

УДК 004.93:519.6

DOI: 10.20998/2411-0558.2022.02.10

В. В. ЛАРИКОВА, магістр, ОНУ ім. І. І. Мечникова, Одеса,

В. В. МОРОЗ, канд. техн. наук, доц., ОНУ ім. І. І. Мечникова, Одеса

ВИДАЛЕННЯ ІМПУЛЬСНОГО ШУМУ З АУДІОСИГНАЛУ

У роботі розглянуто методи, які використовуються для вирішення поставленої задачі. Проаналізовано їх архітектуру, принципи роботи, конкретні приклади, у яких вони використовувалися, їх переваги та недоліки. Описана задача виявлення та видалення імпульсного шуму в аудіосигналах з використанням методів машинного навчання. На основі проведеного дослідження запропоновано програмне рішення з використанням алгоритму спектрального віднімання, яке дозволяє виявляти та видаляти імпульсний шум в аудіосигналах з метою покращення якості сприйняття аудіо. Іл.: 8. Табл.: 1. Бібліогр.: 10 назв.

Ключові слова: аудіосигнал; імпульсний шум; спектральне віднімання; машинне навчання; видалення шуму.

Постановка проблеми. Проблема ізоляції джерел звуку та виділення корисного сигналу з акустичної суміші [1 – 3] залишається складною проблемою в області цифрової обробки сигналів. Алгоритми цифрової обробки переходять від вузькоспеціалізованих до широко використовуваних. Раніше вони використовувалися для радарів, професійного звукозапису та обробки звуку. У сучасному мультимедійному інтерфейсі "людина-машина", який включає мікросхеми цифрової камери, що обробляють зображення, мобільні телефони, які кодують і обробляють звук, і персональні комп'ютери, які можуть функціонувати як центр домашніх розваг завдяки своїм широким можливостям обробки звуку зображень і відео, алгоритми цифрової обробки звукових даних використовуються все частіше. Унаслідок чого постає питання розробки якісного алгоритму для вирішення задачі виявлення та видалення імпульсних шумів в аудіосигналах.

Огляд методів. В [4] розглянуто метод, який можна використовувати для нормалізації аудіосигналів, що були пошкоджені імпульсним шумом, наприклад, клацанням, сплеском або подряпинами. Коли на вхід подається зашумлений аудіосигнал, система автоматично знаходить погіршені зразки та замінює їх більш відповідними значеннями. Припущення, що сигнал може бути локально описаний як реалізація процесу авторегресії лежить в основі обох процесів: виявлення та інтерполяції. Метод демонструє, що повністю автоматизована система з точно визначеним набором параметрів

може досягти кращої продуктивності на звукових сигналах із погіршенням якості.

Було проаналізовано наступні типи автоматизованих систем для розпізнавання мовлення, які використовуються для роботи з аудіосигналами: виокремлене слово, безперервні слова, комбіновані слова та спонтанне мовлення. Охарактеризовано методи розпізнавання голосу, які включають фундаментальні підходи до розпізнавання мовлення: акустично-фонетичний, розпізнавання образів (HMM, SWM, DWT, VQ), на основі штучного інтелекту [5], на базі штучних нейронних мереж [6, 7], прихованої моделі Маркова [8] та моделі гаусової суміші. Результати аналізу демонструють, що розробка методів видалення шумів в аудіосигналах з метою покращення сприйняття аудіо, зараз є актуальною.

Було зроблено висновок, що для вирішення задачі виявлення та видалення імпульсного шуму в аудіосигналах найкраще використовувати алгоритми спектрального віднімання.

Мета роботи. Аналіз існуючих методів видалення імпульсних шумів в аудіосигналах та розробка більш ефективного алгоритму.

Розробка алгоритму видалення імпульсного шуму з аудіосигналів. Були проаналізовані основні алгоритми глибинного навчання для виявлення та видалення імпульсних шумів з аудіосигналів. Серед них перевагу отримав метод спектрального віднімання (англ. Spectral Subtraction). Даний метод виконує відновлення потужності або спектра величини сигналу, що спостерігається в адитивному шумі, шляхом віднімання оцінки середнього спектру шуму від спектру шумового сигналу. Спектр шуму оцінюється та оновлюється з періодів, коли сигнал відсутній і присутній лише шум.

Припускається, що шум є стаціонарним або повільно змінним процесом, і що спектр шуму не змінюється суттєво між періодами оновлення. Для відновлення сигналів у часовій області оцінка спектра миттєвої величини поєднується з фазою шумового сигналу, а потім перетворюється за допомогою дискретного перетворення Фур'є в часову область

Алгоритми, що ґрунтуються на даному підході, діляться на дві основні категорії: стаціонарні (зберігають розрахунковий поріг шуму на одному рівні для всього сигналу) та нестаціонарні (постійно оновлюють розрахунковий поріг шуму з часом).

Стаціонарне виявлення та видалення шумів. Одним із підходів до стаціонарного зменшення шуму є алгоритм спектрального віднімання. На вході подається сигнальний кліп, що містить сигнал і шум, який необхідно

видалити. Для кожного частотного компонента сигналу будь-який частотно-часовий компонент нижче порогового значення відкидається як шум. Алгоритм обчислює середнє значення та стандартне відхилення кожного частотного каналу короткочасного перетворення Фур'є (STFT) [9] сигналу (наприклад, спектрограми). Порогове значення для кожного частотного компонента частоти встановлюється на деякому рівні вище середнього (наприклад, три стандартних відхилення), щоб визначити, чи вважається частотно-часовий компонент у спектрограмі сигналом чи шумом. Потім спектрограма маскується на основі цього порогу та інвертується (з інверсним STFT) назад у часову область.

Нестаціонарне виявлення та видалення шумів. Хоча стаціонарні алгоритми зменшення шуму можуть розглядати шум як будь-який стаціонарний акустичний сигнал, нестаціонарні методи відрізняються тим, як вони розрізняють сигнал і шум. Нестаціонарний шум важче усунути, оскільки алгоритмічно важко охарактеризувати різницю між сигналом і шумом. Один із підходів для визначення межі між сигналом і нестаціонарним шумом полягає у визначенні часової шкали, на якій діє сигнал, і розглядання будь-чого поза цією шкалою часу як шуму. Будь-яка акустична подія, що виходить за межі цього діапазону часу, може вважатися шумом. Спектральне стробування можна розширити до нестаціонарного зменшення шуму шляхом обчислення змінного стробування на основі поточної оцінки фонового шуму.

Основні етапи алгоритму нестаціонарного виявлення та видалення імпульсного шуму в аудіосигналах:

1. По сигналу розраховується спектрограма;
2. Згладжена в часі версія спектрограми обчислюється за допомогою ІІР-фільтра, застосованого вперед і назад на кожному частотному каналі;
3. На основі цієї згладженої в часі спектрограми обчислюється маска;
4. Маска згладжується за допомогою фільтра за частотою та часом;
5. Маска застосовується до спектрограми сигналу та інвертується.

Переваги нестаціонарного видалення імпульсних шумів не є унікальними для акустичного шуму: коли відомий часовий масштаб сигналу, можна використовувати ті самі нестаціонарні принципи для видалення шуму, що виникає в різних часових масштабах. Наприклад, у безперервному записі нейронних даних потенціали дії виникають у діапазоні 1 мс. Тому події, що відбуваються протягом десятків або сотень мілісекунд, можна розглядати як шум.

Дослідження ефективності розробленого алгоритму. З використанням вищезазначених технологій та мови програмування Python

[10] було реалізовано застосунок для виявлення та видалення імпульсних шумів. У роботі використовуються аудіофайли з ресурсу WavSource.

Для тестування алгоритму були обрані наступні параметри:

1. Частотний діапазон для згладжування маски = 500 Гц;
2. Діапазон часу для згладжування маски = 50 мс;
3. Довжина віконного сигналу після доповнення нулями = 1024 Гц;
4. Пропорція зменшення шуму = 1.0 (100 %);
5. Стала часу в секундах для обчислення рівня шуму в нестационарному алгоритмі = 2.0.

На рис. 1 наведена графічна візуалізація обраного аудіофрагмента довжиною 8 секунд.

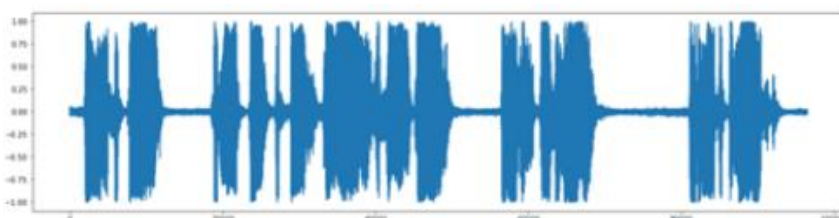


Рис. 1. Графічна візуалізація обраного аудіофрагмента.

Виконуємо накладання випадкового шуму на вхідний аудіосигнал, а також будемо графічну візуалізацію після накладання.

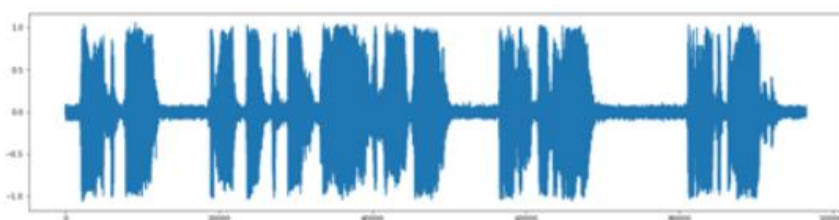


Рис. 2. Накладання шуму та побудова графічної візуалізації.

Далі за допомогою стаціонарного та нестационарного алгоритмів здійснюємо виявлення та видалення імпульсного шуму в аудіосигналах.

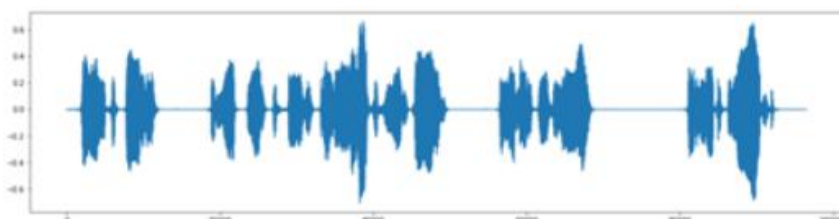


Рис. 3. Використання стаціонарного алгоритму для виявлення та видалення імпульсних шумів.

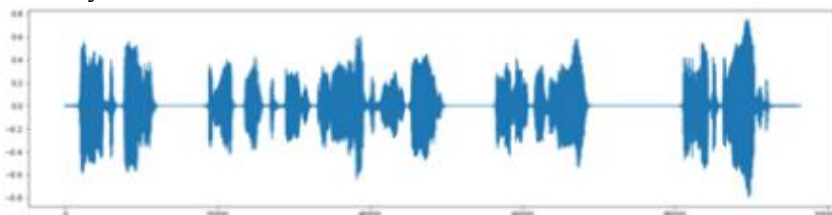


Рис. 4. Використання нестаціонарного алгоритму для виявлення та видалення імпульсних шумів.

Як видно з рис. 3 та рис. 4 після застосування стаціонарного та нестаціонарного алгоритму видалення імпульсних шумів відповідно, відбувається повноцінне видалення шуму та користувач отримує очищений аудіосигнал з чітким відтворенням вхідного аудіо. Можна побачити, що нестаціонарне видалення форматує зашумлені дані трохи краще.

Розглянемо приклад з довшим вхідним аудіофайлом (рис. 5), спробуємо додати зашумлення та порівняти результати різних алгоритмів.

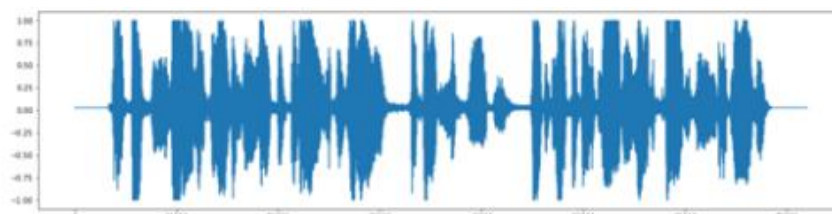


Рис. 5. Графічна візуалізація обраного аудіофрагмента.

На рис. 6 наведена графічна візуалізація аудіофайлу після накладання шуму.

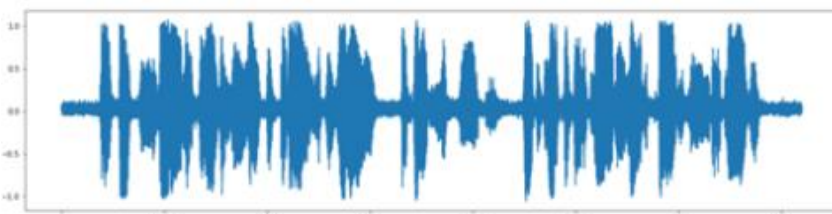


Рис. 6. Зашумлені дані.

Використовуємо стаціонарний та нестаціонарний алгоритм для виявлення та видалення імпульсного шуму з аудіосигналу.

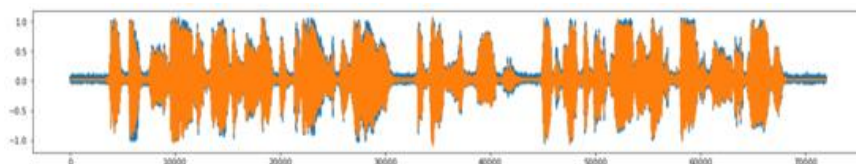


Рис. 7. Використання стаціонарного алгоритму для виявлення та видалення імпульсних шумів.

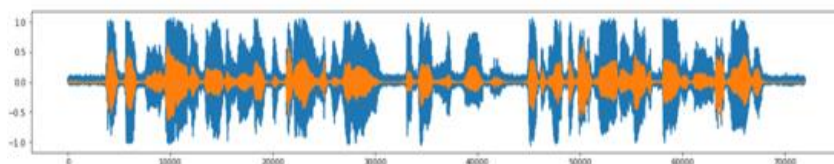


Рис. 8. Використання нестаціонарного алгоритму для виявлення та видалення імпульсних шумів.

Як видно з рис. 7 та рис. 8, у випадку з довшим вхідним аудіо файлом, нестаціонарне видалення форматує зашумлені дані трохи краще.

Представимо порівняльний аналіз розроблених алгоритмів виявлення та видалення імпульсних шумів у табл. 1.

Таблиця 1

Порівняльний аналіз розроблених алгоритмів

Стаціонарний алгоритм виявлення та видалення імпульсних шумів	Нестаціонарний алгоритм виявлення та видалення імпульсних шумів
Середня швидкість обробки аудіоданих.	Висока швидкість обробки аудіоданих.
Вміє працювати з великими аудіо файлами.	Вміє працювати з великими аудіо файлами.
Існує вірогідність появи "артефактів" при обробці аудіо.	Відсутня вірогідність появи "артефактів" при обробці аудіо.
Основна ідея полягає в тому, що статистика обчислюється для кожного частотного каналу, щоб визначити гейт шуму. Потім цей гейт застосовується до сигналу.	Є обгорткою над стаціонарним алгоритмом, але дозволяє змінювати шумовий гейт з часом.
Обробляє локалізовані та глобальні деградації аудіофайлів.	Обробляє локалізовані та глобальні деградації аудіофайлів.

Можна побачити, що стаціонарні та нестаціонарні алгоритми можна використовувати при обробці аудіофайлів будь-якої довжини. Нестаціонарний алгоритм швидше справляється з виявленням та видаленням імпульсного шуму.

Висновки. На основі проведеного дослідження встановлено, що для вирішення задачі виявлення та видалення імпульсного шуму в аудіосигналах найкраще використовувати алгоритми спектрального віднімання. За допомогою експериментів було встановлено, що стаціонарні та нестаціонарні алгоритми можна використовувати при обробці аудіофайлів будь-якої довжини. Якщо для користувача важлива швидкість обробки даних, то необхідно вибрати нестаціонарне виявлення та видалення імпульсних шумів. При використанні нестаціонарного виявлення та видалення імпульсних шумів відсутня вірогідність появи "артефактів" при обробці аудіо, а при використанні стаціонарного алгоритму така вірогідність існує.

Отримані результати можуть бути використані для створення ефективних систем виявлення та видалення імпульсного шуму в аудіосигналах різноманітного природного походження. Наприклад, для очищення аудіо від зайвих звукових подій для повторного відтворення, видалення шуму для кращого розпізнавання мовлення, створення аудіоредакторів. Крім того, видалення шуму буде корисним під час монтажу фільмів, музики, подкастів та інших медіа.

Для покращення роботи застосунку в майбутньому необхідно додати можливість мультипроцесинга, для того, щоб у користувача була можливість працювати ще з більшим об'ємом даних. Разом з тим буде реалізовано GUI інтерфейс з інтерактивними елементами для зручності користувачів та при дослідженнях, а також API інтерфейс для роботи з ендпоінтами та підвищення швидкості виконання запитів.

Список літератури:

1. *Yoshioka T.* Making machines understand us in reverberant rooms: robustness against reverberation for automatic speech recognition / *T. Yoshioka, A. Sehr, M. Delcroix, K. Kinoshita, R. Maas, T. Nakatani and W. Kellermann* // *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 114–12
2. *Chen L.- H.* Adaptive postfiltering for quality enhancement of coded speech / *J.-H. Chen, A. Gersho* // *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1. – 1995. – P. 59–71.
3. *Cho K.* On the properties of neural machine translation: Encoder-decoder approaches / *K. Cho, B. Van Merriënboer, D. Bahdanau, Y. Bengio* // in *Proc. Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8)*, – 2014.
4. *Oudre L.* Automatic Detection and Removal of Impulsive Noise in Audio Signals / *L. Oudre* // *Image Processing On Line*, 5, Nov. 2015. – P. 267 – 281.
5. *Boll S.* Suppression of acoustic noise in speech using spectral subtraction / *S. Boll* // *IEEE Transactions on acoustics, speech, and signal processing*, vol. 27, no. 2 – 1979. – P. 113–120.
6. *Wijoyo S.* Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot / *S. Wijoyo* // *International Conference on Information and Electronics Engineering IPCSIT vol.6, IACSIT Press, P.179-183, Singapore, 2011.*

7. Karpathey A. Understanding the unreasonable effectiveness of recurrent neural networks / A. Karpathey // <http://karpathey.github.io/2015/05/21/rnn-effectiveness>
8. Patel I. Speech recognition using hidden Markov model with MFCC-subband technique / I. Patel, Y. S. Rao // Proc. ITC, Mar. 2010, P. 168–172.
9. Groecheing K. The Short-Time Fourier Transform / K. Groecheing // Foundations of Time-Frequency Analysis, Birkhauser Boston, MA. 2001 – P. 37 – 58.
10. Hunt J. Advanced Guide to Python 3 Programming (Undergraduate Topics in Computer Science) / Hunt J. – Firts Edition, Springer, 2019 – 523 p.

References:

1. Yoshioka T., Sehr A., Delcroix M., Kinoshita K., Maas R., Nakatani T. and Kellermann W., "Making machines understand us in reverberant rooms: robustness against reverberation for automatic speech recognition", *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 114–120
2. Chen L. - H., Gersho A. (1995), "Adaptive postfiltering for quality enhancement of coded speech", *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 59–71.
3. Cho K., Van Merriënboer B., Bahdanau D., Bengio Y. (2014), "On the properties of neural machine translation: Encoder-decoder approaches", in Proc. Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8).
4. Oudre L. (2015), "Automatic Detection and Removal of Impulsive Noise in Audio Signals", *Image Processing On Line*, 5, Nov., pp. 267 – 281.
5. Boll S. (1979), "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Transactions on acoustics, speech, and signal processing*, vol. 27, no. 2, pp. 113–120.
6. Wijoyo S. (2011), "Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot", International Conference on Information and Electronics Engineering IPCSIT vol.6, IACSIT Press, pp.179-183, Singapore.
7. Karpathey A., "Understanding the unreasonable effectiveness of recurrent neural networks", available at: <http://karpathey.github.io/2015/05/21/rnn-effectiveness>
8. Patel I., Rao Y. S. (2010), "Speech recognition using hidden Markov model with MFCC-subband technique", Proc. ITC, Mar., pp.168–172.
9. Groecheing K. (2001), "The Short-Time Fourier Transform", *Foundations of Time-Frequency Analysis*, Birkhauser Boston, MA, pp. 37 – 58.
10. Hunt J. (2019), *Advanced Guide to Python 3 Programming (Undergraduate Topics in Computer Science)*, Firts Edition, Springer, 523 p.

Статтю представив доктор техн. наук, проф. Гунченко Ю.А..

Надійшла (received) 10.11.2022

Volodymyr Moroz, PhD, Associate Professor
Odesa I. I. Mechnikov National University
Str. Dvoryanska, 2, Odesa, Ukraine, 65000
Tel.: +380 93-273-4555, e-mail: v.moroz@onu.edu.ua
ORCID ID: 0000-0002-3240-4590

Valeriia Larikova, master

Odesa I. I. Mechnikov National University

Str. Dvoryanska, 2, Odesa, Ukraine, 65000

Tel.: +380 95-720-1002, e-mail: valeria-larikova@stud.onu.edu.ua

УДК 004.93:519.6

Видалення імпульсного шуму з аудіосигналу / Ларікова В. В., Мороз В. В. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2022. – № 1 – 2 (7 – 8). – С. 107 – 116.

У роботі розглянуто методи, які використовуються для вирішення поставленої задачі. Проаналізовано їх архітектуру, принципи роботи, конкретні приклади, у яких вони використовувалися, їх переваги та недоліки. Описана задача виявлення та видалення імпульсного шуму в аудіосигналах з використанням методів машинного навчання. На основі проведеного дослідження запропоновано програмне рішення з використанням алгоритму спектрального віднімання, яке дозволяє виявляти та видаляти імпульсний шум в аудіосигналах з метою покращення якості сприйняття аудіо. Іл.: 8. Табл.: 1. Бібліогр.: 10 назв.

Ключові слова: аудіосигнал; імпульсний шум; спектральне віднімання; машинне навчання; видалення шуму.

УДК 004.93:519.6

Removal of impulse noise from an audio signal / Larikova V. V., Moroz V. V. // Herald of the National Technical University "KhPI". Series of "Informatics and Modeling". – Kharkov: NTU "KhPI". – 2022. – № 1 – 2 (7 – 8). – P. 107 – 116.

The work examines the methods used to solve the given problem. Their architecture, principles of operation, specific examples in which they were used, their advantages and disadvantages were analyzed. The task of detecting and removing impulse noise in audio signals using machine learning methods is described. On the basis of the conducted research, a software solution using the spectral subtraction algorithm is proposed, which allows detecting and removing impulse noise in audio signals in order to improve the quality of audio perception. Figs: 3. Tabl.: 1. Refs: 10 titles.

Keywords: audio signal; impulse noise; spectral subtraction; machine learning; recognition; noise removal.