УДК 534.8 DOI https://doi.org/10.32782/1995-0519.2023.3.18

ОСОБЛИВОСТІ ДЕРЕВЕРБЕРАЦІЇ МОВНИХ СИГНАЛІВ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ

Гліб Борисов
аспірант кафедри акустичних та мультимедійних електронних систем
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», просп. Берестейський, 37, Київ, Україна, 03056, borusov5364@gmail.com
ORCID: 0000-0003-2780-2700

Кирило Трапезон
кандидат технічних наук, доцент,
доцент кафедри акустичних та мультимедійних електронних систем
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», просп. Берестейський, 37, Київ, Україна, 03056, kirill.trapezon@gmail.com
ORCID: 0000-0001-5873-9519

Запис акустичної інформації з використанням «живої» мови не завжди можна реалізувати в спеціально обладнаному приміщенні з врахуванням усіх вимог архітектурної акустики. В цьому випадку на якість записаного аудіоконтенту впливає і реверберація приміщення, і оточуючі сторонні фонові шуми. Для проведення деверберації мовних сигналів та вирішення задачі зменшення адитивного шуму різної природи пропонується нейронна мережа згорткового типу, яка побудована на основі архітектури U-Net. В роботі наведено алгоритм використання нейронної мережі та проведено практичний експеримент з перевірки ефективності використання цієї мережі для вирішення задач деверберації та боротьби з шумом, який було додано при записі мовних сигналів. Показано, що значний вплив на ефективність роботи нейронної мережі відіграє кількість згорткових шарів та час, який виділено на навчання цієї мережі. Навчання і тренування нейронної мережі виконано на основі інструментів та засобів візуалізації програмного середовища Matlab. В якості основного алгоритму навчання для нейронної мережі обрано метод Адама (метод стохастичної оптимізації). З'ясовано, що задля досягнення більшої точності роботи нейронної мережі на деяких часових інтервалах в області початку впливу адитивного білого шуму (в часовому інтервалі від 0,7 сек до 1 сек) необхідно або провести додаткове навчання нейронної мережі, або внести структурні зміни у функціонування цієї мережі, тобто збільшити розмірність ядра згортки та рецепторне поле нейрону. Наведені практичні рекомендації з використання нейронної мережі можуть бути розповсюджені і на вирішення інших прикладних задач машинного навчання, а саме при організації дистанційного спілкування в конференц-залах, при проектуванні удосконалених слухових апаратів, а також в системах розпізнавання мовних сигналів та системах виявлення голосової активності у додатках та пристроях Інтернету речей.

Ключові слова: деверберація, шум, сигнал, акустика, нейронна мережа, приміщення, алгоритм, ефективність.

Вступ, актуальність роботи. Критерії забезпечення якісного звучання та високого рівня розбірливості слів обов'язково враховуються при розробленні слухових апаратів [1–3], функціонуванні інтерактивних голосових помічників, організації та налагодженні систем розумного будинку в рамках концепції Інтернету речей. В останньому випадку мікрофони використовуються як елементи бездротової акустичної сенсорної мережі. При цьому якісне та чітке передавання сигналів команд через мікрофони має бути забезпечено при керуванні ключових систем розумного будинку [1; 4]. Інший аспект, пов'язаний з оцінкою якості мовних сигналів, можна виділити при розгляді систем запису аудіо інформації в приміщенні. В цьому випадку необхідно враховувати реверберацію приміщення та оточуючі фонові шуми. Ці фактори особливо впливають на якість аудіозаписів, коли основним елементом в них є мовна інформація. В літературі відомі різні підходи та способи, які дозволяють оцінити якість записаної аудіоінформації. Наприклад, в [5] задля покращення розбірливості мовного запису автори пропонують використовувати спеціальні дифузійні генеративні моделі або ймовірнісні моделі, на основі статистичних фреймворків. При цьому в роботі не визначено обмеження до застосування цих моделей на практиці та не повністю наведено параметри самого мовного сигналу. В статті [6] для відокремлення мовних складових сигналу від шуму розглянуто підходи машинного навчання та описано стандартні алгоритми деверберації. Робота [7] містить цікавий підхід до розв'язання задачі зменшення шуму в записаних мовних сиг-

налах. Так, в дослідженні аналізуються особливості людської мови, а точніше наголошується, що між кожним словом розмови є коротка пауза. При записі ці паузи створюють серію часових періодів, в інтервалі яких присутній лише чистий шум. За цими інтервалами тиші проведено аналіз динаміки шуму, і на основі методу спектрального віднімання та підходів машинного навчання приймається рішення щодо зменшення рівня цього шуму. Натомість недоліком такого підходу можна вважати, що він не зовсім ефективний, якщо при розмові присутня фонова музика.

Для вирішення задачі видалення шуму та деревербераці записаних мовних сигналів в роботі [8] пропонується метод, який об'єднує локальну зважену регресію та зважені помилки прогнозування. Разом з тим цей метод не апробовано для ситуації, коли при записі мовних сигналів є стороннє джерело шуму різної природи. В дослідженні [9] для оцінки якості деревербераці мовних сигналів пропонується практичний метод на основі статистичної моделі.

Ряд наукових публікацій пов'язаний з використанням нейронних мереж [2] як основного інструменту для вирішення задач деревербераці та зменшення шуму. Так, в роботі [10] для видалення спотворень, викликаних впливом фонового оточуючого шуму та реверберації приміщення, пропонується використати глибоку нейронну мережу. При цьому використання мережі передбачається на двох окремих етапах – етап зменшення шуму та етап деревербераці. В дослідженні [11] зазначається, що нейронна мережа широко використовується для автоматичного розпізнавання мовних сигналів та її продуктивність значно більша за основними показниками, аніж у випадку використання традиційних систем на основі прихованих марківських моделей. Додатковим прикладом застосування нейронних мереж, як зазначено в [12], є їх використання для процедур розпізнавання домінуючих інструментів в поліфонічний музичній композиції. Таким чином, з огляду наведеної літератури слід відмітити, що на сьогодні недостатня кількість публікацій, які дозволяють розв'язати комплексну проблему видалення шуму з різною інтенсивністю та знайти прості ефективні шляхи розв'язання задачі деревербераці при аналізі записаного мовного сигналу, використовуючи при цьому виключно інструменти та можливості нейронної мережі.

Постановка задачі, мета, задачі дослідження. При організації та проведенні запису різнопланового аудіоконтенту, насамперед «живої» мови, не завжди технічно є можливість для цього підібрати спеціальне приміщення з необхідною акустикою. Тобто у звичайних умовах на якість записаних мовних сигналів їх розбірливість може впливати як реверберація приміщення [13], так і оточуючий шум, який знаходиться в приміщенні або проникає в це приміщення. З теорії акустики приміщень відомо, що реверберація як процес багатократного відбиття сигналів від стін та предметів у приміщенні погіршує розбірливість мови, особливо для людей з порушенням слуху. Відомо багато методів та способів реалізації мовної деревербераці на основі одно- та багатомікрофонних систем [14]. Це і використання схем інверсної фільтрації [11; 15] та моделей багатокрокового лінійного прогнозування [13], залучення схем автоматичних мовних кодерів та декодерів [7], оброблення спектральної області мовного сигналу на основі методу найменших квадратів [7]. Проте головним недоліком перелічених способів та підходів є те, що вони переважно у своїй більшості не повністю використовують спектральну структуру мови, адже в спектрограмі мовного сигналу є чіткі часово-частотні закономірності. І внаслідок використання зазначених підходів ці закономірності можуть бути втрачені. Задачею деревербераці мовлення є збереження глобальної часової та спектральної інформації сигналу [14], що можливо забезпечити на основі використання нейронних мереж [10]. Ідея використання нейронних мереж для оброблення записаних мовних сигналів виникла через те, що ці мережі успішно дозволяють розв'язувати різноманітні прикладні задачі, а саме управління безпілотними транспортними засобами, пошук місця розташування об'єкта в кадрі зображення, розпізнавання голосу та машинний переклад. І всі ці застосування, так або інакше, пов'язані з обробленням зображень. В нашому ж випадку, якщо записаний мовний сигнал з реверберацією та шумом перевести в частотну форму, то на основі часово-частотного представлення маємо спектрограму, яку можна представити у формі зображення. І задачею нейронної мережі вже є аналіз та корекція цього зображення. Спектрограма дозволяє показати типові паттерни (форманти, області клацання, висоту тону сигналу, тощо) а також характерні особливості, які привносить реверберація та адитивний шум [16].

Метою статті є розроблення алгоритму на основі підходів та особливостей нейронної мережі, який дозволить ефективно та швидко

видаляти шум різної природи та проводити деревербераційю записаних мовних сигналів. Для досягнення поставленої мети сформульовано такі задачі:

– визначити вихідні дані та розробити алгоритм застосування нейронної мережі для розв'язання задач деревербераційї та зменшення шуму в записаних наперед мовних сигналах;

– розробити та провести практичний експеримент з перевірки ефективності застосування нейронної мережі з різною кількістю згорткових шарів для досягнення зменшення рівня реверберації та адитивного білого шуму на прикладі записаного тестового словосполучення тривалістю до 5 сек. Для цього попередньо провести навчання нейронної мережі.

Матеріал, результати досліджень. Реверберацію мовних сигналів можна описати на основі співвідношення [9]

$$y(t) = h(t) * s(t) + n(t) = \int_0^t x(\tau) h(t-\tau) d\tau + n(t), \quad (1)$$

де

$*$ – операція згортки функцій;

$h(t)$ – імпульсна характеристика приміщення, яке створює реверберацію;

$s(t)$ – чистий мовний сигнал;

$n(t)$ – адитивний шум, який присутній при записі з мікрофону.

Метою деревербераційї сигналу є отримання максимально наближеної оцінки сигналу $x(t)$ на основі нелінійної функції, форма та вигляд якої визначається на основі роботи нейронної мережі:

$$\hat{y}(t) \approx F\left(y(t)\right).$$

Імпульсна характеристика приміщення або імпульсна характеристика створеного акустичного каналу між джерелом сигналу та мікрофоном може бути визначена шляхом поділу на дві частини:

$h_{пряме}(t)$ – для прямого сигналу від джерела та деякі ранні відбиття;

$h_{відб}(t)$ – для сигналів з пізніми відбиттями;

$t = 0$ – час надходження прямого сигналу.

Зазначимо, що ранні відбиття мають менш негативний вплив на розбірливість мови аніж пізні відбиття. Тоді, можна записати

$$h(t) = \begin{cases} h_{пряме}(t), & 0 \le t \le T \\ h_{відб}(t), & t > T, \end{cases}$$

де T – час поділу для імпульсної характеристики приміщення.

Разом з тим формула (1) є теоретичною і на практиці її використання може супроводжуватись певними труднощами. Тому для кількісної оцінки реверберації в приміщенні можна використати енергетичний підхід [9]. На основі імпульсного відгуку можна виміряти відношення потужності сигналу, обумовленого прямою хвилею до потужності сигналу відбитого (реверберацийного):

$$SRR = \frac{E_{пряме}}{E_{відб}} = \frac{\int_0^{T_{пряме}} h^2(\tau) d\tau}{\int_{T_{пряме}}^{\infty} h^2(\tau) d\tau}, \quad (2)$$

де

$E_{пряме}, E_{відб}$ – енергія прямого та реверберацийного сигналів;

$T_{пряме}$ – час приходу прямого звуку в мікрофон. З [17] значення $T_{пряме}$ знаходиться в межах 8–16 мс. Для розрахунків приймемо $T_{пряме} = 12$ мс.

В даній статі для вирішення задач деревербераційї та зменшення шуму в записаних мовних сигналах використаємо згорткову нейронну мережу за архітектурою U-Net. З математичної точки зору нейронна мережа – спосіб розв'язку нелінійної задачі оптимізації, де вхідні дані (мовний сигнал) являють собою дискретні послідовності амплітуд звуку (семплів), які зареєстровані в певні моменти часу. Рецепторне поле нейрона моделі нейронної мережі описується кількістю пікселів зображення спектрограми і для даної задачі приймемо цю область розмірністю у 150 пікселів (15 пікселів для частотної області та 10 пікселів для часової області). Передбачається, що на вхід кожного з нейронів мережі надходить лише частина інформації для аналізу і попередньо сам мовний сигнал після семплування проходить через одностороннє швидке перетворення Фур'є, яке дозволяє розбити спектр на 512 частин. Ці частини по-елементно помножуються на ядро згортки нейронної мережі і отриманий результат підсумовується. Для реалізації ядра згортки проводиться розрахунок імпульсної характеристики приміщення. Тобто, виконується операція згортки, де на вхід подається сигнал, а на виході маємо сформовану карту ознак. Виберемо розмірність ядра згортки 5×5 і математично це є фільтр, який утворює згортковий шар нейронної мережі. На рисунку 1 наведено алгоритм оброблення записаного мовного сигналу.

Практична частина дослідження. В рамках дослідження поділимо експериментальну частину на 3 частини. Спочатку визначимо рівень реверберації обраного приміщення і проведемо

Рис. 1. Структурна схема деревербераці мовного сигналу

запис тестового словосполучення «Перший тестовий сигнал» тривалістю 5 сек. Далі проведемо запис того ж самого тестового словосполучення, але за умови, що при записі в кімнаті підключається генератор шуму з інтенсивністю 30 дБ і тривалістю не менше 2,5 сек. Генератор шуму при цьому дозволяє обрати 4 типи шуму: коричневий (червоний), рожевий (фліккер-шум), білий та сірий. На наступному етапі підключимо нейронну мережу для вирішення задачі деревербераці та зменшення шуму в записаному мовному сигналі. Причому навчання і тренування нейронної мережі будемо проводити на основі інструментів та засобів візуалізації програмного середовища Matlab. В якості вхідних даних для функціонування нейронної мережі задамо такі параметри: розмір нейронної мережі 512×512; розмір ядра згортки 5×5. Для того щоб отримати більшу точність роботи мережі при навчанні, визначимо три рівні ланок (кількість згортувальних шарів) за архітектурою – 6,8,12.

На початку експерименту визначимо приміщення, де буде проводитись запис сигналів, і воно має розміри: 2.1 м × 4.4 м × 2.5 м, площа 9,24 м^2, об'єм 23,1 м^3. Використовуючи формулу Себіна [18] і прийнявши, що коефіцієнт поглинання на частоті 4 кГц дорівнює 0,15, отримаємо, що час реверберації дорівнює 2,47 сек. Додатково для запису мовних сигналів розташуємо в приміщенні мікрофон USB BOYA BY-M100UA на відстані 50 см від джерела сигналу. Мікрофон має такі характеристики: робочий частотний діапазон – 50 Гц–18 кГц; частота дискретизації – 48 кГц. Потім підключимо програмний генератор шуму, який може створювати 4 види шуму. Схема розташування об'єктів для запису мовної інформації показана на рисунку 2. Таким чином, створено один акустичний канал для запису.

Обговорення отриманих результатів. Спочатку запустимо нейронну мережу на навчання. При цьому в якості тестових сигналів використаємо набір простих слів з різною кількістю літер.

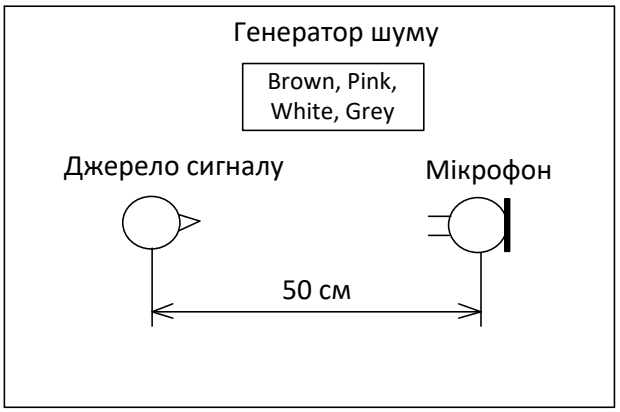
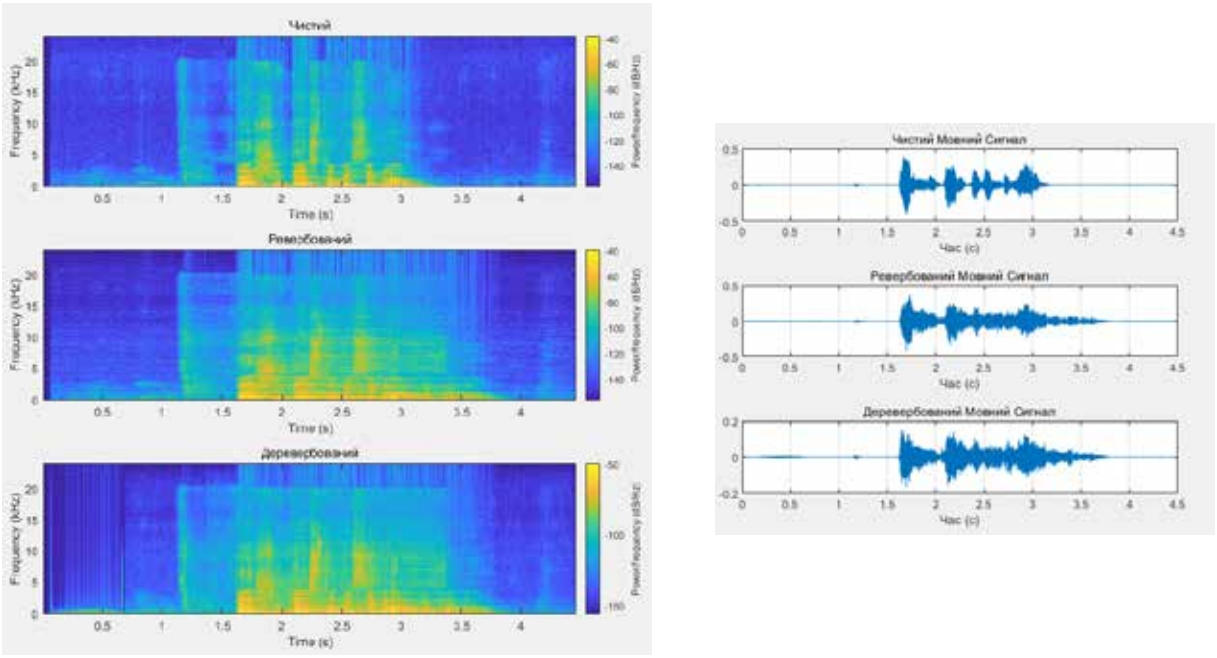


Рис. 2. Схема для запису мовних сигналів

Для навчання задамо кількість ітерацій за цикл – 7; кількість циклів – 50; поріг навчання – 0,0008. В якості основного алгоритму навчання для нейронної мережі оберемо метод Адама (метод стохастичної оптимізації) [19]. При навчанні мережі будемо поступово змінювати і кількість ланок мережі (кількість загорткових шарів) за послідовністю – 6,8,12. При досягненні значення функції втрат на рівні 0,0008 будемо вважати, що мережа успішно пройшла етап навчання. Наприклад, для кількості шарів 6 нейронна мережа пройшла навчання за 35 хв 50 сек.

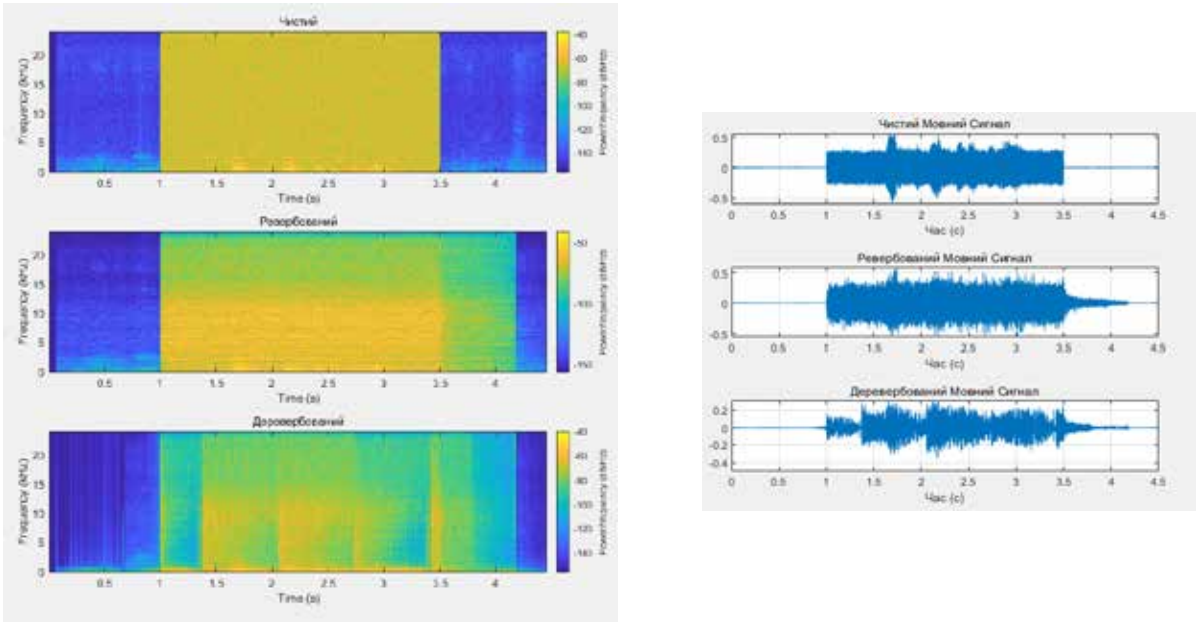
В режимі тренування нейронної мережі підключимо генератор шуму і перевіримо, як мережа вирішує одночасно і задачу деревербераці, і задачу зменшення шуму. Спочатку проаналізуємо, як мережа вирішує задачу деревербераці, коли кількість ланок складає 8. На рисунку 3 наведено сигналограми та спектрограми.

Аналізуючи отримані результати, можна сказати, що нейронна мережа зменшує реверберацію в сигналі (рис. 3,б), але при цьому на початку сигналограми в інтервалі від 0 до 0,5 с проявляється незначний рівень побічного сигналу, який показано зеленою областю на спектрограмі (рис. 3,а). Як вихід з такої ситуації можна збільшити розмірність фільтру ядра згортки або змен-



а) б)

Рис. 3. Графічне зображення результатів роботи мережі: а) спектрограми; б) сигналограми

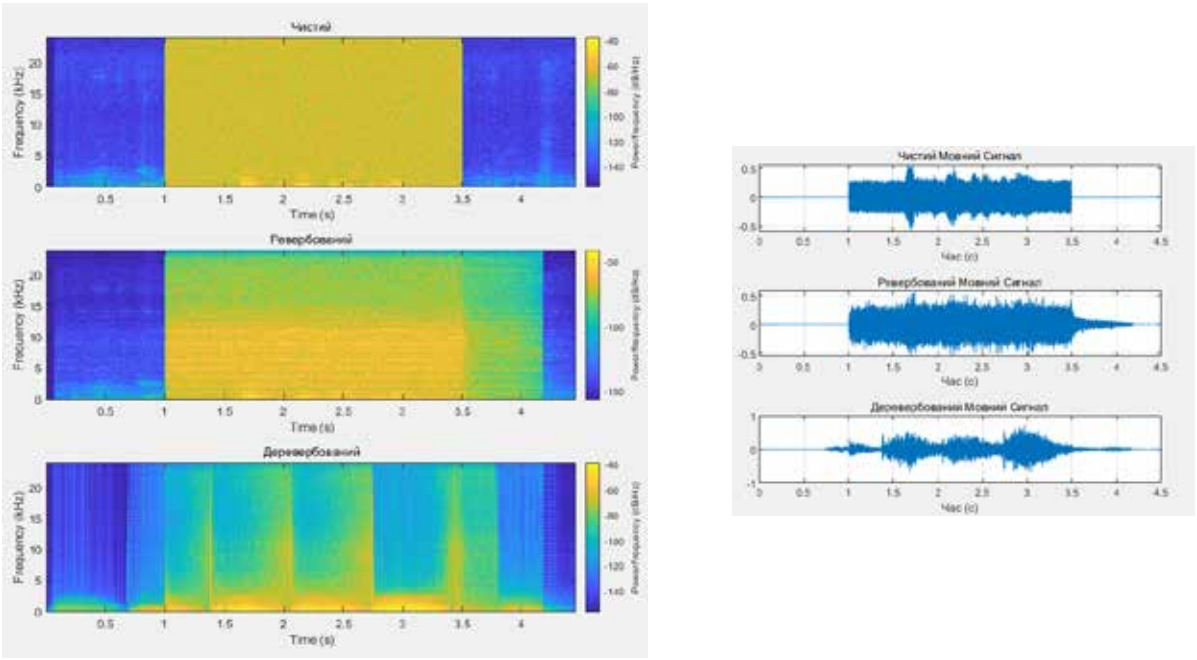


а) б)

Рис. 4. Графічне зображення результатів роботи мережі при кількості ланок 8, та наявності білого шуму: а) спектрограми; б) сигналограми

шити рецепторне поле нейрона моделі нейронної мережі. Порівнюючи деревербероваий та чистий сигнали, також варто відмітити, що деревербероваий сигнал за потужністю буде мати трохи більші значення, ніж вихідний. Така особливість обумовлена особливостями роботи самої нейронної мережі на основі архітектури U-Net.

Далі на вхід нейронної мережі подамо записаний сигнал (рис. 1), до якого додано адитивний шум. На рисунках 4 та 5 наведено результати



а) б)

Рис. 5. Графічне зображення результатів роботи мережі при кількості ланок 12, та наявності білого шуму: а) спектрограми; б) сигналограми

роботи мережі при 8 та 12 ланках та підключеному білому шуму з інтенсивністю 30 дБ.

Аналізуючи отримані результати з рисунків 4 та 5, можна відмітити характерні особливості: 1) якість деревербераційї під впливом одного і того ж шуму підвищується при збільшенні кількості ланок мережі (рис. 4,б та рис. 5,б); 2) спектрограма сигналу при 12 ланках характеризується меншим значенням енергії сигналу, аніж це можна побачити для 8 ланок; 3) середній рівень сигналу в результаті роботи нейронної мережі зменшується, і крім цього, досягається значний ефект зі зменшення рівня шуму (рис. 5,б перша та третя сигналограми), як для 8 так і для 12 ланок мережі. Крім цього, з порівняння сигналограм на рисунках 3,б та 5,б для деревербероваого мовного сигналу можна відмітити, що нейронна мережа навіть за наявності білого шуму намагається зменшити присутність шуму, за виключенням часового інтервалу з 1 с до 1,5 с. Цю особливість можна виправити, якщо в структурі нейронної мережі на етапі попередньої обробки збільшити розмірність одностороннього швидкого перетворення Фур'є і включити в структуру додаткові пропускні з'єднання.

Висновки. В роботі розроблено практичний алгоритм для зменшення негативного впливу реверберації приміщення при записі мовних сигналів. Визначено практичні особливості для відновлення чистих мовних сигналів з аудіозаписів, на які впливають акустичні шуми та реверберація приміщення, де проводиться запис. Зокрема, на основі використання одного мікрофону побудовано схему організації акустичного каналу для розв'язання задачі деревербераційї мовних сигналів. Показано, що на основі використання нейронної мережі та відповідного аналізу зображень спектрограм можна отримати значне зменшення реверберації та рівня шуму, який додається до мовного сигналу ззовні. Знайдено, що на деяких часових інтервалах в області початку впливу адитивного шуму виникла необхідність або провести додаткове навчання нейронної мережі, або внести структурні зміни в архітектуру цієї мережі, тобто збільшити розмірність ядра згортки та рецепторне поле нейрону.

ЛІТЕРАТУРА

1. Lemercier, J-M., Thiemann, J., Koning, R. et al. A neural network-supported two-stage algorithm for lightweight dereverberation on hearing devices. *EURASIP Journal on Audio, Speech, and Music Processing* 18, 1–12 (2023). URL: https://doi.org/10.1186/s13636-023-00285-8.

2. Oo, Z., Wang, L., Phapatanaburi, K. et al. Phase and reverberation aware DNN for distant-talking speech enhancement. *Multimed Tools Appl* 77, 18865–18880 (2018). URL: https://doi.org/10.1007/s11042-018-5686-1.

3. Kinoshita, K., Delcroix, M., Gannot, S. et al. A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research. *EURASIP J. Adv. Signal Process*. 2016, 7 (2016). URL: https://doi.org/10.1186/s13634-016-0306-6.

4. Han, Z., Ke, Y., Li, X. et al. Parallel processing of distributed beamforming and multichannel linear prediction for speech denoising and deverberation in wireless acoustic sensor networks. *J AUDIO SPEECH MUSIC PROC*. 2023, 25 (2023). URL: https://doi.org/10.1186/s13636-023-00287-6.

5. Richter, J., Welker, S., Lemercier, L-M. et al. Speech Enhancement and Dereverberation with Diffusion-based Generative Models. *arXiv:2208.05830,* 1–12 (2023). URL: https://doi.org/10.48550/arXiv.2208.05830.

6. Nercessian, S., et al. Speech dereverberation using recurrent neural networks. *Proceedings of the 22nd International Conference on Digital Audio Effects (DAFx-19)*, Birmingham, UK, September 2–6, DAFX-1 – DAFX-5 (2019).

7. Xu, R., Wu, R., Ishiwaka, Y. et al. Listening to sounds of silence for speech denoising. *arXiv:2010.12013*, 1–7 (2020). URL: https://doi.org/10.48550/arXiv.2010.12013.

8. Sheeja, J.J.C., Sankaragomathi, B. Speech dereverberation and source separation using DNN-WPE and LWPR-PCA. *Neural Comput & Applic* 35, 7339–7356 (2023). URL: https://doi.org/10.1007/s00521-022-07884-0.

9. Berkun, R., Cohen, I. Microphone array power ratio for quality assessment of reverberated speech. *EURASIP J. Adv. Signal Process*. 2015, 49 (2015). URL: https://doi.org/10.1186/s13634-015-0233-y.

10. Routray, S., Mao, Q. A context aware-based deep neural network approach for simultaneous speech denoising and dereverberation. *Neural Comput & Applic* 34, 9831–9845 (2022). URL: https://doi.org/10.1007/s00521-022-06968-1.

11. Zheng, N., Shi, Y., Rong, W. et al. Effects of Skip Connections in CNN-Based Architectures for Speech Enhancement. *J Sign Process Syst* 92, 875–884 (2020). URL: https://doi.org/10.1007/s11265-020-01518-1.

12. Solanki, A., Pandey, S. Music instrument recognition using deep convolutional neural networks. *Int. j. inf. tecnol*. 14, 1659–1668 (2022). URL: https://doi.org/10.1007/s41870-019-00285-y.

13. Ren, B., Wang, L., Lu, L. et al. Combination of bottleneck feature extraction and dereverberation for distant-talking speech recognition. *Multimed Tools Appl* 75, 5093–5108 (2016). URL: https://doi.org/10.1007/s11042-015-2849-1.

14. P. A. Naylor, *Speech Dereverberation*, Springer, London, 388 P. (2010). URL: https://doi.org/10.1007/978-1-84996-056-4.

15. Dong, HY., Lee, CM. Speech intelligibility improvement in noisy reverberant environments based on speech enhancement and inverse filtering. *J AUDIO SPEECH MUSIC PROC*. 2018, 3 (2018). URL: https://doi.org/10.1186/s13636-018-0126-8.

16. Ernst, O., Shlomo, E., Gannot, S. et al. Speech dereverberation using fully convolutional networks. *arXiv:1803.08243,* 1–5 (2019). URL: https://doi.org/10.48550/arXiv.1803.08243.

17. EAP Habets, Single- and multi-microphone speech dereverberation using spectral enhancement. *PhD thesis*, Technische Universiteit Eindhoven (2007).

18. Abdullah, R., Imail, S., Dzulkefli, N. et al. Potential acoustic treatment analysis using sabine formula in unoccupied classroom. *JICETS 2019* 1529, 1–7 (2020). URL: https://doi.org/10.1088/1742-6596/1529/2/022031.

19. Lan, R., Zou, H., Pang, C. et al. Image denoising via deep residual convolutional neural networks. *SIViP* 15, 1–8 (2021). URL: https://doi.org/10.1007/s11760-019-01537-x.

FEATURES OF SPEECH SIGNAL DERVERBERATIONS USING NEURAL NETWORKS

Glib Borisov
Postgraduate Student at the Department of Acoustic and Multimedia Electronic Systems
National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, 37 Beresteyskyi ave., Kyiv, Ukraine, 03056, borusov5364@gmail.com
ORCID: 0000-0003-2780-2700

Kyrylo Trapezon
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Acoustic and Multimedia Electronic Systems
National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, 37 Beresteyskyi ave., Kyiv, Ukraine, 03056, kirill.trapezon@gmail.com
ORCID: 0000-0001-5873-9519

Purpose. The aim of the article is to develop an algorithm based on the approaches and features of a neural network that will allow efficient and fast removal of noise of various nature and dereverberation of recorded speech signals. **Methodology**. The analytical relations for determining speech reverberation and quantifying room reverberation presented in this paper are basic in the theory of signal processing and contain provisions of the theory of physical acoustics. The

acquisition of signograms and spectrograms of recorded speech signals was realized on the basis of special frameworks and applications of the Matlab software environment. **Findings.** A practical algorithm has been developed to reduce the negative impact of room reverberation when recording speech signals. Practical features for restoring clean speech signals from audio recordings affected by acoustic noise and reverberation of the recording room are determined. **Originality**. The considered approach using a neural network allows simultaneously solving two practical problems: the problem of speech dereverberation and the problem of noise reduction in recorded speech signals. Moreover, the developed approach has the features of universality, since it can be used to analyze noise of different nature and with different parameters. **Practical value**. The presented practical provisions and the corresponding developed algorithm, once finalized, can be extended to solve other applied problems of machine computer vision, namely, in the organization of remote communication in conference rooms, in the design of advanced hearing aids, as well as for speech recognition systems and voice activity recording systems for control in applications and devices of the Internet of Things. **Conclusions**. Based on the single-microphone method, an acoustic channel organization scheme is constructed to solve the problem of speech signal dereverberation. It is shown that, based on the use of a neural network and the analysis of spectrogram images, a significant reduction in reverberation and noise level added to the speech signal from the outside can be obtained.

Key words: dereverberation, noise, signal, acoustics, neural network, room, algorithm, efficiency.

REFERENCES

1. Lemercier, J-M., Thiemann, J., Koning, R. et al. A neural network-supported two-stage algorithm for lightweight dereverberation on hearing devices. *EURASIP Journal on Audio, Speech, and Music Processing* 18, 1–12 (2023). https://doi.org/10.1186/s13636-023-00285-8

2. Oo, Z., Wang, L., Phapatanaburi, K. et al. Phase and reverberation aware DNN for distant-talking speech enhancement. *Multimed Tools Appl* 77, 18865–18880 (2018). https://doi.org/10.1007/s11042-018-5686-1

3. Kinoshita, K., Delcroix, M., Gannot, S. et al. A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research. *EURASIP J. Adv. Signal Process*. 2016, 7 (2016). https://doi.org/10.1186/s13634-016-0306-6

4. Han, Z., Ke, Y., Li, X. et al. Parallel processing of distributed beamforming and multichannel linear prediction for speech denoising and deverberation in wireless acoustic sensor networks. *J AUDIO SPEECH MUSIC PROC*. 2023, 25 (2023). https://doi.org/10.1186/s13636-023-00287-6

5. Richter, J., Welker, S., Lemercier, L-M. et al. Speech Enhancement and Dereverberation with Diffusion-based Generative Models. *arXiv:2208.05830,* 1–12 (2023). https://doi.org/10.48550/arXiv.2208.05830

6. Nercessian, S., et al. Speech dereverberation using recurrent neural networks. *Proceedings of the 22nd International Conference on Digital Audio Effects (DAFx-19)*, Birmingham, UK, September 2–6, DAFX-1 – DAFX-5 (2019).

7. Xu, R., Wu, R., Ishiwaka, Y. et al. Listening to sounds of silence for speech denoising. *arXiv:2010.12013*, 1–7 (2020). https://doi.org/10.48550/arXiv.2010.12013

8. Sheeja, J.J.C., Sankaragomathi, B. Speech dereverberation and source separation using DNN-WPE and LWPR-PCA. *Neural Comput & Applic* 35, 7339–7356 (2023). https://doi.org/10.1007/s00521-022-07884-0

9. Berkun, R., Cohen, I. Microphone array power ratio for quality assessment of reverberated speech. *EURASIP J. Adv. Signal Process*. 2015, 49 (2015). https://doi.org/10.1186/s13634-015-0233-y

10. Routray, S., Mao, Q. A context aware-based deep neural network approach for simultaneous speech denoising and dereverberation. *Neural Comput & Applic* 34, 9831–9845 (2022). https://doi.org/10.1007/s00521-022-06968-1

11. Zheng, N., Shi, Y., Rong, W. et al. Effects of Skip Connections in CNN-Based Architectures for Speech Enhancement. *J Sign Process Syst* 92, 875–884 (2020). https://doi.org/10.1007/s11265-020-01518-1

12. Solanki, A., Pandey, S. Music instrument recognition using deep convolutional neural networks. *Int. j. inf. tecnol*. 14, 1659–1668 (2022). https://doi.org/10.1007/s41870-019-00285-y

13. Ren, B., Wang, L., Lu, L. et al. Combination of bottleneck feature extraction and dereverberation for distant-talking speech recognition. *Multimed Tools Appl* 75, 5093–5108 (2016). https://doi.org/10.1007/s11042-015-2849-1

14. P. A. Naylor*, Speech Dereverberation*, Springer, London, 388 P. (2010). https://doi.org/10.1007/978-1-84996-056-4

15. Dong, HY., Lee, CM. Speech intelligibility improvement in noisy reverberant environments based on speech enhancement and inverse filtering. *J AUDIO SPEECH MUSIC PROC*. 2018, 3 (2018). https://doi.org/10.1186/s13636-018-0126-8

16. Ernst, O., Shlomo, E., Gannot, S. et al. Speech dereverberation using fully convolutional networks. *arXiv:1803.08243,* 1–5 (2019). https://doi.org/10.48550/arXiv.1803.08243

17. EAP Habets, Single- and multi-microphone speech dereverberation using spectral enhancement. *PhD thesis*, Technische Universiteit Eindhoven (2007)

18. Abdullah, R., Imail, S., Dzulkefli, N. et al. Potential acoustic treatment analysis using sabine formula in unoccupied classroom. *JICETS 2019* 1529, 1–7 (2020). https://doi.org/10.1088/1742-6596/1529/2/022031

19. Lan, R., Zou, H., Pang, C. et al. Image denoising via deep residual convolutional neural networks. *SIViP* 15, 1–8 (2021). https://doi.org/10.1007/s11760-019-01537-x

Стаття надійшла 06.06.2023