

УДК 004.934

## ЗАСТОСУВАННЯ ТРІАДНИХ ВЕЙВЛЕТІВ ДЛЯ СЕГМЕНТАЦІЇ МОВНОГО СИГНАЛУ

Косоков С.О., Мальчиков В.В.

### APPLICATION OF TRIADIC WAVELETS FOR SPEECH SEGMENTATION

Kosogov S.O., Malchikov V.V.

*В роботі показано застосування тріадних вейвлетів до проведення сегментації мовного сигналу. Обґрунтовано вибір біортогональних вейвлетів для обробки мовлення. Виконано адаптацію існуючого алгоритму сегментації на основі вейвлет-перетворення до використання тріадних вейвлетів. Проведено порівняльний аналіз результатів сегментації окремих слів та цілих фраз російської мови, показано переваги застосування тріадних вейвлетів та недоліки використаного алгоритму. Визначені напрямки проведення подальших досліджень.*

**Ключові слова:** сегментація мовлення, вейвлет-перетворення, тріадні вейвлети

**Вступ** Сучасний етап науково-технічного прогресу неможливо уявити без використання мовних технологій. Мовний – природний для людини – інтерфейс дозволить полегшити виконання низки різноманітних завдань, від пошуку необхідної інформації до керування складними системами. Саме тому важливим напрямком досліджень є розробка інтелектуальних систем образного сприйняття мовної інформації, серед яких значну роль відіграють систем розпізнавання мови.

Якість розпізнавання істотно залежить від попередньої обробки сигналу, одним з етапів якої є сегментація – розбиття мовного сигналу на фрагменти.

Одним з підходів до сегментації мовного сигналу є використання вейвлет-перетворення для визначення міжфонемних переходів. На таких переходах вхідний сигнал сильно змінюється, що спричиняє зростання вейвлет-коефіцієнтів на декількох рівнях деталізації [1], що дає змогу виділити межі фонем.

На практиці в алгоритмах сегментації на основі вейвлетів застосовуються діадні вейвлет-перетворення, що обумовлене простотою їх

програмної реалізації. Водночас у багатьох галузях (наприклад, у обробці зображень) більш ефективними є алгоритми, що базуються на недіадних вейвлет-перетвореннях, коефіцієнт масштабування яких не дорівнює двом [2, 3]. На сьогоднішній день існує багато методів виконання недіадного вейвлет-перетворення, проте відсутні їх програмні реалізації із довільним коефіцієнтом масштабування. Найлегше отримати коефіцієнти фільтрів недіадного вейвлет-перетворення у випадку, коли масштабуючий коефіцієнт дорівнює трьом (тріадні вейвлети).

Мета даної роботи – провести порівняльний аналіз застосування тріадних та діадних вейвлетів до сегментації мовних сигналів.

**Вибір вейвлетного базису** При виконанні сегментації потрібно, щоб використовуваний вейвлетний базис був достатньо чутливим до локальних змін вхідного мовного сигналу. З іншого боку, алгоритм сегментації не повинен знаходити межі фонем у шумі, тобто потрібно забезпечити невеликі коефіцієнти вейвлет-розкладення при плавній зміні вхідного сигналу. Згідно з [4], таким вимогам відповідають біортогональні вейвлетні базиси. Візуальне порівняння зображень біортогонального сплайн-вейвлета та ділянки звукового файлу з записом мовлення російською дає можливість стверджувати, що біортогональний вейвлетний базис з урахуванням зсуву та масштабування краще «вкладається» у локальні ділянки російськомовного сигналу, ніж вейвлетні базиси Хаара та Добеші.

Таким чином, для проведення дослідження доцільно обрати біортогональні діадний та тріадний базиси. Коефіцієнти фільтрів тріадного біортогонального вейвлет-перетворення обчислені в роботі [3].

**Адаптація алгоритму сегментації** В якості алгоритму сегментації на основі вейвлет-перетворення був обраний алгоритм, описаний в [5]. Він характеризується простотою програмної реалізації та відносно невеликою обчислювальною складністю. Проте він потребує адаптації для використання із триадними вейвлетами.

В першу чергу модифікації потребує розмір ковзного вікна. Використаний у [5] розмір вікна у 512 відліків зумовлений необхідністю проведення децимації при виконанні вейвлет-перетворення сигналу і є найближчим ступенем двійки, що більше за середню довжину фонем (при обраній частоті дискретизації).

У випадку триадних вейвлетів можна обрати наступні варіанти:

- розмір вікна буде представляти собою мінімальний степінь трійки, який перевищує середню довжину фонем у відліках;
- половина вікна буде представляти собою мінімальний степінь трійки, який перевищує середню довжину фонем у відліках.

При обраній частоті дискретизації довжина вікна в першому варіанті складає 729 відліків. Оскільки у вікно такого розміру можуть потрапити одночасно дві фонем, то в роботі використовується другий варіант, згідно з яким довжина напіввікна складає 243 відліки.

При використанні діадного вейвлет-перетворення для кожного рівня розкладення обчислюється два набори коефіцієнтів – апроксимуючі та деталізуючі. На відміну від діадного, на кожному рівні триадного вейвлет-розкладення формуються три набори коефіцієнтів – один апроксимуючий та два деталізуючих. Відповідно сегментацію можна проводити наступними способами:

- враховується тільки один із двох наборів деталізуючих коефіцієнтів;
- враховуються обидва набори деталізуючих коефіцієнтів, при цьому границя сегменту проставляється, коли відповідна умова виконується для обох наборів;
- враховуються обидва набори деталізуючих коефіцієнтів, при цьому границя сегменту проставляється, коли відповідна умова виконується принаймні для одного з наборів.

В двох останніх способах порогове значення може як співпадати, так і відрізнитися для обох наборів деталізації.

В даному дослідженні використано третій спосіб при однаковому пороговому значенні.

Внаслідок високої чутливості біортогонального вейвлету до локальних змін у досліджуваному сигналі і базовий алгоритм [5], і запропонована модифікація можуть визначати «зайві» сегменти, тобто близькорозташовані сегменти, довжина яких

дорівнює напіввікну. В роботі пропонується у таких випадках об'єднувати такі сегменти з попереднім.

**Результати** Для проведення порівняльного аналізу в середовищі комп'ютерної математики Matlab було розроблено програмне забезпечення (ПЗ) для виконання сегментації звукових файлів. Створене ПЗ виконує відкриття звукових файлів у форматі «wav» із частотою дискретизації 22050 Гц та розрядністю 16 біт, визначення меж фонем із використанням обраних діадних і триадних вейвлетних базисів та відображає амплітудно-часове представлення досліджуваного сигналу з проставленими межами сегментів. Для проведення вейвлет-розкладення у діадному та триадному випадку використовувалося ПЗ, що розроблене викладачами кафедри прикладної математики НТУУ «КПІ».

Для порівняння використовувалися окремі слова та фрази, взяті з аудіокниг та (у разі потреби) переконвертовані у потрібний формат.

Процес порівняння проводився аудіовізуальним шляхом. Для результатів сегментації спочатку виконувалося візуальне порівняння розставлених меж сегментів, а потім отримані сегменти прослуховувалися в аудіоредакторі.

На рис. 1 наведено результат сегментації слова «ТЯЖКО» за допомогою діадного (верхня частина) та триадного (нижня частина) вейвлетів. При використанні діадного вейвлету виділяється п'ять сегментів, а триадного – шість. Обидва алгоритми виділяють фонем *т'*, *а*, *ш*, причому остання поділяється на два сегменти. Проте фонем *к* та *а* в кінці слова виділяються лише за допомогою триадного вейвлету, у той час як діадний поєднує ці дві фонем в один сегмент.

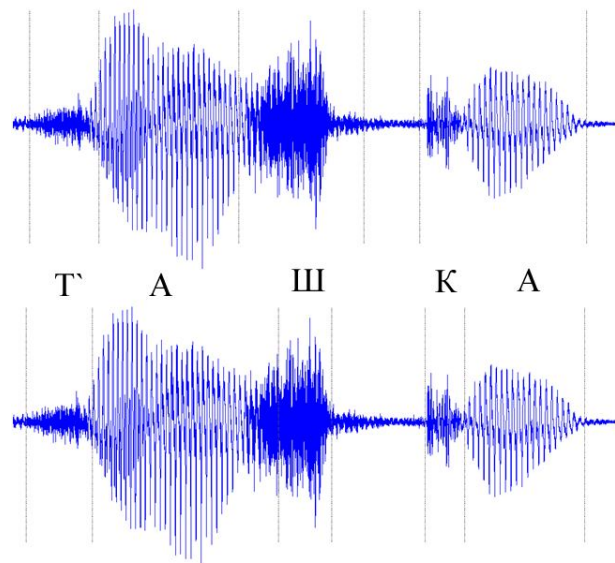


Рис. 1. Сегментація слова «ТЯЖКО»

На рис.2 зображено результат сегментації слова «ПОКРАСНЕЛ». В обох випадках виділяються

фонем *n* та *a*. Остання в діадному випадку представлена двома сегментами. Чітко відокремлюються наступні фонем *к*, *р*, *а* та *с*. У діадному випадку отри останні фонем *н*, *е* та *л* представлені одним сегментом, в той час як у триадному випадку вони виділені в окремий сегмент кожна.

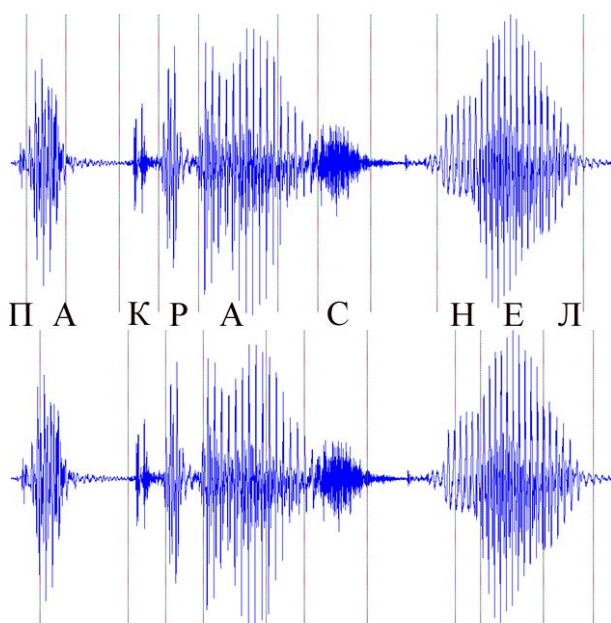


Рис. 2. Сегментація слова «ПОКРАСНЕТЬ»

На рис. 3 представлені результати сегментації слова «ПРОГОНИТЕ». Триадний вейвлет виділив в окремі сегменти фонем *п*, *р*, *а*, *г*, *о* та *н*. Фонем *и* та *т'* представлені одним сегментом. Остання фонема *э* розбита на два сегменти. В діадному випадку не розділені фонем *р* та *а*; *о*, *н* та *и*. При цьому фонем *т'* та *э* поділені на велику кількість сегментів.

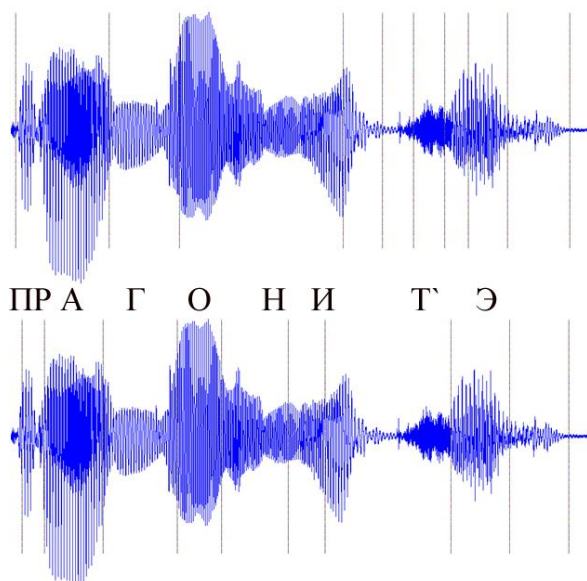


Рис. 3. Сегментація слова «ПРОГОНИТЕ»

На рис. 4 представлені результати сегментації фрази «ЭДУАРД УСПЕНСКИЙ» за допомогою триадного біортогонального вейвлету на другому (верхня частина) та третьому (нижня частина) рівнях розкладення.

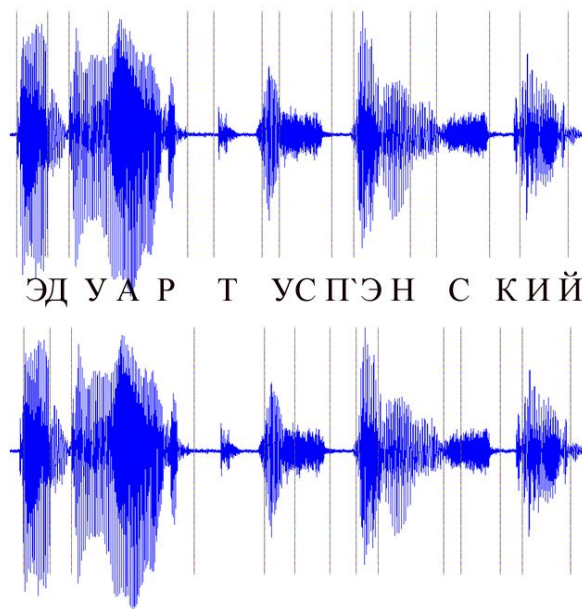


Рис. 4. Сегментація фрази «ЭДУАРД УСПЕНСКИЙ»

В першому випадку у слові «ЭДУАРД» у відділені від *а*, тоді як у другому цього відділення немає. Натомість при роботі на третьому рівні розкладення в слові «УСПЕНСКИЙ» чітко виділяється фонема *э*, чого не відбувається при роботі на другому рівні.

Виходячи з цього, доцільним можна вважати не виділяти при сегментації один з рівнів вейвлет-розкладення як інформативний, а працювати на декількох рівнях. Крім того, в подальшому можна провести дослідження для пошуку залежності сегментації конкретних фонем або їх видів від рівня вейвлет-розкладення.

**Висновки.** В роботі авторами виконано порівняльний аналіз застосування для сегментації фонем діадних та триадних біортогональних вейвлетів. Сегментація виконувалася як для окремих слів, так і для цілих фраз.

Отримані результати показують, що в більшості випадків застосування триадного вейвлет-перетворення дозволяє відокремити голосну від приголосної. Проте іноді діадне вейвлет-перетворення показує себе з кращого боку.

В процесі аналізу була виявлена залежність сегментації сусідніх фонем в окремі сегменти від рівня, який обрано в якості інформативного.

В подальшому доцільно модифікувати алгоритм для одночасного використання декількох рівнів вейвлет-розкладення з метою сегментації фонем різних класів. Іншим напрямком досліджень може бути комплексне застосування діадного та триадного вейвлет-перетворень.

## Література

1. Каркульовський В.І., Ткаченко В.С. Особливості методів сегментації мовленнєвих сигналів // Вісник Національного університету "Львівська політехніка". Комп'ютерні системи проєктування. Теорія і практика. – 2009. – № 651. – с.144-148.
2. Pollock D.S.G. Non-dyadic wavelet analysis / D.S.G. Pollock, I.L. Cascio // Optimization, Econometric and Financial Analysis: Advances in Computational Management Science, Kontoghiorghes E.J. and Gatu C. (eds.). — Springer Verlag. — 2007. — Vol. 9. — P. 167–204.
3. Чертов О.Р. Недіадні одновимірні вейвлет-перетворення / О.Р. Чертов // Наук. вісті НТУУ "КПІ" . — 2010. — № 2. — С. 63-73.
4. Малла С. Вэйвлеты в обработке сигналов. – М.: Мир, 2005. – 671 с.
5. Ермоленко Т., Шевчук В. Алгоритмы сегментации с применением быстрого вейвлет-преобразования [Электронный ресурс] // Труды международной конференции "Диалог 2003", М., 2003. – Режим доступа: <http://www.dialog-21.ru/Archive/2003/Ermolenko.htm>

## References

1. Karkul'ovs'kij V.I., Tkachenko V.S. Osoblivosti metodiv segmentacii movlennjevih signaliv // Visnik Nacional'nogo universitetu "L'vivs'ka politehnika". Komp'juterni sistemi proektuvannja . Teorija i praktika. – 2009. – № 651. – s.144-148.
2. Pollock D.S.G. Non-dyadic wavelet analysis / D.S.G. Pollock, I.L. Cascio // Optimization, Econometric and Financial Analysis: Advances in Computational Management Science, Kontoghiorghes E.J. and Gatu C. (eds.). — Springer Verlag. — 2007. — Vol. 9. — P. 167–204.
3. Chertov O.R. Nediadni odnovimirmi vejvlet-peretvorennja / O.R. Chertov // Nauk. visti NTUU "KPI" . — 2010. — № 2. — S. 63-73.
4. S. Mallat A wavelet tour to of signal processing, 2nd Edition, Academic Press, 1999
5. Ermolenko T., Shevchuk V. Algoritmy segmentacii s primeneniem bystrogo vejvlet-preobrazovanija [Elektronnij resurs] // Trudy mezhdunarodnoj konferencii "Dialog 2003", M., 2003. – Rezhim dostupu: <http://www.dialog-21.ru/Archive/2003/Ermolenko.htm>

## Косоков С.А., Мальчиков В.В. Применение триадных вейвлетов для сегментации речевого сигнала

*В работе рассматривается применение триадных вейвлетов к сегментации речевого сигнала. Обосновано использование биортогональных вейвлетов. Произведена адаптация существующего сегментации на основе вейвлет-преобразования для использования триадных вейвлетов. Проведен сравнительный анализ результатов сегментации отдельных слов и целых фраз русской речи, показаны преимущества применения триадных вейвлетов и недостатки использованного алгоритма. Определены направления для дальнейших исследований.*

**Ключевые слова:** сегментация речи, вейвлет-преобразование, триадные вейвлеты

## Kosogov S.O., Malchikov V.V. Application of triadic wavelets for speech segmentation

*This article deals with application of triadic wavelets for the speech segmentation problem. All wavelet-based methods usually used dyadic wavelet. But application of non-dyadic can give better results. Using of biorthogonal wavelets is substantiated. An algorithm based on dyadic wavelets was adopted for application of triadic wavelets. A comparative analysis of the segmentation of the single words and whole phrases is made. Advantages of using triadic wavelets and disadvantages of algorithm are shown. The directions of the future researchs are determined.*

**Keywords:** speech segmentation, wavelet transform, triadic wavelet

**Косоков С.О.** – магістр кафедри прикладної математики НТУУ «КПІ»

**Мальчиков В.В.** – старший викладач кафедри прикладної математики НТУУ «КПІ»

Рецензент: **Молчанов О.А.** – д.т.н., професор, завідувач кафедри прикладної математики НТУУ «Київський політехнічний інститут».

Стаття подана 13.04.2013