DOI 10.34185/1562-9945-3-128-2020-05
УДК 004.4'277.2

В.Ю. Царик, Вікт.В. Гнатушенко

# ДОСЛІДЖЕННЯ МЕТОДІВ ВИЛУЧЕННЯ ВОКАЛУ У ЗМІКСОВАНИХ ЗАПИСАХ

*Анотація. Розглянута задача сліпого поділу сигналу, а саме, виділення вокальної доріжки з готового зміксованого запису. Метою дослідження є виділення характеристик вокального сигналу на підставі існуючих методів і програмних засобів. Проаналізовані існуючі методи виділення вокалу: методи частотної фільтрації, фазового віднімання та методи на основі систем штучного інтелекту. Проведено порівняльний аналіз роботи програмних засобів для ізоляції вокалу та методу фазового віднімання, що дозволило зробити висновки про недостатню ефективність існуючих методів ізоляції вокалу у зв'язку з неврахуванням особливостей тембру голосу в конкретній музичній композиції.*
*Ключові слова: сліпий поділ сигналу, деміксування, цифрова обробка сигналів.*

**Постановка проблеми.** У сучасному світі для музикантів і працівників звукової індустрії є актуальною задачею сліпий поділ сигналу. Вона полягає у виділенні початкового сигналу з суміші, тобто, розглядаючи область музики – це виділення доріжки одного інструмента з готового міксу. Подібна задача виникає в різних ситуаціях: для вокалістів – при необхідності виконати пісню, для якої відсутня інструментальна версія, для музикантів – при необхідності підібрати на слух партію конкретного інструменту, який зливається з іншими в міксі, для звукорежисера – під час запису живого колективу, коли в мікрофонах музичних інструментів присутні звуки сусідніх інструментів, які перешкоджають якісній обробці звукової доріжки і т.д.

Незважаючи на наявність великої кількості методів обробки сигналів, завдання деміксування на сьогоднішній день не вирішена, а спроби її вирішення дають на виході сигнали з великою кількістю спотворень, що робить неможливим їх подальше використання. Нижче наведені деякі причини, які перешкоджають вирішенню даної задачі:




ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

1. Кожен музичний інструмент і кожен голос має свій унікальний тембр, який може змінюватися в процесі гри, коли виконавець застосовує різні техніки звуковидобування.

2. У процесі зведення звукова доріжка конкретного інструменту / голосу піддається безлічі обробок: динамічним (компресія звуку), частотним (еквалізація), фазовим (накладення звукових ефектів). Так само, за задумом виконавців або звукорежисерів, дані обробки можуть змінюватися протягом композиції. Крім цього, вищевказані обробки застосовуються і до зведеного міксу на етапі мастерингу композиції.

3. Кожен музичний інструмент звучить в своєму діапазоні частот, але часто для декількох інструментів ці діапазони перетинаються або співпадають, що робить неможливим виділити інструмент, використовуючи частотну фільтрацію.

Таким чином, для ефективного вирішення задачі деміксування необхідно виділити характеристики, які притаманні конкретному інструменту або голосу в даній композиції. В даному дослідженні розглянуто методи ізоляції вокальної доріжки, так як це є найбільш затребуваною варіацією вищеописаної задачі. Метою дослідження є виділення характеристик вокального сигналу на підставі існуючих методів і програмних засобів.

**Основна частина**. *Акустичні характеристики голосу.* Людський голос складається із сукупності різноманітних за своїми характеристиками звуків, що утворюються за участю голосового апарату. У голосовому апараті людини виникають і тонові (що породжуються періодичними коливаннями джерела звуку з певною частотою), і шумові звуки (з'являються при безладних коливаннях різної фізичної природи). Всі голосні мають тоновий характер, а глухі приголосні – шумовий. Чим частіше відбуваються періодичні коливання, тим вище сприймається нами звук. Таким чином, висота звуку – це суб'єктивне сприйняття органом слуху частоти коливальних рухів. Частота основного тону вимірюється в герцах і може в звичайній розмовній мові у чоловіків змінюватися в межах від 85 до 200 Гц, а у жінок – від 160 до 340 Гц. [1]

Для характеристики голосу існує таке поняття, як тоновий діапазон – можливість продукувати звуки в певних межах від найнижчого тону до найвищого. Тонової діапазон співочого голосу індивідуальний, але співак повинен володіти голосом з діапазоном мінімум в дві октави. Відомі співаки, у яких діапазон досягає чотирьох і п'яти октав: вони можуть брати звуки від 43 Гц – найнижчі голоси до 2 300 Гц – високі голоси.

Тембр звуку є суттєвою характеристикою голосу, яка відображає акустичний склад, будову голосу. Кожен звук голосу складається з основного тону, що визначає його висоту, і численних додаткових обертонів більш високої, ніж основний тон, частоти. Частота обертонів в два, три, чотири і так далі раз більше, ніж частота основного тону.

*Методи виділення вокалу.* В силу того, що кожен голос унікальний, на сьогоднішній день не існує універсального способу витягти вокальну доріжку з готового міксу. Залежно від конкретного аранжування і конкретного голосу, різні методи можуть давати різні результати. Розглянемо основні методи, якими користуються на сьогоднішній день.

1. Частотна фільтрація вокалу

Даний метод заснований на пошуку діапазону частот, в якому знаходиться голос і прибиранні рівня гучності цього діапазону за допомогою еквалізації, тобто застосування різних видів фільтрів. [2] Використовуючи даний метод, стикаємося з двома основними проблемами. По-перше, голос може звучати на широкій полосі частот, що залежить як від виконавця, так і від конкретної композиції, тому необхідно визначати полосу в якій знаходиться вокал. А для досягнення кращого результату слід застосовувати різну фільтрацію протягом всієї композиції, адаптуючи межі полоси зрізу під вокальну партію. По-друге, в частотному діапазоні голосу так само знаходяться і інші інструменти, які придушуються разом з голосом. Таким чином, використовуючи даний метод, ми втрачаємо корисну інформацію аудіосигналу у вигляді частини музичних інструментів.

Можливості застосування даного методу. Використовуючи частотну фільтрацію, можна виділити інструменти, які розташовані у вузь-



ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

кому частотному діапазоні, а також не перетинаються по частотах з іншими інструментами. Прикладом таких інструментів може бути великий барабан, бас-гітара (які розташовані на низьких частотах), окремі звуки барабанів, які звучать на певній частоті, барабанне «залізо» (розташоване на високих частотах). Також даний метод застосовується в концертній та студійної звукорежисурі для фільтрації окремих звукових доріжок з метою видалення небажаних шумів і сторонніх звуків.

2. Метод фазового віднімання

На сьогоднішній день переважна більшість аудіозаписів є стереофонічними, тобто мають два канали – лівий і правий. Різні елементи аранжування розташовують в різних місцях панорами для створення більш реалістичного звучання і з метою запобігання перенасичення міксу. Але основні інструменти, такі як вокал, великий і малий барабан, бас-гітара, зазвичай розташовуються в центрі панорами, тобто присутні в обох каналах міксу в рівній мірі. Слід також зауважити, що на відміну від основного вокалу, який розташовується по центру, бек-вокали зазвичай розводять по панорамі, тобто їх звучання в двох каналах відрізняється між собою.

Метод фазового віднімання являє собою видалення центральної складової міксу, де і знаходиться основний вокал. [3] Головним недоліком даного методу є те, що в результаті з стереофонічного запису виходить монофонічна. Крім того, зараз дуже часто на вокальну доріжку накладають звукові стерео-ефекти, такі як реверберація, ділей і інші. Виділення їх даним методом не дає позитивного результату.

Алгоритм цього методу наступний: виконується інверсія фази одного з каналів звукового файлу і складається з іншим каналом. В результаті цього залишається один моно-канал, в якому відсутня загальна, центральна складова стерео-панорами. Недоліком такого методу є зміна вихідного рівня гучності деяких елементів міксу. Багато інструментів, такі як барабани і бас можуть повністю зникнути з міксу. Для уникнення такого ефекту слід комбінувати даний метод з описаним вище методом частотної фільтрації.

3. Методи з використанням штучного інтелекту

Зараз активно ведуться розробки програмного забезпечення для ізоляції вокалу на основі систем штучного інтелекту. Вже існують програмні продукти, які вирішують це завдання з результатами, які в рази перевищують описані вище методи, але все одно, на сьогоднішній день не можуть добитися ідеальної ізоляції вокалу. Розглянемо кілька прикладів програмних продуктів, в основі роботи яких лежать дані методи.

*Spleeter*. Розробка Але Корецького, керівника відділу машинного навчання компанії «Splice.com». Застосовувані розробником алгоритми розкриті в патентах US10014002B2 і US9842609B2. Реалізація представлена з відкритим вихідним кодом, написаному мовою Python. Робота можлива тільки в режимі командного рядка, так як відсутня реалізація графічного інтерфейсу. Дана система була представлена в лютому 2019 року і надавала можливість розділити вокальну та інструментальну доріжки композиції. Але розробки автора на цьому не припинилися і зараз стало можливим розділити мікс на 5 доріжок – вокал, ударні, бас, фортепіано і інші інструменти.

В основі даного алгоритму лежить нейронна мережа, яка визначає наявність голосу на довільному фрагменті звукозапису завдяки детектору голосової активності, який реалізований у вигляді бінарного класифікатора. Для цього застосовується згорточна нейронна мережа. На цьому етапі виявляються часові ділянки, де присутній голос. Наступний етап – на обраних часових ділянках розділити голос від музики. Тобто з спектрограми всього міксу виділити ділянки спектра, в яких лежить голос. Для цього застосовується метод вилучення вокалу з міксу з використанням бінарних масок, в якому вихід представляється як бінарне зображення, де значення “1” вказує на переважну присутність вокального контенту на заданій частоті і часовій області, а значення “0” вказує на переважне присутність музики в даному місці. Для навчання нейронної мережі використовувалося 15 млн зразків по 300 мс міксів і відповідних їм вокальних бінарних масок. [4]

Надалі аналогічний підхід застосовувався для виділення інших інструментів з міксу.



ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

*iZotope RX 7*. Пакет програм від компанії iZotope для реставрації аудіо. Нам цікавий модуль Music Rebalance, який дає можливість змінити баланс рівнів гучності інструментів в міксі. Алгоритм роботи даного модуля схожий з програмним продуктом, описаним вище. Music Rebalance використовує технологію машинного навчання, навчену розділяти різні музичні джерела в міксі. Звук обробляється через кілька нейронних мереж, кожна з яких навчена ідентифікувати та ізолювати певний музичний інструмент (вокал, бас або перкусію). Виходи цих нейронних мереж об'єднуються, щоб повідомити нам кількість конкретного джерела, присутнього в кожен момент часу і для кожної частоти вашого аудіо. Ця інформація використовується для адаптивної фільтрації аудіо таким чином, щоб ізолювати конкретний інструмент. Грунтуючись на цьому поділі, Music Rebalance дозволяє контролювати рівень гучності кожного з музичних джерел поряд із іншими джерелами та відносно їх початкової кількості в міксі. За словами розробників, нейронні мережі навчалися на безлічі міксів з використанням різних голосів, як ідеального вихідного сигналу застосовувався тільки сольний вокал з цього міксу. Крім цього, проводилося навчання на міксах з різним балансом інструментів, наприклад, використання одного і того ж вокального треку на різній гучності відносно рівня основного міксу. [5]

**Порівняльна характеристика програмних засобів для ізоляції вокалу.** З метою оцінки ефективності описаних вище методів був проведений порівняльний аналіз. Підготовлено набір прикладів для аналізу, для чого підібрані композиції в різних стилях музики і з різним наповненням музичних інструментів в аранжуванні. Використовувалася окремо вокальна доріжка і окремо інструментальна з метою подальшого порівняння з отриманими результатами. Один із прикладів обраний з використанням необробленої вокальної доріжки – без використання частотної, динамічної корекції і звукових ефектів. Вихідні звукові доріжки були зведені з метою отримання стереофонических міксів і надалі піддавалися обробці наступними методами: в програмі Spleeter, в програмі iZotope RX 7 і методом фазового віднімання.

Так як дані програмні засоби націлені в першу чергу на працівників звукової індустрії і музикантів, які в ході своєї діяльності спираються головним чином на свій слух, основним критерієм оцінювання результатів дослідження є суб'єктивна оцінка звуку результуючих файлів. Результати досліджень наведені в таблиці 1.

Таблиця 1

Результати порівняльного аналізу методів ізоляції вокалу

| Приклад і його характеристика | Spleeter | iZotope RX 7 | Фазове віднімання |
|---|---|---|---|
| Billie Eilish – Bad Guy<br>Характеристика: жіночий вокал, не насичене аранжування, багато бек-вокалів. | Вокал. Коли є бек-вокали – сильне спотворення, присутні артефакти на низьких частотах, клацання пальців віднесено до вокалу. Інструментал. Наявні залишки бек-вокалів, є спотворення на низьких частотах. | Вокал. Періодично вокал спотворюється, але менше, ніж в першому методі, артефактів на низьких частотах немає, клацання залишилися у вокальній дорожці. Інструментал. Менше залишків вокалу, наявні спотворення на низьких частотах. | Вокал. Вокали майже без спотворень. Частина бек-вокалів залишилося у вокальній доріжці, частина в інструментальній. Артефакти на низьких частотах. Інструментал. Менше спотворень, багато вокалу залишилося. |
| Bruno Mars – Uptown Funk<br>Характеристика: чоловічий вокал, аранжування насичене, багато інструментів в частотному діапазоні вокалу. | Вокал. Майже немає інших інструментів. Тембр спотворений – зрізані високі частоти голосу. Інструментал. В цілому – нормальний, артефактів від вокалу немає. Присутні спотворення на низьких частотах. Менше високих частот в порівнянні з оригіналом. | Вокал. Результат дуже схожий на попередній метод. Інструментал. Є артефакти від вокалу. Спотворення на низьких частотах. Немає провалу високих частот як в попередньому методі. | Вокал. Дуже багато сторонніх інструментів, наприклад, гітари. Інструментал. Чути тихо основний вокал, залишилися бек-вокали. В цілому, звук нормальний. |
| Clean Bandit – Rather Be<br>Характеристики : | Вокал. Дуже сильно спотворений звук. На високих нотах | Вокал. Результат схожий на попередній метод, | Вокал. Голос без спотворень, багато зайвих інструментів – |



ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

| | | | |
|---|---|---|---|
| Жіночий вокал. Насичене аранжування, багато сучасних звуків і стерео-ефектів. | іноді проступає скрипка. Інструментал. Є артефакти від вокалу, бас спотворений. Гучність іноді змінюється. Частина звуків бас-гітари і скрипок віднесло в вокал, тому на доріжці з інструменталом вони спотворені. | іноді вокал взагалі втрачається, голос сильно спотворений. Інструментал. Бас більш чистіший, є артефакти вокалу і ефектів. Ударні спотворені, скрипки іноді губляться. | клавіші, ударні, скрипки. Інструментал. Провал на низьких частотах, артефакти вокалу. Барабани спотворені. |
| One Repablic – Counting Stars Характеристики: поп-рок, більш природнє звучання, без безлічі ефектів. Чоловічий вокал. | Вокал. Тембр спотворений. Артефакти від ефектів. Інструменти не присутні. Інструментал. Бракує високих частот. Артефактів від вокалу немає. | Вокал. Тембр нестабільний: то з'являються високі частоти, то пропадають. Багато артефактів. Інструменти не присутні. Інструментал. Є артефакти від вокалу. | Вокал. Вокал не ізольований, дуже багато інструментів, які знаходяться в вокальному діапазоні. Інструментал. Вокал чути, хоч і приглушений. |
| Характеристики: На мінусовку з беками накладена необроблена вокальна доріжка, без ефектів, вокал строго в центрі панорами міксу. | Вокал. Вокальна доріжка спотворена, є артефакти інструментів. Інструментал. Є артефакти вокалу, звук спотворений. | Вокал. Вокальна доріжка спотворена, менше ніж в попередньому методі. Артефакти інструментів. Інструментал. Є артефакти вокалу. | Вокал. Багато залишків інструментів, вокал не ізольований. Інструментал. Частина ударних віднесена до вокальної доріжки, є залишки вокалу. |

**Висновки та перспективи подальших досліджень.** Проаналізовано та досліджено методи ізоляції музичних інструментів, зокрема вокалу, з зміксованих музичних записів, що дозволяє зробити наступні висновки. Оцінка результатів методів з використанням штучного інтелекту дають дуже схожі результати, однак для різних композицій кращими виявляються різні методи. У всіх випадках немає ідеального виділення вокальної лінії – або спотворюється тембр, або присутні призвуки від інших інструментів. Метод фазового віднімання в результаті дає моно-сигнал, що є великим недоліком, та не може відокремити во-

кал від інструментів, які перебувають в одному діапазоні і положенні в панорамі. Загальний недолік всіх методів полягає в тому, що вони не адаптуються під голос в конкретної музичної композиції. У зв'язку з цим, потрібне розроблення методу, який буде визначати характеристики тембру для конкретної композиції і виділяти доріжку з цим тембром.

**ЛІТЕРАТУРА / ЛИТЕРАТУРА**

1. Лаврова Е. В. Логопедия. Основы фонопедии [Текст] : учеб. пособие для студ. вузов, обучающихся по спец. – логопедия / Е. В. Лаврова. – М.: AcademiA, 2007. – 144 с. : ил. – (Высшее профессиональное образование: психология). – Библиогр.: с.139-142 . - ISBN 978-5-7695-3753-0

2. Сергиенко А.Б. Цифровая обработка сигналов. 3-е изд. — СПб.: БХВ-Петербург, 2011. — 768 с.: ил. — (Учебная литература для вузов).

3. В. Попченко. Борьба с фазовыми искажениями при микрофонной записи. [Електронний ресурс]. – Режим доступу: http://prosound.ixbt.com/exp/papchenko-phase.shtml

4. Ale Koretzky. Audio AI: isolating vocals from stereo music using Convolutional Neural Networks. [Електронний ресурс]. – Режим доступу: https://towardsdatascience.com/audio-ai-isolating-vocals-from-stereo-music-using-convolutional-neural-networks-210532383785

5. Hannah Robertson. Exploring the Technology that Makes RX 7 Music Rebalance Possible. [Електронний ресурс]. – Режим доступу: https://www.izotope.com/en/learn/exploring-the-technology-that-makes-rx-7-music-rebalance-possible.html

**REFERENCES**

1. Lavrova E.V. Speech therapy. Basics of phonopaedia [Text]: textbook. manual for university students enrolled in the specialty - speech therapy / E.V. Lavrova. – M.: AcademiA, 2007. – 144 p.: ill. – (Higher vocational education: psychology).– Bibliography: p.139-142. – ISBN 978-5-7695-3753-0

2. Sergienko A.B. Digital signal processing. 3rd ed. – SPb .: BHV-Petersburg, 2011 .– 768 p.: Ill. – (Textbooks for universities).

3. V. Popchenko. Fight phase distortion during microphone recording. [Electronic resource]. – Access mode: http://prosound.ixbt.com/exp/papchenko-phase.shtml

4. Ale Koretzky. Audio AI: isolating vocals from stereo music using Convolutional Neural Networks. [Electronic resource]. – Access mode:



ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

https://towardsdatascience.com/audio-ai-isolating-vocals-from-stereo-music-using-convolutional-neural-networks-210532383785

5. Hannah Robertson. Exploring the Technology that Makes RX 7 Music Rebalance Possible. [Electronic resource]. – Access mode: https://www.izotope.com/en/learn/exploring-the-technology-that-makes-rx-7-music-rebalance-possible.html

Received 12.02.2020.
Accepted 17.02.2020.

***Дослідження методів вилучення вокалу у зміксованих записах***

*Розглянуто існуючі методи ізоляції вокалу з зміксованого запису: метод частотної фільтрації, метод фазового віднімання та методи на основі систем штучного інтелекту. Проведено порівняльний аналіз роботи існуючих програмних засобів для вирішення даної задачі – Spleeter та iZotop RX7. Зроблено висновки про недостатню ефективність існуючих методів ізоляції вокалу.*

***Investigation of methods of vocal extraction in mixed records***

*In the modern world, the blind division of the signal is an urgent task for musicians and the audio industry workers. It is to isolate the source signal from the mix, that is, by looking at the music area, it is the selection of a single instrument track from a finished mix. Despite the presence of a large number of signal processing methods, the problem of demixing has not been solved to date, and attempts to solve it yield signals with many distortions at the output, which makes it impossible to use them further. The purpose of the research is to isolate the characteristics of the vocal signal on the basis of existing methods and software.*

*Due to the fact that each voice is unique, there is no universal way to extract a vocal track from the finished mix. Depending on the particular arrangement and the particular voice, different methods can produce different results.*

*The following methods of vocal isolation are described in this paper:*

*1. Frequency filtering of vocals;*

*2. Phase subtraction method;*

*3. Methods using artificial intelligence.*

*A comparative analysis was conducted to evaluate the effectiveness of these methods. A set of examples has been prepared for analysis, for which compositions in different styles of music and with different filling of musical instruments in arrangement are selected. Selected compositions were subjected to phase subtraction processing and two software products that operate on the basis of artificial intelligence systems: Spleeter and iZotope RX7.*

*Evaluating the results of methods using artificial intelligence give very similar results, but different methods are better for different compositions. In all cases, there is no perfect vocal line distortion – either distorted timbre or tones from other instruments. As a result, the phase subtraction method produces a mono-signal, which is a major drawback and cannot separate vocals from instruments in the same range and position in the panorama. A common disadvantage of all methods is that they do not adapt to the voice in a particular musical composition. In this regard, we need to develop a method that will determine the timbre characteristics for a particular composition and highlight the track with that timbre.*

**Царик Владислав Юрьевич** – аспирант, ассистент кафедры информационных технологий и систем Национальной металлургической академии Украины.
**Гнатушенко Виктория Владимировна** – д.т.н., профессор кафедры информационных технологий и систем Национальной металлургической академии Украины.

**Царик Владислав Юрійович** – аспірант, асистент кафедри інформаційних технологій і систем Національної металургійної академії України.
**Гнатушенко Вікторія Володимирівна** – д.т.н., професор кафедри інформаційних технологій і систем Національної металургійної академії України.

**Tsaryk Vladyslav** – postgraduate student, assistant professor, department of information technologies and systems, National Metallurgical Academy of Ukraine.
**Hnatushenko Viktoriia** – doctor of engineering's sciences, professor, department of information technologies and systems, National Metallurgical Academy of Ukraine.



ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)