

# Апаратно-програмний комплекс «Штучна голова». Частина 2

## Оцінювання розбірливості мови в аудиторіях

Дворник<sup>f</sup> О. О., ORCID [0000-0003-4735-2225](https://orcid.org/0000-0003-4735-2225)

Моторнюк Д. Є., ORCID [0000-0001-9027-5259](https://orcid.org/0000-0001-9027-5259)

Дідковська М. В., к.т.н. доц, ORCID [0000-0003-0818-2008](https://orcid.org/0000-0003-0818-2008)

Продеус<sup>s</sup> А. М., д.т.н. проф., ORCID [0000-0001-7640-0850](https://orcid.org/0000-0001-7640-0850)

Національний технічний університет України

"Київський політехнічний інститут імені Ігоря Сікорського" ROR [00syn5v21](https://ror.org/00syn5v21)

Київ, Україна

**Анотація**—Виконано експериментальні дослідження можливості використання розробленого апаратно-програмного комплексу «Штучна голова» для двоканального оцінювання розбірливості мови, спотвореної реверберацією. На першому етапі такого оцінювання здійснюють запис відгуку приміщення на тестовий сигнал у вигляді *mls*-послідовності. На другому етапі оцінюють імпульсну характеристику приміщення шляхом обчислення взаємнокореляційної функції відгуку із тестовим сигналом. На третьому етапі розраховують модуляційні коефіцієнти за формулою Шредера, використовуючи оцінку імпульсної характеристики приміщення. На четвертому, останньому етапі, оцінюють розбірливість мови модуляційним або формантно-модуляційним методом. Результати проведених досліджень узгоджуються із результатами попередніх досліджень, де показано, що розбірливість мови в середині приміщення може бути меншою за таку біля стінки приміщення.

**Ключові слова** — розбірливість мови; реверберація; тестовий сигнал; імпульсна характеристика; модуляційний метод; формантно-модуляційний метод.

### I. ВСТУП

В першій частині даної роботи [1] розглянуто питання корекції частотної характеристики розробленого апаратно-програмного комплексу «Штучна голова». Штучну голову (ШГ) при цьому планувалося використати для запису відгуку приміщення на тестовий сигнал для подальшого обчислення імпульсної характеристики (ІХ) приміщення та обчислення розбірливості мови в різних точках приміщення [2]. Перевага використання ШГ [3] при таких вимірюваннях полягає в можливості виконувати двоканальні (бінауральні) вимірювання із врахуванням екрануючого впливу голови людини на звукові сигнали, що сприймаються [4].

Представлена в [5] та [6] база даних, що містить записи бінауральних імпульсних характеристик кількох приміщень різного призначення, на жаль, є досить обмеженою. Зокрема, вона містить записи бінауральних ІХ, зроблених із використанням штучної голови, лише для одного лекційного приміщення середнього розміру (приблизно 370 м<sup>3</sup>). Тому використання зазначених записів не дозволяє зробити певне узагальнення акустичних властивостей лекційних приміщень й, таким чином, зробити ще один крок в напрямку паспортизації приміщень навчальних закладів [7]. Крім цього, є низка інших задач, розв'язання яких потребує наявності бази бінауральних ІХ. Серед таких задач, зокрема, дослідження просторових властивостей слуху [2], створення ефекту «акустичне дзеркало» [8], оцінювання розбірливості

мови [9], дослідження алгоритмів дереверберації [10]. Оскільки під дереверберацією розуміють часткову нейтралізацію дії так званих пізніх відбить, неминуче виникає питання щодо корисності ранніх (до 50-80 мс) відбить. Таку корисність доведено, зокрема, в [11] та [12], хоча результати подальших досліджень показали, що цінність ранніх відбить не слід перевищувати [13]. Практична корисність подібних досліджень є очевидною, з огляду на різницю розуміння мови школярами різного віку [14].

Метою даної частини роботи є експериментальна перевірка спроможності розробленої ШГ розв'язувати поставлене завдання (вимірювання бінауральних ІХ та розбірливості реверберованої мови) із достатньою для подальшого використання точністю. Зрештою, така перевірка має показати доцільність використання електроакустичної апаратури побутового, не професійного, рівня для розв'язання вимірювальних завдань.

### II. ПОСТАНОВКА ЗАДАЧІ

При вимірюваннях ІХ приміщення, тестовий звуковий сигнал  $x(t)$  випромінюють гучномовцем, розміщеним в точці приміщення, де найбільш імовірним є знаходження диктора. Відгук приміщення  $y(t)$  на стимул  $x(t)$  записують за допомогою пари мікрофонів, розташованих поблизу вушних раковин штучної голови, розміщеної в точці приміщення, де можуть знаходитися слухачі (Рис. 1).



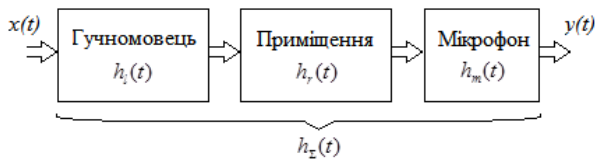


Рис. 1 Спрощена схема процедури оцінювання ІХ приміщення [1]

Якби частотні характеристики гучномовця та мікрофона були близькими до ідеальних (рівномірність амплітудної частотної характеристики та лінійність фазової частотної характеристики) в широкій смузі частот, а також за умови, що сигнал  $x(t)$  є стаціонарним випадковим процесом, спектр якого є постійним в межах широкої смуги частот, імпульсну характеристику приміщення  $h_r(t)$  можна було б оцінювати, обчислюючи взаємно-кореляційну функцію  $K_{XY}(\tau)$  випадкових процесів  $x(t)$  та  $y(t)$  [15]:

$$K_{XY}(\tau) = D_X \cdot h_r(\tau),$$

де  $D_X$  - дисперсія процесу  $x(t)$ .

З огляду на наявність апаратурних спотворень (Рис. 1), замість ІХ приміщення  $h_r(t)$  насправді буде оцінено згортку

$$h_{\Sigma}(t) = h_l(t) \otimes h_r(t) \otimes h_m(t) = h_r(t) \otimes h_m(t), \quad (1)$$

де  $\otimes$  - символ згортки;  $h_l(t)$  - ІХ гучномовця;  $h_m(t)$  - ІХ одного із мікрофонів;  $h_{lm}(t) = h_l(t) \otimes h_m(t)$  - ІХ системи «гучномовець-мікрофон» (ГМ).

В [1] запропоновано оцінювати ІХ приміщення  $h_r(t)$  із використанням виразу

$$h_r(t) = \mathbb{F}^{-1} \{H_r(f)\} = \mathbb{F}^{-1} \left\{ \frac{H_{\Sigma}(f)}{|H_{lm}(f)|} \cdot M_R(f) \right\}, \quad (2)$$

де  $H_{\Sigma}(f) = \mathbb{F}\{h_{\Sigma}(t)\}$  - частота характеристика системи Рис. 1;  $\mathbb{F}$  - символ прямого перетворення Фур'є;  $|\cdot|$  - символ модуля;  $M_R(f)$  - множник регуляризації [16], призначений для зменшення дисперсії оцінки ІХ. В [1] запропоновано в якості такого множника використовувати вікно Ханна, зсунутого по осі частот таким чином, що його максимум знаходиться на частоті  $F_s/2$ :

$$M_r(f) = 0,5 \left[ 1 - \cos(\pi(f - F_s/2)/\Delta F) \right], \quad (3)$$

$$f = 0, \dots, F_s,$$

де  $\Delta F$  - параметр регуляризації;  $F_s$  - частота дискретизації (для  $F_s = 44,1$ кГц в [1] рекомендовано прийняти  $\Delta F = 18$ кГц).

Після обчислення ІХ приміщення стає можливим оцінювання розбірливості мови модуляційним [17], [18] або формантно-модуляційним [19], [20] методом. Першим кроком такого оцінювання є обчислення за

формулою Шредера [21], так званих коефіцієнтів передачі модуляції:

$$m_{ki} = \frac{\left| \int_0^{\infty} h_k^2(t) \exp(-j2\pi F_i t) dt \right|}{\int_0^{\infty} h_{ki}^2(t) dt}, \quad (4)$$

де  $h_k(t)$  - результат фільтрації функції  $h_r(t)$  смуговим фільтром (на практиці використовують 7 октавних фільтрів із центральними частотами від 125 Гц до 8 кГц);  $F_i$  - частота модуляції (на практиці використовують 14 значень  $F_i$ , в діапазоні від 0,63 Гц до 12,5 Гц).

Наступними кроками є обчислення ефективних відношень сигнал-шум для кожної смуги та кожної модуляційної частоти:

$$E_{ki} = 10 \lg \frac{m_{ki}}{1 - m_{ki}}, \quad (5)$$

середнього індексу передачі модуляції

$$MT_k = \frac{1}{14} \sum_{i=1}^{14} T_{ki},$$

$$T_{ki} = \begin{cases} 0, & E_{ki} < -15; \\ (E_{ki} + 15)/30, & -15 \leq E_{ki} \leq +15; \\ 1, & E_{ki} > +15. \end{cases}$$

Останнім кроком, згідно модуляційного методу [18], є обчислення індексу передачі мови:

$$STI = \sum_{k=1}^7 \alpha_k \cdot MT_k - \sum_{k=1}^6 \beta_k \cdot \sqrt{MT_k \cdot MT_{k+1}},$$

де  $\alpha_k$  - вагові коефіцієнти,  $\beta_k$  - коефіцієнти надмірності, значення яких для октавних смуг із центральними частотами  $f_0$  наведено в табл. 1.

Згідно формантно-модуляційного методу [19], на останньому кроці обчислюють словесну розбірливість:

$$W = \begin{cases} 1,54 \cdot A^{0,25} [1 - \exp(-11 \cdot A)], & A < 0,15; \\ 1 - \exp\left(\frac{11 \cdot A}{1 + 0,7 \cdot A}\right), & A \geq 0,15; \end{cases}$$

ТАБЛИЦЯ 1 ВАГОВІ КОЕФІЦІЄНТИ ТА КОЕФІЦІЄНТИ НАДМІРНОСТІ ДЛЯ ОКТАВНИХ СМУГ

$f_{0k}$ , Гц	125	250	500	1000	2000	4000	8000
$\alpha_k$	0,085	0,127	0,230	0,233	0,309	0,224	0,173
$\beta_k$	0,085	0,078	0,065	0,011	0,047	0,095	-



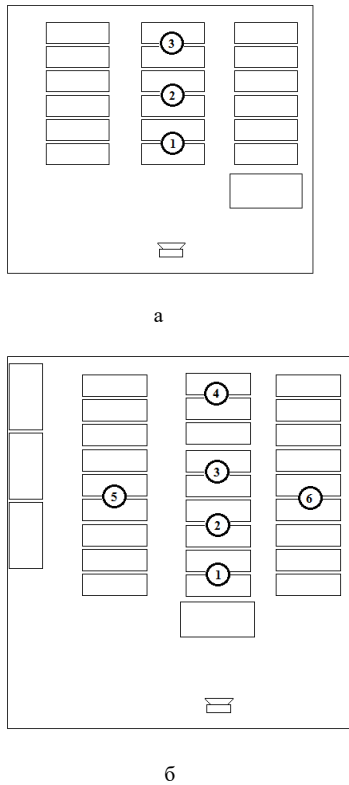


Рис. 2 Плани малої аудиторії 438 (а) та середньої аудиторії 209 (б)

$$A = \sum_{k=1}^7 p_k \cdot P_k(E_k),$$

де  $A$  – формантна розбірливість;  $p_k$  – імовірність знаходження формант в  $k$ -й смузі частот;  $P_k(E_k)$  – коефіцієнт сприйняття мови. В працях Н. Б. Покровського [22] та М. А. Сапожкова [23] коефіцієнти сприйняття відрізняються та вважаються незалежними від смуги частот:  $P_k(E_k) = P(E_k)$ . В [24] викладено результати досліджень причин такої ситуації та показано доцільність використання уточнених коефіцієнтів сприйняття, форма яких залежить від смуги частот.

Метою даної роботи є оцінювання індексу STI та словесної розбірливості  $W$  для різних точок кількох лекційних приміщень та зіставлення отриманих результатів між собою, а також із результатами суб'єктивного оцінювання розбірливості мови. Такі порівняння дозволять зробити висновок щодо достовірності отриманих результатів та спроможності розробленого комплексу «Штучна голова» [1] до об'єктивного вимірювання розбірливості мови.

### III. ОРГАНІЗАЦІЯ ДОСЛІДЖЕНЬ

На відміну від німецького вимірювального комплексу, де використовувалася професійна аудіоапаратура [5], розроблений комплекс «Штучна голова» [1] містить електроакустичну апаратуру різної якості, а саме: побутову активну колонку Genius SP-HF 2.0 500 (14 Вт, 65 Гц – 20 кГц, сигнал-шум 73 дБ), всепрямовані вимірювальні конденсаторні мікрофони

Superlux ECM-999 (20 Гц – 20 кГц, динамічний діапазон 106 дБ, відношення сигнал-шум 70 дБ), зовнішню звукову карту Steinberg UR242 (24 біт, 192 кГц).

Запис сигналів виконувався із частотою дискретизації 44,1 кГц та глибиною квантування 24 біти. Два мікрофони розташовувалися в зонах вушних раковин штучної голови на відстані 1 см від місцезнаходження вушного каналу.

Основою тестового сигналу слугувала mls-послідовність [25] із ступенем 16 (тобто mls-послідовність містила  $2^{16}$  вибірок), що відповідає довжині сигналу 1,49 с при частоті дискретизації 44,1 кГц. Ця mls-послідовність при випромінненні повторювалася 17 разів, що дозволяло при обчисленні ІХ приміщення усереднювати 16 останніх сплесків оцінки ВКФ для підвищення відношення сигнал-шум на 12 дБ.

Для вимірювань було обрано 2 аудиторії, малої та середнього розміру (Рис. 2), що розташовані в навчальному корпусі №12 Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Характеристики малої аудиторії №438: розміри 6,2x8,9x3,1 м, два вікна, двері, задня стінка приміщення застелена, три ряди із 6 парт в кожному ряді та стіл викладача, наповненість приміщення під час запису сигналів становила 11 студентів та викладач. Дистанції від гучномовця до номерів 1, 2 та 3 точок, де розміщувалася ШГ – 2,4 м, 4,8 м та 6,9 м, відповідно.

Характеристики середньої аудиторії №209: розміри 7,4x11,4x3,1 м, чотири вікна, двері, дві книжних шафи, шафа-гардероб, три ряди із 9 парт в кожному ряді, стіл викладача та 2 додаткових вільних столи, наповненість приміщення під час запису сигналів становила 9 студентів та викладач. Дистанції від гучномовця до номерів 1, 2, 3, 4, 5 та 6 точок, де розміщувалася ШГ – 3,5 м, 5,2 м, 8,6 м, 10,3 м, 6,4 м та 6,4 м, відповідно.

### IV. РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ

На Рис. 3 наведено графіки ІХ розглянутих приміщень із результатами вимірювання часу реверберації. Графіки наведено лише для лівого каналу, оскільки для правого каналу вони є подібними. Обвідну ІХ, побудовану за методом Шредера [26], використано для обчислення часу реверберації, оцінки якого представлено на Рис. 3.

Порівнюючи ІХ аудиторій 438 (Рис. 3а) та 209 (Рис. 3б), бачимо, що в аудиторії 209 спостерігається біля 3-х потужних ранніх відбиттів на інтервалі часу до 30 мс та ще один виразний сплеск в момент часу 0,27 с, тоді як спад енергії в аудиторії 438 є більш плавним. Час реверберації в обох приміщеннях виявився досить великим та близьким до 1 с, що у випадку аудиторії 209 пояснюється слабким заповненням порівняно великого приміщення, а у випадку аудиторії 438 пояснюється наявністю скляної задньої стінки.



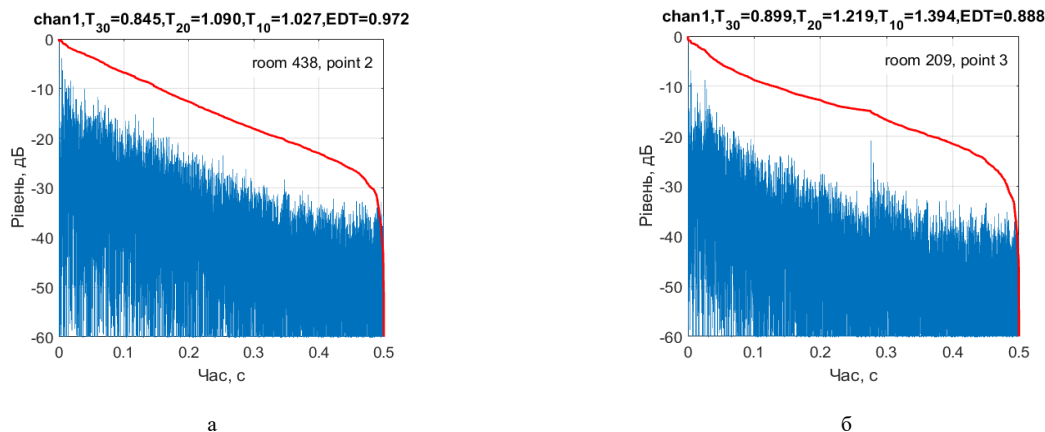


Рис. 3 ІХ та час реверберації для аудиторій 438 (а) та 209 (б)

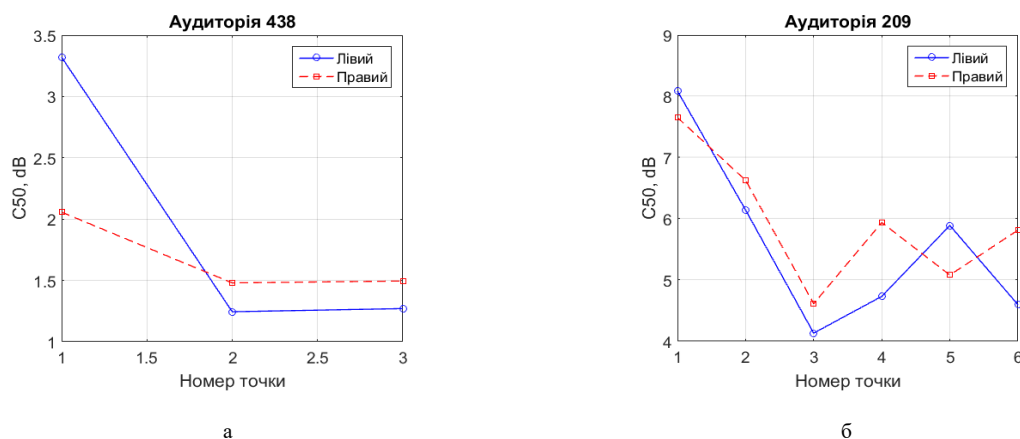


Рис. 4 Оцінки параметру  $C_{50}$  для аудиторій 438 (а) та 209 (б)

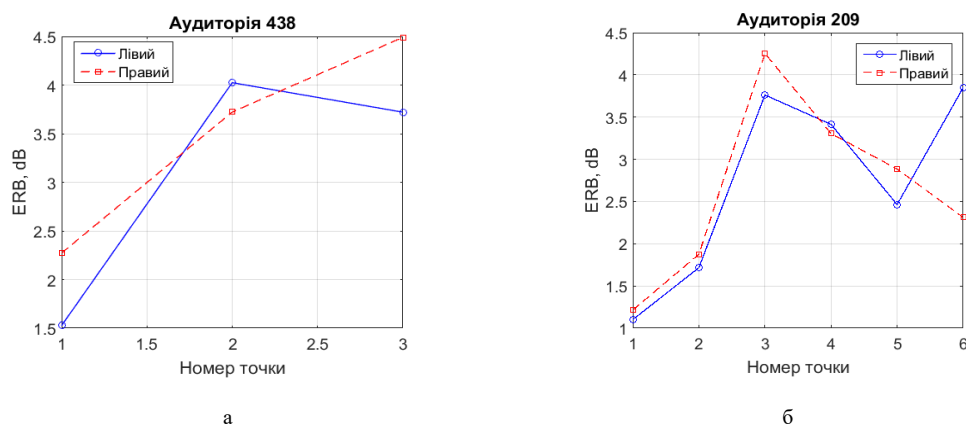


Рис. 5 Оцінки параметру  $ERB$  для аудиторій 438 (а) та 209 (б)

Результати оцінювання характеристик  $C_{50} = 10 \lg(E_0^{50} / E_{50}^\infty)$  та  $ERB = 10 \lg(E_0^{50} / E_0^{10})$  наведено на Рис. 4 та Рис. 5, відповідно, де  $E_a^b = \int_a^b h^2(t) dt$ ,  $a$  та  $b$  - час в мілісекундах.

Оскільки для  $t > 0,5$  с рівень шумової завади переважає над рівнем реверберації (Рис. 3), при обчисленнях  $C_{50}$  в якості  $\infty$  використано значення 0,5 с.



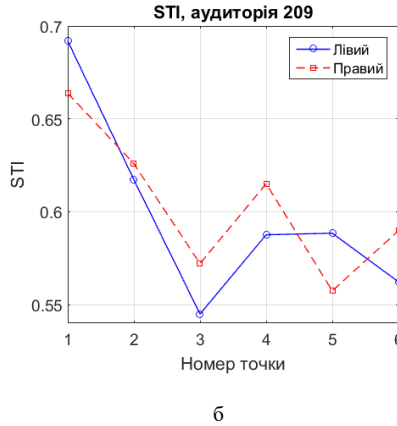
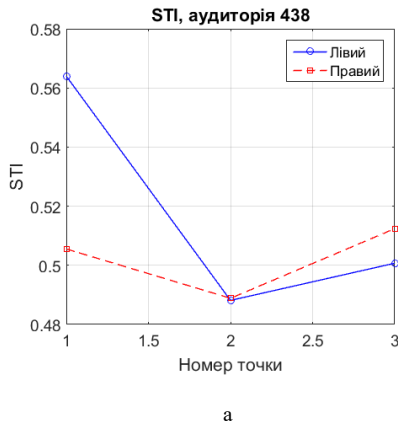


Рис. 6 Оцінки індексу STI для аудиторій 438 (а) та 209 (б)

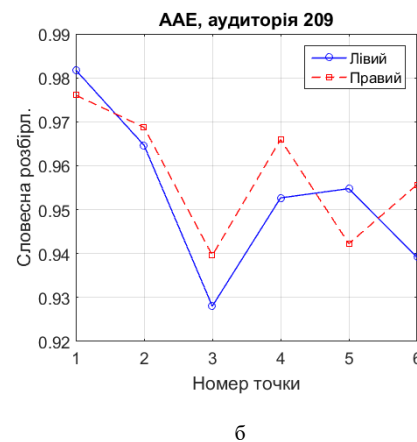
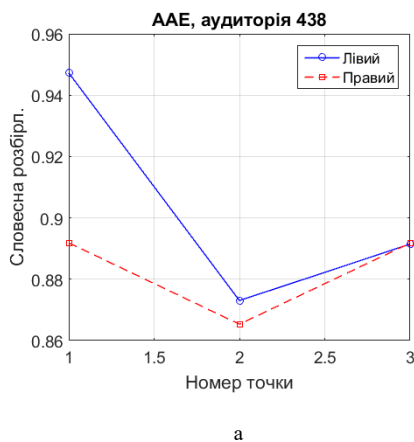


Рис. 7 Оцінки словесної розбірливості для аудиторій 438 (а) та 209 (б)

Згідно отриманим оцінкам  $C_{50}$  (Рис. 4), чіткість мовлення в аудиторії 209 має бути помітно вищою, ніж в аудиторії 438, причому ця чіткість має швидко спадати із наближенням до середини кімнати, після чого має стабілізуватися на певному рівні. Цікаво, що, згідно оцінкам  $C_{50}$ , в аудиторії 209 чіткість мови біля задньої стінки (точка 4) є помітно вищою, ніж в середині кімнати (точка 3). В аудиторії 438 подібний ефект також спостерігається, але є ледь помітним.

Поведінка показника  $ERB$  (Рис. 5), що характеризує співвідношення енергій ранніх відбиттів та прямого звуку, також підтверджує існування зазначеного ефекту. Дійсно, якщо вважати що прямий звук є кращим з точки зору розбірливості мови, порівняно із ранніми відбиттями, тоді, згідно Рис. 5а, бачимо що в аудиторії 438 розбірливість в лівому каналі є вищою біля задньої стінки (точка 3), порівняно із серединою приміщення (точка 2). Проте результати оцінювання  $ERB$  в правому каналі свідчать про протилежне. В аудиторії 209 (Рис. 5б) ситуація є не такою суперечливою: в обох каналах, розбірливість мови біля задньої стінки (точка 4) мала би покращитися, порівняно із серединою приміщення (точка 3).

Зважаючи на певну неоднозначність отриманих оцінок  $C_{50}$  та  $ERB$ , які є непрямими мірами розбірливості мови, цікаво оцінити розбірливість мови у вказаних аудиторіях модуляційним та формантно-модуляційним методами із використанням (1)–(5).

Результати оцінювання розбірливості мови модуляційним методом представлено на Рис. 6, а формантно-модуляційним методом – на Рис. 7, при цьому оцінки, отримані формантно-модуляційним методом, були розраховані із використанням уточнених коефіцієнтів сприйняття, форма яких залежить від смуги частот [24].

Неважко бачити хороше узгодження оцінок розбірливості мови, отриманих модуляційним та формантно-модуляційним методами, а також узгодження їх із оцінками параметру  $C_{50}$  (Рис. 4). Слід, однак, зазначити, що ефект підвищення розбірливості мови біля задньої стінки аудиторії 438, порівняно із серединою приміщення, є значно виразнішим для оцінок розбірливості мови (Рис. 6а та Рис. 7а), ніж для коефіцієнту  $C_{50}$  (Рис. 4а). Що стосується аудиторії 209, де було виконано вимірювання ще й біля бокових стінок (точки з номерами 5 та 6), бачимо, що біля бокових стінок аудиторії розбірливість мови є також



вищою, ніж в центрі приміщення. Зазначимо, що ефект підвищення розбірливості мови біля стінок приміщення було виявлено також в [14] шляхом суб'єктивного оцінювання розбірливості мови.

Даному ефекту можна дати досить просте фізичне пояснення. Дійсно, до вух людини, що сидить біля стінки, майже одночасно (в наших випадках різниця в часі становила 3-5 мс) приходять прямий та відбитий від стінки звук. При цьому потужність відбитого від стінки звуку є ненабагато меншою, порівняно із потужністю прямого сигналу. На низьких частотах (100-200 Гц) сигнали є практично когерентними, тому при близьких рівнях прямого та відбитого звуку на цих частотах можна очікувати підвищення рівня результуючого звуку до 6 дБ, а на високих частотах, де сигнали є практично некогерентними, рівень звуку може підвищитися до 3 дБ. Зауважимо, що хоча ці оцінки є дуже приблизними, проте вони добре узгоджуються із результатами, наведеними в [11], [12]. Що стосується середини приміщення, відбитий від стінки звук приходиться із більшою затримкою відносно прямого звуку (в наших випадках 15-20 мс) та є послабленим за рахунок розповсюдження в просторі. Оскільки про когерентність прямого та відбитого звуків в даному випадку вже не йдеться, то очевидно, що рівень сумарного сигналу має бути меншим, ніж біля стінки. Крім того, слід вказати ще й на те, що через помітну затримку між прямим та відбитим сигналом кімната фактично перетворюється в гребінчастий фільтр (комб-фільтр) [27] із суттєво нерівномірною амплітудно-частотною характеристикою. Таким чином, сумарний звук є ще й спотвореним, що також не може сприяти хорошій розбірливості мови в середині кімнати.

Узагальнюючи, можна сказати, що в аудиторіях, подібних розглянутим в даній роботі, користь ранніх відбиттів, за критерієм розбірливості мови, є різною в різних частинах приміщення, й біля стін приміщення вона є більшою, ніж в середині приміщення.

Оскільки в даній роботі, за браком місця, не приділено уваги аналізу співвідношень розбірливості мови в лівому та правому каналах слухової системи людини, що сидить на невеликій (1-2 м) відстані від бокових стінок приміщення, в подальшому доцільно ліквідувати зазначений пробіл.

#### ВИСНОВКИ

Отримані результати загалом добре узгоджуються із результатами попередніх досліджень, що свідчить про можливість використання розробленого апаратно-програмного комплексу «Штучна голова» для двоканального оцінювання розбірливості мови.

Зокрема, отримані результати кількісно підтвердили відомий феномен зростання розбірливості мови при наближенні слухача до відбиваючої поверхні. Як наслідок, розбірливість мови в середині аудиторій може бути помітно меншою за таку біля стінок приміщення.

В подальшому доцільно приділити увагу аналізу іншого феномену, що спостерігається слухачами на

невеликій (1-2 м) відстані від бокових стін й полягає в зниженні розбірливості мови в слуховому каналі, ближчому до стіни, порівняно із розбірливістю мови в протилежному слуховому каналі.

#### ПЕРЕЛІК ПОСИЛАНЬ

- [1]. O. O. Dvornyk, D. I. Motorniuk, M. V. Didkovska, and A. M. Prodeus, "Artificial Software Complex 'Artificial Head'. Part 1 Adjusting the Frequency Response of the Path," *Microsystems, Electron. Acoust.*, vol. 25, no. 1, pp. 56-64, 2020, DOI: [10.20535/2523-4455.me.198431](https://doi.org/10.20535/2523-4455.me.198431).
- [2]. J. Blauert, Ed., *The Technology of Binaural Listening*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, ISBN: 978-3-642-37761-7.
- [3]. "HEAD acoustics - Telecom Division - Binaural Recording Systems - Artificial Head Measurement System HMS II.3 - Overview." [Online]. Available: [https://www.head-acoustics.com/eng/telecom\\_hms\\_II\\_3.htm](https://www.head-acoustics.com/eng/telecom_hms_II_3.htm). [Accessed: 23-Mar-2020].
- [4]. P. L. Søndergaard and P. Majdak, "The auditory modeling toolbox," in *The Technology of Binaural Listening*, Springer Berlin Heidelberg, 2013, pp. 33-56.
- [5]. M. Jeub, M. Schafer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *2009 16th International Conference on Digital Signal Processing*, 2009, pp. 1-5, DOI: [10.1109/ICDSP.2009.5201259](https://doi.org/10.1109/ICDSP.2009.5201259).
- [6]. "IKS: Aachen Impulse Response Database." [Online]. Available: <https://www.iks.rwth-aachen.de/en/research/tools-downloads/databases/aachen-impulse-response-database/>. [Accessed: 23-Mar-2020].
- [7]. Y. S. Kostyuchok, L. S. Martynovych, D. Y. Motorniuk, V. O. Nechytailo, A. V. Hrapachevskiy, and A. M. Prodeus, "Acoustic certification of classrooms," *Electron. Commun.*, vol. 21, no. 2, pp. 63-70, Nov. 2016, DOI: [10.20535/2312-1807.2016.21.2.82217](https://doi.org/10.20535/2312-1807.2016.21.2.82217).
- [8]. "SOUND FORGE Audio Studio 14 – Домашняя студия редактирования звука." [Online]. Available: <https://www.magix.com/ru/muzyka/sound-forge/sound-forge-audio-studio/>. [Accessed: 23-Mar-2020].
- [9]. A. Prodeus, V. Didkovskiy, M. Didkovska, I. Kotvytskyi, and D. Motorniuk, "Automated Subjective Assessment of Speech Intelligibility Under Diotic and Dichotic Listening," in *Data-Centric Business and Applications*, Springer, Cham, 2020, pp. 1-16, URL: [http://link.springer.com/10.1007/978-3-030-43070-2\\_1](http://link.springer.com/10.1007/978-3-030-43070-2_1), DOI: [10.1007/978-3-030-43070-2\\_1](https://doi.org/10.1007/978-3-030-43070-2_1).
- [10]. E. A. P. Habets, N. D. Gaubitch, and P. A. Naylor, "Temporal selective dereverberation of noisy speech using one microphone," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 4577-4580, DOI: [10.1109/ICASSP.2008.4518675](https://doi.org/10.1109/ICASSP.2008.4518675).
- [11]. G. A. Soulodre, N. Popplewell, and J. S. Bradley, "Combined effects of early reflections and background noise on speech intelligibility," *J. Sound Vib.*, vol. 135, no. 1, pp. 123-133, Nov. 1989, DOI: [10.1016/0022-460X\(89\)90759-1](https://doi.org/10.1016/0022-460X(89)90759-1).
- [12]. J. S. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms," *J. Acoust. Soc. Am.*, vol. 113, no. 6, p. 3233, 2003, DOI: [10.1121/1.1570439](https://doi.org/10.1121/1.1570439).
- [13]. I. Arweiler, J. Buchholz, and T. Dau, "Speech intelligibility enhancement by early reflections | Proceedings of the International Symposium on Auditory and Audiological Research," in *Speech processing and perception under adverse conditions*, 2009, pp. 289-298, URL: <https://proceedings.isaar.eu/index.php/isaarproc/article/view/2009-29>.
- [14]. H. Sato and J. S. Bradley, "Evaluation of acoustical conditions for speech communication in working elementary school classrooms," *J. Acoust. Soc. Am.*, vol. 123, no. 4, pp. 2064-2077, 2008, DOI: [10.1121/1.2839283](https://doi.org/10.1121/1.2839283).
- [15]. V. S. Didkovskiy, M. V. Didkovskaya, and A. M. Prodeus, *Komp'yuterna obrobka akustychnykh sygnaliv. Navchal'nyy*



posibnyk. [Computer processing of acoustic signals. Tutorial.]. Kyiv: Imecs-LTD, 2010.

- [16]. A. N. Tikhonov, "O nekorrektnykh zadachakh lineynoy algebrы i ustoychivom metode ikh resheniya [On ill-posed problems of linear algebra and a stable method for solving them]," *DAN USSR*, vol. 163, no. 3, pp. 591–594, 1965.
- [17]. H. J. M. Steeneken and T. Houtgast, "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am.*, vol. 67, no. 1, pp. 318–326, 1980, DOI: [10.1121/1.384464](https://doi.org/10.1121/1.384464).
- [18]. H. J. . Steeneken and T. Houtgast, "Validation of the revised STIR method," *Speech Commun.*, vol. 38, no. 3–4, pp. 413–425, Nov. 2002, DOI: [10.1016/S0167-6393\(02\)00010-9](https://doi.org/10.1016/S0167-6393(02)00010-9).
- [19]. A. M. Prodeus, L. B. Dronzhevskaya, V. A. Klimkov, and D. A. Shagitova, "Formantnyy i formantno-modulyatsionnyy metody otsenki razborchivosti rechi. Chast' 1. Unifikatsiya algoritmov," *Electron. Commun.*, vol. 15, no. 6 (part 2), pp. 117–124, 2010.
- [20]. A. N. Prodeus, L. B. Dronzhevskaya, V. A. Klimkov, and D. A. Shagitova, "Formantnyy i formantno-modulyatsionnyy metody otsenki razborchivosti rechi. Chast' 2. Tochnost' i skorost' izmereniy [Formant and formant-modulation methods for assessing speech intelligibility. Part 2. Accuracy and speed of measurements.]," *Electron. Commun.*, vol. 16, no. 6, pp. 16–24, 2011.
- [21]. M. R. Schroeder, "Modulation Transfer Functions: Definition and Measurement," *Acta Acust. united with Acust.*, vol. 49, no. 3, pp. 79-182(4), 1981, URL: <https://www.ingentaconnect.com/content/dav/aaua/1981/00000049/00000003/art00004?crawler=true&mimetype=application/pdf>
- [22]. N. B. Pokrovskiy, *Raschet i izmerenie razborchivosti rechi [Prediction and measurement of speech intelligibility]*. Moscow, USSR: Svyazizdat, 1962.
- [23]. M. A. Sapozhkov, *Rechevoy signal v kibernetike i svyazi [Speech signal in cybernetics and communication]*. Moscow: Svyazizdat, 1963.
- [24]. A. M. Prodeus, V. S. Didkovskiy, and M. V. Didkovskaya, *Akusticheskaya ekspertiza i korrektsiya kommunikatsionnykh kanalov [Acoustic examination and correction of communication channels]*. LAP LAMBERT Academic Publishing, 2017, ISBN: 978-3-330-04591-0.
- [25]. V. E. Gantmakher, N. E. Bystrov, and D. V. Chebotarev, *Shumopodobnyye signaly. Analiz, sintez, obrabotka [Noise-like signals. Analysis, synthesis, processing]*. St. Petersburg: Nauka i tekhnika, 2005.
- [26]. M. R. Schroeder, "New Method of Measuring Reverberation Time," *J. Acoust. Soc. Am.*, vol. 37, no. 6, pp. 1187–1188, Jun. 1965, DOI: [10.1121/1.1939454](https://doi.org/10.1121/1.1939454).
- [27]. U. Zölzer, Ed., *DAFX: Digital Acoustic Effects*, 2nd ed. John Wiley & Sons, 2011, ISBN: 978-0-470-66599-2.

**Надійшла до редакції 14 серпня 2020 року**

UDC 534.84

# Hardware and Software System "Artificial Head". Part 2

## Evaluation of Speech Intelligibility in Classrooms

O. O. Dvornyk<sup>f</sup>, ORCID [0000-0003-4735-2225](https://orcid.org/0000-0003-4735-2225)

D. Ie. Motorniuk, ORCID [0000-0001-9027-5259](https://orcid.org/0000-0001-9027-5259)

M. V. Didkovska, PhD Assoc.Prof., ORCID [0000-0003-0818-2008](https://orcid.org/0000-0003-0818-2008)

A. M. Prodeus<sup>s</sup>, Dr.Sc.(Eng.) Prof., ORCID [0000-0001-7640-0850](https://orcid.org/0000-0001-7640-0850)

National technical university of Ukraine "Igor Sikorsky Kyiv polytechnic institute" ROR [00syn5v21](https://ror.org/00syn5v21)  
Kyiv, Ukraine

DOI: [10.20535/2523-4447.209928](https://doi.org/10.20535/2523-4447.209928)

**Abstract**—Experimental studies of the use of the developed hardware and software tool “Artificial Head” for two-channel estimation of the intelligibility of the speech distorted by reverberation have been performed in this paper. The peculiarity of this complex is that it contains electro-acoustic equipment of different quality, including household appliances with mediocre quality features.

In the first stage of this evaluation, the response of the room to the test signal is recorded. The test signal was based on the maximum-length sequence (mls) which is a periodic two-level signal of length  $2^{16}$ , which corresponds to a signal length of 1.49 s at a sampling rate of 44.1 kHz. This mls-sequence was repeated 17 times during radiation and averaging the last 16 bursts of the cross-correlation estimates have been made to increase the signal-to-noise ratio by 12 dB in the calculation of the room impulse response.

The calculation of the cross-correlation function of the response and the test signal was performed in the second stage, while the correction of the frequency characteristic of the measuring path was performed by it dividing by the amplitude-frequency characteristic of the loudspeaker-microphone subsystem. To improve the accuracy of the calculations, the result of such division was multiplied by the regularization factor in the form of a Hann window.

In the third stage, the modulation coefficients were calculated according to the Schroeder formula, using the room impulse response estimate. In the fourth, last step, speech intelligibility was evaluated by a modulation or formant-modulation method.

The results of the speech intelligibility evaluation in two lecture rooms of small and medium size showed that the speech intelligibility in the middle of the room can be less than that near the wall of the room. These results are in line with the results of previous studies, where speech intelligibility has been evaluated by objective and subjective methods for another lecture room. It should be noted that although the estimates of the widely used  $C_{50}$  coefficient are consistent with the estimates of speech intelligibility, the phenomenon of increasing speech intelligibility near the walls of the room is more pronounced when applying the speech intelligibility estimates.

The results presented in this paper indicate that the usefulness of early reflections in the room is different in different parts of the room. This important fact must be taken into account both in the design and the reconstruction of the lecture rooms.

**Keywords** — *speech intelligibility; reverberation; test signal; impulse response; modulation method; formant-modulation method.*

