УДК.004.934(045)

О. Д. Любімов, канд. техн. наук, доц.
В. М. Сігнаєвський,

МЕТОД АНАЛІЗУ ТА ОБРОБКИ МОВНИХ СИГНАЛІВ

Інститут електроніки та систем управління НАУ, e-mail: iesy@nau.edu.ua

Здійснено підхід до обробки та аналізу мовних сигналів звукового діапазону на прикладі мовної інформації. Наведено приклад дослідження розпізнавання окремо взятих звуків алфавіту української мови. Висвітлено основні прийоми підходу до розпізнавання спектру звуків мови.

Ключові слова: розпізнавання, перетворення Фур'є, динамічний діапазон, віконна функція.

Вступ та постановка задачі. Розпізнавання мовних сигналів являє собою окремий випадок розпізнавання образів, у яких об'єктами розпізнавання є акустичні сигнали, що створюються людиною під час вимови окремо взятих звуків чи слів загалом. Вирішення цієї перманентної проблеми дозволить уводити мовну інформацію в комп'ютерні системи без клавіатури, дасть змогу керувати технічними комплексами з використанням робототехнічних систем за допомогою мовних команд, а також розв'язувати питання несанкціонованого доступу до інформації.

Натепер немає достатнього програмного забезпечення, яке б відповідало вимогам розпізнавання мовної інформації українською мовою. Тому за основу дослідження взято українську мову як базову. Для дослідження спектральних особливостей звуків українського алфавіту було створено спеціальне програмне забезпечення "SWIM", для опрацювання отриманих даних (рис. 1).

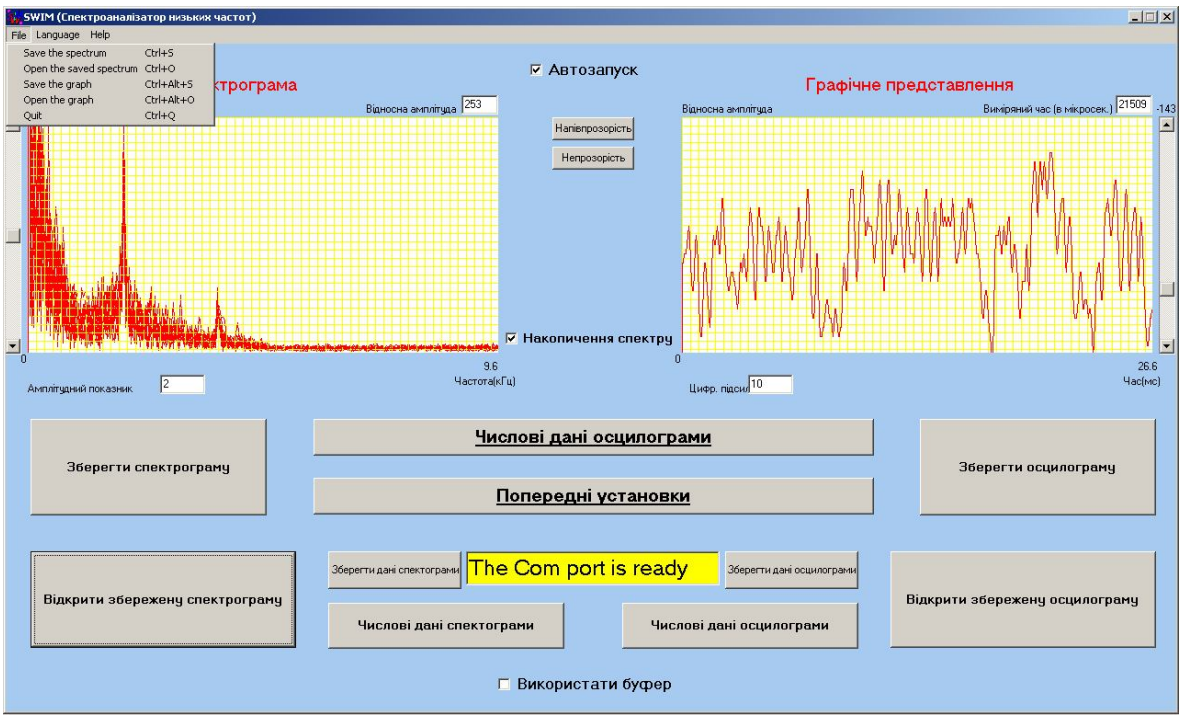


Рис. 1. Програмне забезпечення SWIM

Програмне забезпечення виконує такі операції. У вибраному режимі роботи «Автозапуск» цифровий масив заповнюється числовими даними з мікрофонного входу звукової карти ПК, необхідною кількістю значень (кількість значень наперед обумовлюється

дискретним перетворення Фур'є) після чого числовими методами виконується саме дискретне перетворення Фур'є. Графічний результат осцилограми звуків відображається в правому вікні, спектральний – у лівому. Оскільки основною метою є аналіз мовного сигналу, то цілком можна обмежитись частотою дискретизації 19200 Гц, обумовленою тактовою частотою аналогово-цифрового перетворювача. На рис.2 продемонстровано можливості програми опрацьовувати архіви спектрів звуків. Програма також має інтерфейс англійської мови. На рис. 2 знизу показано реалізації архівів мовних сигналів.

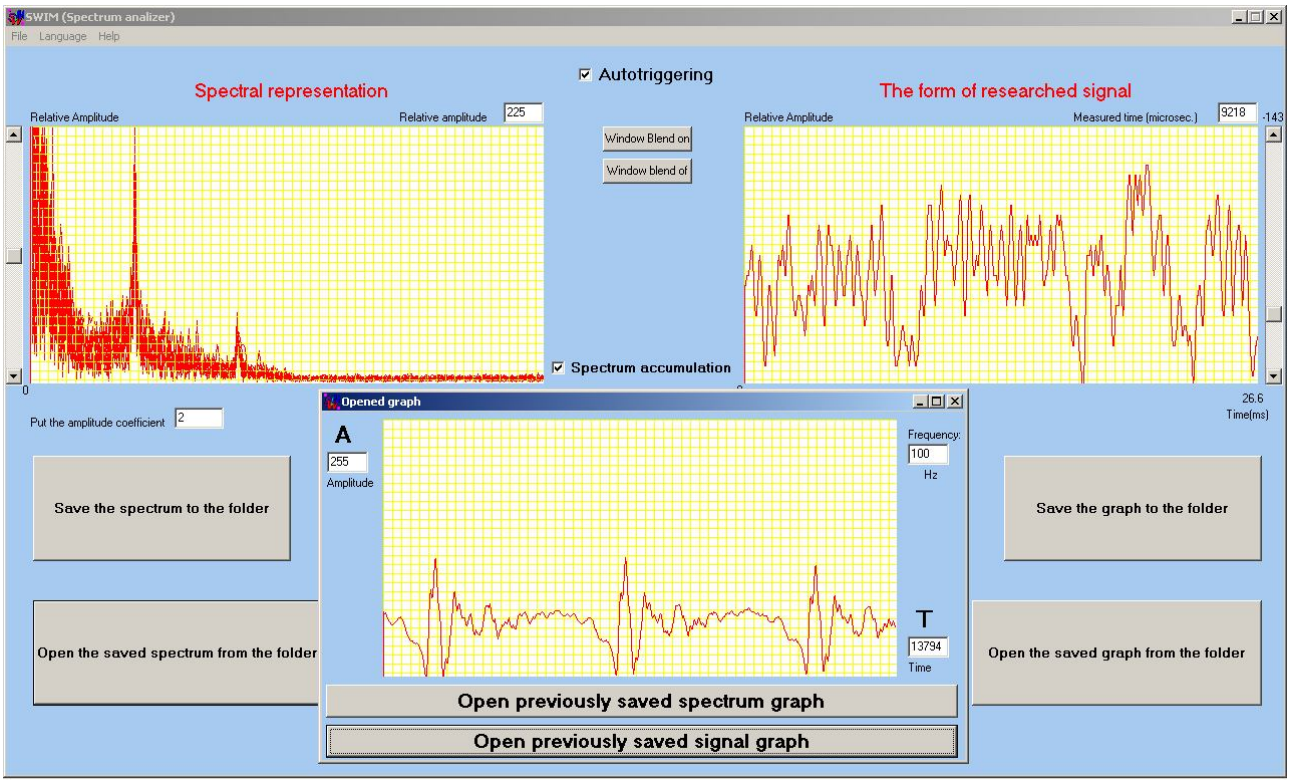


Рис. 2. Робота з архівами спектрів звуків

Спектрограми окремо взятих звуків українського алфавіту показано на рис. 3. Гранична частота для всіх зображених графіків становить 4800 Гц.

З отриманих спектральних графіків можна отримати необхідну інформацію про спектральні складові того чи іншого звукового сигналу. Більш того, користуючись спеціальним програмним забезпеченням можна значно полегшити роботу над ними. Зокрема для перетворення мови в текстову форму необхідно виконати такі кроки:

1) створити моделі (спектральні масиви) звуків, для подальшої можливості порівнювати їх з отриманими зі звукової карти даними. Для прикладу, створити 257-байтні масиви даних окремо взятих звуків, запам'ятати їх і послідовно порівнювати з отриманою в реальному часі спектрограмою звуку.

2) забезпечити високу швидкість обробки за рахунок вибору оптимального розміру масивів для збереження даних.

3) підрахувати оптимальну швидкість оновлення програми, враховуючи час дискретизації сигналу і час виконання перетворення Фур'є та інших функцій програми, що безпосередньо беруть участь у циклі опрацювання отриманої зовні інформації.

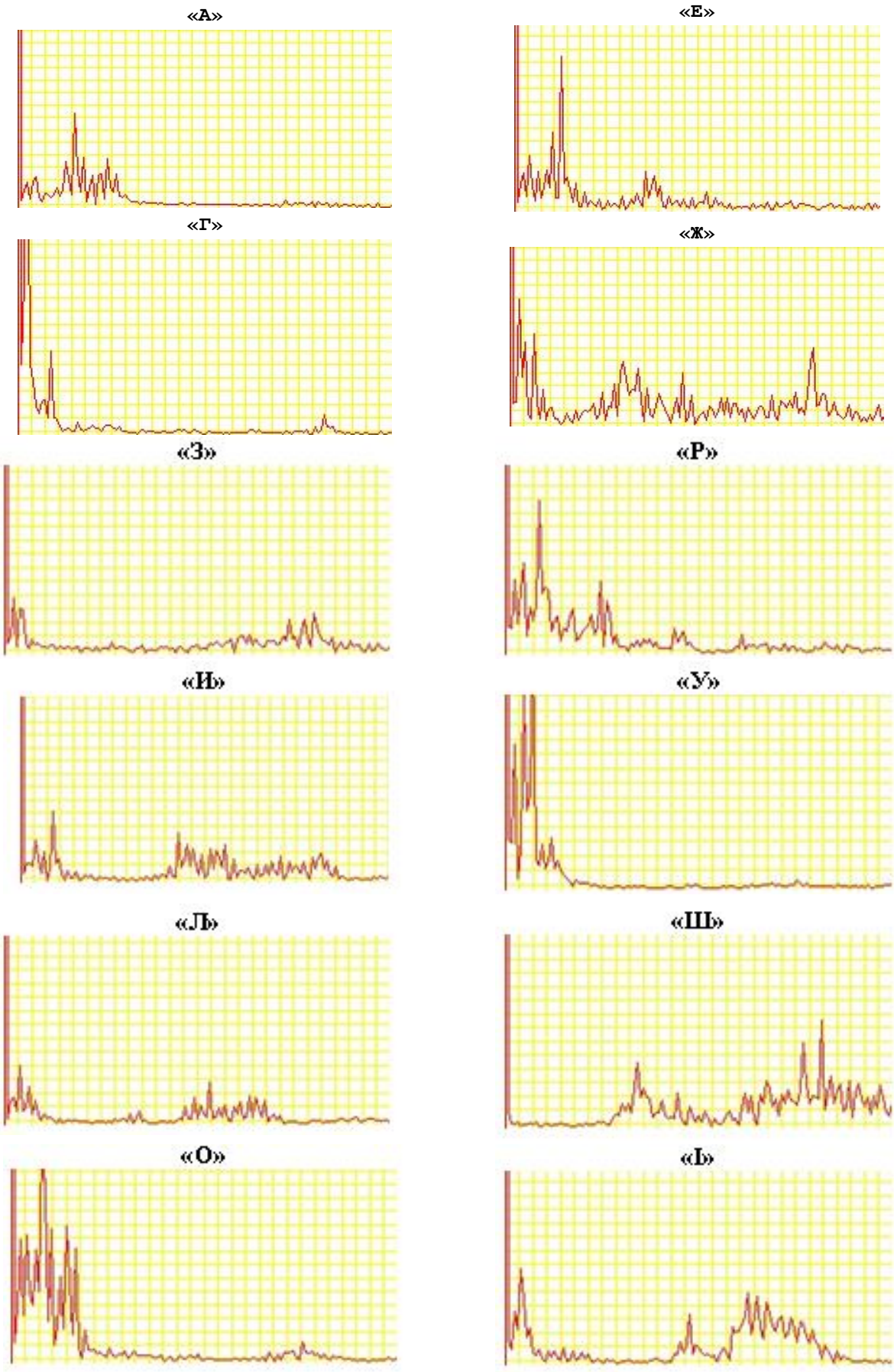


Рис. 3. Результат спектрального перетворення для деяких літер алфавіту

Зручним в аналітичному розумінні способом обробки звукових сигналів є використання програмного пакета MathCad. Хоча програма не дає змоги обробки сигналів у реальному часі, але дозволяє відслідковувати основні моменти, користуючись записами wav-файлів. Нижче наведено результати дослідження спектральних складових літери «А» українського алфавіту (рис. 4). На рис. 4, *а* показано числові значення масиву wav-файлу, на рис. 4, *б* – осцилограму сигналу, побудовану за даними з отриманого масиву, а на рис. 4, *в* – спектрограму, що отриману в результаті перетворення Фур'є.

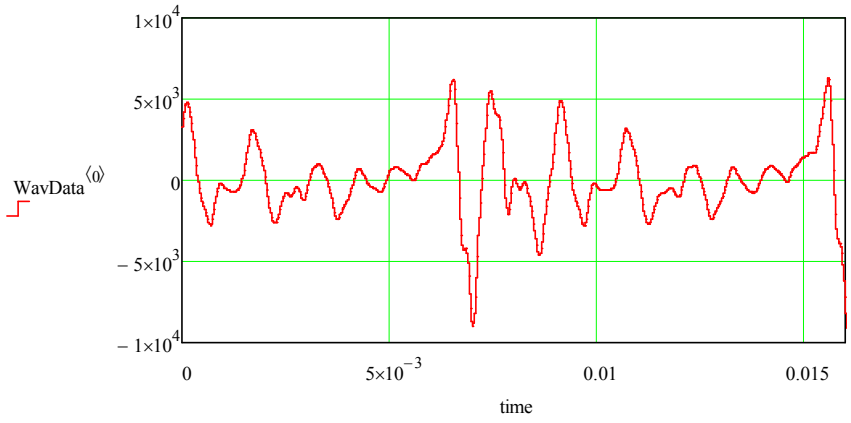
WavData =

	0	1
0	$3.276 \cdot 10^3$	$3.276 \cdot 10^3$
1	$3.812 \cdot 10^3$	$3.812 \cdot 10^3$
2	$4.25 \cdot 10^3$	$4.25 \cdot 10^3$
3	$4.564 \cdot 10^3$	$4.564 \cdot 10^3$
4	$4.737 \cdot 10^3$	...

$$time_n := \frac{n}{SampleRate}$$

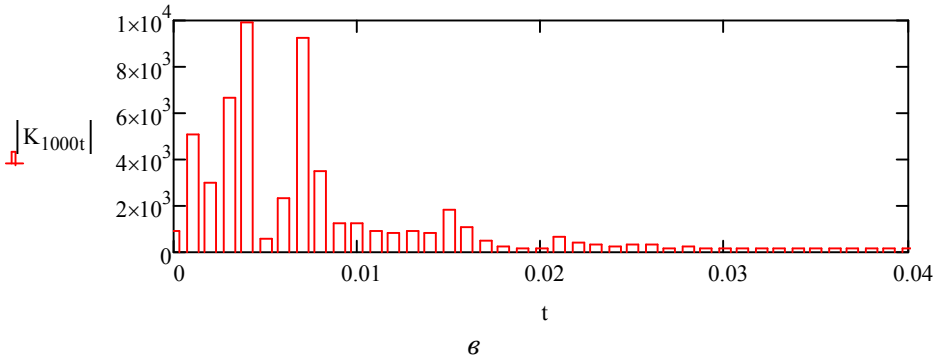
$$Data_{1000 \cdot t} := \left(WavData^{\langle 0 \rangle}\right)_{1000 \cdot t}$$

а



б

$$K := fft(Data)$$



в

Рис. 4. Приклад дослідження звуків мови з використанням програмного пакета MathCad

У разі застосування цифрової електроніки важливим є також врахування динамічного діапазону мовленнєвого сигналу, оскільки це обладнання має скінченну розрядність. Динамічний діапазон рівня розмови становить 35...45 дБ. Незайвим є і врахування тривалості голосних та приголосних звуків, що становить 0,15 та 0,08 с відповідно.

Наступною потребою є фільтрація досліджуваного сигналу, наприклад шумів мікрофону, напівпровідникового тракту підсилювача тощо. Один з класів шумів – адитивно-стаціонарний шум. Для зниження такого типу шумів використовують алгоритм спектральної різниці. Він полягає в такому [1]:

1) короткочасному перетворенні Фур'є (SHFT) для визначення енергії сигналу;
2) аналізі спектру шуму та отриманні різниці між амплітудами складових спектру шуму та амплітудами спектру самого сигналу;
3) зворотному перетворенні (SHFT) – синтезі результуючого сигналу.

Для адекватного розпізнавання мови із заданою точністю недостатньо використовувати наведені методи. Додатковою умовою отримання істинного спектру звуку мови є застосування віконної функції (наприклад, «вікна Хеммінга»). «Вікно Хеммінга» (рис. 5) застосовують для отримання більш точного спектрального подання при перетворенні Фур'є [2]. Інформацію про дискретне перетворення Фур'є із застосуванням цифрової електроніки можна отримати з праці [3].

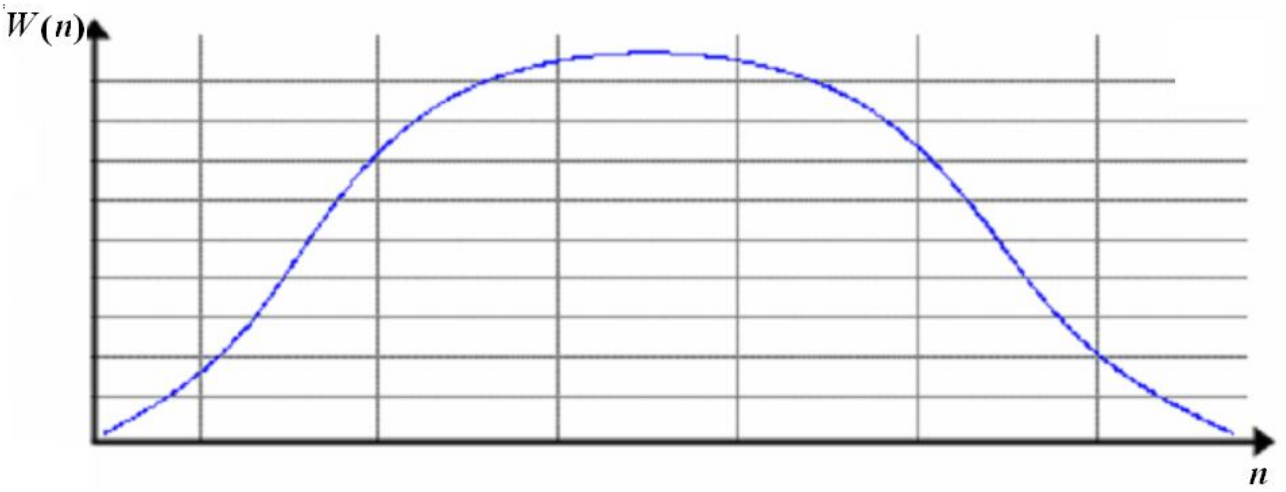


Рис. 5. Віконна функція Хеммінга

У загальному випадку віконну функцію Хеммінга можна виразити як залежність:

$$w(n) = \frac{1}{\beta}\left(\alpha - (1-\alpha)\cos\left(\frac{2\pi n}{N-1}\right)\right),$$

де α – константа в діапазоні 0 – 1, на практиці $\alpha = 0.54$; N_s – розмір вікна у відліках;

$\beta = \sqrt{\frac{1}{N}\sum w(n)}$ – нормалізаційна константа.

Висновок

1. Здійснено підхід до розроблення програмного забезпечення розпізнавання звуків української мови.

2. Програмне забезпечення дозволяє розпізнавати вказані звуки в реальному масштабі часу.

3. Використання віконної функції Хеммінга дозволяє підвищити якість розпізнавання мовної інформації.

Список літератури

1. *Gaskill* Linear Systems, Fourier Transforms, and Optics / D. Gaskill, Jack – New York: John Wiley & Sons, 1978. – 554 p.
2. *Sen M Kuo* Real-Time Digital Signal Processing Implementations and Applications / Sen M Kuo, Bob H Lee, Wenshun Tian, USA, 646 p.
3. *Wolf J. J.* and *Woods W. A.*, The HWIM speech understanding system, Hartford, 1977. – 787 p.
4. *Stephen E. Levinson*, Mathematical models for speech technology / Stephen E. Levinson University of Illinois at Urbana-Champaign, USA, 2005. – 284 p.
5. *Tou J. T.* Pattern recognition principles / J. T. Tou and R. C.Gonzalez, MA, 1974. – 521 p.
6. *Steve White*, Digital Signal Processing – Filtering Approach, 2004, 228 p.
7. *Arndt Jorg*, FFTs for programmers (algorithms and source code), 2001. – 87 p.
8. *Лукин Алексей*, Введение в цифровую обработку сигналов (математические основы), 2007. – 68 с.
9. *Aldrich Nika*, The Audio Engineeris Approach to Understanding Digital Filters, 2002. – 45 p.
10. *Arndt Jorg*, Algorithms for programmers (ideas and source code), 2006. – 760 p.

А. Д. Любимов, В. Н. Сигнаевский

Некоторые методы анализа и обработки речевых сигналов

Осуществлен подход к обработке и анализу речевых сигналов звукового диапазона на примере речевой информации. Приведен пример исследования для некоторых звуков алфавита украинского языка. Высветлено некоторые приемы подхода к распознаванию спектра речи.

A. D. Lyubimov, V. M. Signaevskyi

Some methods of the analysis and speech signals processing

The approach to speech signals processing and analysis on an example of the speech information is carried out. The example of research for some sounds of the Ukrainian language alphabet is resulted. Enlighten some technique to the approach of spectrum speech recognition.