

DOI: <https://doi.org/10.36910/6775-2524-0560-2021-44-20>

УДК 004.89

Проніна Ольга Ігорівна, к. т. н., доцент

<https://orcid.org/0000-0001-7085-8027>

Дегтяр Валерія Віталіївна, магістр

<https://orcid.org/0000-0002-2756-3618>

Державний вищий навчальний заклад «Приазовський державний технічний університет», м. Маріуполь, Україна

ВИКОРИСТАННЯ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ КОРИГУВАННЯ ДЕФЕКТІВ МОВЛЕННЯ У ДІТЕЙ

Проніна О. І., Дегтяр В. В. Використання нейронних мереж для коригування дефектів мовлення у дітей. Визначення дефектів мовлення у дитини є актуальною науково-практичною задачею, оскільки своєчасне коректування мови дозволяє покращити комунікативні здібності дитини в майбутньому. Грунтуючись на аналізі наведеному в роботі логічним є використання згорткових нейронних мереж як інструменту виявлення дефекту і його виду. Робота присвячена розробці алгоритму для виявлення дефекту мови, що включає в себе підготовку даних для навчання моделі і використання згорткової нейронної мережі. Описана архітектура згорткової нейронної мережі, що була розроблена. Проведено експерименти для перевірки точності і адекватності розробленої моделі нейронної мережі, результати наведені у вигляді табличних даних.

Ключові слова: мова, дефекти мовлення, звуки, згорткова нейронна мережа.

Пронина О. И., Дегтярь В. В. Использование нейронных сетей для корректирования дефектов речи у детей.

Определение дефектов речи у ребенка является актуальной научно-практической задачей, поскольку своевременная корректировка речи позволяет улучшить коммуникативные способности ребенка в будущем. Основываясь на анализе приведенном в работе логичным является использование сверточных нейронных сетей как инструмента выявления дефекта и его вида. Работа посвящена разработке алгоритма для выявления дефекта речи, что включает в себя подготовку данных для обучения модели и использование сверточной нейронной сети, что была разработана. Описана архитектура сверточной нейронной сети. Проведены эксперименты для проверки точности и адекватности разработанной модели нейронной сети, результаты приведены в виде табличных данных.

Ключевые слова: язык, дефекты речи, звуки, сверточная нейронная сеть.

Pronina O., Dehtiar V. Using neural networks to correct speech defects in children. Determination of speech defects in a child is an urgent scientific and practical task, since timely correction of speech can improve the child's communication skills in the future. Based on the analysis given in the work, it is logical to use convolutional neural networks as a tool to identify a defect and its type. The work is devoted to the development of an algorithm for the identified speech defect, which includes the preparation of data for training the model and the use of a convolutional neural network that was developed. The architecture of a convolutional neural network is described. Experiments were carried out to verify the accuracy and adequacy of the developed neural network model, the results are presented in the form of tabular data.

Keywords: language, speech defects, sounds, convolutional neural network.

Постановка наукової проблеми. Правильна мова необхідна для того, щоб успішно побудувати свою кар'єру, розвинути хороші комунікативні навички, вписатися спочатку в дитячий колектив, потім у дорослий. Дефекти мови можуть стати причиною розвитку у людини комплексів. Тому вкрай важливо їх усунути в ранньому віці. Крім того, дітям звукова вимова важлива для того, щоб правильно писати, читати, не пропускати букви при письмі, не робити помилок в аналізі звуків та літер.

У дошкільному віці гра – провідний вид діяльності. Це означає, що через гру дитина висловлює емоції, отримує знання і навіть буде відносини. Для дітей дошкільного віку в 21 столітті гра на різних девайсах займає більше місця в житті, ніж будь-яка інша діяльність, тому інформаційні технології є важливим інструментом у розвитку дитини. Інформаційні технології активно допомагають сфері логопедії допомогти виявити і підібрати необхідні вправи для профілактики і корекції дефектів мовлення. Завдяки сучасним технологіям дитина втягується в роботу і освоює потрібні мовленнєві навички.

У дітей спостерігається підвищення стомлюваності, зниження працездатності при роботі зі звичайними методами виправлення недоліків мови. Тому класичні методи і прийоми часто не спрацьовують, а значить повинні бути сучасні, інноваційні методи, здатні зацікавити дитину, розкрити його потенціал і досягти позитивних результатів в корекційній роботі. Система для корекції дефектів мовлення допоможе зацікавити дитину і виявити всі проблеми. Так само це велика допомога для батьків, так як без фахівця часом важко самостійно підібрати необхідні вправи і виявити більш приховані неточності в мові дитини. Для зручності і мобільності було вирішено створити систему у вигляді мобільного додатка. Його перевагою перед

традиційними методами є те, що додаток завжди буде під рукою і в будь-який момент часу можна виділити час для занять.

Мета дослідження. Метою даної роботи є проектування, розробка і перевірка адекватності роботи систем дефектів мовлення дітей, реалізованої на основі згорткової нейронної мережі.

Аналіз досліджень. В наш час є досить багато досліджень і публікацій, які присвячені темам нейронних мереж і їх впровадження в розвиток мовлення у дітей.

У роботі М. Алама, М. Д. Самада «Опитування про глибокі нейронні мережі в мові та системи зору» [1] представляється огляд найсучасніших архітектури глибоких нейронних мереж, алгоритмів та систем у зорі та мовленнєві програми. З цієї роботи можна зробити висновок, що останні досягнення в галузі глибоких штучних нейронних мережевих алгоритмів та архітектура стимулюють швидкі інновації та розвиток інтелектуального зору та мовних систем.

Сіддік Латіф і Раджіб Рана Глибоке у роботі «Репрезентативне навчання в мовленнєвій обробці: виклики, останні досягнення та тенденції майбутнього» [2] розглядають проблеми та ключові характеристики моделей навчання репрезентації та обговорення, а також вказують на майбутні тенденції.

Гендрік Пурвінз у «Глибоке навчання для обробки звукових сигналів» [3] розглядає сучасний стан методів глибокого навчання для обробки звукових сигналів та моделі глибокого навчання, включаючи згорткові нейронні мережі.

У своїй роботі «Автоматизований мовленнєвий скринінг депресії за допомогою глибокого згорткової нейронні мережі» [4] Кароль Хласта пропонує новий підхід до автоматизованого виявлення депресії в мовленні згорткова нейронна мережа (CNN) та багатостороннє інтерактивне навчання.

Робота «Машинне навчання для ідентифікації заїкання: огляд, виклики та майбутні напрямки» [5] автора Шейкл А. Шейха всебічно розглядає акустику, особливості, методи класифікації заїкання/розбіжності на основі статистичного та глибокого навчання.

У роботі «FluentNet: Наскрізне виявлення мовлення. Нездатність до глибокого навчання» [6] Тедд Куркунакіс запропонував наскрізну глибоку нейронну мережу, FluentNet, здатну виявити ряд різних типів розбіжностей.

Фаді Медхат у «Класифікація музичних жанрів за допомогою маскованих умовних нейронних мереж» [7] оцінили виконання MCLNN (умовні нейронні мережі в масках), використовуючи набори даних музичних жанрів Ballroom та Homburg.

У своїй роботі «Контрольоване розділення мовлення на основі глибокого навчання: огляд» [8] ДеЛян Ван сформулював більш новий підхід розділення мовлення як контрольовану навчальну проблему, де дискримінаційні моделі мовлення, ораторів та фоновий шум отримуються з даних навчання.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження. При різноманітності соціальних мереж, інтернет-магазинів, навчальних майданчиків і літературних порталів не існує такого, який би дозволив за допомогою нейронної мережі максимально точно виявити конкретні недоліки та порушення мови і підібрати відповідні тренувальні вправи. Це свідчить про актуальність роботи.

Основною ідеєю є надання можливості батькам та дітям, використовуючи технології згорткової нейронної мережі з зручним й інтуїтивно зрозумілим інтерфейсом, визначити недоліки та неточності в вимові окремих звуків та допомогти підібрати необхідні вправи для того, щоб виправити конкретні дефекти мовлення. Мова - особлива і найбільш досконала форма спілкування, притаманна тільки людині. У процесі комунікації люди обмінюються думками і взаємодіють один з одним. Мова - найважливіший засіб зв'язку між дитиною і навколишнім світом. Таким чином комунікативна функція мови сприяє розвитку навичок спілкування з однолітками, розвиває можливість спільної гри. Це має неоціненне значення для формування адекватної поведінки, емоційно-вольової сфери та особистості дитини [9].

Мовленнєві порушення в дитячому віці - це великий обсяг порушень, що розрізняються за механізмом освіти, за структурою і за способом їх подолання. Усі порушення діляться на свої класи. Найчастіше зустрічаються такі дефекти:

- дислалія – порушення, при якому малюкові складно вимовляти шиплячі звуки;
- ріносалія – при такій патології голос дитини звучить гугняво;
- заїкання – дана патологія пов'язана з порушенням ритму мови, розтягуванням окремих звуків;
- дисфонія – дефект, пов'язаний з порушеннями в голосовому апараті.

Причинами поганої дикції у дітей можуть стати такі сукупності факторів, які можна поділити на 2 групи:

- Біологічні - до них відносяться порушення мови, викликані патогенними факторами, які діяли на організм ще в період внутрішньоутробного розвитку і в дитячому віці, наприклад важкі інфекції, травми. Сюди ж відносяться порушення, викликані генетичною схильністю. Прояви біологічних причин - заїкання, гугнявість голосу, загальмованість мови, проковтування окремих звуків.

- Соціальні - до соціальних причин, що призвели до порушення дикції, відносяться проблеми в спілкуванні з оточуючими, особисті комплекси та невпевненість у собі. Дитина, що сумнівається в своїх здібностях, говорить тихо і невиразно, «ковтає» звуки, намагаючись швидше закінчити розмову і піти в тінь.

Для реалізації майбутньої системи аналізу мовлення був розроблений алгоритм, представлений на рисунку 1. З рисунку 1 можна побачити, що на першому етапі підготовуються дані, з якими система буде працювати. Це один з найдовгіших етапів, оскільки від якості даних залежить ступень навчання розроблюваної математичної моделі майбутньої системи. Ці дані можна отримати з глобальної мережі, а саме з відео, в якому логопеди розмовляють або займаються з дітьми, у яких є порушення у мовленні.

Після того як підготовлені дані в черновому вигляді, їх потрібно очистити від зайвого, а саме залишити тільки фрази, які говорить дитина в аудіо форматі. Всі аудіофайли повинні зберігатися в одній частоті дискретизації, а саме оптимальна частота 44,1 кГц.

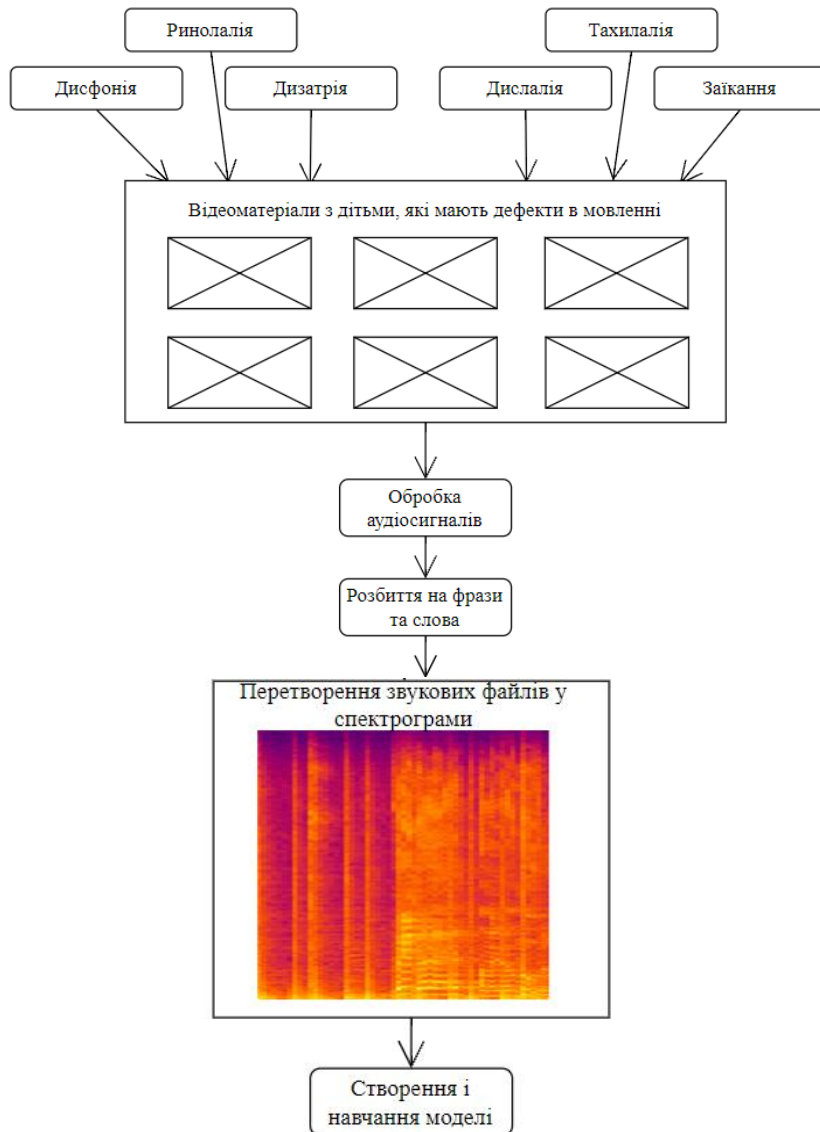


Рис.1. Загальний алгоритм обробки дефектів

Вхідними даними для системи, що розробляється є відеоматеріали з вадами мовлення у дітей. Сигнали з цих відеоматеріалів обробляються і розрізаються на фрази та слова. В подальшому ці аудіосигнали

перетворюються в спектрограму. Спектрограма являє собою зображення на якому можна бачити різницю спектра в усьому звуковому уривку відразу [10]. По вертикалі відкладається частота, по горизонталі час, а амплітуда зазвичай відображається яскравістю або кольором.

Після того як всі дані були підготовлені починається навчання математичної моделі. В якості математичного апарату виступає згортоква нейронна мережа. Оскільки саме вона показує найкращі результати при розпізнаванні. Оскільки моделі згорткових нейронних мереж можуть застосовуватися для розпізнавання зображення, тексту, відео або звуків. Можна виділити дві особливості, за якими часто вибирають саме згорткові нейронні мережі: CNN самі витягують ознаки; CNN можуть бути перенавчені для виконання нових завдань розпізнавання. Архітектура згорткової нейронної мережі, що була розроблена для поставленої задачі визначення виду мовного порушення приведена на рисунку 2.

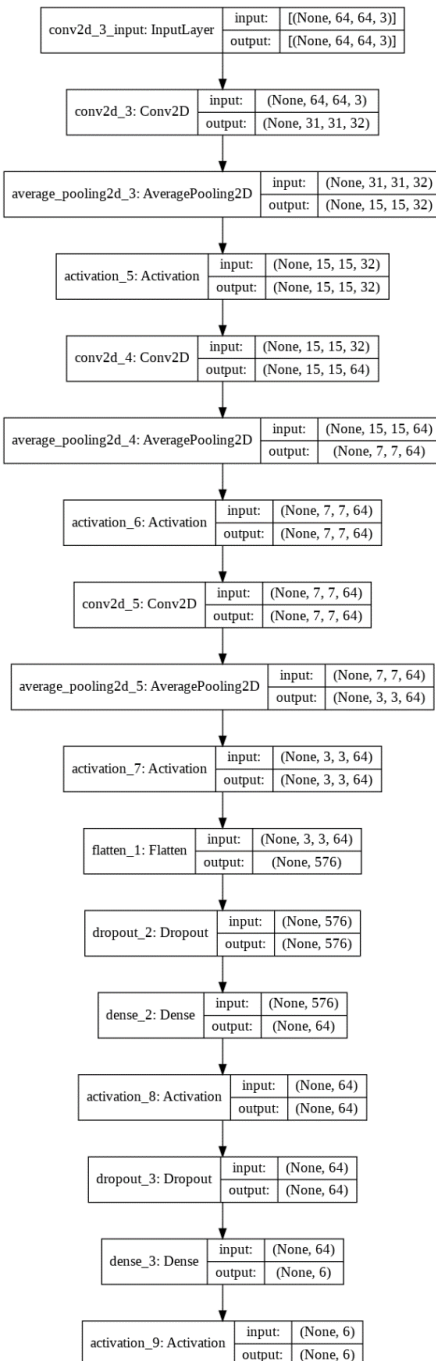


Рис. 2. Архітектура згорткової нейронної мережі

На рисунку 2 можна побачити, що архітектура складається з шести шарів. Перший шар виконує згортку Conv2D. За допомогою input_shape вказано, що вхід буде розміром 64 на 64 пікселі і в трьох каналах. Даний процес відбувається на наступних двох шарах. Для другого і третього прихованого шару вказано padding рівне значенню same, це було зроблено для того щоб згортка захопило повністю все зображення, тому на всі боки зображення додаються нулі в разі потреби.

У кожного шару відбувається процес зменшення розміру введення і витяг важливої інформації AveragePooling2D. Після відбувається обчислення виведення за допомогою активаційної функції «relu». Четвертий шар конвертує вхідні дані в меншу розмірність. Виконується Dropout для обнулення ваг моделі і передається значення 0.5, тобто половина ваг обнуляється. На п'ятому шарі знаходиться повнозв'язний шар Dense, який отримує всю інформацію з минулих шарів і потрібен для повного з'єднання шарів один з одним, на виході – Dropout, який потрібен для зменшення ймовірності перенавчання моделі. Останній шар є вихідним, у ньому використовується функція активації softmax для класифікації мовних порушень.

Так як використовується згорткова нейронна мережа, яка може добре виділяти ознаки на зображеннях, тому використовується перетворення аудіо сигналів в спектрограми.

Після підготовки dataset відбувається процес навчання моделі. Було проведено 60 різних експериментів, з різною кількістю епох навчання, вибірка з них наведено у таблиці 1. Метою експерименту біло визначити оптимальну кількість епох та оптимальний час, що буде витрачено на навчання моделі.

Таблиця 1. Результати дослідження нейронної мережі

№ експерименту	Час навчання моделі, с	Загальна точність моделі, %	Кількість епох навчання
1	3600,4	64,50	50
2	3450	63,92	60
3	3821	67,00	70
4	4150	69,46	80
5	4050	68,65	90
6	4450	71,32	100
7	4952,4	70,68	110
8	4950	67,87	120
9	4253,9	65,30	130
10	5250	68,61	140
11	5301	74,85	150
12	5700,6	67,52	160
13	5404	68,65	170
14	6402,8	65,12	180
15	6007	68,65	190
16	6707	63,76	200
17	7354,1	64,32	210
18	7754	67,12	220
19	7954	69,54	230
20	8154,5	70,12	240

З таблиці 1 видно, що найкращий експеримент йде під номером 11, так як вдалося натренувати модель на 74,85% точності, що для поставленої задачі є дуже хорошим результатом. Розглядачі інші експерименти можна побачити, що зі збільшенням кількості епох точність навчання не збільшується, але часу буде витрачено більше. Тобто експериментальним шляхом було встановлено, що 150 епох для навчання розробленої моделі бути оптимальними.

Наступний експеримент був проведений для вже навченої нейронної мережі. Він був проведений на реальних даних, які не були використані при навчанні моделі. Для цього дані були взяті в Інтернеті та відібрані відеозаписи із записами дітей, у яких різні порушення мови. Після чого використовуючи алгоритм представлений на рисунку 1 дані були перетворені в спектрограми.

Для кожного порушення мови було відібрано по 20 записів. Записи були для різної статі дітей. Тим самим експеримент також виявляв залежність розпізнавання дефекту мовлення від статі дитини. Результати представлені в таблиці 2.

Таблиця 2. Результат тестування моделі

Стать дитини	Вид порушення мовлення					
	Дисфонія	Ринолалія	Дизартія	Дислалія	Тахилалія	Заїкання
Жіноча	78%	73%	76%	77%	75%	76%
чоловіча	74%	75%	74%	75%	74%	74%

Як можна побачити з результатів таблиці 2, навчену модель можна використовувати незалежно від статі дитини. Оскільки для кожної статі точність розпізнавання мовного дефекту змінюється не суттєво. При цьому різницю точності можна віднести до похибки вимірювання, оскільки не всі обрані дані були в якісному виді та у високої якості. Тим самим експеримент підтвердив, що розроблений алгоритм, та математична модель що лягла в його основу є працездатною та може бути використана в розпізнавання мовного дефекту будь якого виду.

Висновки та перспективи подальшого дослідження. У результаті роботи була розроблена модель згорткової нейронної мережі та був продемонстрований алгоритм, яких буде покладено в основу системи, що дозволить коригувати дефекти мовлення дітей. Розроблений алгоритм містить в собі етапи обробки даних. Процес обробки полягав у пошуку і очищення звукових даних, які містять в собі цікаві звукові коливання вимови слів з порушеннями в мові. Наступний етап полягав у побудові згорткової нейронної мережі і навчанні моделі. Експериментальні дослідження для виявлення оптимальності архітектури згорткової нейронної мережі показав що для отримання більш хорошої якості навченої моделі потрібно було навчати модель на 150 епохах, в такому випадку точність моделі складає 74,85%.

Розроблена модель дозволяє визначити який саме дефект мовлення присутній є у дитини, коли вона розмовляє, та наскільки це порушення серйозне. В подальшому розроблений алгоритм та математичну модель планується реалізувати у вигляді системи підтримки коригування дефектів мовлення у дітей. В системі буде використовуватися відеоматеріали з вадами мовлення у дітей, а далі ці сигнали будуть оброблятися і розрізатися на фрази та слова Для корегування дефектів в мовленні дитини система буде підбирати необхідні вправи для вирішення конкретної проблеми. Таким чином правильний розвиток комунікативної функції мови буде сприяти розвитку навичок спілкування з оточуючими, що в свою чергу має велике значення для формування адекватної поведінки, емоційно-вольової сфери та особистості дитини.

Список бібліографічного опису

1. Survey on deep neural networks in speech and vision systems. M Alam, MD Samad, L Vidyaratne, A Glandon, KM Iftekharuddin. *Neurocomputing* 417, 302-321, 2020.
2. Deep Representation Learning in Speech Processing: Challenges, Recent Advances, and Future Trends. Siddique Latif1, Rajib Rana, Sara Khalifa, 2020, 25 pages.
3. Hendrik Purwins, Bo Li, Tuomas Virtanen, Jan Schlüter, Shuoyin Chang, Tara Sainath. Deep Learning for Audio Signal Processing. In *Journal of Selected Topics of Signal Processing*, Vol. 13, No. 2, May 2019, pages 206–219.
4. Automated speech-based screening of depression using deep convolutional neural networks. Karol Chlastaa,b, Krzysztof Wołka, Izabela Krejtz, 2019, 11 pages.
5. Shakeel A. Sheikh, Md Sahidullah, Fabrice Hirsch, and Slim Ouni. 2021. Machine Learning for Stuttering Identification: Review, Challenges & Future Directions. under review in *ACM Comput. Surv.* 54, 4 (July 2021), 27 pages.
6. FluentNet: End-to-End Detection of Speech Disfluency with Deep Learning. Tedd Kourkounakis, Amirhossein Hajavi, and Ali Etemad, 2020, 13 pages.
7. Music Genre Classification using Masked Conditional Neural Networks. Fady Medhat, David Chesmore, and John Robinson, 2018, 11 pages.
8. Supervised Speech Separation Based on Deep Learning: An Overview. DeLiang Wang, Fellow, IEEE, Jitong Chen, 2018, 27 pages.
9. Матвеев Н. Чинники впливу на мовленнєві порушення дітей молодшого шкільного віку / Н. Матвеева // Освітній простір України, 2019, 159 с.
10. S.Haykin. *Neural Networks and Learning Machines*. 3rd Edition. Pearson, 2018.

References

1. Survey on deep neural networks in speech and vision systems. M Alam, MD Samad, L Vidyaratne, A Glandon, KM Iftekharuddin. *Neurocomputing* 417, 302-321, 2020.
2. Deep Representation Learning in Speech Processing: Challenges, Recent Advances, and Future Trends. Siddique Latif1, Rajib Rana, Sara Khalifa, 2020, 25 pages.

3. Hendrik Purwins, Bo Li, Tuomas Virtanen, Jan Schlüter, Shuoyiin Chang, Tara Sainath. Deep Learning for Audio Signal Processing. In Journal of Selected Topics of Signal Processing, Vol. 13, No. 2, May 2019, pages 206–219.
4. Automated speech-based screening of depression using deep convolutional neural networks. Karol Chlastaa,b, Krzysztof Wolka, Izabela Krejtz, 2019, 11 pages.
5. Shakeel A. Sheikh, Md Sahidullah, Fabrice Hirsch, and Slim Ouni. 2021. Machine Learning for Stuttering Identification: Review, Challenges & Future Directions. under review in ACM Comput. Surv. 54, 4 (July 2021), 27 pages.
6. FluentNet: End-to-End Detection of Speech Disfluency with Deep Learning. Tedd Kourkounakis, Amirhossein Hajavi, and Ali Etemad, 2020, 13 pages.
7. Music Genre Classification using Masked Conditional Neural Networks. Fady Medhat, David Chesmore, and John Robinson, 2018, 11 pages.
8. Supervised Speech Separation Based on Deep Learning: An Overview. DeLiang Wang, Fellow, IEEE, Jitong Chen, 2018, 27 pages.
9. Matveyeav N. Chynnyky vplyvu na movlennyevi porushennya ditey molodshoho shkil'noho viku / N. Matveyeva // Osvitniy prostir Ukrainy, 2019, 159 s.
10. S.Haykin. Neural Networks and Learning Machines. 3rd Edition. Pearson, 2018.