

## ОЦІНКА РІВНЯ ЗАХИЩЕНОСТІ МОВНОЇ ІНФОРМАЦІЇ ВІД ВИТОКУ ВІБРОАКУСТИЧНИМИ КАНАЛАМИ ОБ'ЄКТИЗОВАНИМ МЕТОДОМ ЗА КРИТЕРІЄМ ЗАЛИШКОВОЇ РОЗБІРЛИВОСТІ МОВИ З ВИКОРИСТАННЯМ АРТИКУЛЯЦІЙНИХ ТАБЛИЦЬ УКРАЇНСЬКОЇ МОВИ

К. т. н. С. М. Нужний, П. В. Заноскіна

Національний університет кораблебудування імені адмірала Макарова

Україна, м. Миколаїв

s.nuzhniy@gmail.com; lilianrhade@gmail.com

*Оцінка захищеності мови дикторів на межі контрольованої зони є основою для побудови систем захисту і головним критерієм довіри до системи безпеки в цілому. Запропоновано об'єктизований метод оцінювання захищеності мовної інформації за критерієм її залишкової розбірливості. Для використання методу необхідні спеціалізовані артикуляційні таблиці української мови. В Україні відсутній науково-обґрунтований метод їхньої розробки. Проведено дослідження по вибору критеріїв формування артикуляційних таблиць. Таблиці також необхідні для криміналістичних фоноскопичних експертиз, криптоаналізу тощо.*

*Ключові слова: оцінка захищеності мови дикторів, об'єктизований метод, залишкова розбірливість, артикуляційні таблиці української мови.*

Сучасний етап розвитку людства визначається впровадженням цифрових технологій у всі сфери життя людини та суспільства. Такий підхід не є виключенням й для задач отримання та обробки акустичної інформації. Особливої актуальності це питання набуло для систем технічного захисту інформації від витоку акустичними та вібраційними каналами. Ще до недавнього часу технічні засоби розвідки будувались по принципу безпосередньої первинної обробки перехопленої інформації на місці її отримання або інформація передавалась на аналогових носіях в центри обробки, що призводило до значних витрат часу. Широке запровадження мережевих технологій дозволило впровадити методи дистанційної обробки, які в режимі on-line забезпечують миттєву передачу та дослідження перехопленої мовної інформації [1]. Це призвело до суттєвого відставання методів та технологій протидії витоку мовної інформації від можливостей технічних засобів розвідки. Головним критерієм такого відставання є неможливість достовірно визначити рівень захищеності мовної інформації при використанні співвідношення сигнал/завада за умов застосування зловмисниками сучасних методів обробки перехопленої інформації.

Таким чином, без вирішення задачі достовірного оцінювання рівня захищеності мовної інформації від витоку акустичними та вібраційними каналами, які мають відповідати сучасному рівню розвитку технологій цифрової обробки мовної інформації, неможливо побудувати дійсно надійну систему захисту.

Аналіз існуючих методів оцінювання захищеності мовної інформації показує, що їх можна умовно розділити на три групи.

**1 група – енергетичні методи.** Методи з цієї групи відносяться до категорії опосередкованих об'єктивних методів і надають можливість оцінити рівень захищеності мовної інформації за результатами розрахунку відношення сигнал/завада (Signal-to-noise ratio – SNR) на межі контрольованої зони. Такий метод застосовується на пострадянських територіях, в тому числі й в Україні. Виміри проводяться для діапазону частот від 90 до 5600 Гц або 1100 Гц в 18 або 21 смузді. Такий підхід збільшує достовірність оцінки рівня захищеності мовної інформації, однак не враховує можливості фільтрації шумової завади, кореляційного аналізу та інших методів цифрової обробки.

Варіантом підвищення достовірності є оцінювання рівня захищеності мовної інформації за критерієм «клас захищеності мови» (Speech Privacy Class – SPC) та відношення сигнал/завада в мовному діапазоні частот (uniform-weighted signal-to-noise ratio – SNR<sub>uni32</sub>), які використовуються в Канаді [2]. Головними змінами в методі є використання як тест-сигналів фонограм дикторської мови та ви-

значення співвідношення сигнал/завада за відношенням середніх значень амплітуд сигналів без врахування пауз та рівнів при яких відношення сигнал/завада має значення менші за –32 дБ.

**2 група – дослідження частотного спектру.** Їх можна розділити на декілька підгруп, які дозволяють визначати:

- SII (Speech Intelligibility Index – коефіцієнт (індекс) розбірливості мови) (KPM)) [3, 6];
- STI (Speech Transmission Index – індекс передачі мови (ІПМ)); RASTI (Room Acoustics STI – ІПМ для приміщень); STIPA (STI for Public Address systems – ІПМ для систем звукопідсилення); STITEL (STI for Telecommunication Systems – ІПМ для телекомунікаційних систем); ALcons (Percentage Articulation Loss of Consonants – оцінка втрат артикуляції приголосних) [4, 5].

**3 група – контроль розбірливості мовного сигналу.** Оцінка рівня захищеності мовної інформації здійснюється за розбірливістю мовного сигналу. Використовуються критерії  $W$  (словесна розбірливість мови (СРМ)) [6] та SII (Speech Intelligibility Index) [3, 6].

Коефіцієнт  $SII$  може бути віднесений до всіх трьох груп. Визначається згідно з виразом [3]

$$S = \sum_{i=1}^n I_i A_i,$$

де  $n$  – кількість смуг, по яких розраховується коефіцієнт розбірливості мови (зазвичай  $n$  дорівнює 6 чи 7 при використанні октавних смуг або 18 чи 21 при використанні  $1/3$ -октавних смуг);  $I$  – коефіцієнт вагомості інформації в смузі;  $A$  – коефіцієнт якості розпізнавання інформації в смузі.

Недоліком вказаного методу залишається залежність від шумової завади – коефіцієнт визначається тільки до співвідношення сигнал/завада –10 дБ.

На основі SII в роботі пропонується введення нового критерію оцінювання – коефіцієнту RII (Residual intelligibility index – коефіцієнту залишкової розбірливості мови), який визначається як [6]

$$S^R = \frac{1}{m} \sum_{y=1}^m \left[ R_y^a H_y \sum_{i=1}^n I_i A_i \right],$$

де  $m$  – загальна кількість алофонів в слові;  $R^a$  – коефіцієнт важливості розпізнавання алофону для розпізнавання слова (фрази, речення);  $H$  – коефіцієнт частоти вживання даного алофону в мові.

Введення коефіцієнтів  $H$  та  $R^a$  дозволяє враховувати як частотно-часове розподілення формант в фонемах вплив на їхню форму та величину шумової завади та залишкових викривлень внаслідок проведення фільтрації і цифрової обробки фонограми, так і фонетико-лінгвістичні характеристики.

Основою проведення фонетико-лінгвістичного аналізу є артикуляційні таблиці слів та коротких фраз. Аналіз особливостей мовних діалогів, які відбуваються в виділених приміщеннях (захищених зонах) показав необхідність врахування таких вагомих для розпізнавання мови дикторів факторів, як:

- спеціалізація мови дикторів за професійною спрямованістю з використанням «сленгових» виразів, умовних скорочень, назв тощо;
- фактична мультимовність дикторів;
- вплив діалектів.

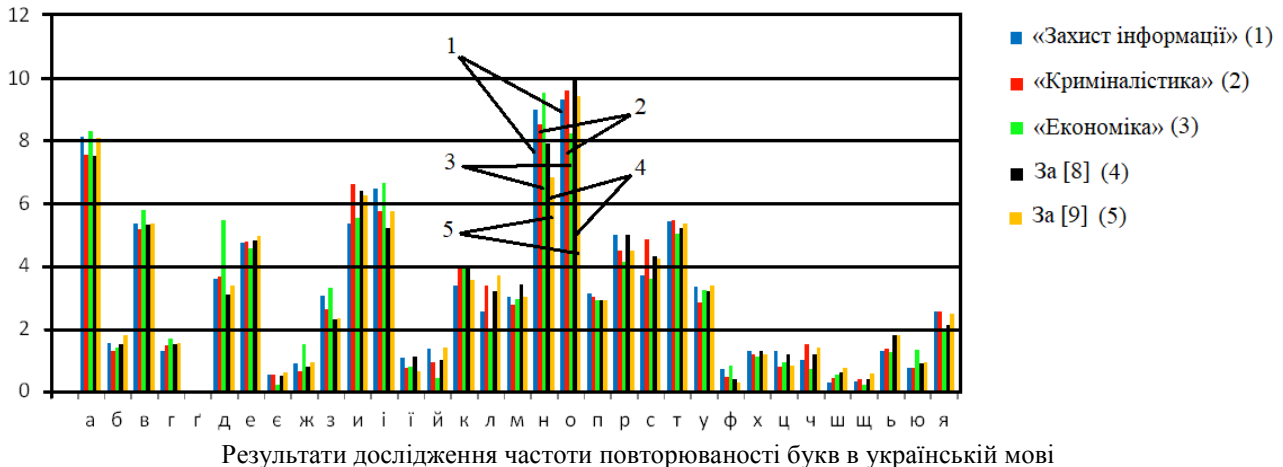
Аналіз літератури показав широку різноманітність артикуляційних таблиць для задач педагогіки і медицини (біоакустики) та їх відсутність для спеціалізованих задач обробки та захисту інформації [7]. Більш того, в Україні навіть відсутні й науково-обґрунтовані методи їхньої розробки.

Проведено аналіз частоти повторюваності букв українського алфавіту в спеціалізованих за фахом напрямках (див. рисунок). Розглянуто три фахові групи («Захист інформації», «Криміналістика» та «Економіка»). Для порівняння результатів використані роботи [8, 9]. Встановлено наявність незначного загального відхилення частоти повторюваності букв від попередніх досліджень. Однак за найбільш вживаними буквами [а], [д], [н] та [о] для напрямку «Економіка» є значне відхилення значень по відношенню до інших напрямків та з [8] і [9]. Так, для букви [д] середнє значення частоти повторюваності склало 5,45%, а взяте з [14] — 3,38 %, що зумовлює відносну розбіжність більше 40%.

Проведено аналіз частоти застосування дифонів (див. таблицю).

Частота вживання деяких дифонів з фоном [а]

Дифон	[ац]	[ца]	[ас]	[са]	[ас]	[са]	[аї]	[їа]	[аї]	[їа]	[ао]	[оа]
Частота вживання, рази	2250	0	823	247	330	0	641	0	0	494	3	7
Відносна частота вживання, %	8,23	0	3,01	0,9	1,21	0	2,34	0	0	1,81	0,01	0,03



Таким чином, проведено аналіз існуючих методів оцінювання рівня захищеності мовної інформації та вперше запропоновано введення нового критерію — коефіцієнту залишкової розбірливості мови, який враховує як частотно-часові характеристики фонем та їхній формант, так і фонетико-лінгвістичні характеристики алофонів і слів української мови.

Проведено дослідження частоти вживання букв та дифонів в українській мові в текстах, спеціалізованих за фаховими призначеннями «Захист інформації», «Криміналістика» та «Фінанси», що є основою для створення науково-обґрунтованого методу розробки артикуляційних таблиць.

#### ВИКОРИСТАНІ ДЖЕРЕЛА

1. Модель технічних розвідок (ТР-2030).— Київ: ДССЗЗІ України, 2015.
2. ASTM E2638-10 ASTM Intl., West Conshohocken, PA.
3. ANSI/ASA S3.5-1997 (R2017) American National Standard Methods for Calculation of the Speech Intelligibility Index / STANDARD by American National Standards of the Acoustical Society of America, 1997.01.01.
4. IEC 60268-16:2011 Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index. Available at: <https://webstore.iec.ch/publication/1214>.
5. ISO 3382 (2012), “Acoustics – Measurement of room acoustics parameters – Part 3: Open plan offices”
6. Blintsov V., Nuzhniy S., Parkhuts L., Kasianov Y. The objectified procedure and a technology for assessing the state of complex noise speech information protection // Eastern-European Journal Of Enterprise Technologies.— 2018.— Vol. 5, N 9 (95).— P. 26–34. <http://dx.doi.org/10.15587/1729-4061.2018.144146>
7. Нужний С.М., Заноскіна П.В. Загальні підходи до формування артикуляційних таблиць української мови для оцінки стану захисту складнозашумленої мовної інформації // Сучасна спеціальна техніка.— 2018.— № 4(55).
8. Сушко С.О., Фомичова Л.Я., Барсуков Є.С. Частоти повторюваності букв і біграм у відкритих текстах українською мовою // Захист інформації.— 2010.— Т. 12, № 3. <http://dx.doi.org/10.18372/2410-7840.12.1968>
9. Архипова О., Журавльов В., Кумейко В. Артикуляційні таблиці слів української мови // Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні : науково-технічний збірник. — 2009.— Вип. 2(19).— С. 13—17.

S. M. Nuzhniy, P. V. Zanoskina

#### Estimating the level of protection of speech from leakage by vibroacoustic channels using the objectified method by the criterion of residual legibility of the language with the use of articulation tables of the Ukrainian language

*Estimating the protection of speech information on the security boundary is the basis for building protection systems and the main criterion for trust in the security system as a whole. The authors propose an objectivized method for estimating the security level of the speech information by the criterion of its residual legibility. To use the method, specialized articulation tables of the Ukrainian language are required. In Ukraine there is no scientifically substantiated method of their development. The research on the selection of criteria for the generation of articulation tables was carried out. Such tables are also required for forensic phonoscopic examinations, cryptanalysis, etc.*

*Keywords: estimation of speaker language security, objectivized method, residual legibility index, articulation tables of the Ukrainian language.*