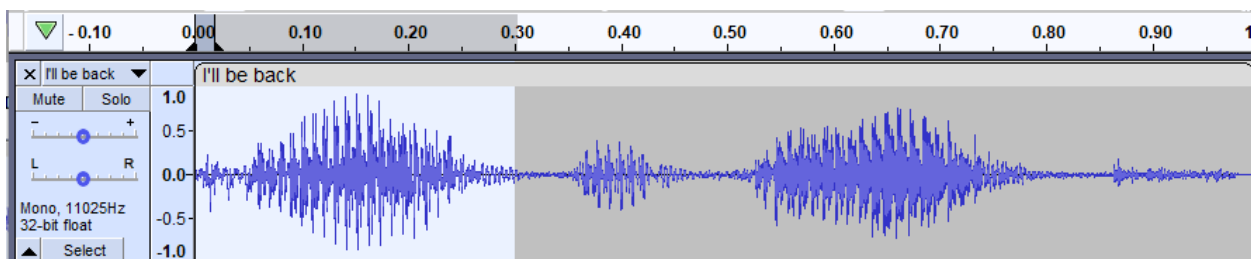


Рис. 4. Фрагмент сигналу, на який будуть впливати шуми

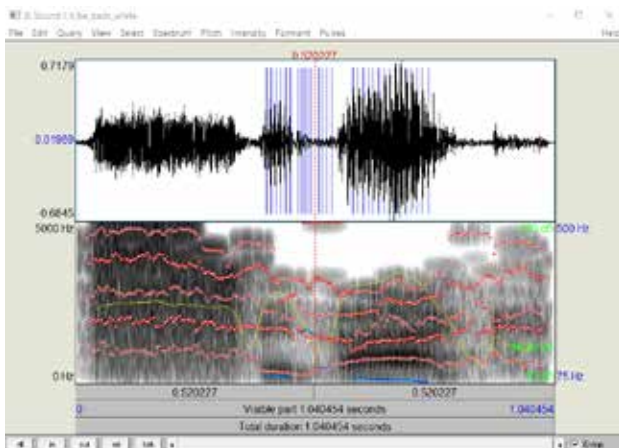


Рис. 5. Результат впливу білого шуму

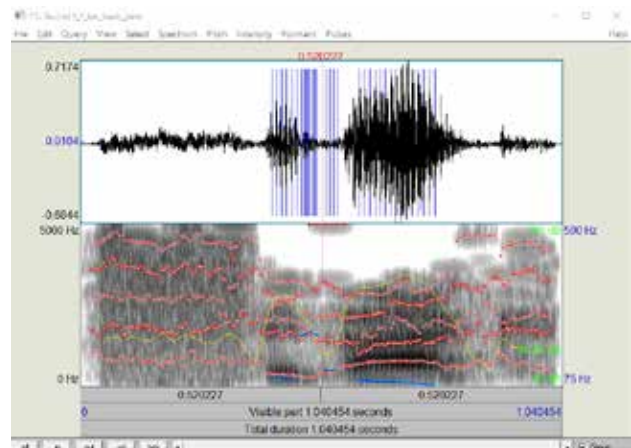


Рис. 6. Результат впливу рожевого шуму

Таблиця 1

Енергетичні параметри сигналу

Тип	Енергія, Па ² ×сек	Потужність, Па ²	Середня інтенсивність, дБ
Сигнал без шуму	0.029155827141758725	0.0295037167726379	78.67816738961012
Шум № 1	0.016964538640653327	0.016304946256926415	76.10259380329269
Шум № 2	0.013904316909971848	0.013363707953311797	75.23866984793247
Шум № 3	0.014149027497899638	0.013598904033156962	75.3144391766493

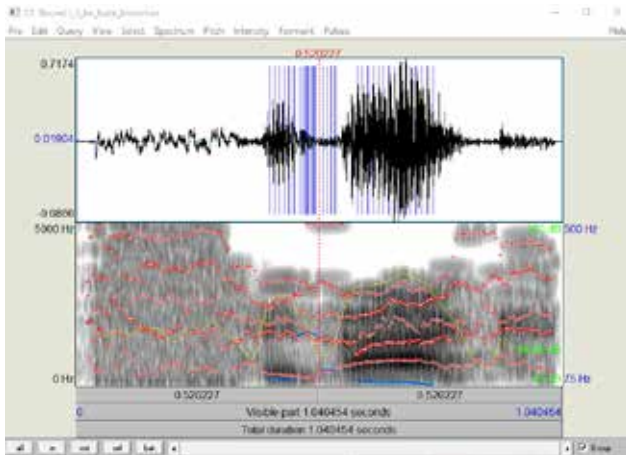


Рис. 7. Результат впливу броунівського шуму

енергетичних параметрів цього сигналу. Причому така тенденція спостерігається більше для шуму №2 і менше для шуму під номером №1. Зазначимо, що при додаванні шуму зменшується і середня інтенсивність мовного сигналу (в середньому на 2-3 дБ), що може бути перепоною при проведенні розбірливості мовного сигналу на приймальному кінці лінії зв'язку. Крім цього, при додаванні шуму на інтервалі від 0 до 0,3 с при загальній тривалості сигналу 0,98 с (величина 1,040454 с обрана програмою з певним запасом відповідно до алгоритмів оброблення) середня енергія зменшується майже у 2 рази, що також впливає на розбірливість окремих букв при відтворенні, особливо це стосується шиплячих звуків.

Далі в рамках дослідження, отримуємо коефіцієнти LPC для сигналу та трьох його копій з шумами. Оскільки ці коефіцієнти визначають

форманти сигналу і відповідні полюса формантного фільтру, то необхідно визначити, як модель голосового тракту трансформується у випадку впливу шумової складової. Вважається, що порядок формантного фільтру дорівнює 16, вікно аналізу 0,025 сек, крок зсуву 0,005 сек, частота передспотворень 50 Гц. В таблиці 2 наведено розраховані коефіцієнти за методом LPC на основі автокореляційного алгоритму розрахунку.

Аналізуючи отримані коефіцієнти з таблиці 2 приходимо до висновку, що наявність шуму призводить до того, що полюса передаточної функції формантного фільтру повністю змінюються, а отже змінюються і формантні частоти моделі мовного тракту. При цьому така закономірність спостерігається для усіх отриманих коефіцієнтів за номерами. Це може свідчити про те, що метод LPC є адаптивним до появи шумових складових і при наявності шуму модель мовного тракту повністю перебудовується. Можна припустити, що в даному випадку така поведінка розрахункового алгоритму спрямована на те, щоб забезпечити певний рівень розбірливості сигналу за умови стиснення сигналу з шумом за правилами методу і враховуючи прийоми кодування.

Висновки. Визначено передумови, які підкреслюють необхідність використання методу лінійного кодування з прогнозуванням в системах передавання мовної інформації. Зроблено припущення, що даний метод є зручним при створенні узагальненої дискретної моделі мовного сигналу в умовах зашумленого середовища. Розглянуто випадок, коли частина мовного сигналу знаходиться під окремою дією трьох різних видів шуму

Таблиця 2

Коефіцієнти LPC (полюса фільтру)

Номер	Сигнал без шуму	Шум № 1	Шум № 2	Шум № 3
1	-7.679748873591303	-5.64749051665389	-5.983858524269302	-6.392913346365203
2	30.45567564142604	18.232707055378953	19.81434940565449	22.088542875918197
3	-82.67230389773415	-42.531088888574025	-47.08348260255059	-53.8229150831381
4	171.4476791728815	78.86521331519664	88.71160707018083	102.53383575538665
5	-287.3776602224649	-121.64742887370451	-139.16914522737875	-160.84680741658207
6	-287.3776602224649	160.25472320558333	186.7542866158488	213.92686635145185
7	-478.909995620459	-182.90329827672934	-217.6246538568542	-245.31991623962784
8	489.8867749950707	182.21701228359981	221.94506686612988	244.68422750210738
9	-431.41704135119164	-158.71599635094026	-198.4901747008896	-212.78825237887776
10	325.5643211444223	120.48386792552004	155.17880980186817	160.78082566262037
11	-208.0404077359442	-78.98014815745658	-105.10255317263896	-104.47903961697219
12	110.24592013620382	43.98247347378441	60.66818165190962	57.28231684879698
13	-46.81505474354785	-20.21952318358637	-29.01206031814341	-25.65194953076383
14	15.028931809498154	7.309955075031905	10.936795669634435	8.876171547695526
15	-3.2650429521771023	-1.887717308060039	-2.9432311670509637	-2.133674563627802
16	0.36334390868451527	0.27396634715036117	0.43729864561822496	0.2746517922324624

з однаковою амплітудою. Отримано за допомогою інструментів програмного забезпечення PRAAT, характеристики мовного сигналу без шуму та з шумом, і на основі методу лінійного кодування з прогнозуванням знайдено коефіцієнти формантного фільтру. На основі отриманих результатів виявлено, що присутність шуму призводить до зменшення усереднених значень енергетичних параметрів мовного сигналу. Причому, більше це помітно для рожевого шуму і менше для білого шуму. Крім цього знайдено, що при додаванні шуму зменшується і середня інтенсивність мов-

ного сигналу (в середньому на 2-3 дБ), що може бути причиною зниження розбірливості мовного сигналу на приймальному кінці лінії зв'язку.

Додатково отримано підтвердження того, що присутність шумових компонент у мовному сигналі є причиною зміни значень формантних частот моделі мовного сигналу. І ця перебудова не залежить від природи шуму. Тобто, можна зробити припущення, що на основі отриманих результатів для обраного тестового мовного сигналу метод лінійного кодування з прогнозуванням є адаптивним до появи шумових складових.

Список літератури:

1. Thimmaraja Y.G., Nagaraja B.G., Jayanna H.S. Speech enhancement and encoding by combining SS-VAD and LPC. *International Journal of Speech Technology*. 2021. No 24. P. 165–172.
2. Rabiner L. R., Schafer R.W. Digital processing of speech signal. New Jersey : Prentice-Hall, 1978. 496 p.
3. Singh P., Mukul M. K., Prasad R. Bone Conducted Speech Signal Enhancement Using LPC and MFCC. *IHCI. 2018. Lecture Notes in Computer Science*. 2018. Vol. 11278. P. 148–158.
4. Sadeghi M., Marvi H., Ali M. The effect of different acoustic noise on speech signal formant frequency location. *International Journal of Speech Technology*. 2018. Vol. 21. P. 741–752.
5. Moriya T., Sugiura R. Progress in LPC-based frequency-domain audio coding. *SIP*. 2016. Vol. 5, No 11. P. 1–10
6. Tagomori T., Tsuruda R., Matsuo K. Speaker verification from mixture of speech and non-speech audio signals via using pole distribution of piecewise linear predictive coding coefficients. *Journal of Ambient Intelligence and Humanized Computing*. 2020. 1–11

Svitlovskiy Ye.V., Trapezon K.O. ANALYSIS OF SPEECH ACOUSTIC SIGNALS IN COMMUNICATION SYSTEMS WITH PARTIAL NOISE

The method of linear coding with prediction in practical application in communication systems is an integral part in the creation of algorithms for recognizing a noisy speech signal. Based on the analysis of literary sources, it was confirmed that the method of linear coding with prediction is convenient for creating a generalized discrete model of the speech signal in noisy environments. We consider the case when a part of the speech signal is under the separate action of three different types of noise with the same amplitude. Using the PRAAT software tools, the energy characteristics of the signal without noise and with noise were obtained, and based on the linear predictive coding method, the formant filter coefficients were found based on the use of the autocorrelation function. According to the obtained results, it was found that the presence of noise leads to a decrease in the averaged values of the energy parameters of the speech signal. Moreover, it is more noticeable for pink noise and less for white noise. In addition, it was found that when noise is added to the signal, the average intensity of the signal decreases (on average by 2-3 dB), which can be an obstacle in obtaining the necessary quality in terms of intelligibility of the speech signal at the receiving end of the communication line.

In addition, it was confirmed that the presence of noise components in the speech signal is probably the reason for the change in the values of the formant frequencies of the speech signal model. And this reconstruction does not depend on the nature of the noise and its individual characteristics. That is, it can be assumed that based on the obtained results, for the selected test speech signal with a duration of about 1 sec. the method of linear coding with prediction is fully adaptive to the appearance of noise components.

Key words: vocoder, acoustics, signal, processing, method, noise, prediction, restoration.