НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

Факультет електроніки
(повна назва інституту/факультету)

Кафедра акустичних та мультимедійних електронних систем
(повна назва кафедри)

«До захисту допущено»
Заступник завідувача кафедри
_____ О.І.Дрозденко
(підпис) (ініціали, прізвище)

«01» червня 2022 р.

Дипломна робота
на здобуття ступеня бакалавра

зі спеціальності (спеціалізації) 171 Електроніка
(код та назва спеціальності)

на тему: Об'єктивне оцінювання розбірливості мови в університетських лекційних приміщеннях формантним та модуляційним методами

Виконав: студент 4 курсу, групи ДГ-81
(шифр групи)

Найда Антон Сергійович _____
(прізвище, ім'я, по батькові) (підпис)

Керівник проф.каф.АМЕС, д.т.н, проф. Продеус А.М. _____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Консультант _____ _____ _____
(назва розділу) (посада, вчене звання, науковий ступінь, прізвище, ініціали) (підпис)

Рецензент доцент кафедри ЕІ, к.т.н., доцент О.П. Шуляк _____
(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище та ініціали) (підпис)

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студент _____
(підпис)

Київ – 2022 року

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Факультет електроніки
(повна назва)

Кафедра акустичних та мультимедійних електронних систем
(повна назва)

Рівень вищої освіти – перший (бакалаврський)

Спеціальність (спеціалізація) 171 Електроніка
(код і назва)

ЗАТВЕРДЖУЮ

Заступник завідувача кафедри

_____ О.І.Дрозденко
(підпис) (ініціали, прізвище)

«01» червня 2022 р.

ЗАВДАННЯ

на дипломну роботу студенту

Найді Антону Сергійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Об'єктивне оцінювання розбірливості мови в університетських лекційних приміщеннях формантним та модуляційним методами

керівник роботи Продеус А.М., д.т.н., проф.
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від «06» червня 2022 р. №911-с

2. Строк подання студентом роботи 01 червня 2022 року

3. Вихідні дані до роботи вид простору, відстань до диктора.

4. Зміст (дипломної роботи) пояснювальної записки (перелік завдань, які потрібно розробити) 1. Огляд сучасних об'єктивних методів вимірювання розбірливості мови. 2. Розрахунок розбірливості мови, що маскується шумовою та реверберaційною перешкодами. 3. Розробка Андроїд-додатку для вимірювання розбірливості мови. 4. Вимірювання розбірливості мови в лекційних приміщеннях факультету електроніки КПІ ім.Ігоря Сікорського.

5. Перелік графічного (ілюстративного) матеріалу (із зазначенням обов'язкових креслеників, плакатів, презентацій тощо) Презентація

6. Дата видачі завдання 31 січня 2022 року.

7. Орієнтовний перелік публікацій:

публікації у виданнях, які включені до наукометричних базах Scopus або Web of Science - 4.

Календарний план

№ з/п	Назва етапів виконання дипломного проекту (роботи)	Строк виконання етапів проекту (роботи)	Примітка
1	Огляд сучасних об’єктивних методів вимірювання розбірливості мови.	29.02.2022р.	Виконано
2	Розрахунок розбірливості мови, що маскується шумовою та реверберaційною перешкодами	31.03.2022р.	Виконано
3	Розробка Андроїд-додатку для вимірювання розбірливості мови.	30.04.2022р.	Виконано
4	Вимірювання розбірливості мови в лекційних приміщеннях факультету електроніки КПІ ім.Ігоря Сікорського.	25.05.2022р.	Виконано
5	Оформлення пояснювальної записки до дипломної роботи та графічного матеріалу	01.06.2022р.	Виконано

Студент _____ А.С. Найда

(підпис) (ініціали, прізвище)

Керівник роботи _____ А.М. Продеус

(підпис) (ініціали, прізвище)

УДК 534.4

РЕФЕРАТ

Найда, А. С. Об'єктивне оцінювання розбірливості мови в університетських лекційних приміщеннях формантним та модуляційним методами : дипломна робота бакалавра : 171 Електроніка / Найда Антон Сергійович. – Київ, КПІ ім. Ігоря Сікорського, 2022. – 86 с.

Обсяг дипломної роботи складає 86 сторінок, 10 таблиць, 35 рисунків та 54 літературних джерел.

Бакалаврська робота виконана на актуальну тему об’єктивного оцінювання розбірливості мови в університетських лекційних приміщеннях. В роботі зроблено огляд моделей мовного тракту людини та розглянуто спектральний склад звуків розмовної мови. Проаналізовані переваги та недоліки сучасних об’єктивних методів оцінювання розбірливості мови, а саме, формантного та модуляційного. Проведено розрахунок розбірливості мови, спотвореної шумом, на основі алгоритмів прогнозування розбірливості зашумленої мови. В середовищі Android розроблено калькулятор розбірливості мови. Крім того, виконано експериментальне оцінювання розбірливості мови, спотвореної реверберацією, в різних точках лекційної аудиторії (№209, корпус №12) кафедри акустичних та мультимедійних електронних систем факультету електроніки КПІ ім. Ігоря Сікорського.

Публікації:

1. N.Y. Filipova, O.V. Korzhik, A.S. Chayka, S.A. Naida, M.O. Korzhik, A.S. Naida. Dynamics of Receiving Electroelastic Spherical Shell with a Filler. *J. Nano- Electron. Phys.* 12 No4, 04034 (2020). DOI: **https://doi.org/10.21272/jnep.12(4).04034**

2. Korzhyk, O., Naida, S., Kurdiuk, S., Nizhynska, V., Korzhyk, M., & Naida, A. Use of the pass-through method to solve sound radiation problems of a

spherical electro-elastic source of zero order. EUREKA: Physics and Engineering, (5), 133-146. (2021). DOI: https://doi.org/10.21303/2461-4262.2021.001292

3. S.A. Naida, T.M. Zheliaskova, A.S. Naida, H.A. Kliushnichenko, A.V. Damarad. Methods for Calculating the Transfer Functions of Broadband Plate Piezoelectric Transducers with Transition Layers // *J. Nano- Electron. Phys.* , pp. 1-6, V.13, №6, 2021. DOI: https://doi.org/10.21272/jnep.13(6).06029

4. Sergey Naida, Oleksii Korzhyk, Ivan Lastivka, Olha Pavlenko, Tetiana Zheliaskova, Maxim Korzhyk, Anton Naida, Nikita Naida, Oleksander Chaika. Mode-Matching Method Applied to Solving Sound Reception Problem Using Helmholtz Resonator. Journal of Nano- and Electronic Physics, Vol. 14, № 1, 01035(7pp) (2022). DOI: https://doi.org/10.21272/jnep.14(1).01035

Ключові слова: формантний метод, модуляційний метод, формантний аналіз, шум, реверберація, індекс розбірливості мови, індекс передачі мови

UDC 534.4

ABSTRACT

Naida, A. S. Objective assessment of speech intelligibility in university classrooms by formant and modulation methods : bachelor's thesis : 171 Electronics / Naida Anton. – Kyiv, Igor Sikorsky Kyiv Polytechnic Institute, 2022. – 86 p.

The volume of the thesis is 86 pages, 10 tables, 35 figures and 54 references.

The bachelor's thesis was performed on the topical issue of objective assessment of speech intelligibility in university lecture halls. The paper reviews the models of the human speech tract and considers the spectral composition of the sounds of spoken speech. The advantages and disadvantages of modern objective methods of speech intelligibility assessment, namely, formant and modulation ones, are analyzed. The intelligibility of speech distorted by noise and reverberation was calculated on the basis of algorithms for predicting intelligibility of noisy speech and reverberated speech. A speech intelligibility calculator has been developed for Android operational system. In addition, an experimental assessment of speech intelligibility at different points of the lecture hall (№209, building №12 of the Department of Acoustic and Multimedia Electronic Systems, Faculty of Electronics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute".

Publications:

1. Korzhyk, O., Naida, S., Kurdiuk, S., Nizhynska, V., Korzhyk, M., & Naida, A. Use of the pass-through method to solve sound radiation problems of a spherical electro-elastic source of zero order. EUREKA: Physics and Engineering, (5), 133-146. (2021). DOI: https://doi.org/10.21303/2461-4262.2021.001292

2. S.A. Naida, T.M. Zheliaskova, A.S. Naida, H.A. Kliushnichenko, A.V. Damarad. Methods for Calculating the Transfer Functions of Broadband Plate Piezoelectric Transducers with Transition Layers // *J. Nano- Electron. Phys.* , pp. 1-6, V.13, №6, 2021. DOI: https://doi.org/10.21272/jnep.13(6).06029

3. Korzhyk, O., Naida, S., Zheliaskova, T., Chaika, O., & Naida, N. (2021). Application of partial areas method in the problem of sound radiation by a sphere in a waveguide with soft acoustically boundaries. Eastern-European Journal of Enterprise Technologies, 5(5 (113), 62–79. DOI: https://doi.org/10.15587/1729-4061.2021.243161

4. Sergey Naida, Oleksii Korzhyk, Ivan Lastivka, Olha Pavlenko, Tetiana Zheliaskova, Maxim Korzhyk, Anton Naida, Nikita Naida, Oleksander Chaika. Mode-Matching Method Applied to Solving Sound Reception Problem Using Helmholtz Resonator. Journal of Nano- and Electronic Physics, Vol. 14, № 1, 01035(7pp) (2022). DOI: https://doi.org/10.21272/jnep.14(1).01035

Keywords: formant method, modulation method, formant analysis, noise, reverberation, speech intelligibility index, speech transmission index.

ЗМІСТ

ВСТУП.. 9
РОЗДІЛ 1. ОГЛЯД СУЧАСНИХ ОБ'ЄКТИВНИХ МЕТОДІВ ОЦІНЮВАННЯ РОЗБІРЛИВОСТІ МОВЛЕННЯ 10
1.1 Модель мовного тракту людини 10
1.2 Спектральний склад звуків мови 16
1.3 Формантний аналіз мови 18
1.4 Методи оцінки разбірливості мови 23
1.4.1 Формантний метод 23
1.4.2 Модуляційний метод 29
1.5 Висновки до Розділу 1 31
РОЗДІЛ 2. РОЗРАХУНОК РОЗБІРЛИВОСТІ МОВИ, СПОТВОРЕНОЇ ШУМОМ ТА РЕВЕРБЕРАЦІЄЮ 32
2.1 Алгоритм прогнозування розбірливості зашумленої мови 37
2.2 Алгоритм прогнозування розбірливості реверберованого мовлення 38
2.3 Висновки до Розділу 2 41
РОЗДІЛ 3. КАЛЬКУЛЯТОР РОЗБІРЛИВОСТІ МОВИ НА ПЛАТФОРМІ ANDROID 43
3.1. Розробка алгоритму розрахунку 43
3.2 Реалізація алгоритму у середовищі Android 45
3.3 Висновки до Розділу 3 48
РОЗДІЛ 4. ЕКСПЕРИМЕНТАЛЬНЕ ВИМІРЮВАННЯ РОЗБІРЛИВОСТІ МОВЛЕННЯ СПОТВОРЕНОГО РЕВЕРБЕРАЦІЄЮ 49
4.1 Запис та обробка тестових сигналів для оцінювання приміщення 49
4.2 Оцінювання розбірливості мовлення в різних точках університетської аудиторії 50
4.3 Висновки до Розділу 4 62
ВИСНОВКИ 64
ПЕРЕЛІК ЛІТЕРАТУРНИХ ДЖЕРЕЛ 66
ДОДАТОК А. КАЛЬКУЛЯТОР РОЗБІРЛИВОСТІ МОВЛЕННЯ В СЕРЕДОВИЩІ ANDROID STUDIO 72
ДОДАТОК Б. ПРОГРАМИ ДЛЯ ПОБУДОВИ ГРАФІКІВ ОЦІНОК РОЗБІРЛИВОСТІ МОВЛЕННЯ 75

ВСТУП

Задача розрахунку і вимірювання розбірливості мови виникла досить давно, ще на початку XX століття, при налаштуванні трансатлантичних телефонних ліній зв'язку [53], і сфера застосувань її постійно розширюється. Сьогодні методи оцінки розбірливості мови застосовуються не лише для атестації ліній зв'язку, але й для оцінки якості акустики приміщень, визначення ступеня захищеності мовленнєвої інформації при передачі акустичними каналами та ін.

Технічні засоби інженерів постійно змінюються і удосконалюються, а це означає, що зростає перелік чинників, які треба враховувати при оцінюванні розбірливості мови. Це потребує постійного оновлення математичного і програмного забезпечення. Таким чином, тема бакалаврської роботи є новою та актуальною.

В теперішній час найбільшого поширення отримали версії формантного методу оцінювання розбірливості мови, такі як індекс артикуляції AI (Articulation Index) та індекс розбірливості мови SII (Speech Intelligibility Index), а також модуляційний метод, в якому мірою розбірливості мови служить індекс передачі мови STI (Speech Transmission Index). Існує необхідність зіставлення потенційних можливостей цих методів в конкретних практичних ситуаціях, наприклад, коли дія шуму переважає над дією реверберації.

В роботі зроблено огляд моделей мовного тракту людини та розглянуто спектральний склад звуків розмовної мови. Проаналізовані переваги та недоліки сучасних об'єктивних методів оцінювання розбірливості мови, а саме, формантного та модуляційного. Проведено оцінювання розбірливості мови, спотвореної шумом та реверберацією, на основі алгоритмів прогнозування розбірливості зашумленої мови та вимірювання розбірливості реверберованого мовлення. З цією метою в середовищі Android Studio розроблено калькулятор розбірливості мови. Крім того, виконано експериментальне оцінювання розбірливості мови в різних точках лекційної аудиторії (№209, корпус №12) кафедри акустичних та мультимедійних електронних систем факультету електроніки НТУУ «КПІ ім. Ігоря Сікорського».

РОЗДІЛ 1
ОГЛЯД СУЧАСНИХ ОБ'ЄКТИВНИХ МЕТОДІВ ОЦІНЮВАННЯ РОЗБІРЛИВОСТІ МОВЛЕННЯ

1.1 Модель мовного тракту людини

Під мовним трактом людини мається на увазі акустична система, утворена фарингальною і ротовою порожнинами [1-4]. У вимовленні деяких звуків мови бере участь також носова порожнина, яка паралельна ротовій порожнині. Збудження мовного тракту здійснюється потоком повітря, що проходить через голосові зв'язки. Артикуляція - зміна конфігурації мовного тракту з метою отримання членороздільної мови - проводиться за допомогою язика, губ, м'якого неба і нижньої щелепи.

Мовний тракт (МТ) є дуже складною по конфігурації акустичною системою [1]. Одним з найбільш зручних засобів опису форми МТ є функція площі. Функція площі A(x) показує зміну площі поперечного перерізу МТ уздовж його довжини. Функція площі не дає вичерпної інформації про конфігурацію МТ, оскільки важлива не лише площа поперечного перерізу, але і форма перерізу. Як правило, форму перерізу МТ наближено вважають круглою [1-3].

Вплив складності і неоднорідності МТ на його поведінку не вдається описати за допомогою єдиного диференціального рівняння. Для побудови моделі МТ в [1,2] запропоновано використовувати наступну методику:

1) складна і неоднорідна акустична система разбивається на частини, які є досить простими за формою і неоднорідні за властивостями;

2) для кожної з таких частин складається "елементарна" модель;

3) враховуючи фізичний сенс процесів, що відбуваються в системі, визначаються умови сполучення "елементарних" моделей між собою, і будується модель усієї системи в цілому.

Побудована таким чином модель є імітаційною. При правильно обраних початкових положеннях вона здатна описувати усі істотні процеси, що мають місце в системі, а також стати зручним засобом дослідження такої системи.

Розіб'ємо МТ людини на N відрізків достатньо малої довжини Δl: $\Delta l = \frac{l_r}{N}$, де l_r – загальна довжина МТ [1,2]. Будемо вважати, що в межах кожного з відрізків площа його поперечного перетину й властивості залишаються незмінними. Таким чином, ми робимо ступінчасту апроксимацію неперервної функції площі МТ. Хвильовий процес в трубі постійного перетину описується рівнянням:

$$\frac{d^2P}{dt^2} = \frac{1}{c}\frac{d^2P}{dx^2}. \qquad (1.1)$$

де $p(x,t)$ - тиск звукової хвилі в МТ, який є функцією часу t та координати x; c - швидкість розповсюдження звукових хвиль в МТ.

Слід зауважити, що рівняння (1.1) описує тільки випадок плоских акустичних хвиль в МТ. Така ситуація має місце у тому випадку, коли довжина хвилі набагато більше поперечних розмірів МТ. При цьому напрям поширення хвиль співпадає з подовжньою віссю МТ, а наявністю поперечних хвиль можна знехтувати. Такий підхід є зручним засобом побудови моделей МТ. Надалі перед собою ставимо завдання побудови вдосконаленої моделі, яка здатна врахувати ефекти виникнення і поширення поперечних хвиль. Розв'язок рівняння (1.1) зазвичай знаходиться у вигляді:

$$P(x,t) = P_1(x)P_2(t), \qquad (1.2)$$

де $P_2(t)$ - функція, яка описує коливальні властивості системи; $P_1(x)$ - функція, яка описує зміну амплітуди коливального процеса вздовж довжини труби (тобто розповзюдження хвиль вздовж труби).

При цьому розв'язок (1.1) зводиться до розв'язку системи рівнянь:

$$P_1''(x) + \lambda^2 P_1(x) = 0,\ P_2''(t) + \frac{\lambda^2}{c^2}P_2(t) = 0. \qquad (1.3)$$

Якщо відрізок труби має малу довжину Δl, то:

$$P_1(0) = P_1(\Delta l) = \text{const}. \qquad (1.4)$$

Отже

$$P_1(x,t) = k\,P_2(t), \qquad (1.5),$$

тобто ми зводимо систему з розподіленими параметрами, враховуючи її дуже малі розміри, до системи із зосередженими параметрами. В цьому випадку поведінка відрізка труби малої довжини описується рівнянням:

$$P_2''(t) + \frac{\lambda^2}{c^2} P_2(t) = 0. \qquad (1.6)$$

У зв'язку з тим, що ми будуємо модель МТ, призначену для програмної реалізації, складемо різницеву схему розв’язку рівняння (1.6):

$$P(nT) = KP(nT\text{-}2T),\ n = 0,1,2\ldots, \qquad (1.7)$$

де $P(nT)$ – звуковий тиск, який є функцією дискретного часу nT; T – період квантування.

Період квантування (проміжок часу, через який проводяться відліки звукового тиску) знаходиться із співвідношення:

$$T = \frac{\Delta l}{c}\ . \qquad (1.8)$$

Критерій вибору такого T, і, відповідно, Δl, яка задовольняє співвідношенню (1.3), розглянемо нижче. Коефіцієнт K в цій схемі опущено, оскільки він не впливає принципово на динаміку її роботи. При подачі на вхід цієї схеми сигнала типу δ-функції, вона генерує коливання, аналогічні розв’язку (1.7).

Для того, щоб правильно описати поведінку відрізку труби, в неї необхідно ввести граничні умови. Для цього скористаємося наступними міркуваннями. Аналітичний розв’язок рівняння (1.1) може бути записаний у вигляді:

$$P(x,t) = P_+ e^{i(x\text{-}ct)} + P_-\, e^{i(x+ct)}\ . \qquad (1.9)$$

Вираз (1.9) являє собою хвильовий процес у відрізку труби у вигляді суми двох хвиль, одна з яких, $P_+ e^{i(x\text{-}ct)}$, є прямою (падаючою), а друга, $P_- e^{i(x+ct)}$- зворотньою (відбитою) хвилею. Падаюча і відбита хвилі виникають у відрізку труби внаслідок наявності відбиттів до заломленої хвилі на його кінцях. В цьому сенсі затримки можна інтерпретувати наступним чином:

- затримка 1 описує час проходження прямої хвилі, відбитої від початку відрізку труби, уздовж його довжини;

- затримка 2 описує час проходження зворотної хвилі, від кінця відрізку труби, уздовж його довжини.

Тепер перейдемо до побудови загальної моделі МТ на основі раніше розробленої елементарної моделі відрізку труби. Представимо МТ у вигляді послідовності відрізків труб (секцій), що мають площу поперечного перерізу $A_n \leq n \leq N$ та довжину Δl. Площа поперечного перерізу відрізків труб міняється відповідно до тієї функції площі, яку вони апроксимують. Хвилевий процес в такій ступінчастій трубі зручно представити у вигляді суперпозиції прямої і зворотної хвиль. У місцях зчленувань відрізків труб різної площі виникатимуть відбиття і заломлення звукових хвиль. Величини відбитої і заломленої складових звукової хвилі визначаються коефіцієнтами відбиття і заломлення. Ці коефіцієнти можуть бути розраховані по формулах:

$$\left.\begin{aligned}
\alpha^{+}_{пвідб} &= \frac{P^{+}_{пвідб}}{P^{+}_{n}} = \frac{Z_{n+1} - Z_n}{Z_{n+1} + Z_n} \\
\alpha^{+}_{пвідб} &= \frac{P^{+}_{пвідб}}{P^{+}_{n}} = \frac{2Z_{n+1}}{Z_{n+1} + Z_n} \\
\alpha^{-}_{пвідб} &= \frac{P^{-}_{пвідб}}{P^{-}_{n}} = \frac{Z_n - Z_{n+1}}{Z_{n+1} + Z_n} \\
\alpha^{-}_{пвідб} &= \frac{P^{-}_{nnp}}{P^{-}_{n}} = \frac{2Z_n}{Z_{n+1} + Z_n}
\end{aligned}\right\} \qquad (1.10)$$

де: P^{+}_{n} - звуковий тиск прямої хвилі в n –ому відрізку труби; P^{-}_{n} - звуковий тиск зворотньої хвилі в n–ому відрізку труби; $P^{+}_{пвідб}$, $P^{+}_{пвідб}$ – відбита і заломлена складові прямої хвилі; $P^{-}_{пвідб}$, $P^{-}_{пвідб}$ – відбита і заломлена складові зворотньої хвилі; Z_{n+1}, Z_n - характеристичний опір (n+1)-го і n-го відрізків труб.

Характеристичні опори відрізків труб дуже малої довжини визначаються тільки площею і формою їх поперечного перерізу. Зокрема, для випадку перерізу круглої форми:

$$Z_n = (\rho c)/A_n, \qquad (1.11),$$

где: ρ - густина повітря; c – швидкість звуку.

Поширення звукових хвиль, навіть в ідеальних трубах з жорсткими стінками, пов'язане з втратами. У МТ мають місце втрати на тертя, на теплопровідність, втрати внаслідок коливань стінок порожнин. В якості першого наближення введемо загасання $b(b \leq 1)$, яке враховує зменшення амплітуди звукової хвилі при її проходженні через елементарну секцію довжиною Δl.

Для правильної роботи моделі потрібно врахувати граничні умови. На початку МТ, в місці розташування голосових зв'язок має місце ситуація, близька до випадку повністю закритого кінця труби. Площа голосової щілини, навіть при повному її розкритті, не перевищує 0,2 см^2, в той час, як площа поперечного перерізу фарингальної порожнини на відстані не більше 0,5 см від голосових зв'язок складає близько 3 см^2. У зв'язку з цим коефіцієнт відбиття звукової хвилі від початку МТ $\alpha^{-}_{пвідб} \approx 0{,}9$.

Більш складна ситуація має місце у кінці МТ, у ротовому отворі. Тут потрібно врахувати характеристики випромінювання, які залежать від площі і форми ротового отвору, а також від геометричних розмірів і форми голови людини. Досить добре наближення до реальної ситуації дає теорія поршня, що коливається в сферичному екрані. На основі цієї теорії можливе обчислення кінцевої поправки, яка дає довжину додатковій секції МТ. Ця додаткова секція придає моделі додаткові коливальні властивості, що імітують вплив маси повітря у ротового отвору. Якщо відкритий кінець МТ має сферичний екран радіусом 9 см, то кінцева поправка обчислюється за формулою:

$$l_e = 0.8\sqrt{\frac{A_n}{\pi}}, \qquad (1.12)$$

де: A_n - площа ротового отвору (см^2).

Одна з основних умов адекватності імітаційної моделі МТ - знаходження такого Δl, яке забезпечує виконання співвідношення (1.3). При визначенні величини Δl необхідно враховувати наступні чинники:

а) довжину звукових хвиль, що поширюються в МТ;

б) зміну форми і властивостей МТ уздовж його довжини.

Спектр частот, генерований МТ, обмежений 20000 Гц.

Мінімальна довжина звукових хвиль в МТ складає: $\lambda_{min}=c/f_{max}$; $\lambda_{min}\approx 1{,}72$ см. Тоді згідно теоремі відкликів (теоремі Котельникова) довжина елементарної секції визначається співвідношенням:

$$\Delta l \leq \lambda_{min}/2, \qquad (1.13),$$

тобто $\Delta l \leq 0{,}86$ см.

Для точної апроксимації форми і властивостей МТ величина Δl повинна знаходитися в межах 0,5-1 см. Таким чином, вибір Δl відбувається згідно нерівності 0,5 см $\leq \Delta l \leq$ 0,86 см.

Для експериментальної перевірки моделі МТ був складений комплекс програм для ПК. Перевірка проводилася шляхом обчислення спектральних характеристик для акустичних систем різних конфігурацій. Спектральна характеристика акустичної системи є найбільш компактним засобом опису її властивостей.

Спектральний аналіз здійснювався на основі вдосконаленого алгоритму швидкого перетворення Фур'є (1.8) в межах смуги 17250 Гц з максимальною роздільною здатністю 67,5 Гц.

Спочатку розрахунки виконувалися для найбільш простого випадку прямої труби постійної площі. Результати розрахунків порівнювалися з аналітичними розв’язками, які в цьому випадку можливо отримати. Досліджувався вплив загасання на форму спектральних характеристик. Для врахування ефективності випромінювання вводилася кінцева поправка відповідно до формули (1.12).

Збільшення загасання не змінює максимум в спектрі, а впливає тільки на огинаючу, яка стає більш плавною. Максимуми спектру, які мають назву формантних частот, є резонансами системи, і залежать тільки від її форми і геометричних розмірів. У зв'язку з цим, формантні частоти використовуються в дослідженнях мови, як найбільш зручна і компактна характеристика спектру довільного мовленнєвого звуку.

Труба постояної площі має резонанси на частотах: $F_n = \frac{(2n-1)c}{4l}$, n =1,2,3, де: $l=l_t+l_e$ - довжина труби з врахуванням кінцевої поправки.

Далі були отримані спектральні характеристики для труб складніших конфігурацій. Отримані дані порівнювалися з результатами Г.Фанта, який проводив дослідження процесу мовотворення за допомогою аналога мовного тракту - моделі, складеної з RLC- ланок по методу електроакустичних аналогій. Параметри аналога мовного тракту підбиралися вручну, щоб як можна точніше апроксимувати певну функцію площі і властивості труби.

Таким чином, експериментальна перевірка показала працездатність імітаційної моделі МТ і можливість її використання для дослідження в області мовотворення. Крім того, побудована модель досить компактна і не потребує великих часових витрат для своєї реалізації.

1.2 Спектральний склад звуків мови

Як відомо, звуки мови, складаються з основного тону, що визначає їх висоту, і великої кількості (до 20) обертонів - звуків вищої, ніж основний тон, частоти. Кількістю цих обертонів, більшою силою одних і меншою силою інших, великою виразністю тих або інших обертонів, визначається не лише тембр мовних звуків, але і приналежність їх до тих або інших голосних або приголосних.

Ще в минулому столітті знаменитий фізик і лікар Герман Гельмгольц за допомогою своїх скляних резонаторів виявив, що для кожного голосного звуку характерні свої специфічні обертони, які він назвав характеристичними. Надаючи тому або іншому звуку обертон, характеристичний для певної голосної, за допомогою своїх резонаторів Гельмгольц ще в шестидесятих роках минулого століття намагався створити машину, здатну видавати звуки, схожі із звуками мови.

Детальному аналізу резонатора Гельмгольця методом часткових областей присвячені роботи [5-8]. Привабливість вказаного метода полягає в можливості розбиття досліджуваного акустичного об`єкту на канонічні за формою області

та використання відповідних розв`язків для складових полів. При цьому поєднання окремих елементів областей в одне ціле, що утворює та описує досліджувану конструкцію, вимагає залучення не тільки традиційних граничних умов типу умов Дирихле, Неймана, умов спряження кінематичного та силового типу для всіх обраних прямокутних (кубічних) областей, але й умов на кутових точках (ребрах), які утворені в перетинах переходу від однієї області до іншої [8].

Було встановлено, що виконання умов на ребрі є достатньо впливовим чинником формування акустичного поля в замкнених об`ємах типу резонаторів Гельмголця, а обрання типів граничних умов на елементах його порожнин дозволяє виконувати амплітудно – фазове корегування розподілень коливальних швидкостей, потенціалів і тисків.

В останні роки вивчення спектрального складу звуків мови ведеться на дуже високому науково-технічному рівні, що дало можливість збагатити наші знання з цього питання новими відомостями [2, 9-12]. Виявлено, що для кожного голосного звуку характерний не один або два характеристичні обертона, які тепер прийнято іменувати формантами, а три, чотири, і навіть п'ять: втім, найбільш важливими, надаючими голосній особливу визначеність звучання, являються перші 3 обертона. Такими формантами для різних голосних є:

а) для голосної "У" — 300 Гц, 625 Гц, 2500 Гц;

б) для голосної "О" — 535 Гц, 780 Гц, 2500 Гц;

в) для голосної "А" — 700 Гц, 1080 Гц, 2600 Гц;

г) для голосної "Е" — 440 Гц, 1800 Гц, 2550 Гц;

д) для голосної "І" — 240 Гц, 2250 Гц, 3200 Гц;

е) для голосної "И" — 300 Гц, 1480 Гц, 2230 Гц.

Одночасно встановлено, що у різних людей форманти одних і тих же голосних по своїй висоті, силі вираженості, і по ширині, займаній ними на спектрограмі, різняться між собою. У дитячих і жіночих голосах ці форманти виявляються вищими, ніж у чоловіків. Форманти голосних дещо різняться

залежно від положення голосної в слові, від ударної або ненаголошеності її, від висоти основного тону, на якій вимовляється голосна, і навіть від особливостей психічної настроєності людини, спектральний склад звуків якої досліджується.

За допомогою спектрометричних досліджень вдалося виявити дуже важливу відмінність, яка існує в спектрі розмовних і в спектрі вокальних звуків. Виявляється, що в спектрі розмовних голосних максимум спектральної енергії припадає на низькі звуки в діапазоні 300 Гц – 400 Гц.

1.3 Формантний аналіз мови

Форматний аналіз мови можна розглядати як особливий вид спектрального аналізу [9-12]. Завдання полягає у визначенні комплектних частот формант мовоутворюючого механізму в процесі їх зміни в часі. Ці зміни, очевидно, обумовлені перебудовою органів артикуляцій голосового тракту. Один з можливих підходів до такого роду аналізу полягає в з'ясуванні того, в якому виді проявляються в миттєвому спектрі сигналу резонансні максимуми системи мовного тракту.

Втрати в голосовому тракті відносно сталі і можуть бути розраховані заздалегідь. Зважаючи на це, зазвичай інтерес представляють часові зміни не дійсних, а уявних частин комплексних формантних частот. Проте для оцінок процесів сприйняття і для спектрального аналізу важливі відповідні відомості і про дійсні частини або смуги частот формантних областей. Підхід до аналізу мови з позицій розгляду «функціонування системи» дозволяє представити мовний сигнал за допомогою передавальної функції тракту і функції збудження. Якщо конфігурація голосового тракту відома, то можна розрахувати спектральні максимуми, які їй відповідають, і отримати вихідний відклик на задане збудження. У разі автоматичного аналізу, що проводиться в цілях кодування і передачі мови, зазвичай має місце зворотна ситуація. У розпорядженні є тільки акустичний сигнал, за результатами аналізу якого вимагається виявити властивості джерела і резонанси мовоутворюючої системи. Основна проблема при цьому полягає в незнанні того, як однозначно розділити властивості джерела і системи.

Резонансні максимуми мовоутворюючої системи з плином часу постійно переміщаються, але вони не завжди можуть бути виявлені, наприклад, по миттєвому спектру сигналу. Окремі полюси можуть на короткі проміжки часу виявитися прихованими, а також пригніченими нулями джерела або нулями системи, що утворюються із-за елементів бічних відгалужень (таких, як носова порожнина). У миттєвому спектрі досить добре проявляються високі резонансні максимуми, але іноді все ж важко з упевненістю визначити місце розташування полюсів з малою амплітудою або суттєвих пар полюсів - нулів.

Подальше ускладнення ситуації полягає в тому, щоб сигнал на виході мовного тракту, як правило, не представляє собою мінімально фазову функцію (тобто не всі його нулі можуть знаходитися в лівій напівплощині комплексних частот). Якщо функція була б мінімально фазовою, то її фазовий спектр повністю визначався б її амплітудним спектром. Передатна функція мовного тракту буде мінімально фазовою в тих випадках, коли випромінювання відбувається тільки з одної точки, тобто з рота або з ніздрей. Ця умова не виконується при одночасному випромінюванні з обох точок. Можна показати, що якщо в деякі моменти часу за період коливання голосових зв'язок величина потоку, який створюється джерелом, дорівнює нулю, то останній не має полюсів і характеризується тільки нулями з кінцевими значеннями своїх частот. Крім того, можна показати, що ці нулі можуть знаходитися як в правій, так і лівій напівплощинах або одночасно в обох. Внаслідок вказаних труднощей численні зусилля, направлені на створення точно працюючого автоматичниго формантного аналізатора, не досягли цілей.

У першому наближенні вокалізоване збудження резонансної області мовного тракту можна звести до збудження короткими періодичними імпульсами одиночного резонансного контура. Вихідний сигнал в цьому випадку набуває форму затухаючих гармонічних коливань, які слідують з частотою повторення імпульсів. На частоті, яка дорівнює уявній складовій комплексної частоти полюса, обвідна амплітудного спектру має максимум. Цю формантну частоту можна знайти вимірюванням або частоти переходів через

нуль часового сигналу, або частоти максимуму обвідної спектру. Якщо ширина резонансної області порівняно невелика, то уявну складову частоти полюса можна оцінити по моменту першого порядку амплітудного спектру: $\bar{f} = \frac{\int fA(f)df}{\int A(f)df}$.

Мовний тракт є багаторезонансною системою. Тому часовий сигнал на його виході є результатом накладання великої кількості затухаючих гармонійних коливань, а спектр амплітуд характеризується безліччю максимумів. Результати вимірювання частоти переходів через нуль, максимуми спектру і моментів дозволяють визначити частоту форманти тільки у тому випадку, якщо кожна резонансна область може бути досліджена в ізольованому виді. Якщо треба виявити тонші властивості джерела і системи (скажімо, нулі, які створюються голосовими зв'язками або резонатором бічного відгалуження в спектрі), необхідно використовувати відповідно складніші способи вимірювання спектральних максимумів. Один з таких способів полягає в точній підгонці гіпотетичної моделі спектру до реального мовного спектру, для чого вигідно використовувати ПК.

Згідно акустичної теорії мовотворення сприйняття звуків визначається областями максимальної концентрації енергії - формантами *F*. У сучасних дослідженнях мови широко застосовуються різні автоматизовані електроакустичні вимірювальні прилади, проте отримані результати і їх точність в деякій мірі ще залежать від майстерності експериментатора.

Форманта характеризується (рис.1.1) амплітудою A_i і частотою F_i, для голосних добротністю Q_i або шириною смуги пропускання B_i, а для глухих приголосних - моментами перших трьох порядків M_i. Для опису формант дзвінких приголосних можуть знадобитися усі перераховані параметри.

Часто область максимальної концентрації енергії може бути представлена не однією максимальною складовою спектру, а декількома, причому положення форманти по частоті може не співпасти з ними. Такий же результат отримують

при асиметричній обвідній форманти. Крім того, перша, друга і третя форманти мають значні області перекриття.

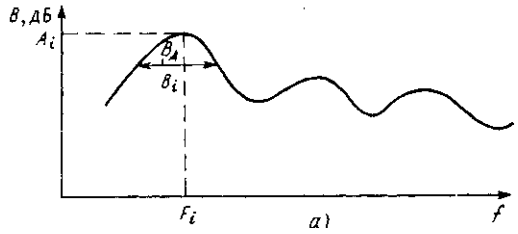


Рис.1.1 – Характеристики форманти

Наведемо деякі результати вимірювання частотного розподілу формант:

- форманти дзвінких звуків займають смуги частот 150-900 Гц (F_1), 550-2800 Гц (F_2), і 1500-3400 Гц (F_3);

- форманти глухих звуків 1000-3500 Гц (F_1), 2500-6000 Гц (F_2), і 1500-4000 Гц (антиформанта).

З йовірністю 0,98 частотний діапазон F_1 дорівнює 200-850 Гц, F_2 – 850-2550 Гц, і F_3 – 2100-3300 Гц.

Формування дзвінких звуків мови в голосовому апараті людини визначається коливаннями голосових зв'язок, які задають основний тон мови. Для опису властивостей основного тону використовується ряд параметрів: частота, частотний діапазон його можливих значень, мелодія, яка являє собою усереднені за деякий інтервал значення частоти основного тону, та ін. На результати статистичних вимірювань параметрів основного тону помітний вплив чинить прийнята методика: об'єм і характер мовної вибірки, представництво голосів дикторів і техніка вимірювань. Оскільки перші два чинники досить очевидні, зупинимося детальніше на питаннях техніки вимірювань.

Статистичні вимірювання параметрів основного тону можна виконати по осцилограмах мовного сигналу на виході високоякісного мікрофону і по осцилограмах коливань голосових зв'язок, отриманих при промовлянні диктором окремих звукосполучень і фраз.

При проведенні статистичних вимірювань необхідно враховувати деякі особливості основного тону. Інтервали дотримання імпульсів голосових зв'язок безперервно міняються і притому в значних межах.

Вимірювання частоти форманти по частоті переходів через нуль. Одна з перших спроб автоматичного стеження за частотами формант полягала в підрахунку середньої кількості переходів через нуль. В якості частот першої і другої формант приймалися величини, пропорційні середній щільності переходів через нуль мовного сигналу і його похідної в часі відповідно. Спосіб грунтувався на тому факті, що на вокалізованих ділянках нефільтрованої мови перша форманта є найбільш чітко виступаючою спектральною складовою. Тому очікувалося, що вона чинитиме найбільший вплив на величину частоти переходів через нуль. З іншого боку, в диференційованому сигналі перша форманта виявляється пригніченою і домінуючу роль набуває друга форманта. Проте отримані результати експериментів виявилися недостатньо хорошими, і був зроблений висновок, що такий спосіб не забезпечує прийнятної точності вимірювань.

Були запропоновані удосконалення методу, заснованого на вимірюванні частоти переходів через нуль. Одне з них полягало в тому, що мовний сигнал заздалегідь розфільтровувався на ряд частотних діапазонів, що відповідають окремим формантам. У кожному з цих діапазонів вимірювалися частота переходів сигналу через нуль і амплітуда. Проте і тут мали місце помилки, що виникають внаслідок взаємного перекриття діапазонів формантних частот. Та ж ідея була реалізована більш ретельно з деякими особливостями, що дозволили звести до мінімуму небажані взаємні перекриття. Принцип дії пристрою полягав в застосуванні повторного вимірювання середньої швидкості переходів сигналу через нуль в цьому частотному діапазоні і в істотному звуженні цього частотного діапазону на основі вимірювання частоти до її істинного значення. За допомогою фільтрів з фіксованими смугами пропускання сигнал заздалегідь розбивається на два діапазони, що приблизно відповідають першим двом формантам. У низькочастотному діапазоні вимірюється частота переходів

через нуль, і її величина використовується для автопідстроювання більше вузькосмугового фільтру з переналаштованою смугою пропускання. Частота переходів через нуль, виміряна на виході вузькосмугового фільтру, приймається, у свою чергу, за частоту першої форманти (F_1). Її величина використовується для встановлення частоти зрізу фільтру верхніх частот зі змінною смугою пропускання. Середнє значення частоти переходів через нуль на виході останнього береться в якості оцінки частоти другої форманти ($F2$).

Якщо спектральний розподіл сигналу має безперервний характер, як у разі невокалізованих звуків, середня частота переходів через нуль для цього спектрального елементу дорівнює приблизно подвоєному моменту першого порядку для цієї ділянки спектру.

1.4 Методи оцінки разбірливості мови

Відомий ряд показників і методів оцінки розбірливості мови, які вживаються на практиці для оцінки аудиторій, кіно-театральних залів, тощо, ліній зв'язку та їх захищеності [13]. Одні методи спочатку «пристосовані» до умов шумової перешкоди - до таких можна віднести метод парціальних співвідношень сигнал-шум і формантный метод. Інші методи і показники орієнтовані на умови переважаючого впливу реверберaційної перешкоди - серед таких можна назвати метод %Alcons і коефіцієнти чіткості C50 і C80 [14-16]. Значний інтерес викликають «універсальні» методи, що дозволяють одночасно враховувати вплив як шумових, так і реверберaційних, перешкод - це група модуляційних методів STI, RASTI, STITEL і STIPA [17, 18].

Розглянемо два об'єктивних методи оцінки разбірливості мови: формантний метод і модуляційний метод.

1.4.1 Формантний метод

Формантний метод в ролі показника якості використовує скалярну величину - формантну розбірливість, що є середньою ймовірністю сприйняття формант на фоні шуму :

$$A = \sum_{k=1}^{K} p_k \cdot w_k(E_k), \tag{1.14}$$

де p_k - ймовірність перебування формант в k-тій смузі частот з граничними частотами $f_{нk}$ и $f_{вk}$, $k = 1,\ldots,K$:

$$p_k = F_1(f_{вk}) - F_1(f_{нk}), \tag{1.15}$$

$F_1(f)$ - функція розподілу ймовірностей формант по частоті; $w_k(E_k)$ - коефіцієнт сприйняття (ймовірність сприйняття формант на фоні шуму в k-тій смузі частот); E_k - рівень сприйняття (відношення сигнал-шум) в k-тій частотній смузі.

Порівнюючи формантний метод з методом парціальних співвідношень сигнал-шум, можна побачити аналогію з задачею синтезу оптимального виявника сигналів на фоні завад [19], де замість розподілу ймовірностей:

$$P(a_i, b_j), \quad i, j = 0, 1, \tag{1.16}$$

який характеризує якість розв'язку задачі виявлення сигналу на фоні завад (a_0 - сигнал відсутній; a_1 - сигнал присутній; b_0 - розв'язок про відсутність сигналу; b_1 - розв'язок про наявність сигналу), використовують показник, що має назву "середнього ризику":

$$R_c = \sum_{i=0}^{1}\sum_{j=0}^{1} R_{ij} P(a_i, b_j), \tag{1.17}$$

де R_{ij} - величина (ризик), яка характеризує втрату від неправильного розв'язку [19].

Наведемо ще один приклад: задачу оптимального кодування повідомлення, яке представляється елементами s_i [20]. Таку задачу розв'язують використовуючи прказник у вигляді середньої довжини коду:

$$H = -\sum_i p(s_i) \cdot \log_2 p(s_i), \tag{1.18}$$

де $p(s_i)$ - ймовірність появи елементу s_i.

Зіставляючи співвідношення (1.14), (1.15) та (1.16), неважко побачити принципову подібність підходів до ефективного розв'язку різних задач, а саме: показником якості виступає середнє значення деякої скалярної величини.

Важливою особливістю форматного методу є те, що показником якості є ймовірність. Використовуючи інформацію про статистичні властивості мови,

середню ймовірність сприйняття формант A потім можна перерахувати у ймовірність правильного розпізнавання звуків, складів, слів і навіть фраз. Це надзвичайно зручно, оскільки при використанні суб'єктивних методів оцінки розбірливості мови показником якості також служить ймовірність правильного розпізнавання звукосполучень, складів, слів або фраз. Отже, для зіставлення якості розв’язку однієї і тієї ж задачі об'єктивним і суб'єктивним методами досить порівняти дві оцінки ймовірністі.

Наприкінці 50-х років минулого століття можна було виділити 3 підходи до оцінки розбірливості мови, за Ю.С. Биковим, М.А. Сапожковим, та Н.Б. Покровським. Ці підходи реалізовували формантний метод [21-23].

Відмінність версій полягала, по-перше, у відмінності трактувань поняття ‘формантний спектр”. По-друге, відрізнялася форма коефіцієнтів сприйняття. Нарешті, не співпадали думки авторів версій і з приводу форми функції розподілу вірогідності формант по частоті $F_1(f)$.

Ю.К. Калинцев в роботі [24] співставив існуючі коефіцієнти сприйняття на одному графіку. При цьому були допущені помилки, які були виправлені в роботах [25-27]. На рис.1.2 наведено результати такого правильного співставлення коефіцієнтів сприйняття Н.Б. Покровського $P(E'_k)$, М.А. Сапожкова $w(E)$, Ю.С. Бикова $P_i(\Delta\lambda_i)$, а також їх аналога $P_{AI}(\Delta L)$ - коефіцієнта сприйняття американської версії формантного метода, який має назву індексу артикуляції (Articulation Index - AI).

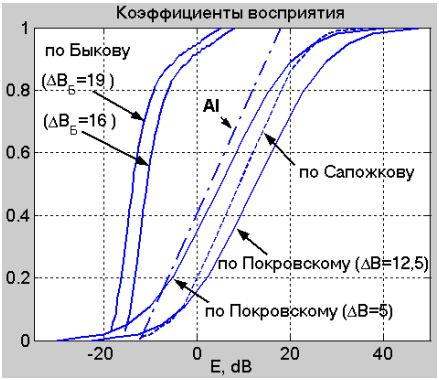


Рис.1.2 – Коефіцієнти сприйняття $P_{AI}(\Delta L)$, $w(E)$, $P_i(\Delta\lambda_i)$ и $P(E'_k)$ [27]

Зіставлення коефіцієнтів сприйняття на одному графіку свідчить про велику проблему трьох вище зазначених версій формантного методу, а саме: неоднозначність визначення коефіцієнта сприйняття для російської мови.

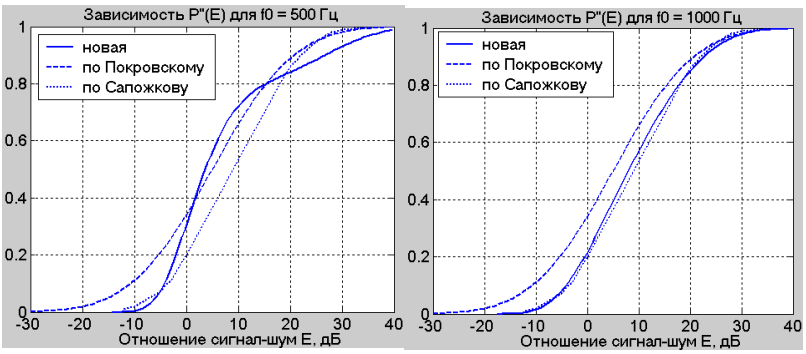
Аналіз причин неоднозначності свідчить, що виною тому відмінність методик формування коефіцієнта сприйняття. У роботах [13, 25-17] показано, що найбільш коректна методика М.А. Сапожкова, згідно якої коефіцієнт сприйняття $w(x)$ визначається співвідношенням:

$$w(x) = 1 - F(-x), \tag{1.19}$$

де $F(x)$ - функція розподілу рівнів мовного сигналу. У роботі [13] представлені результати оцінювання коефіцієнтів сприйняття російської мови з використанням співвідношення (1.19), а саме: сімейство нових коефіцієнтів сприйняття для 5 октавних смуг частот, з центральними частотами 250 Гц, 500 Гц, 1000 Гц, 2000 Гц, 4000 Гц. Деякі коефіцієнти цього сімейства показані на рис.1.3, де також для порівняння приведені коефіцієнти сприйняття по М.А. Сапожкову і Н.Б. Покровскому.

Слід звернути увагу, що для смуги частот з центральною частотою 1000 Гц новий коефіцієнт сприйняття практично співпав з коефіцієнтом сприйняття по М.А. Сапожкову (рис.1.3,б). Це можна трактувати як ознаку достовірності нових коефіцієнтів сприйняття. У інших смугах частот (рис.1.3,а) нові коефіцієнти сприйняття, що відрізняються від коефіцієнта сприйняття по М.А. Сапожкову, слід розглядати як уточнення результатів, отриманих М.А. Сапожковим.

Зазначимо, що сучасні російські дослідники [28-30] використовують коефіцієнт сприйняття, запропонований Н.Б. Покровским, що є некоректним, оскільки методика формування коефіцієнта сприйняття Н.Б. Покровского містить ряд недостатньо обгрунтованих припущень і навіть помилок [13, 25-27].



(а) (б)

Рис.1.3 - Зспівставлення коефіцієнтів сприйняття: (а) – по М.А. Сапожкову, (б) – по Н.Б. Покровському

Відсутність в літературі відомостей про формантні властивості української мови означала, по суті, неможливість отримання коректних оцінок розбірливості української мови. Цей пропуск був заповнений в роботах [31, 32], де показано, що функції розподілу формант по рівню для української і російської мови при великих рівнях мовного сигналу (вище за середній рівня мови) практично однакові. При малих рівнях мовного сигналу функції розподілу української мови майже скрізь перевищують такі для російської мови (рис.1.4).

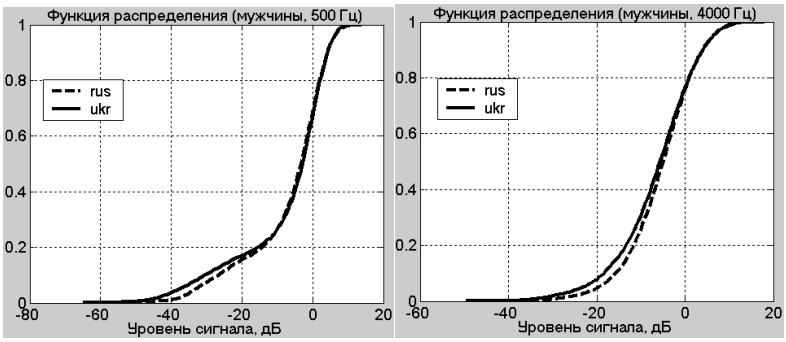


Рис.1.4 - Коефіцієнти сприйняття української та ромійської мови

Даний факт, з урахуванням співвідношення (1.19), дозволяє дійти висновку, що при малих відношеннях сигнал-шум розбірливість російської і української мови практично однакова, а при великих відношеннях сигнал-шум розбірливість української мови може бути дещо нижчою.

М.А. Сапожков відмічав практично лінійний характер коефіцієнта сприйняття в межах деякого інтервалу значень відношення сигнал-шум [10]. Ідея лінеаризації коефіцієнта сприйняття доведена до логічного завершення в сучасній американській версії формантного методу, іменованій «індексом артикуляції» (AI - articulation index), де замість «гладкого» коефіцієнта сприйняття використовують його кусочно-лінійну апроксимацію [13, 28].

При цьому, позитивний ефект полягає у тому, що вдається уникнути: 1) неоднозначності результатів, викликаної відмінністю методик оцінювання коефіцієнта сприйняття (рис.1.2); 2) погрішностей результатів, обумовлених відмінністю умов вимірювання функції $F(x)$. А негативний ефект полягає в тому, що тепер мірою якості є не ймовірність, а індекс. Як наслідок, стає важче зіставляти результати об'єктивних і суб'єктивних оцінок. Крім того, через злами кусочно-лінійної характеристики можуть виникнути проблеми з трактуванням результатів вимірювань при малих відношеннях сигнал-шум, типових для завдання захисту акустичних каналів.

При оцінюванні розбірливості мови, відповідно до співвідношення (1.14), передбачається, що мова сприймається на фоні шумової завади. Проте, вплив реверберaційної перешкоди можна врахувати. Поскільки реверберaційна і шумова завади статистично незалежні, загальна потужність завад дорівнює сумі їх потужностей, і результуюче відношення сигнал-шум дорівнює:

$$SNR = 10\lg(10^{0,1\cdot SNR_N} + 10^{0,1\cdot SNR_R}), \qquad (1.20)$$

де SNR_N, і SNR_R - відношення сигнал-шум для шумової і реверберaційної завад відповідно.

Таким чином, при спільному впливі шумові і реверберaційної завад у співвідношенні (1.14) в якості E_k слід використовувати відношення сигнал-шум, яке визначається виразом (1.20).

Значення SNR_R можна розрахувати по відомому часу реверберації T_{60}:

$$SNR_R = 10\lg(e^{\frac{0,69}{T_{60}}} - 1). \qquad (1.21)$$

Співвідношення (1.21) отримано в наближенні $SNR_R = C_{50}$, де C_{50} - індекс чіткості, який характеризує відношення енергії прямого та відбитого звуку [30]: $C_{50} = 10\lg\frac{\int_0^{0,05} p^2(t)dt}{\int_{0,05}^{\infty} p^2(t)dt}$, де $p^2(t) = p^2 \cdot e^{-\alpha t}$ - квадрат імпульсної характеристики приміщення; $\alpha = \frac{13,8}{T_{60}}$; 0,05 – момент часу в секундах. Слід зауважити, що в роботі [16] вказано інше значення $t = 0,095c$, тобто в два рази більше. В цьому випадку $SNR_R = 10\lg(e^{\frac{1,31}{T_{60}}} - 1)$. Наприклад, при $T_{60} = 1\,c$ в першому випадку отримуємо $SNR_R \approx 0\ дБ$, а у другому $SNR_R \approx 4,3\ дБ$.

1.4.2 Модуляційний метод

Модуляційний метод [17, 18] базується на ідеї використання тестового шумового сигналу, потужність якого змінюється за гармонійним законом. Показником якості при цьому служить глибина модуляції потужності сигналу в точці спостереження, що дозволяє врахувати вплив шумової і реверберaційної завади.

У випадку шумової завади, залежність глибини модуляції m_N від відношення сигнал-шум SNR, в децибелах, має вигляд:

$$m_N = \left(1 + 10^{-0,1 \cdot SNR}\right)^{-1}. \tag{1.22}$$

У випадку реверберaційної завади глибина модуляції залежить від частоти модуляції F і визначається співвідношенням:

$$m_R(F) = \frac{1}{\sqrt{1 + \left(\frac{2\pi FT}{13,8}\right)^2}}, \tag{1.23}$$

де T - час реверберації.

Об’єднавши співвідношення (1.22) і (1.23), отримуємо остаточний вираз для глибини модуляції (або функції передачі модуляції):

$$m(F) = \frac{1}{\sqrt{1 + \left(\frac{2\pi FT}{13,8}\right)^2}} \cdot \frac{1}{1 + 10^{-0,1 \cdot SNR}} = m_R(F) \cdot m_N(F). \tag{1.24}$$

Співвідношення (1.24) є теоретичною основою модуляційного методу. При практичних вимірюваннях множина значень $m(F)$ усереднюється по частотах F та по смузі частот k, після чого, множина значень $m(F)$ перетворюється в індекс передачі мови STI (Speech Transmission Index).

Розглянемо повну версію модуляційного методу, в якій перетворення множини значень $m(F)$ в індекс STI виконується в наступній послідовності:

1. Оцінюється 98 значень коефіцієнту $m_k(F_i)$, які отримуюються для 14 значень частот модуляції F_i з діапазону значень 0,63-12,5 Гц в кожній з 7 октавних смуг з центральними частотами від 125 Гц до 8 кГц (k - номер смуги частот).

2. Кожна з 98 оцінок коефіцієнту $m_k(F_i)$ переводиться в ефективне відношення сигнал-шум: $(S/N)_{еф\,k,i} = 10\log\frac{m_k(F_i)}{1-m_k(F_i)}$. Якщо $(S/N)_{dB\,еф\,k,i} < -15\,дБ$, приймають $(S/N)_{dB\,еф\,k,i} = -15\,дБ$, а якщо $(S/N)_{dB\,еф\,k,i} > +15\,дБ$, приймають $(S/N)_{dB\,еф\,k,i} = +15\,дБ$.

3. Виконується усереднення оцінок ефективних відношень сигнал-шум для кожної з октавних смуг: $(S/N)_{еф\,k} = \frac{1}{14}\sum_{i=1}^{14}(S/N)_{еф\,k,i}$.

4. Розраховується середнє взважене значення: $\left(\overline{S/N}\right)_{эф} = \frac{1}{14}\sum_{k=1}^{7}(S/N)_{эф\,k}\cdot w_k$, де w_k - вагові коефіцієнти ($\sum_k w_k = 1$).

5. Розраховується індекс STI, з використанням формули (1.25) або рис.1.5:

$$STI = \frac{\left(\overline{S/N}\right)_{еф} + 15}{30}. \qquad (1.25)$$

Можна побачити суттєву подібність кусочно-лінійних характеристик, які використовуються в методах AI и STI (рис.1.2 і рис.1.5).

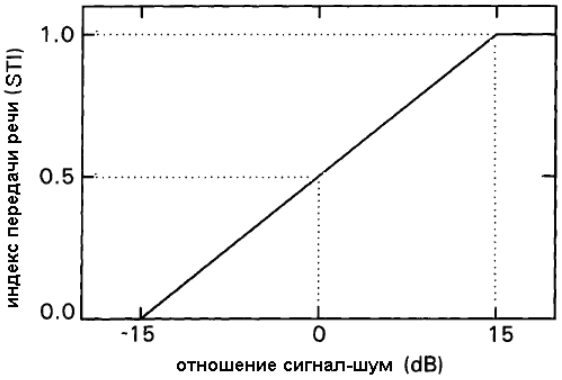


Рис.1.5 - Залежність між STI і ефективним відношенням сигнал-шум [29]

1.5 Висновки до Розділу 1

1. Виконано огляд моделей мовного тракту людини. Для побудови моделі мовного тракту вибрано наступну методику: складна і неоднорідна акустична система розбивається на частини, які є досить простими за формою і неоднорідні за властивостями; для кожної з таких частин складається "елементарна" модель; враховуючи фізичний сенс процесів, що відбуваються в системі, визначаються умови сполучення "елементарних" моделей між собою, і будується модель усієї системи в цілому.

2. Розглянуто спектральний склад звуків розмовної та мови. Встановлено, що у різних людей форманти одних і тих же голосних по своїй висоті, силі вираженості, і по ширині, займаній ними на спектрограмі, різняться між собою. У дитячих і жіночих голосах ці форманти виявляються вищими, ніж у чоловіків. Форманти голосних дещо різняться залежно від положення голосної в слові, від ударної або ненаголошеності її, від висоти основного тону, на якій вимовляється голосна, і навіть від особливостей психічної настроєності людини, спектральний склад звуків якої досліджується.

3. Форматний аналіз мови розглянуто як особливий вид спектрального аналізу. Перевагою модуляційного методу над формантним є легкість врахування впливу реверберaційної завади. Експериментальні дослідження свідчать, що по точності експертизи приміщень, де превалює реверберaційна завада, модуляційній метод переважає як коефіцієнт чіткості C50, так і метод %Alcons [31].

РОЗДІЛ 2
РОЗРАХУНОК РОЗБІРЛИВОСТІ МОВИ, СПОТВОРЕНОЇ ШУМОМ ТА РЕВЕРБЕРАЦІЄЮ

Завдання розрахунку і вимірювання розбірливості мови не є новим, її вік сьогодні складає 93 років, якщо відлік вести від піонерської роботи [33]. Проте, сфера застосувань оцінювання розбірливості мови постійно розширюється, технічні засоби інженерів змінюються і удосконалюються, зростає перелік чинників, що враховуються при оцінюванні розбірливості мови. Як наслідок, виникає необхідність в постійному оновленні відповідного математичного і програмного забезпечення [34].

За кордоном найбільше поширення отримали дві версії формантного методу: індекс артикуляції AI (Articulation Index) [35], та індекс розбірливості мови SII (Speech Intelligibility Index) [36]. У СРСР до кінця 50-х років сформувалося декілька наукових шкіл, як вже зазначалося вище, очолюваних Н.Б. Покровским, М. А. Сапожковим і Ю.С. Биковим, де розвивалися свої версії формантного методу [21-23].

У 1973 році з'явився модуляційний метод, де мірою розбірливості мови служить індекс передачі мови STI (Speech Transmission Index) [18]. Оскільки модуляційний метод має здатність враховувати вплив не лише шуму, але і реверберації на розбірливість мови, в деяких роботах навіть з'явилися висловлювання про «застарілість» формантного методу [37]. В той же час, ретельне зіставлення потенційних можливостей цих двох методів розбірливості мови показує переваги формантного методу над модуляційним по точності і швидкості обчислень, у випадку, коли дія шуму переважає над дією реверберації [38].

Алгоритм комп'ютерного моделювання оцінювання розбірливості зашумленной мови формантным методом описаний в [21], [36]. Структура цього алгоритму представлена на рис.2.1. Перший етап полягає у формування первинних моделей мовного сигналу і шуму у вигляді масивів вибірок стаціонарних випадкових процесів із заданими спектральними

характеристиками. Наступним кроком коректується дисперсія даних модельних процесів, для забезпечення необхідного інтегрального відношення сигнал-шум SNR_0. Після такої корекції оцінюються парціальні стосунки сигнал-шум Ek. На завершальному етапі обчислюються заходи розбірливості мови : формантная розбірливість A і словесна розбірливість W.

Рис.2.1 - Структура алгоритму комп'ютерного моделювання

Формантний метод оцінювання розбірливості мови заключається в тому, що область частот мовленнєвого сигналу розбивається на K суміжних смуг частот, із центральними частотами f_{0k} і граничними частотами $f_{нk}$ і $f_{вk}$, в межах кожної смуги спектри мови і шуму вважаються незмінними [21]. Словесну розбірливість W розраховують через формантную розбірливість A [29]:

$$W = \begin{cases} 1.54 \cdot A^{0.25}[1 - exp(-11 \cdot A)], & A < 0.15; \\ 1 - exp\left(\dfrac{11 \cdot A}{1 + 0.7 \cdot A}\right), & A \geq 0.15 \end{cases} . \tag{2.1}$$

$$A = \sum_{k=1}^{K} p_k \cdot P(E'_k), \tag{2.2}$$

де p_k - ймовірність перебування формант в k-тій смузі частот:

$$p_k = F(f_{вk}) - F(f_{нk}), \tag{2.3}$$

$$F(f) = \begin{cases} 2.57 \cdot 10^{-8} \cdot f^{2.4}, & 100 < f \leq 400\,Гц, \\ 1 - 1.074 \cdot exp(-10^{-4} \cdot f^{1.18}), & 400 < f \leq 10000\,Гц \end{cases}, \tag{2.4}$$

$P(E'_k)$ - коефіцієнт сприйняття мови.

У відповідності з методикою Н.Б. Покровского [21]:

$$P(E'_k) = \begin{cases} \dfrac{0.78 + 5.46\,exp[-4.3 \cdot 10^{-3} \cdot (27.3 - |E'_k|)^2]}{1 + 10^{0.1 \cdot |E'_k|}}, & E'_k \leq 0, \\ 1 - \dfrac{0.78 + 5.46\,exp[-4.3 \cdot 10^{-3} \cdot (27.3 - |E'_k|)^2]}{1 + 10^{0.1 \cdot |E'_k|}}, & E'_k \leq 0 \end{cases}, \tag{2.5}$$

E'_k - ефективний рівень відчуття формант в k-тій смузі частот:

$$E'_k = E_k - \Delta B(f_{0k}), \tag{2.6}$$

E_k - ефективний рівень відчуття мовленнєвого сигналу в k частотній смузі, який дорівнює (при достатньо великих рівнях шуму) відношенню сигнал-шум q_k в ції смузі частот:

$$E_k = q_k = 10\,lg\frac{D_{sk}}{D_{nk}}, \tag{2.7}$$

де D_{sk} і D_{nk} - дисперсії сигналу і шуму в k-тій смузі частот; $\Delta B(f)$ - різниця між усередненими спектрами мови і формант:

$$\Delta B(f) = \begin{cases} 200/f^{0.43} - 0.37, & f \le 1000\,Гц, \\ 1.37 + 1000/f^{0.69}, & f > 1000\,Гц \end{cases}. \tag{2.8}$$

Згідно методиці М.А. Сапожкова, спектр формант можна вважати таким, що практично співпадає із спектром мови, тобто $\Delta B(f)$=0, звідки слідує $E'_k = E_k$ [22]. Крім того, як показано в [26], залежність $P(E'_k)$ М.А. Сапожкова відрізняється від такової у Н.Б. Покровского.

В роботі [39] оцінка за Сапожковим доповнена тим, що враховуються залежності коефіцієнтів сприйняття від частотної смуги. Тоді формула (2.1) замінюється наступною:

$$A = \sum_{k=1}^{K} p_k \cdot P_k(E_k), \tag{2.9}$$

де коефіцієнти сприйняття $P_k(E_k)$ можна описати поліноміальними залежностями.

Коефіцієнти сприйняття описуються виразами:

$$P'_i(x) = \begin{cases} 0, \quad x \le x'_{\min}; \\ \sum_{n=0}^{N'} a'_n x^n, \quad x'_{\min} < x \le 0 \\ \sum_{m=0}^{M'} b'_m x^m, \quad 0 < x \le x'_{\max}; \\ 1, \quad x > x'_{\max}. \end{cases} \tag{2.10}$$

де значення коефіцієнтів a'_n и b'_m наведені в табл.2.1, 2.2. Дані коефіцієнти сприйняття відрізняються від коефіцієнтів, які наводяться в роботі [39], наявністю семи октавних смух частот, замість п’яти.

Таблиця 2.1 - Коефіцієнти a'_n [54]

n $f_{серед}$, Гц	0	1	2	3	4	5
125	0.324508	0.086098	0.004887	-0.00053	-6.56E-05	-1.87E-06
250	0.358141	0.070497	-0.00767	-0.00311	-0.00029	-8.76E-06
500	0.309301	0.054827	-0.00207	-0.00104	-7.77E-05	-1.82E-06
1000	0.19779	0.028848	0.000603	-0.00011	-7.30E-06	-1.45E-07
2000	0.238407	0.042094	0.000906	-0.00028	-2.29E-05	-5.43E-07
4000	0.231306	0.040168	0.001327	-0.00014	-1.07E-05	-2.07E-07
8000	0.178871	0.01705	-0.00064	-6.50E-05	2.91E-06	1.91E-07

Таблиця 2.2 – Коефіцієнти b'_m [54]

m $f_{сер}$, Гц	0	1	2	3	4	5
125	0.315151	0.057481	-0.00116	-4.45E-05	2.10E-06	-2.22E-08
250	0.35763	0.088259	-0.00646	0.000232	-3.89E-06	2.46E-08
500	0.301964	0.07124	-0.00382	9.63E-05	-8.28E-07	-9.67E-10
1000	0.195479	0.032135	0.000349	-2.04E-05	-1.22E-07	5.59E-09
2000	0.238336	0.042175	0.002185	-0.00025	7.53E-06	-7.75E-08
4000	0.233574	0.037324	0.003033	-0.0003	8.87E-06	-9.05E-08
8000	0.185163	0.01171	0.002126	-2.81E-05	-2.24E-06	4.55E-08

В теперішній час існує тенденція до часткової уніфікації формантного і модуляційного [18] методів оцінювання розбірливості мови. Так, наприклад, згідно із спрощеною методикою оцінювання розбірливості мови, представленої в [42], розбірливість мови оцінюють за допомогою індексу STI:

$$STI = \sum_{k=1}^{7} \alpha_k \cdot T_k - \sum_{k=1}^{6} \beta_k \cdot \sqrt{T_k \cdot T_{k+1}}, \tag{2.11}$$

$$T_k = \begin{cases} 0, \ E_k < -15; \\ (E_k + 15)/30, -15 < E_k < +15; \\ 1, \ E_k > +15; \end{cases}, \tag{2.12}$$

де α_k – вагові коефіцієнти, β_k – коефіцієнти надлишковості, значення яких для октавниxx смуг з центральними частотами f_0 приведені в табл.2.3.

Таблиця 2.3 – Вагові коефіцієнти і коефіцієнти надлишковості для октавних смуг [42]

f_{0k}, Гц	125	250	500	1000	2000	4000	8000
α_k	0.085	0.127	0.230	0.233	0.309	0.224	0.173
β_k	0.085	0.078	0.065	0.011	0.047	0.095	-

Можна зробити висновок, що співвідношення (2.2), (2.9), (2.11), схожі. При цьому, друга складова в (2.11) є поправкою, яка враховує корелюваність мовних сигналів в суміжних смугах частот, а співвідношення (2.12) можна трактувати як результат лінеаризації коефіцієнта сприйняття $P(E'_k)$.

Незважаючи на здатність комп'ютерного моделювання оцінювати працездатність і ефективність прототипів програмного забезпечення реальних цифрових вимірювальних систем, не менш актуальним завданням є прогнозне оцінювання розбірливості мови. Проте вирішувати задачу прогнозування методом комп'ютерного моделювання навряд чи раціонально, враховуючи ресурсоємкість цього методу. Економнішим є метод аналітичного моделювання, згідно з яким моделі мовного сигналу і шуму описують детермінованими функціями у вигляді розподілів потужностей або спектральної щільності потужності.

По суті, етапи аналітичного і комп'ютерного моделювання подібні. Більше того, співвідношення (2.1)-(2.12), які використовуються на фінальному етапі, однакові для обох видів моделювання. Проте питання аналітичного опису початкового і проміжних етапів в літературі освітлене недостатньо добре. Одна з цілей цього розділу полягає в усуненні вказаної прогалини. Інша мета полягає в співставленні результатів оцінювання з використанням співвідношень (2.1), (2.2) і (2.11). Незважаючи на очевидну корисність такого співставлення, воно не було реалізоване до останнього часу.

2.1 Алгоритм прогнозування розбірливості зашумленої мови

Структура пропонованого алгоритму прогнозування розбірливості зашумленної мови аналогічна структурі на рис. 2.1. Відрізняється лише спосіб реалізації окремих етапів алгоритму, заснований на аналітичному описі спектральних властивостей мовного сигналу і шуму. Деталізуємо опис кожного з етапів:

1) Формування вхідних даних:

– задання аналітичної спектральної моделі мовного сигналу в вигляді розподілу дисперсій D_{sk}, де k - номер смуги частот;

– розробка аналітичної спектральної моделі шуму у вигляді розподілу дисперсій D_{nk};

– задання очікуваного інтегрального співвідношення сигнал-шум SNR_0 для сигналу, що прослуховується.

2) Коректування розподілу дисперсій мовного сигналу (або шуму) для забезпечення заданого інтегрального співвідношення сигнал-шум SNR_0:

– розрахунок “первинного” значення інтегрального співвідношення сигнал-шум:

$$SNR = D_s/D_n = \frac{\sum_{k=1}^{K} D_{sk}}{\sum_{k=1}^{K} D_{nk}}; \tag{2.13}$$

– розрахунок поправочного коефіцієнту:

$$T = SNR_0/SNR; \tag{2.14}$$

– коректування розподілу дисперсій мовного сигналу у відповідності до співвідношення:

$$D_{0sk} = TD_{sk}, \tag{2.15}$$

або коректування розподілу дисперсій шуму у відповідності до співвідношення:

$$D_{0nk} = D_{nk}/T. \tag{2.16}$$

3) Розрахунок парціальних співвідношень сигнал-шум для заданого SNR_0:

– розрахунок парціальних співвідношень сигнал-шум з врахуванням (2.15) та (2.16):

$$E_{0k} = 10\lg\frac{D_{osk}}{D_{nk}} = 10\lg\frac{D_{sk}}{D_{onk}} = 10\lg(D_{sk}/D_{nk}) + 10\lg(T), \qquad (2.17)$$

– коректування значень E_{0k} у відповідності з (2.6), якщо використовується методика Покровского.

4) Формування вихідних даних:

– розрахунок артикуляційної розбірливості і розмовної розбірливості у відповідності з (2.1)-(2.2), або розрахунок індексу розбірливості у відповідності з (2.12).

В якості вихідних даних можна використовувати щільності ймовірностей сигналу і завади, якщо врахувати зв’язок між дисперсією D_k випадкового стаціонарного процесу в k-й смузі частот Δf_k і середніи значенням спектральної щільності потужності P_k в межах цієї смуги частот

$$P_k = D_k/\Delta f_k\ . \qquad (2.18)$$

Очевидно, що при розрахунку розбірливості мови в лініях зв'язку, значення очікуваного інтегрального відношення сигнал-шум SNR_0 слід задавати, виходячи з типових умов експлуатації лінії зв'язку. Дещо складніше задавати SNR_0, якщо мова прослуховується в приміщеннях.

При розрахунку розбірливості мови в приміщеннях з відкритим плануванням (стандарт ISO 3382-3-2013), до яких відносяться офіси, бібліотечні зали, учбові приміщення, для невеликих (2-4 м) відстаней R між диктором і слухачем можна припустити, що мовний сигнал ослабляється так само, як у вільному просторі, тобто на 6 дБ з подвоєнням відстані. Зі збільшенням R, проте, ця закономірність порушується, і тоді розрахунок рівня сигналу в точці прослуховування дещо ускладнюється.

2.2 Алгоритм прогнозування розбірливості реверберованого мовлення

Запропонований алгоритм перевірявся для чотирьох видів шумів: білого, рожевого, коричневого і типового для навчальних приміщень (табл.2.4). Помітимо, що типовий розподіл рівнів шуму Dnk,dB по частотним каналам для навчальних приміщень взято з нормативного документу [40], а параметри довготривалого спектру мови з роботи [21].

Таблиця 2.4 - Первинні розподіли дисперсій мови та шуму

Параметри частотних каналів							
f_{0k}	125	250	500	1000	2000	4000	8000
Δf_k	90	175	355	690	1400	2800	5270
Параметри білого шуму							
D_{nk}	0.130	0.254	0.514	1	2.029	4.058	7.638
Параметри рожевого шуму							
D_{nk}	1.043	1.014	1.029	1	1.014	1.014	0.955
Параметри коричневого шуму							
D_{nk}	8.348	4.058	2.058	1	0.507	0.254	0.119
Параметри типової шумової завади							
$D_{nk,dB}$	61	54	49	45	42	40	38
D_{nk}	5.193	2.017	1.284	1	1.023	1.273	1.5471
Параметри довготривалого спектру мови							
$P_{sk,dB}$	60	70	68	59,5	52	46	40,5
D_{sk}	0.146	2.846	3.642	1	0.361	0.181	0.096

Для спрощення розрахунків, наведені в табл.2.4 значення розподілів дисперсій нормовані по значенням дисперсії в четвертому частотному каналі ($f_{04} = 1000$ кГц).

На рис.2.2 наведені результати розрахунків у вигляді залежностей оцінок розбірливості мови від очікуваного відношення сигнал-шум SNR_0 в діапазоні значень SNR_0 = -26…+20 дБ. Залежності на рис.2.2,а,б,в, побудовані з використанням коефіцієнтів сприйняття Н.Б. Покровського, М.А. Сапожкова і коефіцієнтів сприйняття з табл.2.1-2.2. Вони добре узгоджуються з результатами, що наведені в роботах [26], [39]. На рис. 2.2,г наведені оцінки STI.

Провівши порівняння між собою ориманих графіків можна зробити висновки, що графіки, які отримані для коефіцієнтів сприйняття Н.Б. Покровського (рис..2.2,а), принципово відрізняються від інших графіків тим, що в усьому діапазоні розглянутих значень SNR_0, маскуючі властивості

білого шуму виявляються кращими, ніж у коричневого шуму. Між тим, в [26] і [39] вперше була помічена і виправлена некоректність коефіцієнтів сприйняття Н.Б. Покровского.

Залежності на рис.2.2,б і рис.2.2,в, побудовані з використанням відкоректованих коефіцієнтів сприйняття, які свідчать, що для малих значень SNR_0 ($SNR_0 < -8$ дБ) маскуючі властивості білого шуму уступают маскуючим властивостям для коричневого шуму. Цей результат узгоджується з поведінкою залежностей на рис.2.2,г, але в даному випадку порогове значення SNR_0 дорівнює: $SNR_0 =$ -13 дБ.

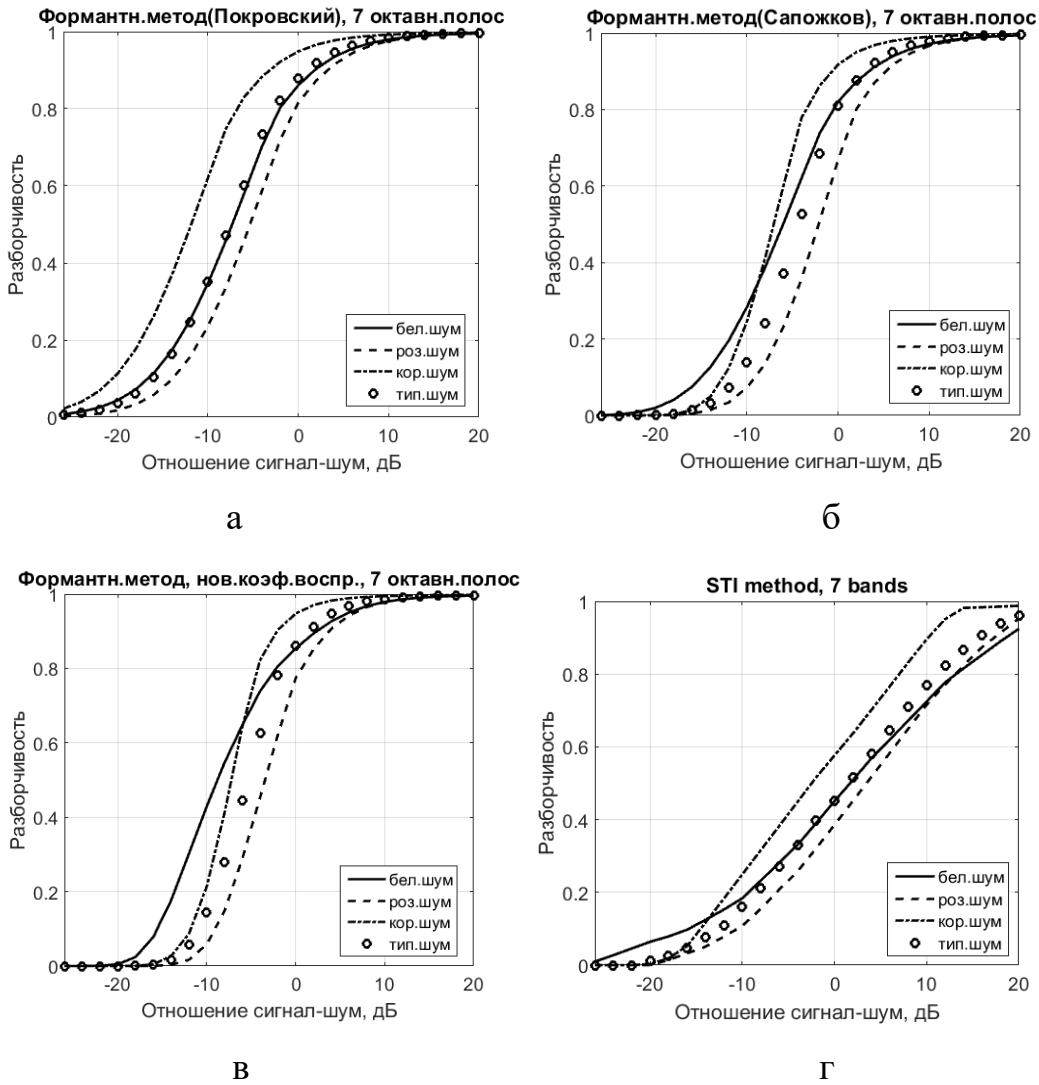


Рис.2.2 - (а, б, в) - Оцінки розмовної розбрливості; (г) - індексу розбірливості

Представлені на рис.2.3 результати суб'єктивних випробувань артикуляцій [41] свідчать про виражену тенденцію до погіршення маскуючих властивостей білого шуму при малих значеннях відношення сигнал-шум SNR_0. Відсутність явного програшу білого шуму в даному випадку може бути пояснена недостатньою коректністю організації випробувань артикуляцій.

Як і очікувалося, час обчислень при аналітичному моделюванні виявився на порядок менше часу, необхідного при комп'ютерному моделюванні, і не перевищував 1 с.

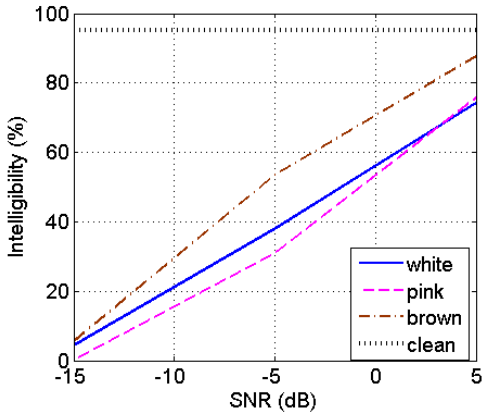


Рис.2.3 – Суб’єктивні оцінки розбірливості за складами [41]

2.3 Висновки до Розділу 2

1. Розглянуто алгоритми прогнозування розбірливості зашумленої мови та прогнозування розбірливості реверберованого мовлення. Перший алгоритм складається з етапів: формування вхідних даних; коректування розподілу дисперсій мовного сигналу (або шуму) для забезпечення заданого інтегрального співвідношення сигнал-шум; розрахунок парціальних співвідношень сигнал-шум для заданого інтегрального співвідношення сигнал-шум; формування вихідних даних. Другий алгоритм перевірявся для чотирьох видів шумів: білого, рожевого, коричневого і типового для навчальних приміщень. Співпадіння отриманих результатів з відомими аналогічними літературними результатами свідчить про коректність запропонованих компонентів аналітичного алгоритму.

2. Зіставлення результатів оцінювання розбірливості мови, отриманих відповідно до усталеного підходу, з результатами оцінювання індексу розбірливості мови STI дозволило підтвердити, маскувальна здатність білого шуму є низькою при малих відношеннях сигнал-шум.

3.Зроблено детальний опис початкового і проміжних етапів алгоритму прогнозування розбірливості мови методом аналітичного моделювання. Працездатність запропонованого алгоритму перевірена для чотирьох видів шумової завади: білої, рожевої, коричневої і типової для навчальних приміщень. Співпадіння отриманих результатів з відомими аналогічними літературними результатами свідчить про коректність запропонованих компонентів аналітичного алгоритму.

РОЗДІЛ 3
КАЛЬКУЛЯТОР РОЗБІРЛИВОСТІ МОВИ НА ПЛАТФОРМІ ANDROID

В даному розділі представлено результати розробки програмно-апаратного калькулятора прогнозування розбірливості мовлення, що маскується шумовою завадою. Розроблений програмний код має бути придатним для розміщення на платформі операційної системи (ОС) Android, що дозволить володарям смартфонів, що працюють на цій ОС, в будь-який момент оцінити очікувану розбірливість мовлення в приміщенні або на вулиці. При цьому розглянуто 4 ситуації:

1) аудиторія;
2) офіс;
3) тиха вулиця;
4) шумна вулиця.

3.1. Розробка алгоритму розрахунку

1) Задання вхідних даних:

– відстані r до диктора;
– вибір виду простору (аудиторія, офіс, тиха вулиця, шумна вулиця).

2) Розрахунок інтегрального рівня сигналу.

У відповідності до [42], інтегральний рівень сигналу на відстані r_0 =1 м від диктору приймається рівним:

$$L_{s0} = 70 \text{ дБ}. \tag{3.1}$$

У відповідності з [43], рівень сигналу на відстані r від диктора:

- для вільного простору (вулиця):

$$L_s = L_{s0} + 20\lg\frac{r_0}{r}. \tag{3.2}$$

- для великого приміщення:

$$L_s = \begin{cases} L_{s0} + 20\lg\frac{r_0}{r}, \ r < 3 \\ L_{s0} + 20\lg\frac{r_0}{r} + 10\lg\left(1 + \frac{r^2}{r^2+9}\right), \ r \geq 3 \end{cases}. \tag{3.3}$$

3) Визначення інтегрального рівня шуму.

У відповідності до [44], введено наступні градації інтегрального рівня шуму:

Таблиця 3.1 – Рівні шуму

Вид простору	**Рівень шуму L_n,** дБ
аудиторія	50
офіс	60
тиха вулиця	60
шумна влиця	70

4) Розрахунок відношення сигнал-шум.

Відношення сигнал-шум SNR_0 в заданій точці:

$$SNR_0 = L_s - L_n. \tag{3.4}$$

5) Розрахунок первинного значення інтегрального відношення сигнал-шум:

$$SNR = 10\lg\left(\frac{\sum_{k=1}^{K} D_{sk}}{\sum_{k=1}^{K} D_{nk}}\right), \tag{3.5}$$

де розподіли D_{sk} і D_{nk} беруться з табл.3.2.

Таблиця 3.2 - Первинні розподіли дисперсій мови і шуму

Параметри частотних каналів							
k	1	2	3	4	5	6	7
D_{sk}	0,146	2,846	3,642	1	0,361	0,181	0,096
D_{nk}	5,193	2,017	1,284	1	1,023	1,273	1,5471

6) Розрахунок парціальних відношень сигнал-шум E_k:

$$E_k = 10\lg(D_{sk}/D_{nk}) + SNR_0 - SNR, \quad k = 1, \dots, 7, \tag{3.6}$$

7) Розрахунок індексу STI.

Згідно методики оцінювання розбірливості мови, яка наведена в [42], розбірливість мови оцінюють за допомогою індексу STI:

$$STI = \sum_{k=1}^{7} \alpha_k \cdot T_k - \sum_{k=1}^{6} \beta_k \cdot \sqrt{T_k \cdot T_{k+1}}, \tag{3.7}$$

$$T_k = \begin{cases} 0, \ E_k < -15; \\ (E_k + 15)/30, -15 < E_k < +15; \\ 1, \ E_k > +15; \end{cases},$$

де α_k - вагові коефіцієнти, β_k – коефіцієнти надлишковості, значення яких для октавних смуг з центральними частотами f_0 приведені в табл.2.3.

8) Конвертування оцінки розбірливості.

Для кращого розуміння користувачами результатів розрахунків розбірливості мови, значення індексу (3.7) доцільно конвертувати в значення 5-бальної шкали оцінок відповідно до табл.3.3 [45].

Таблиця 3.3 – Конвертування значень STI в значення 5-бальної шкали

Розбірливість, R	STI
5 (відмінно)	>0,75
4 (добре)	0,6-0,75
3 (задовільно)	0,45-0,6
2 (незадовільно)	0,3-0,45
1 (дуже погано)	<0,3

9) Формування вихідних даних:

– розбірливість *R*;

– умови вимірювань (вид простору, дистанція до диктора).

3.2 Реалізація алгоритму у середовищі Android

Для розробки калькулятора розбірливості мови, за вище описаним алгоритмом, скористаємося безкоштовним застосунком Android Studio [46]. Це нове і повністю інтегроване середовище розробки застосунків, яке недавно було випущено компанією Google для операційної системи Android. Цей продукт покликаний забезпечити розробників новими інструментами для створення застосунків, а також надати альтернативу Eclipse, що є нині найбільш популярним середовищем розробки.

При створенні нового проекту в Android Studio, доступна структура проекту з усіма файлами, що містяться в каталозі SDK. Цей перехід до системи системі управління Gradle надає процесу розробки ще більшу гнучкість.

Android Studio дозволяє побачити будь-які візуальні зміни, які проводяься в реальному часі. Також можна побачити, як застосунок одночасно виглядатиме на різних пристроях під управлінням Android, з різними налаштуваннями і розподільною здатністю екрану.

Продукт також має нові інструменти для упаковки і маркування коду. Це дозволяє легко орієнтуватися в проекті з великою кількістю коду. У програмі також задіяна функція перетягання, завдяки якій можна переміщати компоненти за допомогою призначеного для користувача інтерфейсу.

На додаток до всього, Android Studio має функцію Google Cloud Messaging, яка дозволяє надсилати дані з сервера на Android-пристрій через хмару.

Таким чином, Android Studio – це надійне і просте середовище розробки, за допомогою якого можна перевірити продуктивність застосунку на різних типах пристроїв. Воно містить зручні помічники і шаблони для загальних елементів програмування для Android, а також, повнофункціональний редактор з множиною додаткових інструментів, сприяючих прискоренню розробки застосунків.

Вікно Android Studio з розробленим кодом калькулятора розбірливості мови показано на рис.3.1. Лістинг програми на мові програмування java наведено в Додадку А.

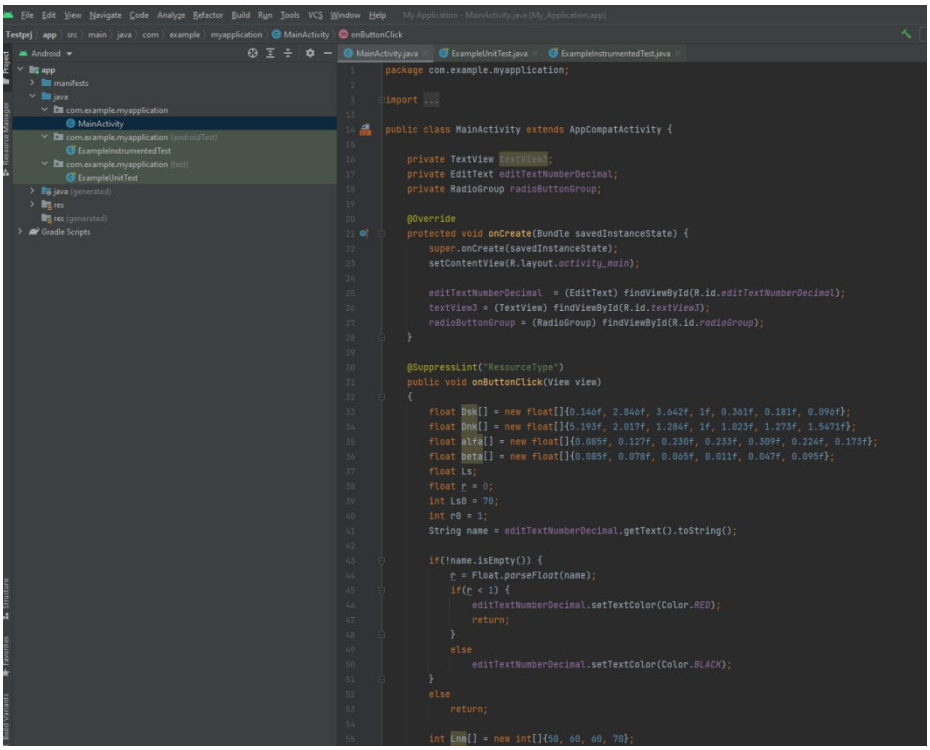


Рис.3.1 – Вікно Android Studio з розробленим кодом калькулятора розбірливості мови

Після запуску розробленого застунку за допомогою, наприклад безкоштовного Android емулятора NoxPlayer [47], з’являється вікно (рис.3.2) в якому пропонується вибрати вид простору (аудиторія, офіс, тиха вулиця, шумна вулиця) та ввести відстань до диктора.

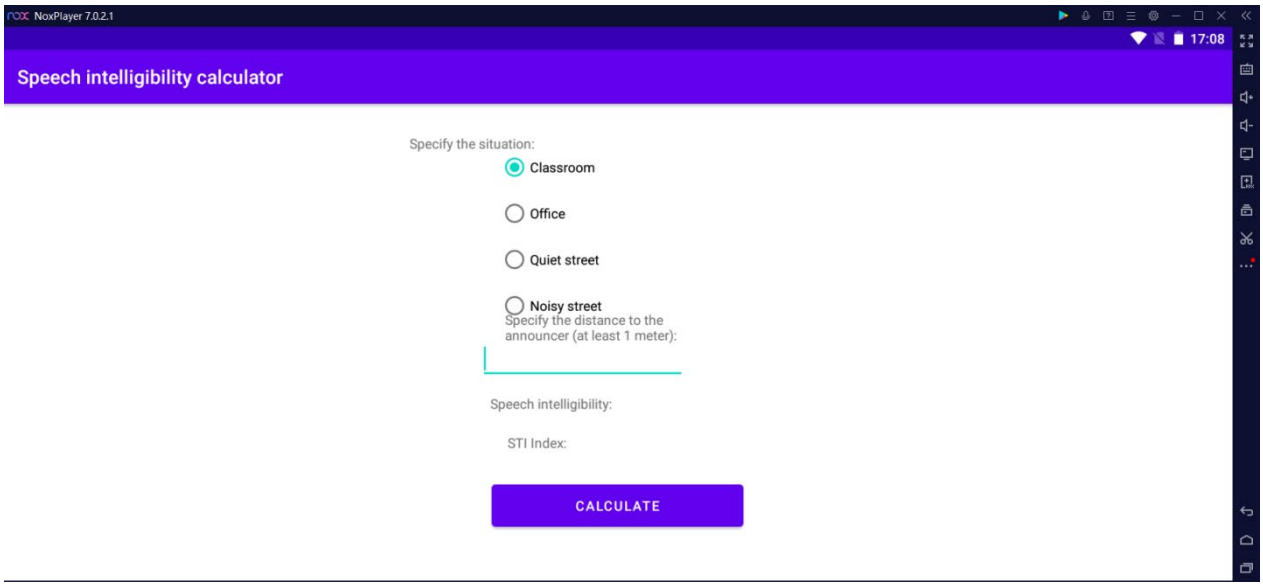


Рис.3.2 – Вікно задання початкових даних в розробленому калькуляторі розбірливості мови

Після введення початкових даних та натискання кнопки “CALCULATE” миттєво з’являються розраховані дані (рис.3.3).

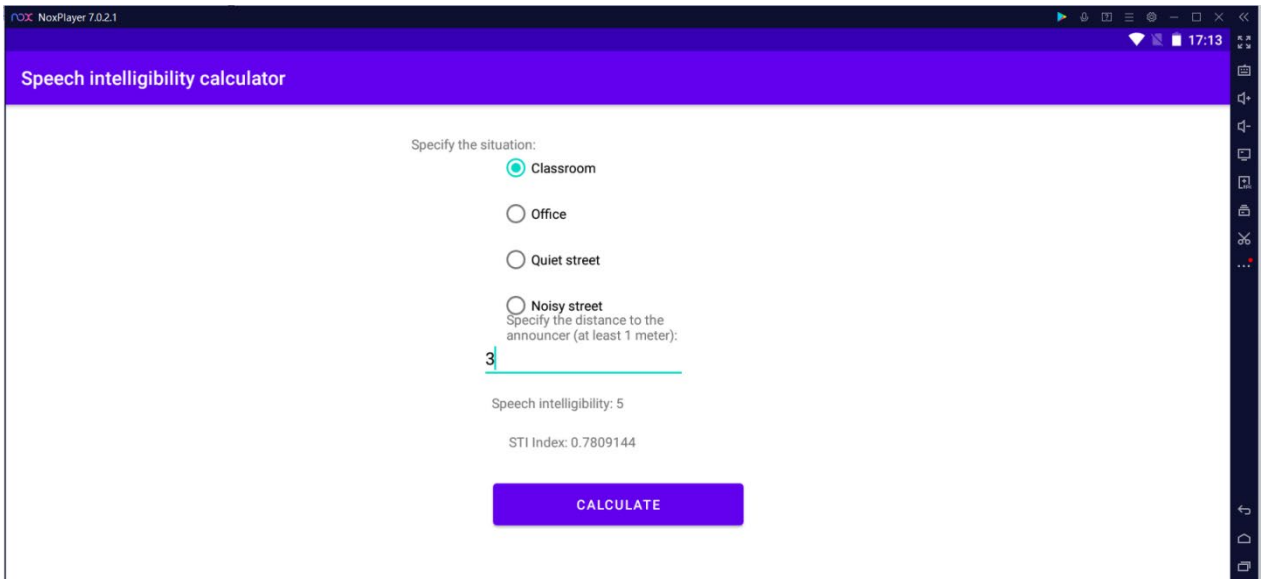


Рис.3.3 – Вікно розрахованих даних в розробленому калькуляторі розбірливості мови

Таким чином, на рис.3.3 ми бачимо розраховану розбірливість мови, та індекс STI.

3.3 Висновки до Розділу 3

На основі існуючого алгоритму, який вже реалізований в прикладному пакеті MATLAB, був розроблений калькулятор розбірливості мови в середовищі Android Studio. Android Studio – це надійне і просте середовище розробки, за допомогою якого можна перевірити продуктивність застосунку на різних типах пристроїв. Воно містить зручні помічники і шаблони для загальних елементів програмування для Android, а також, повнофункціональний редактор з множиною додаткових інструментів, сприяючих прискоренню розробки застосунків.

Перевагами розробленого калькулятора є обгрунтований алгоритм розрахунку розбірливості мови, простота користування, швидкість отримання результатів, та повністю безкоштовне програмне середовище в якому він був розроблений.

РОЗДІЛ 4
ЕКСПЕРИМЕНТАЛЬНЕ ВИМІРЮВАННЯ РОЗБІРЛИВОСТІ МОВЛЕННЯ СПОТВОРЕНОГО РЕВЕРБЕРАЦІЄЮ

В даному розділі представлено результати експериментальних досліджень розбірливості мовлення, спотвореного реверберацією, в одній із аудиторій корпусу №12 НТУУ «КПІ ім. І. Сікорського». Дослідження проводилися із використанням наступного обладнання: активна колонка Genius SP-HF 2.0 500, вимірювальні конденсаторні мікрофони Superlux ECM-999, зовнішня звукова карта Steinberg UR242. Два мікрофони було розміщено біля вушних проходів саморобної штучної голови, яку почергово переміщали в різні точки приміщення (рис. 4.1). Ці точки було обрано таким чином, щоб можна було дослідити:

1) залежність розбірливості мовлення від дистанції між диктором та слухачем;
2) вплив задньої та бокових стін аудиторії на розбірливість мовлення.

4.1 Запис та обробка тестових сигналів для оцінювання приміщення

Вимірювання розбірливості мови проводилося в аудиторії 209, 12 корпусу, НТУУ «КПІ ім.Ігоря Сікорського» (рис.4.1).

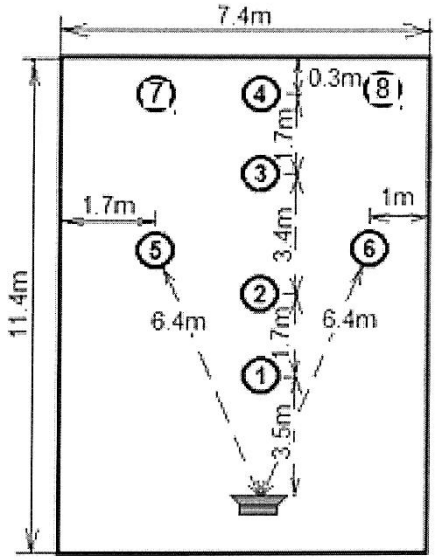


Рис.4.1 – Розміщення 8 точок вимірювання в аудиторії 209-12 НТУУ «КПІ ім. Ігоря Сікорського»

Використосувалися тестові сигнали 4 типів:

1) звук плескання долонями (clap);
2) звук гумової кульки, що лопнула (ball);
3) широкосмуговий сигнал «послідовність максимальної довжини» (mls) тривалістю 1.5 с;
4) гармонічний сигнал, модульований за частотою по квадратичному закону (sweep) тривалістю 4 с.

Запис сигналів виконувався у 8 точках аудиторії, згідно плану на рис.4.1, із частотою дискретизації 44.1 кГц.

Перед та після записаного сигналу витримувалася пауза не менше 2 с, яка потрібна для наступного аналізу фонового шуму (відносно рівня сигналу).

При записах сигналів мікрофони розташовувалися в зонах вушних раковин штучної голови на відстані 1 см від вушного каналу.

Апаратно-програмний комплекс «Штучна голова» [50] містить наступну електроакустичну апаратуру: активну колонку Genius SP-HF 2.0 500, вимірювальні конденсаторні мікрофони Superlux ECM-999, зовнішню звукову карту Steinberg UR242.

Через особливості вимірювального тракту, підсилення в правому каналі під час експерименту було сильнішим, ніж в лівому.

4.2 Оцінювання розбірливості мовлення в різних точках університетської аудиторії

Результати оцінювання параметрів імпульсної характеристики (ІХ) приміщення можуть вважатися достовірними за умови, що відстань U між піковим значенням ІХ та рівнем фонового шуму є величиною, близькою до 30-35 дБ [48]. Тому доцільно виконати попередні вимірювання величини U.

Для автоматизації вимірювання різниці U між піком імпульсної характеристики приміщення та фоновим шумом можна використати два способи:

1) із використанням ефективної напруги фонової завади: $U = R_{\max} - R_{eff\ noise}$;

2) із використанням максимального значення фонової завади: $U = R_{\max} - R_{\max noise}$.

Будемо вважати, що в аудиторіях середнього розміру, до яких, зокрема, відноситься аудиторія 209, фонова завада починається приблизно через 1 с після приходу прямого звуку в дану точку приміщення (максимальне значення $R_{\max}$). Інтервал аналізу фонового шуму виберемо приблизно 0.25 с.

Обидва способи будуть давати подібні результати, хоча слід очікувати, що в першому випадку будуть отримані більші значення, оскільки максимальні значення випадкового сигналу завжди є більшими за ефективне значення.

Оскільки фоновий шум можна вважати розподіленим за нормальним (гаусівським) законом, цю різницю можна зробити малою, якщо після обчислення ефективного значення $\sigma = \sqrt{D}$ фонового шуму (D - дисперсія відрізку фонового шуму) вважати

$$R_{\max noise} = 20\lg r_{\max noise}, \tag{4.1}$$

де $r_{\max noise} \approx 3 \cdot r_{eff\ noise} = 3\sigma$ - максимальне значення фонового шуму в лінійному масштабі.

Виберемо другий спосіб як такий, що гарантовано дає менше значення оцінки *U*, що важливо, наприклад, при оцінюванні часу реверберації пороговим методом. Величину *U* для кожного із каналів будемо брати із знаком «+».

Результати вимірювання представлені у вигляді табл.4.1 (з точністю до 2-х десяткових знаків), та рис.4.2-4.5, для D – параметру, який показує різницю рівнів між піком імпульсної характеристики і початком фонової завади.

На основі проведених вимірювань (рис.4.2-4.5) проведемо обгрунтування використання даних типів сигналів.

З практичної точки зору виявилося цікаво порівняти між собою різні тестові сигнали. Тобто, порівняти результати, які отримуються при вимірювання розбірливості мови, і з’ясувати, чи можна в даній невеликій аудиторії (рис.4.1) обмежитися простими сигналами типу clap, і не

використовувати широкосмугові сигнали, які потребують розрахунку взаємно кореляційної функції.

Таблиця 4.1 – D-параметр

№ точки / **Сигнал**		1	2	3	4	5	6	7	8
ball	Лівий канал	39.97	39.13	34.44	34.36	33.88	32.90	35.40	32.14
	Правий канал	36.05	37.24	33.58	33.23	34.24	35.5011	32.91	35.84
clap	Лівий канал	35.31	32.55	34.19	27.16	32.66	30.6362	29.45	26.67
	Правий канал	33.76	34.63	40.20	31.28	33.57	40.2970	33.11	29.52
mls	Лівий канал	34.73	33.18	29.26	24.93	28.31	25.0106	27.84	22.97
	Правий канал	35.51	31.60	28.26	26.00	28.17	30.4620	28.29	26.78
sweep	Лівий канал	37.80	35.80	21.67	35.22	34.58	25.9570	21.02	19.34
	Правий канал	41.13	38.23	26.36	38.04	34.59	32.9383	27.44	24.15

Порівнюючи графіки рис. 4.2-4.5, бачимо, що для mls сигналу значення *U* є меншим за 30 дБ в більшості точок приміщення. Пояснити це можна тим, що при проведенні експерименту сигнали mls та sweep випромінювалися після сигналів ball та clap. При використанні сигналів ball та clap коефіцієнт підсилення тракту запису потрібно було зменшувати, щоб записаний сигнал не кліпувався (обмежувався за рівнем). Тому при використанні mls та sweep сигналів коефіцієнт підсилення був малим, що призводило до малого рівня записаних сигналів й, як наслідок, до порівняно малих значень параметру *U*.

Цікаво, однак, що в даному експерименті sweep сигнал виявився дещо більш стійким до впливу фонового шуму, порівняно із mls сигналом. Цей факт добре узгоджується із відомим твердженням [48], що sweep сигнал є кращим за mls сигнал в умовах порівняно слабкого фонового шуму.

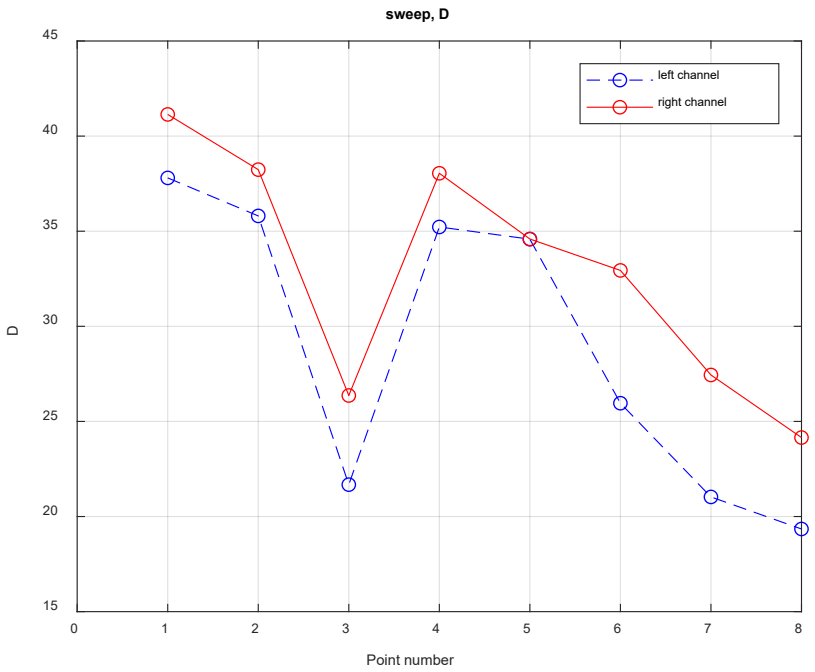


Рис.4.2 – D – параметр для сигналу sweep, 4 с

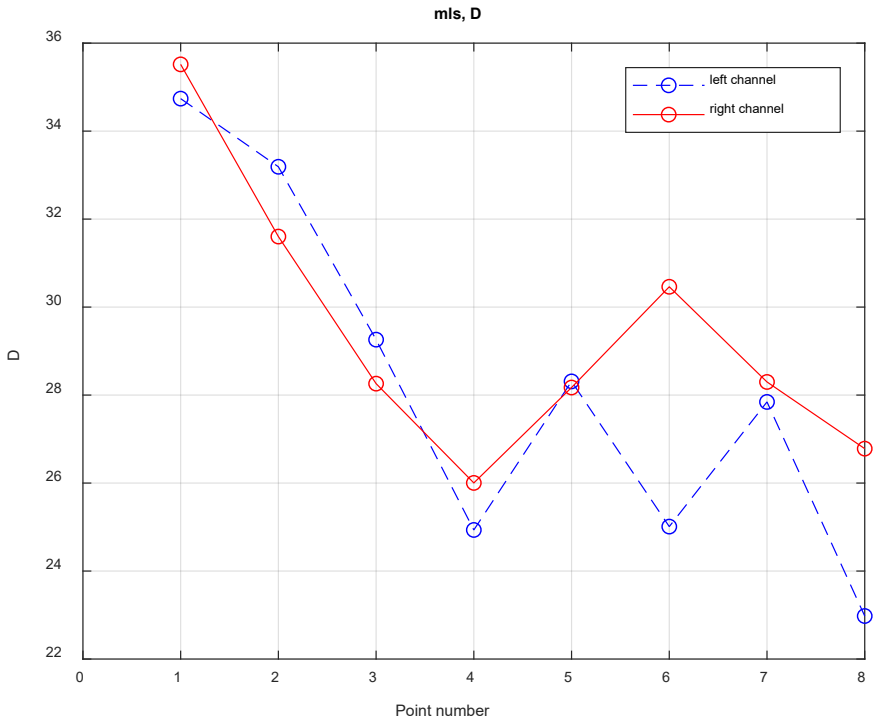


Рис.4.3 – D – параметр для сигналу mls, 1.5 с

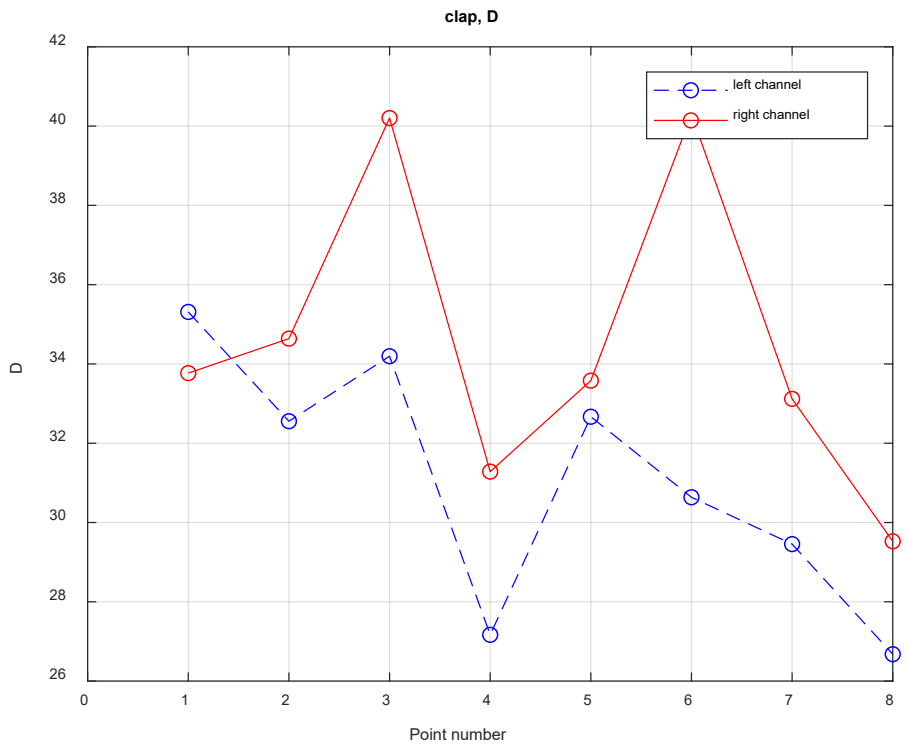


Рис.4.4 – D – параметр для сигналу clap

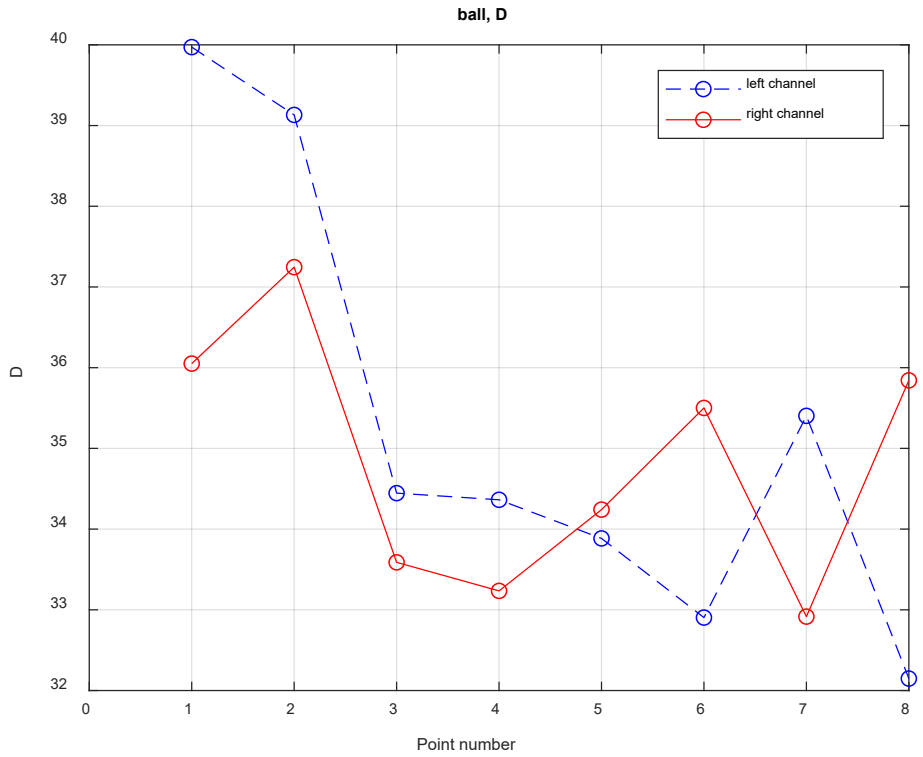


Рис.4.5 – D – параметр для сигналу ball

Табл. 4.1 дозволяє пояснити різницю оцінок розбірливості мовлення, які будуть наведені далі. Графіки на рис.4.4, 4.5, які отримані для сигналів типу clap та ball відповідно, мають перевагу, завдяки вищому динамічному діапазону.

Проблему різниці в каналах для складних сигналів (sweep (рис.4.2) і mls (рис.4.3)) можна пояснити малим підсиленням, яке використовувалося при вимірюваннях.

Для обрахунку розбірливості мови використаємо чотири типи оцінок: індекс розбірливості мови (STI), оцінка розбірливості мови за Покровським, оцінка розбірливості мови за Сапожковим, оцінка розбірливості мови за запропонованими коефіцієнтами сприйняття.

1. Алгоритм розрахунку розбірливості мови за індексом STI.

Слід зазначити, що криві STI(Q) однакові для кажної із 7 смуг частот, де Q - вектор вхідних відношень сигнал-шум.

Для розрахунку скористаємося формулою (3.7), яка записана у вигляді:

$$I = \sum_{k=1}^{7} \alpha_k T_k - \sum_{k=1}^{6} \beta_k \sqrt{T_k T_{k+1}}, \quad (4.2)$$

де k = 1,2..7 – смуги частот; α_k = 0.085, 0.127, 0.230, 0.233, 0.309, 0.224, 0.173; β_k = 0.085, 0.078, 0.065, 0.011, 0.047, 0.095; для відношення Q(k) <(-15)дБ, T(k) = 0; для відношення Q(k) = [-15:15] дБ, T(k) = (Q(k)+15)/30; для відношення Q(k) > (-15)дБ, T(k) = 1.

2. Алгоритм розрахунку формантної та словесної розбірливості мови за нашими коефіцієнтами сприйняття.

Слід зазначити, що криві P(E') різні для кажної із 7 смуг частот.

P(E') – постійна артикуляційна характеристика мови [13].

Якщо відношення сигнал-шум (E) менше за ліву границю значень аргументу поліному, що описує ліву вітку коефіцієнтів сприйняття ($E_{лів.min}$), то:

$$P(i) = 0. \quad (4.3)$$

$P(i)$ – коефіцієнт сприйняття, що враховує втрати розбірливості через наявність заважаючих факторів [13].

Якщо відношення сигнал-шум (E) менше за нуль, то значення коефіцієнту сприйняття P для лівої гілки приймає вигляду:

$$P(i) = P_{лів\ 1} E^n + P_{лів\ 2} E^{n-1} + \cdots + P_{лів\ n} E + P_{лів\ n+1}, \quad (4.4)$$

Якщо відношення сигнал-шум (E) менше за праву границю значень аргументу поліному, що описує ліву вітку коефіцієнтів сприйняття ($E_{прав.min}$), то:

$$P(i) = P_{\text{прав } 1}E^n + P_{\text{прав } 2}E^{n-1} + \cdots + P_{\text{прав } n}E + P_{\text{прав } n+1}. \quad (4.5)$$

Якщо відношення сигнал-шум (E) більше за праву границю значень аргументу поліному ($E_{\text{прав.min}}$), то:

$$P(i) = 1. \quad (4.6)$$

Визначається ймовірність розбірливості формант.

Для $f_0 \leq 400$ формула для ймовірності розбірливості формант приймає вигляд:

$$p_{\text{в}}(i) = 2{,}57 \cdot 10^{-8} \cdot (f_{\text{в}}(i)^{2{,}4} - f_{\text{н}}(i)^{2{,}4}). \quad (4.7)$$

Для $f_0 > 400$:

$$p_{\text{в}}(i) = 1{,}074 \cdot [\exp(-10^{-4} \cdot f_{\text{н}}(i)^{1{,}18}) - \exp(-10^{-4} \cdot f_{\text{в}}(i)^{1{,}18})], \quad (4.8)$$

Далі визначається артикуляційна (формантна) розбірливість мови:

$$A = \sum_{i=1}^{7} p_{\text{в}_i} \cdot P_i. \quad (4.9)$$

Наступном кроком, визначається словесна розбірливість мови.

Для $A < 0{,}15$:

$$W = 1{,}54 \cdot \sqrt[4]{A} \cdot (1 - \exp(-11 \cdot A)). \quad (4.10)$$

Для $A \geq 0{,}15$:

$$W = 1 - \exp\left(\frac{-11 \cdot A}{1 + 0{,}7 \cdot A}\right). \quad (4.11)$$

3. Алгоритм розрахунку розбірливості мови за Сапожковим подібний до алгоритму у пп.2, але у функцію підставляються значення поліному на рівні 1000 Гц.

4. Алгоритм розрахунку словесної розбірливості мови за Покровським.

Рівні відчуття мають вигляд

$$E = Q - \left(\frac{200}{f_0^{0{,}43}} - 0{,}37\right), \text{для } f_0 \leq 1000 \text{ Гц},$$

$$E = Q - \left(1{,}37 + \frac{1000}{f_0^{0{,}69}}\right), \text{для } f_0 > 1000 \text{ Гц}. \quad (4.12)$$

де Q - вектор вхідних значень відношеня сигнал-шум.

Визначаються коефіцієнти сприйняття:

$$P = \frac{0{,}78 + 5{,}46 \cdot \exp[-4{,}3 \cdot 10^{-3} \cdot (27{,}3 - |E|)^2]}{1 + 10^{0{,}1 \cdot |E|}}, \text{для } E \leq 0,$$

$$P = 1 - \frac{0{,}78+5{,}46\cdot\exp[-4{,}3\cdot10^{-3}\cdot(27{,}3-|E|)^2]}{1+10^{0{,}1\cdot|E|}}\text{, для } E > 0. \quad (4.13)$$

Розраховується ймовірність розбірливості формант.

Для $f_0 \leq 400$:

$$p_{\text{в}}(i) = 2{,}57 \cdot 10^{-8} \cdot (f_{\text{в}}(i)^{2,4} - f_{\text{н}}(i)^{2,4}). \quad (4.14)$$

Для $f_0 > 400$:

$$p_{\text{в}}(i) = 1{,}074 \cdot [\exp(-10^{-4} \cdot f_{\text{н}}(i)^{1,18}) - \exp(-10^{-4} \cdot f_{\text{в}}(i)^{1,18})]. \quad (4.15)$$

Далі визначаються артикуляційна (формантна) розбірливість мови за формулою (4.9), та словесна розбірливість мови за формулами (4.10), (4.11).

Розрахунки за алгоритмами пп.1-4, які проводилися за розробленим прграмним забезпеченням (Додаток Б), зведені в табл.4.2, і представлені графічно на рис.4.6-4.9.

Таблиця 4.2 – Значення оцінок розбірливості мови

Тип оцінки розбірливості мови	Номер точки / Тип сигналу	1	2	3	4	5	6	7	8
Оцінка за Покровським, 1 канал	ball	0.9685	0.9680	0.9559	0.9225	0.9642	0.9653	0.9536	0.9583
	clap	0.9720	0.9659	0.9199	0.9640	0.9576	0.9661	0.9588	0.9584
	mls	0.9692	0.9639	0.9568	0.9654	0.9610	0.9586	0.9557	0.9557
	sweep	0.9736	0.9669	0.9603	0.9680	0.9609	0.9619	0.9586	0.9594
Оцінка за Покровським, 2 канал	ball	0.9723	0.9688	0.9656	0.9238	0.9617	0.9627	0.9578	0.9618
	clap	0.9738	0.9678	0.9295	0.9672	0.9570	0.9625	0.9592	0.9565
	mls	0.9736	0.9668	0.9594	0.9655	0.9619	0.9630	0.9646	0.9599
	sweep	0.9744	0.9706	0.9652	0.9669	0.9609	0.9666	0.9666	0.9642
Оцінка за Сапожковим, 1 канал	ball	0.9525	0.9526	0.9307	0.8607	0.9456	0.9476	0.9263	0.9361
	clap	0.9594	0.9477	0.8519	0.9443	0.9325	0.9472	0.9335	0.9323
	mls	0.9537	0.9435	0.9295	0.9472	0.9390	0.9340	0.9278	0.9290
	sweep	0.9614	0.9495	0.9370	0.9518	0.9388	0.9402	0.9337	0.9365
Оцінка за Сапожковим, 2 канал	ball	0.9590	0.9535	0.9487	0.8597	0.9401	0.9436	0.9337	0.9425
	clap	0.9626	0.9513	0.8703	0.9498	0.9324	0.9405	0.9332	0.9290
	mls	0.9616	0.9495	0.9352	0.9472	0.9405	0.9427	0.9459	0.9375
	sweep	0.9632	0.9568	0.9468	0.9498	0.9388	0.9489	0.9495	0.9456
Оцінка за нашими коефіцієнтами сприйняття, 1 канал	ball	0.9739	0.9740	0.9579	0.9037	0.9678	0.9698	0.9537	0.9615
	clap	0.9779	0.9699	0.8847	0.9659	0.9589	0.9688	0.9577	0.9556
	mls	0.9750	0.9672	0.9556	0.9693	0.9635	0.9590	0.9541	0.9553
	sweep	0.9797	0.9709	0.9623	0.9727	0.9618	0.9633	0.9587	0.9617
Оцінка за нашими коефіцієнтами сприйняття, 2 канал	ball	0.9780	0.9742	0.9705	0.9705	0.9011	0.9667	0.9611	0.9658
	clap	0.9789	0.9724	0.9073	0.9073	0.9712	0.9624	0.9584	0.9533
	mls	0.9796	0.9715	0.9597	0.9697	0.9634	0.9661	0.9693	0.9624
	sweep	0.9807	0.9762	0.9685	0.9703	0.9618	0.9705	0.9717	0.9684

STI, 1 канал	ball	0.6367	0.6418	0.5967	0.5115	0.6274	0.6259	0.5876	0.6113
	clap	0.6597	0.6239	0.5100	0.6050	0.5904	0.6171	0.5891	0.5895
	mls	0.6397	0.6083	0.5797	0.6256	0.6026	0.5923	0.5827	0.5923
	sweep	0.6663	0.6300	0.6060	0.6356	0.6249	0.6055	0.5933	0.6096
STI, 2 канал	ball	0.6525	0.6425	0.6332	0.5064	0.6074	0.6229	0.5992	0.6267
	clap	0.6775	0.6315	0.5221	0.6217	0.5931	0.5991	0.5851	0.5849
	mls	0.6593	0.6273	0.5883	0.6227	0.6037	0.6130	0.6240	0.6061
	sweep	0.6750	0.6541	0.6238	0.6262	0.6249	0.6281	0.6318	0.6253

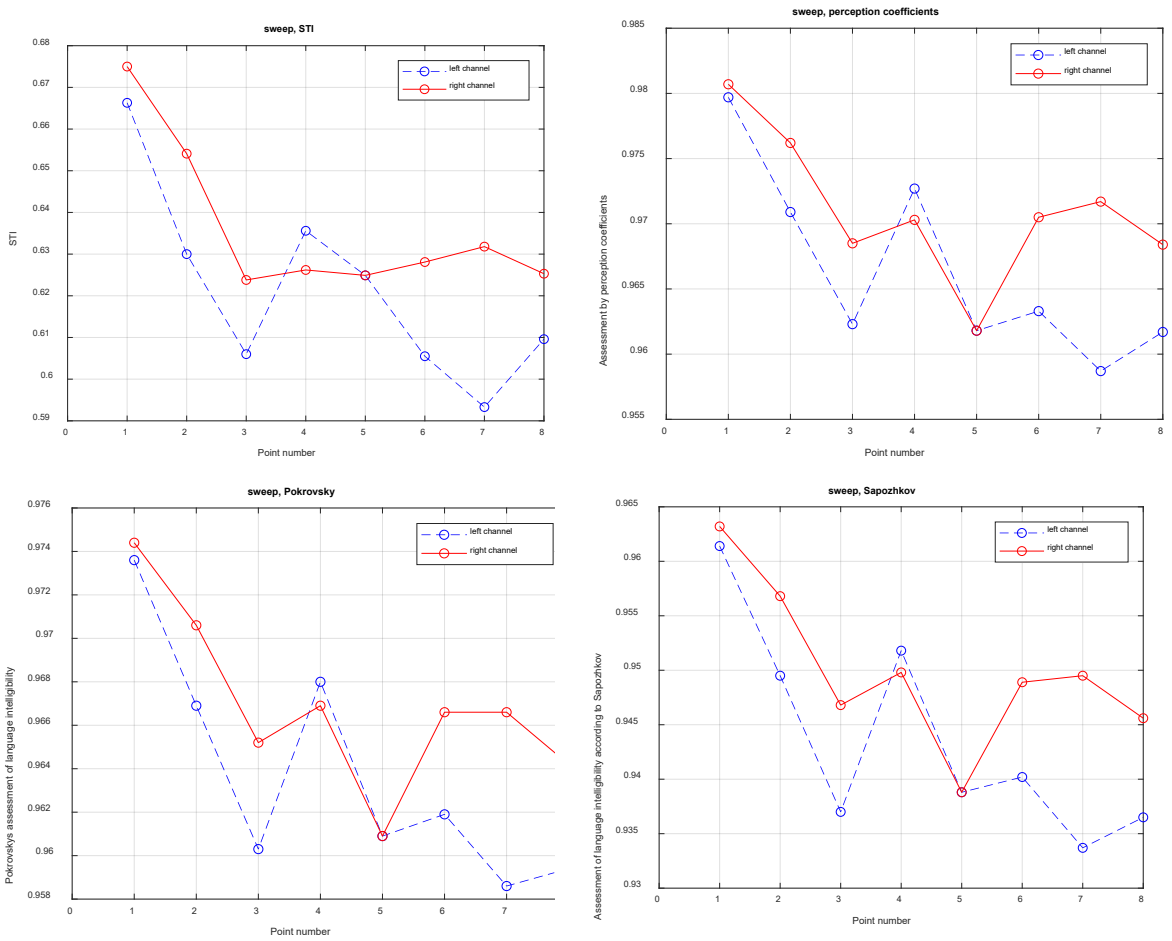
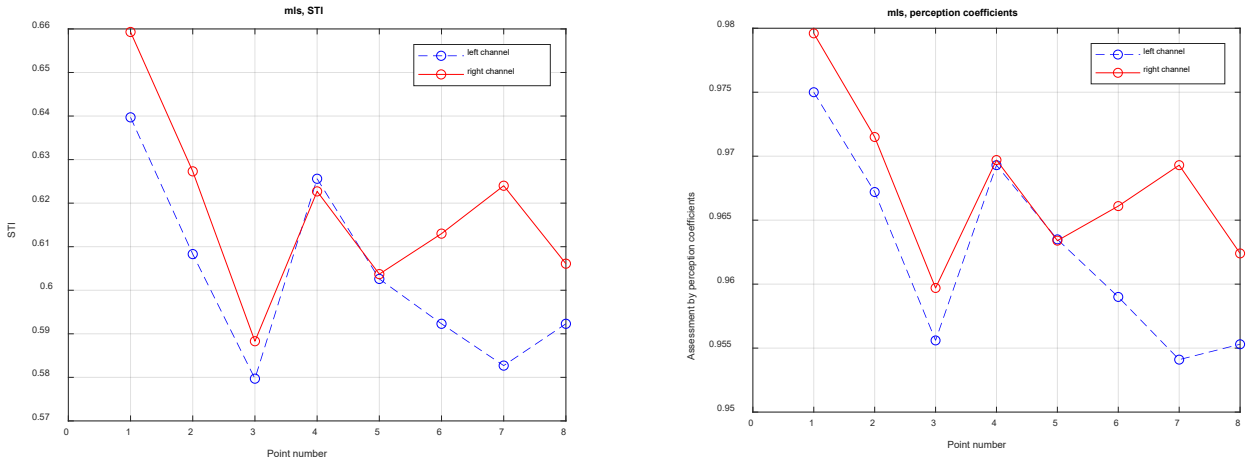


Рис.4.6 – Оцінки для сигналу sweep, 4 с



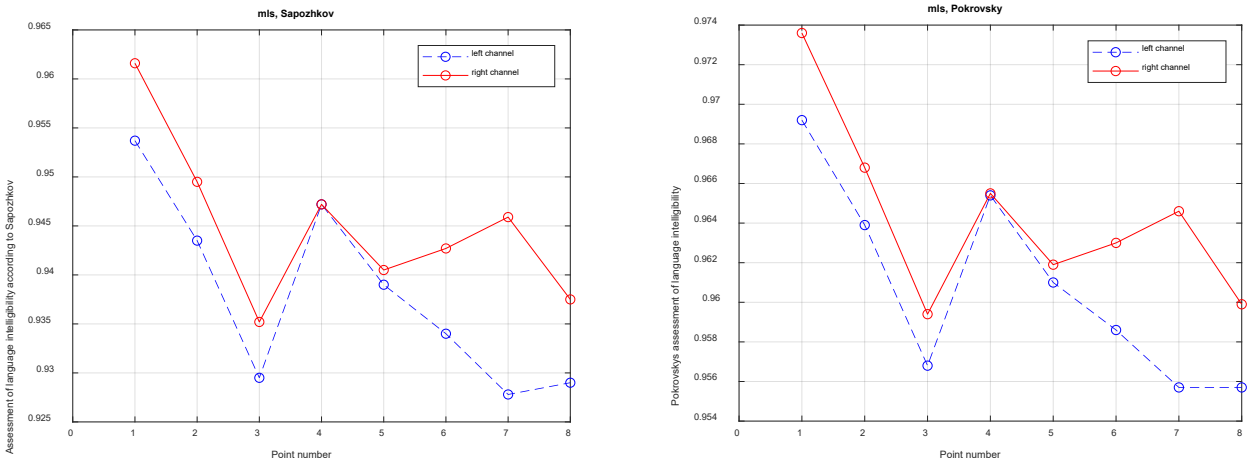


Рис.4.7 – Оцінки для сигналу mls, 1.5 с

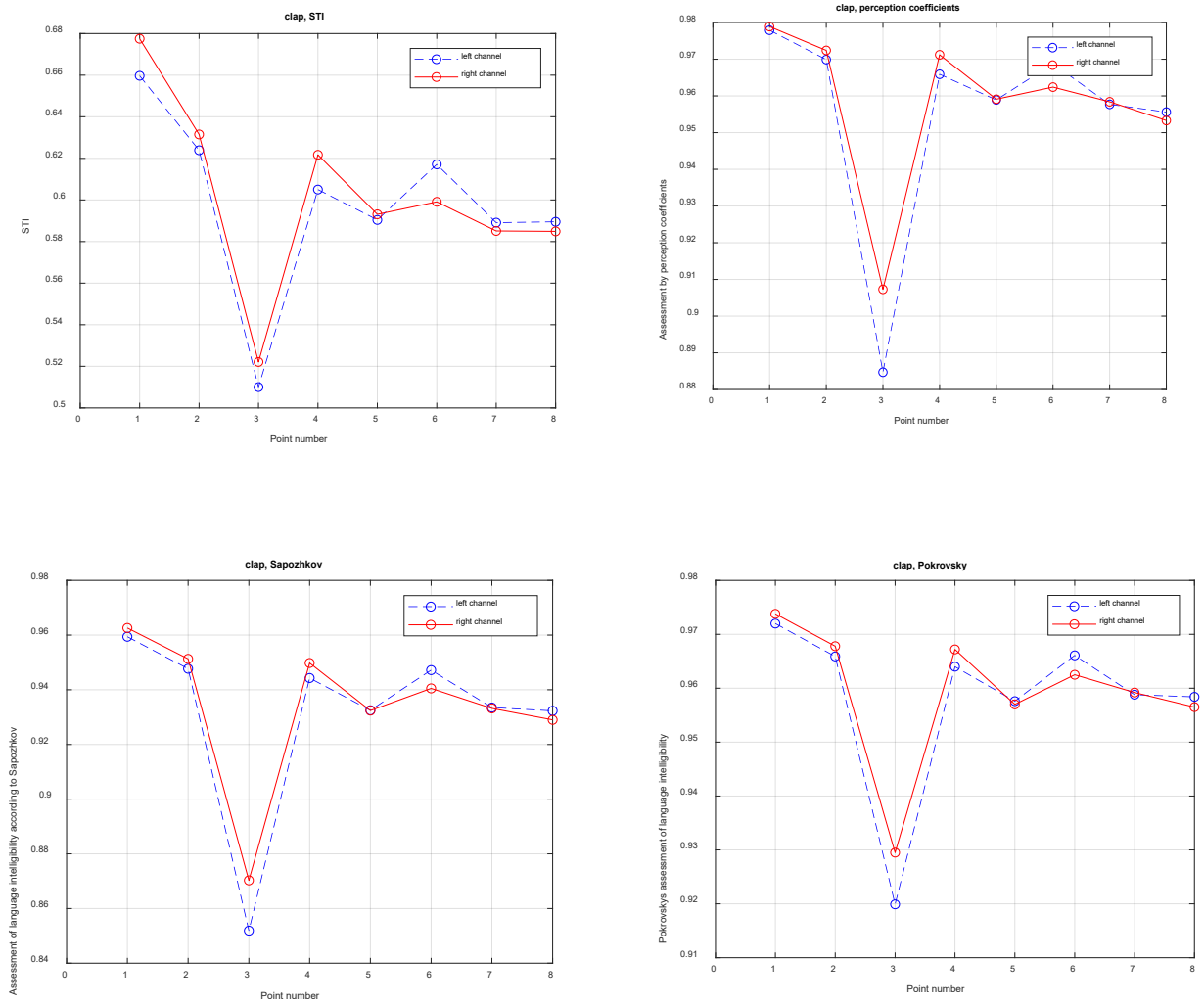


Рис.4.8 – Оцінки для сигналу clap

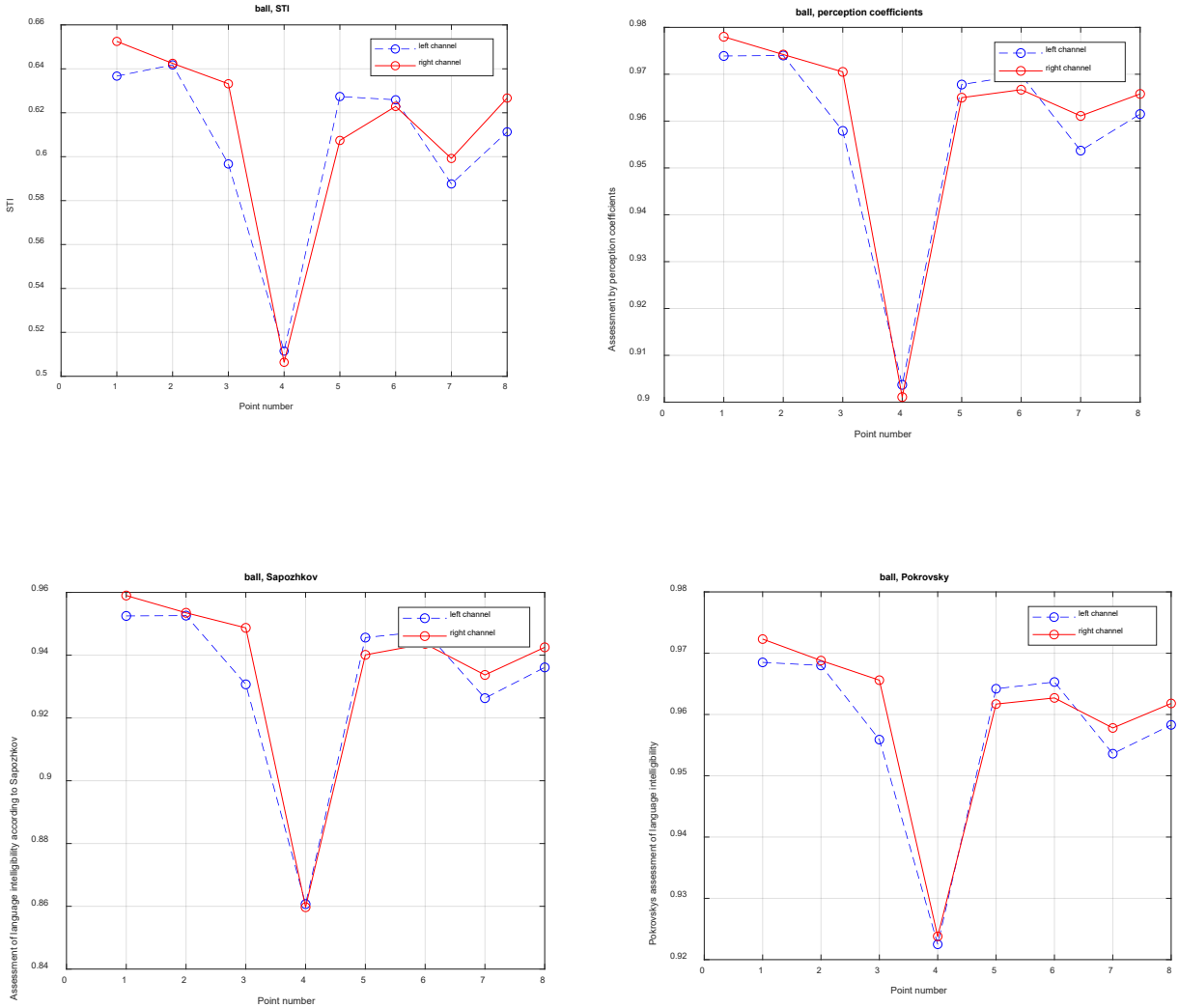


Рис.4.9 – Оцінки для сигналу sweep, 4 с

Підтверджено [48], що перевагою mls-послідовності є стійкість до фонового шуму. Для отримання значного динамічного діапазону відображення спадання звуку необхідно оптимізувати вимірювальну систему. Для вимірювань потрібно максимізувати інтервал спадання звуку – інтервал між максимумом імпульсної характеристики та її значенням в момент, коли спадання зрівнюється із фоновим шумом. Для такої максимізації треба використовувати довгу mls-послідовність та робити усереднення результатів для кількох mls-послідовностей.

Перевагою сигналу типу хлопку в долоні є доступність, на відміну від сигналів типу mls та sweep, для яких отримання імпульсних характеристик потребує подальшої обробки.

Як видно з графіків на рис.4.6-4.9, для всіх сигналів найкращі оцінки отримані в 1 та 2 точках. В цих точках слухач знаходиться ближче всього до диктора, та звук до нього йде прямо.

Знаходження приймача звуку поблизу задньої (точка 4) та бокових (точки 5, 6) стінок підвищує оцінки в порівнянням із випадком, коли приймач знаходився ближче до центру аудиторії (точка 3). Це пов'язано з тим, що в цих точках розбірливість мови підвищується за рахунок відбиття звуку від стін. Як видно з графіків в точці 6 розбірливість мови вище в лівому каналі, оскільки приймач розташовано біля лівої стінки, в 5 точці значення обох каналів рівні, але там відстань від слухача до стінки вища і не так сильно впливає на розбірливість.

Порівняємо отримані результати із рейтингом розбірливості мови і значеннями STI, які наведена табл.4.1 [49].

Таблиця 4.3 - Рейтинг розбірливості мови і значення STI [49]

Qualification	**STI**
Excellent	> 0.75
Good	0.60-0.75
Fair	0.45-0.60
Pour	0.30-0.45
Bad	<0.30

Як видно з табл 4.2 розбірливість мови для сигналу mls є доброю, згідно класифікації табл.4.3, є доброю для точок 1,2,4,5 на обох каналах, для точок 6,7,8 розбірливість є доброю тільки для правого каналу. Для всіх інших випадків значення STI є прийнятним.

Для сигналу sweep значення STI є добрим для всіх точок, окрім лівого каналу точки 7, де воно є прийнятним.

Для сигналу типу clap добра розбірливість тільки в точках 1,2,4 та 6.

Для сигналу ball в точках 1,2,3 (тільки для правого каналу), 5,6,8 значення STI є добрими, значення в усіх інших точках прийнятні. Різке погіршення в точці 4 пов'язане з тим, що при експеременті сигнал було сильно послаблено.

Отже, для всіх сигналів значення STI добрі в точках 1,2,4,5 для обох каналів, та точки 6 (для всіх сигналів, окрім лівого каналу для сигналу mls), тобто для точок, що розміщенні ближче всього до диктора. Для більш віддалених точок розбірливість погіршується, що не дивно, враховуючи зниження рівня сигналу та збільшення впливу реверберації. Дуже цікавим є той факт, що на осьовій лінії аудиторії найгіршою є розбірливість в середині приміщення, а не біля задньої стінки. Цей результат добре узгоджується із попереднім експериментальними дослідженнями [52] й пояснюється тим, що біля задньої стінки до вух слухача практично одночасно приходять прямий звук та його відбиття від задньої стінки, що призводить до суттєвого, до 6 дБ, підвищення відношення сигнал-шум. В середині кімнати так звані «ранні відбиття» надходять із помітними часовими затримками відносно моменту приходу прямого звуку. Це призводить до спотворення форми звукового сигналу (явище інтерференції). Як наслідок, розбірливість мовлення погіршується, незважаючи на те, що рівень сигналу в середині кімнати також дещо підвищується завдяки наявності ранніх відбиттів.

4.3 Висновки до Розділу 4

Досліджено формантний та модуляційний методи розбірливості мови. Вказано на розбіжності між ними, які полягають в різному способі вимірювання відношення сигнал-шум.

При формантному методі сигнал являє собою стаціонарний випадковий процес у вигляді гармоніки або шуму, при модуляційному - це нестаціонарний випадковий процес у вигляді шуму із спектром, який ідентичний спектру довготривалої мови [13].

Зазначено, що достовірність оцінки за Н.Б. Покровським викликає сумніви, оскільки його підхід до побудови коефіцієнта сприйняття P(E) містить надто багато припущень [21], [39].

Оцінка за М.А. Сапожковим [22] має переваги, оскільки коефіцієнт сприйняття P(E), який використовується, співпадає з результатами наших вимірювань [39] на частоті 1 кГц. Недолік оцінки за М.А. Сапожковим в тому, що зроблено припущення про однаковість значень P(E) на всіх частотах.

Недолік індексу STI, з нашої точки зору, полягає в тому, що замість залежності P(E), що плавно змінюється, використовується її лінійна апроксимація, яка до того ж, як і у М.А. Сапожкова, не залежить від частоти. Також, індекс STI є не дуже зручним для використання на практиці, оскільки це індекс, а не розбірливість, внаслідок чого доводиться користуватися таблицею переводу значень STI в оцінки розбірливості за 5-бальною шкалою.

Обчислюючи індекс STI через IX приміщення, ми в практичній частині роботи використали модуляційний метод. Обчислюючи словесну розбірливість W через IX приміщення, ми використали формантно-модуляційний метод.

Формантний метод в “чистому” вигляді в роботі майже (за виключенням Розділу 3) не використовувався, оскільки він придатний тільки для ситуації, коли мова маскується шумом, а реверберації немає.

ВИСНОВКИ

В даній дипломній роботі виконано дослідження, що стосуються об’єктивного оцінювання розбірливості мови в університетських лекційних приміщеннях. Тема роботи є актуальною з огляду на важливість правильної передачі акустичної інформації від викладача до здобувачів освіти.

1. Виконано аналітичний огляд моделей мовленнєвого тракту людини та спектрального складу звуків розмовного мовлення. Показано, що у різних людей параметри формант одних і тих же фонем різняться між собою. У дитячих і жіночих голосах ці форманти виявилися вищими, ніж у чоловіків. Розглянуто форматний аналіз мови як особливий вид спектрального аналізу.

2. Проведено оцінювання розбірливості мови, спотвореної шумом та реверберацією, на основі алгоритмів прогнозування розбірливості зашумленої мови та алгоритмів вимірювання розбірливості реверберованого мовлення. Співпадіння отриманих результатів з відомими аналогічними літературними результатами підтверджує правильність програмної реалізації алгоритмів оцінювання розбірливості мовлення.

3. На основі існуючого алгоритму, який вже реалізований в прикладному пакеті MATLAB, був розроблений калькулятор розбірливості мови в середовищі Android Studio. Перевагами розробленого калькулятора є обґрунтований алгоритм розрахунку розбірливості мови, простота користування, швидкість отримання результатів, а також повністю безкоштовне програмне середовище, в якому він був розроблений.

4. Виконано експериментальне оцінювання розбірливості мови у 8 точках аудиторії (№209, корпус №12) кафедри акустичних та мультимедійних електронних систем факультету електроніки НТУУ «КПІ ім. Ігоря Сікорського».

4. Проаналізовані переваги та недоліки сучасних об’єктивних методів оцінювання розбірливості мови, а саме, формантного та модуляційного. Встановлено розбіжності між ними, які полягають в різному способі

вимірювання відношення сигнал-шум. Крім того, при формантному методі тестовий сигнал являє собою стаціонарний випадковий процес у вигляді гармоніки або шуму, при модуляційному - це нестаціонарний випадковий процес у вигляді шуму із спектром, ідентичним спектру довготривалої мови.

5. Проаналізовані оцінки розбірливості мови за Н.Б. Покровским та М.А. Сапожковим. Оцінка за Н.Б. Покровським викликає сумніви, оскільки підхід до побудови коефіцієнта сприйняття P(E) містить надто багато припущень. Оцінка за М.А. Сапожковим має переваги, оскільки коефіцієнт сприйняття P(E) співпадає з експериментальними результатами, отриманими на кафедрі АМЕС, на частоті 1 кГц. Недолік оцінки за М.А. Сапожковим в тому, що зроблено припущення про однаковість значень P(E) на всіх частотах.

6. Недолік індексу STI полягає в тому, що замість залежності P(E), що плавно змінюється, використовується її лінійна апроксимація, яка до того ж, як і у М.А. Сапожкова, не залежить від частоти. Також, індекс STI є не дуже зручним для використання на практиці, оскільки це індекс, а не розбірливість, внаслідок чого доводиться користуватися таблицею переводу значень STI в оцінки розбірливості за 5-бальною шкалою.

ПЕРЕЛІК ЛІТЕРАТУРНИХ ДЖЕРЕЛ

1. Fant G., Tatham M. A. A. Auditory Analysis and Perception of Speech. Acadedmic Press, 1st Edition, 1975. eBook ISBN: 9780323145480.

2. Fant Gunnar, Liljencrants Johan. Data reduction of LF voice source parameters // Working papers, Lund University, Department of Linguistics and Phonetics, V.43, pp.62-65, 2009. https://scholar.google.com/scholar?oi=bibs&cluster=11162109288768989187&btnI=1&hl=en

3. Flanagan James L. Speech analysis synthesis and perception. Springer, 1972. https://link.springer.com/book/10.1007/978-3-662-01562-9

4. Gelfand Stanley A. Hearing: An Introduction to Psychological and Physiological Acoustics. Informa Healthcare UK. 2010. ISBN-13: 978-1-4200-8865-6.

5. Korzhyk O., Naida S., Kurdiuk S., Nizhynska V., Korzhyk M., Naida A. Use of the pass-through method to solve sound radiation problems of a spherical electro-elastic source of zero order. EUREKA: Physics and Engineering, (5), 133-146. (2021). DOI: https://doi.org/10.21303/2461-4262.2021.001292

6. Naida S. A., Zheliaskova T. M., Naida A. S., Kliushnichenko H. A., Damarad A. V. Methods for Calculating the Transfer Functions of Broadband Plate Piezoelectric Transducers with Transition Layers // J. Nano- Electron. Phys. , pp. 1-6, V.13, №6, 2021. DOI: https://doi.org/10.21272/jnep.13(6).06029

7. Korzhyk O., Naida S., Zheliaskova T., Chaika O., Naida N. Application of partial areas method in the problem of sound radiation by a sphere in a waveguide with soft acoustically boundaries. Eastern-European Journal of Enterprise Technologies, 5(5 (113), 62–79, 2021. DOI: https://doi.org/10.15587/1729-4061.2021.243161

8. Naida Sergey, Korzhyk Oleksii, Lastivka Ivan, Pavlenko Olha, Zheliaskova Tetiana, Korzhyk Maxim, Naida Anton, Naida Nikita, Chaika Oleksander. Mode-Matching Method Applied to Solving Sound Reception Problem

Using Helmholtz Resonator. Journal of Nano- and Electronic Physics, (1), pp. 1-12, 2022.

9. Naida S.A. “On the genetic relationship between the formants of the sounds "a" and "y" and the resonance frequency of the normal middle ear,” Electronics and Communications, No. 17, pp. 7-11, 2002.

10. Naida S.A. “On the acoustic resonances of speech and auditory paths and their correlation,” Electronics and Communications, No. 16, pp. 56-58, 2002.

11. Naida Sergey. “Acoustic Theory Problems Of Speech Production In The Light Of The Discovery Of The Formula For The Middle Ear Norm Parameter,” Proc. of IEEE 35th Int. Sc. Conf. Electronics and Nanotechnology (ELNANO), pp. 347-350, 21-24 April 2015, Kyiv, Ukraine. DOI: https://doi.org/10.1109/ELNANO.2015.7146907

12. Naida Sergey, Naida Nikita, Didkovsky Vitalii, Pavlenko Olha. Spectral Analysis of Sounds by Acoustic Hearing Analyzer // Proc. of IEEE 39th Int. Sc. Conf. Electronics and Nanotechnology (ELNANO), pp. 421-424, 16-18 April 2019, Kyiv, Ukraine. DOI: https://doi.org/10.1109/ELNANO.2019.8783915

13. Дидковский В.С., Дидковская М.В., Продеус А.Н. Акустическая экспертиза каналов речевой коммуникации. Монография. – К: Имэкс-ЛТД, 2008. – 420 с.

14. Гавриленко А.В., Дидковский В.С., Продеус А.Н. Парциальные отношения сигнал-шум в роли критерия, альтернативного разборчивости речи // Электроника и связь, №4, 2006. С.48-53.

15. Дідковський В.С., Луньова С.А. Основи архітектурної та фізіологічної акустики. – К.: 2001. – 422 с.

16. Jacob K.D. Correlation of Speech Intelligibility Tests in Reverberant Rooms with Tree Predictive Algorithmes. // J. Audio Eng. Soc. - 1989. - Vol. 37. – P. 1019-1030.

17. Houtgast T., Steeneken H. J. M. Evaluation of Speech Transmission Channels by using Artificial Signals. Acustica. - 1971. –Vol. 25. - P. 355–367.

18. Steeneken H. J. M., Houtgast T. A physical method for measuring speech-transmission quality. // J.Acoust. Soc. Am. - 1980. - Vol. 67. – P. 318-326.

19. Гуткин Л.С. Теория оптимальных методов радиоприема при флуктуационных помехах. - М.: Сов.радио. – 1972. – 448 с.

20. Ватолин Д, Ратушняк А., Смирнов М., Юткин В. Методы сжатия данных. – М.: Диалог-МИФИ. – 2002. – 381 с.

21. Покровский Н.Б. Расчет и измерение разборчивости речи. – М.: Связьиздат, 1962. – 390 с.

22. Сапожков М. А. Речевой сигнал в кибернетике и связи. – М.: Связьиздат, 1963. – 472 с.

23. Быков Ю. С. Теория разборчивости речи и повышение эффективности радиотелефонной связи. – М.-Л.: Госэнергоиздат, 1959. – 350 с.

24. Калинцев Ю. К. Разборчивость речи в цифровых вокодерах. – М.: Радио и связь, 1991. – 219 с.

25. Гавриленко А. В., Дидковский В. С., Продеус А. Н. Сравнительный анализ некоторых методов оценки разборчивости речи. // Сб.трудов Акустического симпозиума "Консонанс-2007", К., 2007. С.273-278.

26. Гавриленко А. В., Дидковский В. С., Продеус А. Н. Сопоставление версий формантного метода оценки разборчивости речи. // Электроника и связь, тематический выпуск "Проблемы электроники", ч.1, 2008. С.227-231.

27. Гавриленко А. В., Дидковский В. С., Продеус А. Н., Самойленко Н. А. Формантный метод оценки разборчивости речи // Электроника и связь, №5, 2007. С.63-70.

28. Рашевский Я.И., Каргашин В.Л. Обзор зарубежных методов определения разборчивости речи // Специальная техника, №№ 3-6 за 2002 год, № 1 за 2003 год (http://www.detektor.ru/publication/d2003_1.htm).

29. Железняк В. К., Макаров Ю. К., Хорев А. А. Некоторые методические подходы к оценке эффективности защиты речевой информации // Специальная техника. – М.: 2000, № 4. – С. 39-45.

30. Макаров Ю. К., Хорев А. А. К оценке эффективности защиты акустической (речевой) информации // Специальная техника, 2000, № 5, с. 46 – 56.

31. Продеус А. Н. Об измерениях вероятностных характеристик формант украинской и русской речи // Сб.трудов Акустического симпозиума "Консонанс-2009", К., 2009. С.285-291.

32. Дидковский В. С., Продеус А. Н. Сопоставление формантных свойств украинской и русской речи // Электроника и связь, тематический выпуск "Электроника и нанотехнологии ", Ч.2, 2009. №4-5. - С.88-94.

33. Collard J. A. Theoretical Study of the Articulation and Intelligibility of a Telephone Circuit // Electrical Communication. – 1929. -- Vol.7. – P. 168.

34. Harasiuk A. O., Myronov M. V., Lozinsky V. V., Thanh Vy N., Darchuk A. V., Prodeus A. M. Прогнозоване оцінювання розбірливості мови, замаскованої шумовою завадою // Мікросистеми, Електроніка та Акустика, 24(5), 48–55, 2019. https://doi.org/10.20535/2523-4455.2019.24.5.192893

35. Kryter K. D., Ball J. H. "SCIM - A meter for measuring the performance of speech communicationsystems," Techn. Doc. report No. ESD-TDR-64-674, 1964.

36. Rhebergena K. "A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. 117 (4), Pt. 1, April 2005, p.2181-2192.

37. Алдошина И., Приттс Р. Музыкальная акустика. – С-Пб.: Композитор, Санкт-Петербург, 2006. - 717 с.

38. Продеус А. Н., Дронжевская Л. Б., Климков В. А., Шагитова Д. А. Формантный и формантно-модуляционный методы оценки разборчивости речи. Часть 2. Точность и скорость измерений. // Электроника и связь. – 2011. – Т. 64. – №6. – С. 16–24.

39. Дидковский В. С., Дидковская М. В., Продеус А. Н. Акустическая экспертиза каналов речевой коммуникации. Монография. – К., «Имекс-ЛТД», 2008. – 420 с.

40. ДСН 3.3.6.037-99 Державні санітарні норми. Санітарні норми виробничого шуму, ультразвуку та інфразвуку, Київ, 1999.

41. Prodeus A., Bukhta K., Morozko P., Serhiienko O., Kotvytskyi I., Dvornyk O. "Automated Subjective Assessment of Speech Intelligibility in Various Listening Modes," ISSN 2523-4447. Microsystems, Electronics and Acoustics, 2018, vol. 23, no. 3, pp.49-57. DOI: https://doi.org/10.20535/2523-4455.2018.23.3.130367

42. ISO 24504:2014 Ergonomics — Accessible design — Sound pressure levels of spoken announcements for products and public address systems

43. ISO 14257:2001. Acoustics. Measurement and parametric description of spatial sound distribution curves in workrooms for evaluation of their acoustical performance

44. ДСН 3.3.6.037-99 Державні санітарні норми. Санітарні норми виробничого шуму, ультразвуку та інфразвуку, Київ, 1999.

45. Steeneken Herman J. M. Standardisation of performance criteria and assessment methods for speech communication. Online available: http://www.steeneken.nl/wp-content/uploads/2014/04/Standardization_STI_2014-V1.pdf

46. https://developer.android.com/studio

47. https://en.bignox.com/

48. Bradley J. S. Optimizing the decay range in room acoustics measurements using maximum-length-sequence techniques. Journal of the Audio Engineering Society, 44, (4), pp. 266-273, April 01, 1996.

49. Steeneken H. "Forty years of speech intelligibility assessment (and some history)," Proceedings of the Institute of Acoustics, Vol. 36. Pt.3 2014. https://www.researchgate.net/publication/281715024_Forty_years_of_speech_intelligibility_assessment_and_some_history

50. Продеус А., Дідковська М., Дворник О., Моторнюк Д. Апаратно-програмний комплекс «Штучна голова». Частина 1. Коригування частотної характеристики тракту. Мікросистеми, електроніка та акустика, 2020, vol. 22, no. 1, c.56-64, DOI: 10.20535/2523-4455.mea.198431.

51. Захист акустичної інформації: комп'ютерний практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 171 «Електроніка» / КПІ ім. Ігоря Сікорського ; уклад. А. М. Продеус. – Електронні текстові дані (1 файл: 2,8 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2021. – 101 с. –

52. Prodeus A., Didkovska M. "Assessment of speech intelligibility in university lecture rooms of different sizes using objective and subjective methods" Eastern-European Journal of Enterprise Technologies, 3(5) (111), pp. 47–56, 2021. DOI: https://doi.org/10.15587/1729-4061.2021.228405

53. Перший трансатлантичний кабель. http://hi-news.pp.ua/tehnka-tehnologyi/10516-pershiy-transatlantichniy-kabel.html

54. Захист акустичної інформації: комп'ютерний практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 171 «Електроніка» / КПІ ім. Ігоря Сікорського ; уклад. А. М. Продеус. – Електронні текстові дані (1 файл: 2,8 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2021. – 101 с.

ДОДАТОК А

КАЛЬКУЛЯТОР РОЗБІРЛИВОСТІ МОВЛЕННЯ В СЕРЕДОВИЩІ ANDROID STUDIO

```
package com.example.myapplication;
import androidx.appcompat.app.AppCompatActivity;
import android.annotation.SuppressLint;
import android.graphics.Color;
import android.os.Bundle;
import android.view.View;
import android.widget.EditText;
import android.widget.RadioButton;
import android.widget.RadioGroup;
import android.widget.TextView;
public class MainActivity extends AppCompatActivity {
private TextView textView3;
private EditText editTextNumberDecimal;
private RadioGroup radioButtonGroup;
@Override
protected void onCreate(Bundle savedInstanceState) {
super.onCreate(savedInstanceState);
setContentView(R.layout.activity_main);
editTextNumberDecimal  = (EditText) findViewById(R.id.editTextNumberDecimal);
textView3 = (TextView) findViewById(R.id.textView3);
radioButtonGroup = (RadioGroup) findViewById(R.id.radioGroup);
}
@SuppressLint("ResourceType")
public void onButtonClick(View view)
{
float Dsk[] = new float[]{0.146f, 2.846f, 3.642f, 1f, 0.361f, 0.181f, 0.096f};
float Dnk[] = new float[]{5.193f, 2.017f, 1.284f, 1f, 1.023f, 1.273f, 1.5471f};
float alfa[] = new float[]{0.085f, 0.127f, 0.230f, 0.233f, 0.309f, 0.224f, 0.173f};
float beta[] = new float[]{0.085f, 0.078f, 0.065f, 0.011f, 0.047f, 0.095f};
float Ls;
float r = 0;
int Ls0 = 70;
int r0 = 1;
String name = editTextNumberDecimal.getText().toString();
if(!name.isEmpty()) {
r = Float.parseFloat(name);
if(r < 1) {
editTextNumberDecimal.setTextColor(Color.RED);
return;
}
else
editTextNumberDecimal.setTextColor(Color.BLACK);
}
else
return;
int Lnm[] = new int[]{50, 60, 60, 70};
int radioButtonID = radioButtonGroup.getCheckedRadioButtonId();
```

```
View radioButton = radioButtonGroup.findViewById(radioButtonID);
int p = 1;
if(radioButtonGroup.getCheckedRadioButtonId() == findViewById(R.id.radioButton).getId())
p = 1;
else if(radioButtonGroup.getCheckedRadioButtonId() ==
findViewById(R.id.radioButton2).getId())
p = 2;
else if(radioButtonGroup.getCheckedRadioButtonId() ==
findViewById(R.id.radioButton3).getId())
p = 3;
else if(radioButtonGroup.getCheckedRadioButtonId() ==
findViewById(R.id.radioButton4).getId())
p = 4;
if(p < 3)  {
Ls = (float) (Ls0 + (20 * Math.log10(r0 / r)));
}
else  {
if(r < 3)
Ls = (float) (Ls0 + (20 * Math.log10(r0 / r)));
else
Ls = (float) (Ls0 + (20 * Math.log10(r0 / r)) + (10 * Math.log10(1 + Math.pow(r,2) /
(Math.pow(r,2) + 9))));
}
int Ln = Lnm[p-1];
float SNR0 = Ls - Ln;
float SNR, SDsk = 0, SDnk = 0, I1 = 0, I2 = 0, STI;
float Ek[] = new float[7];
float T[] = new float[7];
for(int i=0; i<7; i++) {
SDsk += Dsk[i];
}
for(int i=0; i<7; i++) {
SDnk += Dnk[i];
}
SNR = (float) (10 * Math.log10(SDsk/SDnk));
for(int k=0; k<7; k++) {
Ek[k] = (float) (10 * Math.log10(Dsk[k]/Dnk[k]) + SNR0 - SNR);
}
for(int k=0; k<7; k++) {
if (Ek[k] < (-15))
T[k] = 0;
else if (Ek[k] >= (-15) && Ek[k] <=15)
T[k] = (Ek[k] + 15) / 30;
else
T[k] = 1;
}
for(int k=0; k<7; k++) {
I1 += alfa[k] * T[k];
}
for(int k=0; k<6; k++) {
I2 += beta[k] * Math.sqrt(T[k] * T[k+1]);
}
STI = I1 - I2;
```

```
int _R = 0;
if(STI > 0.75)
_R = 5;
else if (STI <= 0.75 && STI > 0.6)
_R = 4;
else if (STI <= 0.6  && STI > 0.45)
_R = 3;
else if (STI <= 0.45  && STI > 0.3)
_R = 2;
else if (STI <= 0.3)
_R = 1;
TextView tv3 = (TextView)findViewById(R.id.textView3);
TextView tv4 = (TextView)findViewById(R.id.textView4);
tv3.setText(new StringBuilder().append("Speech intelligibility: ").append(_R).toString());
tv4.setText(new StringBuilder().append("STI Index: ").append(STI).toString());
}
```

ДОДАТОК Б

ПРОГРАМИ ДЛЯ ПОБУДОВИ ГРАФІКІВ ОЦІНОК РОЗБІРЛИВОСТІ МОВЛЕННЯ

1. Програма побудови графіків оцінок розбірливості мовлення

```
x= [1 2 3 4 5 6 7 8] %points in the lecture room
y= [0.9685 0.9680 0.9559 0.9225 0.9642 0.9653 0.9536 0.9583] %ball
,1 channel ,Pokrovsky
z= [0.9723 0.9688 0.9656 0.9238 0.9617 0.9627 0.9578 0.9618] %ball
,2 channel ,Pokrovsky
q= [0.9525 0.9526 0.9307 0.8607 0.9456 0.9476 0.9263 0.9361] %ball
,1 channel ,Sapozhkov
w= [0.9590 0.9535 0.9487 0.8597 0.9401 0.9436 0.9337 0.9425] %ball
,2 channel ,Sapozhkov
e= [0.9739 0.9740 0.9579 0.9037 0.9678 0.9698 0.9537 0.9615] %ball
,1 channel ,perception coefficients
r= [0.9780 0.9742 0.9705 0.9011 0.9650 0.9667 0.9611 0.9658] %ball
,2 channel ,perception coefficients
t= [0.6367 0.6418 0.5967 0.5115 0.6274 0.6259 0.5876 0.6113] %ball
,1 channel ,STI
u= [0.6525 0.6425 0.6332 0.5064 0.6074 0.6229 0.5992 0.6267] %ball
,2 channel ,STI
i= [0.9720 0.9659 0.9199 0.9640 0.9576 0.9661 0.9588 0.9584] %clap
,1 channel ,Pokrovsky
o= [0.9738 0.9678 0.9295 0.9672 0.9570 0.9625 0.9592 0.9565] %clap
,2 channel ,Pokrovsky
p= [0.9594 0.9477 0.8519 0.9443 0.9325 0.9472 0.9335 0.9323] %clap
,1 channel ,Sapozhkov
a= [0.9626 0.9513 0.8703 0.9498 0.9324 0.9405 0.9332 0.9290] %clap
,2 channel ,Sapozhkov
s= [0.9779 0.9699 0.8847 0.9659 0.9589 0.9688 0.9577 0.9556] %clap
,1 channel ,perception coefficients
d= [0.9789 0.9724 0.9073 0.9712 0.9591 0.9624 0.9584 0.9533] %clap
,2 channel ,perception coefficients
f= [0.6597 0.6239 0.5100 0.6050 0.5904 0.6171 0.5891 0.5895] %clap
,1 channel ,STI
h= [0.6775 0.6315 0.5221 0.6217 0.5931 0.5991 0.5851 0.5849] %clap
,2 channel ,STI
j= [0.9692 0.9639 0.9568 0.9654 0.9610 0.9586 0.9557 0.9557] %mls
,1 channel ,Pokrovsky
k= [0.9736 0.9668 0.9594 0.9655 0.9619 0.9630 0.9646 0.9599] %mls
,2 channel ,Pokrovsky
l= [0.9537 0.9435 0.9295 0.9472 0.9390 0.9340 0.9278 0.9290] %mls
,1 channel ,Sapozhkov
M= [0.9616 0.9495 0.9352 0.9472 0.9405 0.9427 0.9459 0.9375] %mls
,2 channel ,Sapozhkov
b= [0.9750 0.9672 0.9556 0.9693 0.9635 0.9590 0.9541 0.9553] %mls
,1 channel ,perception coefficients
c= [0.9796 0.9715 0.9597 0.9697 0.9634 0.9661 0.9693 0.9624] %mls
,2 channel ,perception coefficients
v= [0.6397 0.6083 0.5797 0.6256 0.6026 0.5923 0.5827 0.5923] %mls
,1 channel ,STI
```

```
N= [0.6593 0.6273 0.5883 0.6227 0.6037 0.6130 0.6240 0.6061] %mls
,2 channel ,STI
U= [0.9736 0.9669 0.9603 0.9680 0.9609 0.9619 0.9586 0.9594]
%sweep ,1 channel ,Pokrovsky
I= [0.9744 0.9706 0.9652 0.9669 0.9609 0.9666 0.9666 0.9642]
%sweep ,2 channel ,Pokrovsky
O= [0.9614 0.9495 0.9370 0.9518 0.9388 0.9402 0.9337 0.9365]
%sweep ,1 channel ,Sapozhkov
P= [0.9632 0.9568 0.9468 0.9498 0.9388 0.9489 0.9495 0.9456]
%sweep ,2 channel ,Sapozhkov
A= [0.9797 0.9709 0.9623 0.9727 0.9618 0.9633 0.9587 0.9617]
%sweep ,1 channel ,perception coefficients
S= [0.9807 0.9762 0.9685 0.9703 0.9618 0.9705 0.9717 0.9684]
%sweep ,2 channel ,perception coefficients
D= [0.6663 0.6300 0.6060 0.6356 0.6249 0.6055 0.5933 0.6096]
%sweep ,1 channel ,STI
F= [0.6750 0.6541 0.6238 0.6262 0.6249 0.6281 0.6318 0.6253]
%sweep ,2 channel ,STI
plot(x,y,'b--o')
hold on
plot(x,z, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('ball, Pokrovsky')
xlabel('Point number')
ylabel('Pokrovskys assessment of language intelligibility')
figure
%%%%%
plot(x,q, 'b--o')
hold on
plot(x,w, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('ball, Sapozhkov')
xlabel('Point number')
ylabel('Assessment of language intelligibility according to
Sapozhkov')
figure
%%%%%
plot(x,e, 'b--o')
hold on
plot(x,r, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('ball, perception coefficients')
xlabel('Point number')
ylabel('Assessment by perception coefficients')
figure
%%%%%
plot(x,t, 'b--o')
hold on
```

```
plot(x,u, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('ball, STI')
xlabel('Point number')
ylabel('STI')
figure
%%%%%
plot(x,i, 'b--o')
hold on
plot(x,o, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('clap, Pokrovsky')
xlabel('Point number')
ylabel('Pokrovskys assessment of language intelligibility')
figure
%%%%%
plot(x,p, 'b--o')
hold on
plot(x,a, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('clap, Sapozhkov')
xlabel('Point number')
ylabel('Assessment of language intelligibility according to
Sapozhkov')
figure
%%%%%
plot(x,s, 'b--o')
hold on
plot(x,d, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('clap, perception coefficients')
xlabel('Point number')
ylabel('Assessment by perception coefficients')
figure
%%%%%
plot(x,f, 'b--o')
hold on
plot(x,h, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('clap, STI')
xlabel('Point number')
ylabel('STI')
figure
%%%%%
```

```
plot(x,j, 'b--o')
hold on
plot(x,k, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('mls, Pokrovsky')
xlabel('Point number')
ylabel('Pokrovskys assessment of language intelligibility')
figure
%%%%%
plot(x,l, 'b--o')
hold on
plot(x,M, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('mls, Sapozhkov')
xlabel('Point number')
ylabel('Assessment of language intelligibility according to
Sapozhkov')
figure
%%%%%
plot(x,b, 'b--o')
hold on
plot(x,c, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('mls, perception coefficients')
xlabel('Point number')
ylabel('Assessment by perception coefficients')
figure
%%%%%
plot(x,v, 'b--o')
hold on
plot(x,N, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('mls, STI')
xlabel('Point number')
ylabel('STI')
figure
%%%%%
plot(x,U, 'b--o')
hold on
plot(x,I, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('sweep, Pokrovsky')
xlabel('Point number')
ylabel('Pokrovskys assessment of language intelligibility')
```

```
figure
%%%%%
plot(x,O, 'b--o')
hold on
plot(x,P, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('sweep, Sapozhkov')
xlabel('Point number')
ylabel('Assessment of language intelligibility according to
Sapozhkov')
figure
%%%%%
plot(x,A, 'b--o')
hold on
plot(x,S, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('sweep, Assessment by perception coefficients')
xlabel('Point number')
ylabel('Assessment by perception coefficients')
figure
%%%%%
plot(x,D, 'b--o')
hold on
plot(x,F, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('sweep, STI')
xlabel('Point number')
ylabel('STI')
```

2.Програма для побудови графіків D

```
x= [1 2 3 4 5 6 7 8] %points in the lecture room
q= [39.9718 39.1316 34.4456 34.3630 33.8845 32.9037 35.4031
32.1481] %ball, left channel
w= [35.3120 32.5550 34.1948 27.1673 32.6684 30.6362 29.4552
26.6779]%clap, left channel
e= [34.7384 33.1897 29.2600 24.9343 28.3109 25.0106 27.8443
22.9780]%mls, left channel
r= [37.8027 35.8040 21.6753 35.2229 34.5817 25.9570 21.0297
19.3452]%sweep, left channel
t= [36.0505 37.2455 33.5881 33.2344 34.2414 35.5011 32.9157
35.8428]%ball, right channel
u= [33.7684 34.6392 40.2027 31.2835 33.5790 40.2970 33.1189
29.5255]%clap, right channel
i= [35.5173 31.6031 28.2600 26.0034 28.1718 30.4620 28.2983
26.7822]%mls, right channel
```

```
o= [41.1397 38.2369 26.3627 38.0457 34.5904 32.9383 27.4416
24.1517]%sweep, right channel
plot(x,q, 'b--o')
hold on
plot(x,t, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('ball, D')
xlabel('Point number')
ylabel('D')
figure
%%%%%%%%
plot(x,w, 'b--o')
hold on
plot(x,u, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('clap, D')
xlabel('Point number')
ylabel('D')
figure
%%%%%%%%
plot(x,e, 'b--o')
hold on
plot(x,i, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('mls, D')
xlabel('Point number')
ylabel('D')
figure
%%%%%%%%
plot(x,r, 'b--o')
hold on
plot(x,o, 'r-o')
xlim([0 8])
grid on
legend('left channel','right channel')
title('sweep, D')
xlabel('Point number')
ylabel('D')
```

3.Програма для обчислення оцінок розбірливості мовлення

```
function Ws = main_reverb_razb_2chan_final(rirname,T1,avtr,bnd)
% Реальне вимірювання розбірливості мови через IX
%
% ==== вхідні параметри: =====
% rirname: ім’я масиву, що аналізується, IX, якщо live = 0
% T1 - тривалість фрагменту IX (с);
```

```
% = Розрахунки: ===
% -- Частина 1: підготовка до розрахунків
% - 1) задання внутрішніх параметрів
%live=0;
%T=4; % T - тривалість записи (с);
T1=0.2;
avtr = 4;
bnd = 7;
Ws= zeros(1,2);
%- 2) зчитування коефіцієнтів гребінки фільтрів
%load  fltrbnk7_ struct;
%a=fltrbnk7_struct.a;
%len_a=fltrbnk7_struct.len_a;
% - 4) зчитування з диску
[x,Fs]=audioread('8_sweep_rir.wav');
if Fs=44100 % програма октавної фільтрації розрахована на Fs =
22050 Гц
x1 = resample(x,22050,44100);
x = x1;
end
for chan = 1:2
h=x(:,chan)/max(abs(x(:,chan))); % нормування запису по максимуму

% - Частина 2: виділення потрібної інформації з ІХ
habs=abs(h); % модуль ІХ
maxhabs=max(habs); % максимум модулю ІХ
hpor05=sign(abs(h) 0.5*maxhabs)+1; % знак (максимум модулю без
врахування порогового значення 0,5)+1
n0=find(hpor05,1,'first');% знаходимо 1-й ненульовий елемент n0
hc=h(n0:n0+round(T1*Fs)); % вирізаємо T1 с, починаючи з n0
tc=0:1/Fs:(length(hc)-1)/Fs; % масив моментів часу
figure;plot(tc,hc);% графік
xlabel('Час,с');ylabel('Рівень');
grid on ;
disp('для продовження натисніть довільну кнопку');
pause;
% Частина 2: виділення потрібної інформації з ІХ
%hc=h(1:round(T*Fs)); % вирізаємо T1 с, починаючи з n0
% фільтрація запису
%xband7=filter_band_7_func(hc,a,[1 1 1 1 1 1 1]); % фільтрація
гребінкою
xband7=filter_band_7_ideal_func(hc,[1 1 1 1 1 1 1],Fs); %
фільтрація ідеальна
%xband7=xband7(len_a:end,:); % виділяємо неправильні виборки
% квадрат  ІХ в кожній смузі
xband7=xband7.^2;
%  ПФ від квадрата ІХ в кожній смузі
Nfft=size(xband7,1);
ft7=ft(xband7,Nfft);
% розрахунок mk(f)=2*sum(xband7)
mkf=zeros(size(ft7,1),size(ft7,2));
%mkf=2*abs(ft7(:,:))./abs(ft7(1,:));
for k=1:7
mkf(:,k)=abs(ft7(:,k))./abs(ft7(1,k));
```

```
end
% -розрахунок snrk(f)
snrkf = 10*log10(mkf./(1-mkf));

% --- Частина 3: вимірювання Ek
nf=round(14*T1); % 14 Hz
Ek=sum(snrkf(2:nf,:),1)/(nf-1); % для 7 октавних смуг
if bnd=5
Ek=Ek(2:6);
end
% Частина 4: Оцінювання розбірливості мови формантним методом
% 1) розрахунок словесної розбірливості за Покровським
if avtr= 1
if bnd=7
W= azbor_pokr(Ek); % 7 смуг частот
else
W=razb1(Ek); % 5 смуг частот
end
disp(['W=' num2str(W)'; Покровський;' num2str(bnd) 'октавна
смуга'])
end
%- 2) розрахунок словесної розбірливості за Сапожковим
if avtr = 2
if bnd= 7
W=razb_sapoz_7(Ek);   % 7 полос частот
else
W=razb_sapoz_5(Ek);   % 5 полос частот
end
disp(['W=' num2str(W) ';Сапожков; ' num2str(bnd) ' октавна
смуга'])
end
%  3) розрахунок словесної розбірливості за запропонованими
коефіцієнтами сприйняття if avtr==3
if bnd == 7
W=razb_we_short(Ek);
else
W=razb_we_5(Ek);
end
disp(['W = ' num2str(W) '; Запропонована версія;' num2str(bnd)
'октавна смуга'])
end
% -- 10) вычисление STI ---
if avtr == 4
if bnd == 7
W = razb_sti(Ek);
else
W = 0;
end
disp(['sti=' num2str(W)'; 'num2str(bnd) ' октавн. полос'])
end
Ws(1,chan)=W;
End
```

4. Програма для обчислення значень D для сигналів sweep і mls

```
% === sweep_mls_D_Prod.m (rir_crosscorel.m) ===
% програма обчислення імпульсної характеристики приміщення шляхом
% обчислення взаємно-кореляційної функції запису x та еталону y
clear;clc;close all

% --- вибір типу сигналу ---
num_type = input('Вкажіть тип сигналу (1 - mls;  2 - sweep):   ',
's');
if strcmp(num_type,'1')
    type_sig = 'mls';
else
    type_sig = 'sweep';
end


% --- відображення та запис результатів ---
num_zapys = input('Записувати результати на диск? (0 - ні;  1 -
так):   ', 's');

num = 8;
for n = 1:num
    filename_x = [num2str(n) '_' type_sig '.wav'];
    if strcmp(type_sig,'mls')
        filename_y = 'mls_1p5s_44kHz.wav';
    else
        filename_y = 'sweep_4s_65Hz-18kHz_44kHz.wav';
    end
[x,Fs] = audioread(filename_x);
y = audioread(filename_y);
len_y = length(y);
zer_x = zeros(len_y,2);
x = [x; zer_x];

[rir1,lags] = xcorr(x(:,1),y);
rir2 = xcorr(x(:,2),y);
rir = [rir1 rir2];
    maxrir = max(max(abs(rir)));
    rir = rir/maxrir;
t=lags/Fs;

% --- Графіки ---
figure; plot(t,rir(:,1));
title(['Point ' num2str(n) ', ' type_sig]);
xlabel('Time (s)'); ylabel('Level');
legend('left')
ax = gca; ax.FontSize =14;
grid on

figure; plot(t,rir(:,2), 'r');
title(['Point ' num2str(n) ', ' type_sig]);
xlabel('Time (s)'); ylabel('Level');
legend('right')
```

```
ax = gca; ax.FontSize =14;

figure; plot(t,20*log10(abs(rir(:,1))));
ylim([-50,0]); grid on
title(['Point ' num2str(n) ', ' type_sig]);
xlabel('Time, s'); ylabel('Level, dB');
legend('left')
ax = gca; ax.FontSize =14;

figure; plot(t,20*log10(abs(rir(:,2))), 'r');
ylim([-50,0]); grid on
title(['Point ' num2str(n) ', ' type_sig]);
xlabel('Time, s'); ylabel('Level, dB');
legend('right')
ax = gca; ax.FontSize =14;

% === обчислення D ===
rir=abs(rir);
[Ml,Il]=max(rir(:,1));   % === Il - максимум імп.характеристики в
1-му каналі
D(n,1)=20*log10(abs(rir(Il,1)))-
20*log10(abs(max(rir((Il+Fs/2):(Il+1.5*Fs),1))));
[Ml,Il]=max(rir(:,2));   % === Il - максимум імп.характеристики в
2-му каналі
D(n,2)=20*log10(abs(rir(Il,2)))-
20*log10(abs(max(rir((Il+Fs/2):(Il+1.5*Fs),2))));
if num_zapys == 1
    audiowrite([num2str(n) '_' type_sig '_rir.wav'],rir,Fs);
end
%
end
disp(D)
```

5. Програма для обчислення значень D для сигналів ball і clap

```
clear all
clc
close all
num_type = input('Вкажіть тип сигналу (1 - ball;  2 - clap):   ',
's');
if strcmp(num_type,'1')
    type_sig = 'ball';
else
    type_sig = 'clap';
end

num = 8;
for n = 1:8

    filename_x = [num2str(n) '_' type_sig '.wav'];
    [rir,Fs] = audioread(filename_x);

% === обчислення D ===
rir=abs(rir);
```

```
[Ml,Il]=max(rir(:,1));    % === Il - максимум імп.характеристики в
1-му каналі
D(n,1)=20*log10(abs(rir(Il,1)))-
20*log10(abs(max(rir((Il+Fs/2):(Il+1.5*Fs),1))));
[Ml,Il]=max(rir(:,2));    % === Il - максимум імп.характеристики в
2-му каналі
D(n,2)=20*log10(abs(rir(Il,2)))-
20*log10(abs(max(rir((Il+Fs/2):(Il+1.5*Fs),2))));
end
disp(D)
```

6. Програма для побудови імпульсних характеристик сигналів mls і sweep

```
% === rir_crosscorel.m ===
% програма обчислення імпульсної характеристики приміщення шляхом
% обчислення взаємно-кореляційної функції запису x та еталону y
clear;clc;close all

% --- вибір типу сигналу ---
num_type = input('Вкажіть тип сигналу (1 - mls;  2 - sweep):   ',
's');
if strcmp(num_type,'1')
    type_sig = 'mls';
else
    type_sig = 'sweep';
end


% --- відображення та запис результатів ---
num_zapys = input('Записувати результати на диск? (0 - ні;  1 -
так):   ', 's');

num = 8;
for n = 1:num
    filename_x = [num2str(n) '_' type_sig '.wav'];
    if strcmp(type_sig,'mls')
        filename_y = 'mls_1p5s_44kHz.wav';
    else
        filename_y = 'sweep_4s_65Hz-18kHz_44kHz.wav';
    end
[x,Fs] = audioread(filename_x);
y = audioread(filename_y);
len_y = length(y);
zer_x = zeros(len_y,2);
x = [x; zer_x];

[rir1,lags] = xcorr(x(:,1),y);
rir2 = xcorr(x(:,2),y);
rir = [rir1 rir2];
    maxrir = max(max(abs(rir)));
    rir = rir/maxrir;
t=lags/Fs;

% --- Графіки ---
```

```
figure; plot(t,rir(:,1));
title(['Point ' num2str(n) ', ' type_sig]);
xlabel('Time (s)'); ylabel('Level');
legend('left')
ax = gca; ax.FontSize =14;
grid on

figure; plot(t,rir(:,2), 'r');
title(['Point ' num2str(n) ', ' type_sig]);
xlabel('Time (s)'); ylabel('Level');
legend('right')
ax = gca; ax.FontSize =14;

figure; plot(t,20*log10(abs(rir(:,1))));
ylim([-50,0]); grid on
title(['Point ' num2str(n) ', ' type_sig]);
xlabel('Time, s'); ylabel('Level, dB');
legend('left')
ax = gca; ax.FontSize =14;

figure; plot(t,20*log10(abs(rir(:,2))), 'r');
ylim([-50,0]); grid on
title(['Point ' num2str(n) ', ' type_sig]);
xlabel('Time, s'); ylabel('Level, dB');
legend('right')
ax = gca; ax.FontSize =14;

if num_zapys == 1
    audiowrite([num2str(n) '_' type_sig '_rir.wav'],rir,Fs);
end
%
end
```