УДК 004.934

Г.О. Добрушкін, В.Я. Данилов

ЗАСТОСУВАННЯ ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ ДЛЯ СЕГМЕНТАЦІЇ І ВИДАЛЕННЯ ШУМУ З МОВНИХ СИГНАЛІВ

Вступ

Значна увага до розпізнавання природної мови підтверджується великою кількістю публікацій у наукових виданнях і високим рівнем представництва, введенням відповідних рубрик на спеціалізованих науково-технічних конференціях та створенням літніх шкіл. Розпізнавання природної мови є дуже складною технічною задачею, розв'язання якої знаходиться на стику багатьох галузей науки.

У своїх дослідженнях щодо вейвлетів, опублікованих у 1992 р., Інгрід Добеші вважає [1], що людське вухо влаштовано так, що при обробці звукового сигналу передає мозку вейвлет-образ сигналу. Коливання амплітуди тиску передаються від барабанних перетинок на мембрану і далі поширюються по всій довжині завитка внутрішнього вуха. Завиток скручений у вигляді спіралі у внутрішньому вусі. Якщо уявити, що він розпрямлений у певний сегмент, а також і розпрямлена мембрана, то можна побачити, що результуюче перетворення сигналу буде з точністю до константи збігатися з вейвлет-перетворенням. Звичайно, зараз є набагато кращі моделі для моделювання процесу сприймання звукових коливань людським вухом, а найбільш прогресивна на сьогодні модель ґрунтується на моделі мел-частотних кепстральних коефіцієнтів [2, 3], однак це не применшує здатності апарату вейвлет-перетворення до обробки мовних сигналів. Так, в дослідженні [4] було проведено серію експериментів із застосування вейвлет-перетворення до розпізнавання мови. Експерименти проводилися окремо для розпізнавання неперервного мовлення та ізольованих слів. Кількість помилок розпізнавання ізольованих слів була в п'ять-шість разів меншою за кількість помилок при розпізнаванні злитої мови. Цей експеримент, звичайно, наштовхує науковців на пошук розв'язку нової задачі — задачі сегментації неперервного мовлення.

Постановка задачі

Дане дослідження спрямоване на розв'язання задач автоматичної сегментації мовного сигналу на фонеми та автоматичного видалення з нього шумів і артефактів. Метою статті є формалізація методів видалення шумів і сегментації мовлення з використанням потужного апарату вейвлетного перетворення.

Автоматична сегментація мовлення

Алгоритми сегментації мовлення умовно поділяють на три класи [5]. До першого класу належать алгоритми, які сегментують мовний сигнал за умови наявності апріорного знання щодо послідовності фонем, що вимовлені в даному мовному сигналі. До другого класу відносяться алгоритми, які визначають межі сегментів за ступенем зміни акустичних характеристик сигналу і які не використовують апріорну інформацію про послідовність фонем. Третій клас становлять алгоритми, які приймають рішення як на основі апріорної інформації про фонеми, так і на основі аналізу змін акустичних характеристик сигналу.

У системах автоматичного розпізнавання мовлення важливим завданням є сегментація мовного сигналу відповідно до фонетичної транскрипції мови. Для побудови якісної системи розпізнавання мовлення перед етапом власне розпізнавання мовного сигналу необхідно виконати його первинну обробку. Перш за все з мовного сигналу слід видалити ділянки лише з шумом і залишити ділянки, які мають корисний мовний сигнал. Після цього важливо розв'язати задачі розбиття злитої мови на слова, а слова, в свою чергу, сегментувати на фонеми. Відомо, що мовний сигнал складається з квазістаціонарних ділянок, котрі чергуються з ділянками, на яких швидко змінюються спектральні характеристики сигналу [6, 7]. Тому з погляду на сегментації на фонеми задача полягає в пошуку міжфонемних переходів. Традиційними способами (перетворення Фур'є) вирішити цю проблему досить складно, адже перетворення Фур'є відображає мовний сигнал у частотній області і повністю втрачає інформацію про час, що з огляду на нестаціонарність мовного сигналу призводить до значних втрат інформації про сигнал. Вейвлет-перетворення за своєю побудовою відображає сигнал у частотно-часовій області, чим може значно полегшити

задачу пошуку міжфонемних переходів. Суть полягає в тому, що мовний сигнал на стику фонем, які дуже різняться за артикуляцією, зазнає значних змін, що в термінах вейвлет-перетворення характеризується значним підвищенням вейвлет-коефіцієнтів на багатьох масштабах дослідження. В той самий час, зміни коефіцієнтів вейвлет-перетворення на стаціонарних ділянках мовного сигналу згруповані поблизу невеликої кількості масштабів.

Теорія вейвлет-перетворення

Неперервне вейвлет-перетворення формально записується таким чином:

$$\gamma(s,\tau) = \int f(t)\overline{\psi}_{s,\tau}(t)dt .$$

Обернене вейвлет-перетворення, що відновлює початковий сигнал $f(t)$, виглядає так, як

$$f(t) = \iint \gamma(s,\tau)\psi_{s,\tau}(t)d\tau ds ,$$

де $\psi_{s,\tau}(t)$ – дочірні вейвлети, побудовані з материнського вейвлету $\psi(t)$ за допомогою масштабування та зсуву:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right),$$

де s – масштаб; τ – зсув; $\frac{1}{\sqrt{s}}$ – множник, що нормує енергію вейвлету на різних масштабах. Материнським вейвлетом може бути будь-яка функція, не обов'язково аналітично задана, що має вигляд хвилі, є локалізованою за часом і частотою, задовольняє умову *обмеженості* і умову *регулярності* (має нульове середнє і нульові перші m моментів). Для кожної задачі, що планується бути розв'язаною за допомогою вейвлет-перетворення, необхідно підбирати або створювати власний материнський вейвлет.

У реальних задачах для обробки мовних сигналів необхідно застосовувати дискретне вейвлет-перетворення (ДВП):

$$f(t) = \sum_{j,k}\gamma(j,k)\psi_{j,k}(t) . \qquad (1)$$

Але для ідеального розкладу сигналу знадобився б дуже великий набір вейвлетів $\{\psi(t)\}$, тому розклад у вигляді (1) на практиці майже не використовується. Для зменшення набору вейвлетів у базисі застосовуються добре відомі *смугові фільтри*. Завдяки їм можливо замість нескінченної кількості вейвлетів низькочастотну смугу закрити "пробкою" – низькочастотним фільтром, що в ДВП зображається функцією масштабування $\varphi(t)$, яку, в свою чергу, можна розкласти також за вейвлетним базисом (1). Треба зазначити, що коли до спектра вейвлету додати спектр функції масштабування, то отримаємо нову функцію масштабування, спектр якої вдвічі ширший за спектр попередньої функції. Таким чином, першу функцію масштабування можна записати в термінах другої, другу – в термінах третьої і так далі. Формально матимемо

$$\varphi(2^j t) = \sum_k h_{j+1}(k)\varphi(2^{j+1}t-k) .$$

Перша функція масштабування замінює собою набір вейвлетів, і, таким чином, можна виразити вейвлети з цього набору за допомогою функцій масштабування

$$\psi(2^j t) = \sum_k g_{j+1}(k)\varphi(2^{j+1}t-k) .$$

Оскільки вхідний сигнал $f(t)$ може бути виражений за допомогою вейвлетів аж до масштабу $j-1$, то його можна записати за допомогою функції масштабування аж до масштабу j:

$$f(t) = \sum_k \lambda_j(k)\varphi(2^j t-k) .$$

Переходячи в цьому рівнянні до масштабу $j-1$, щоб не втратити рівень деталізації, необхідно додати вейвлети

$$f(t) = \sum_k \lambda_{j-1}(k)\varphi(2^{j-1}t-k) + \\ + \sum_k \gamma_{j-1}(k)\psi(2^{j-1}t-k). \qquad (2)$$

Якщо функції масштабування і вейвлети ортогональні[1], то коефіцієнти $\lambda_{j-1}(k)$, $\gamma_{j-1}(k)$ записуються у вигляді дискретного скалярного добутку:

$$\lambda_{j-1}(k) = \langle f(t),\varphi_{j,k}(t)\rangle = \sum_m h(m-2k)\lambda_j(m),$$

$$\gamma_{j-1}(k) = \langle f(t),\psi_{j,k}(t)\rangle = \sum_m g(m-2k)\gamma_j(m).$$

[1] При використанні ортогонального набору вейвлетів $\{\psi(t)\}$ вейвлет-перетворення стає інваріантним до зсуву, тобто вейвлет-перетворення сигналу, зсунутого в часі, є зсуненим вейвлет-перетворенням вихідного сигналу.

Саме вейвлет-перетворення у вигляді (2) частіше за всі інші застосовується на практиці. Якщо продовжити зменшувати масштаб, додаючи вейвлети, то (2) можна спрощено записати таким чином:

$$f(t) = \sum_k \lambda_i(k)\varphi_{i,k}(t) + \sum_k \sum_{j=i}^{\infty} \gamma_j(k)\psi_{j,k}(t).$$

У такому розкладі $\lambda_j(k)$ називаються коефіцієнтами апроксимації, а $\gamma_j(k)$ – коефіцієнтами деталізації.

Застосування вейвлет-перетворення для очищення мовного сигналу від шумів

Очистити мовний сигнал від шумів за допомогою вейвлет-перетворення нескладно. Основна ідея полягає в нехтуванні невеличкими порівняно з іншими коефіцієнтами вейвлет-перетворення мовного сигналу. На рис. 1, *а* наведено осцилограму слова “вода”, записану на якісному електретному мікрофоні і перетворену в цифрове зображення 16-бітним аналогово-цифровим перетворювачем (АЦП) з частотою дискретизації 11025 Гц.

Візуально помітно, що на сигнал впродовж всієї його довжини накладено низькочастотний шум. Відношення сигнал/шум у цьому прикладі становить 32 дБ. На рис. 1, *б* зображено частину вхідного сигналу (відліки 7500–10100), на якій наведено осцилограму чистого шуму, що накладається на весь сигнал. Цей шум створюється сторонніми звуками, як наприклад, звуками роботи механізмів, що знаходяться в приміщенні поряд з диктором, і не несе жодної інформації про мовний сигнал. Для якісного розпізнавання і сегментації мовлення цей шум необхідно видалити із всього сигналу і з мінімальними спотвореннями корисної частини сигналу.

У наших дослідженнях для очищення сигналу від шуму використовувався вейвлетний базис Добеші 10-го порядку. Заради експерименту також використовувався вейвлет Хаара, але він дав дещо гірший результат, адже за своєю будовою він не цілком підходить для моделювання мовних сигналів. Якщо далі по тексту окремо не вказано, за яким вейвлетним базисом розкладено сигнал, то його розкладено по вейвлетному базису Добеші 10-го порядку.

На рис. 2, *а* наведено коефіцієнти деталізації 1-го рівня розкладу вхідного сигналу по вейвлетах Добеші 10-го порядку. На рис. 2, *б* подано частину вейвлет-коефіцієнтів, якими зображено чистий шум.

Основний етап шумоочищення полягає в знаходженні тих коефіцієнтів деталізації вейвлет-розкладу, модуль яких менше за деякий поріг. Вибір рівня порога є індивідуальним для кожної задачі, але загалом залежить від модуля вейвлет-коефіцієнтів корисного сигналу і рівня деталізації, на якому проводиться пошук. Пошук відбувається окремо для кожного рівня деталізації. Знайдені коефіцієнти деталізації замінюються нулями і мовний сигнал син-

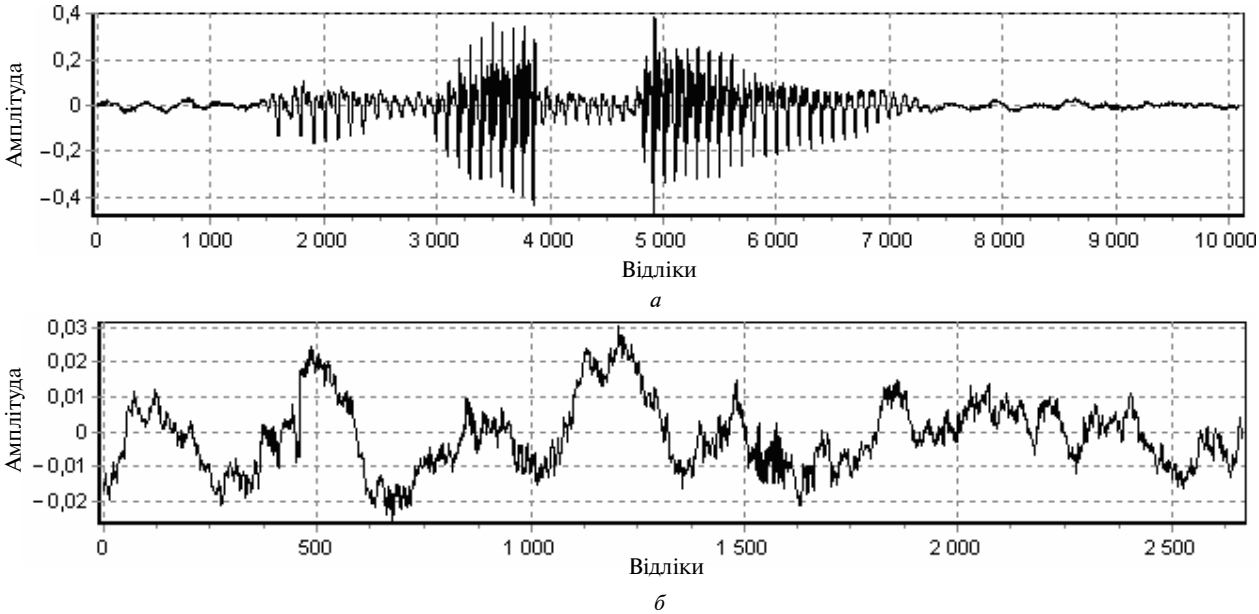

Рис. 1. Осцилограма слова “вода”: *а* – повна осцилограма; *б* – осцилограма чистого шуму

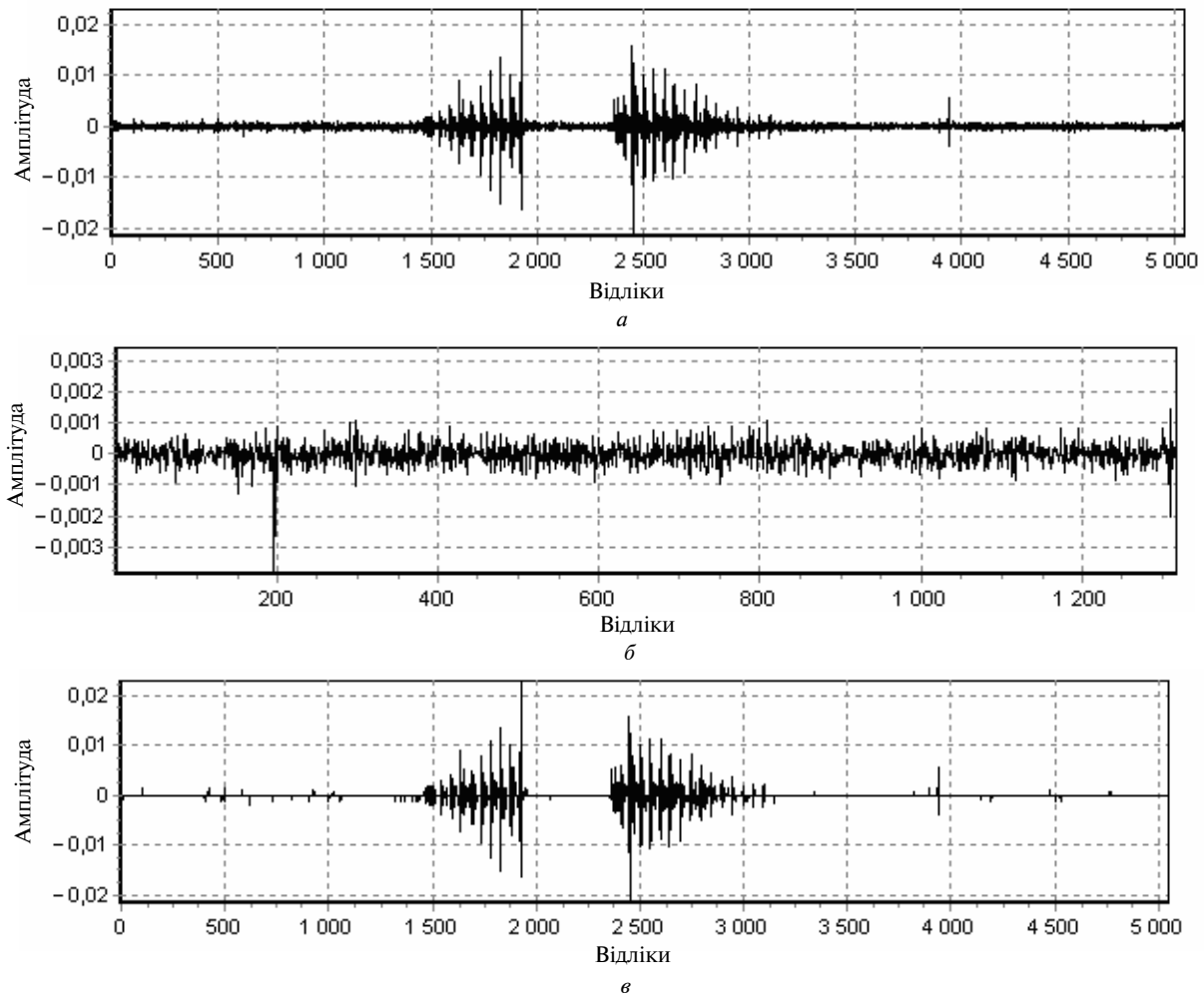

Рис. 2. Коефіцієнти деталізації 1-го рівня вейвлет-перетворення слова “вода”: *а* – весь сигнал; *б* – зона шуму; *в* – весь сигнал після знешумлення

тезується оберненим вейвлет-перетворенням. Коефіцієнти деталізації 1-го рівня вейвлет-перетворення після шумоочищення зображено на рис. 2, *в*. Відношення сигнал/шум після цієї процедури становило 37 дБ. Вибір більшого порога, звичайно, допоміг би збільшити відношення сигнал/шум, але водночас підвищив би ймовірність спотворення корисного сигналу.

Застосування вейвлет-перетворення для очищення мовного сигналу від артефактів АЦП

Електронний шум виникає практично у всіх електронних пристроях. Його джерелами можуть бути неоднорідності в провідному середовищі, генерація і рекомбінація носіїв заряду в транзисторах, неякісний мікрофон та акустичне обладнання, навколишній шум, неякісний аналогово-цифровий перетворювач тощо. Під час розробки системи розпізнавання мовлення ми зіткнулись з тим, що АЦП вносив у сигнал завади, які акустично було чути як “клацання”. Було виявлено, що природа цих артефактів бере початок у конструкції самого АЦП, в якому бракувало додаткового буфера для паралельної обробки даних. В АЦП використовувався єдиний буфер, і під час зчитування з нього даних частина вхідного сигналу, який неперервно надходив на вхід АЦП, спотворювалася. Звичайно, найпростішим методом вирішення цієї проблеми є вибір якісного аналогово-цифрового перетворювача, але метод, розроблений нами, можна з успіхом використати і для очищення сигналу від інших електронних шумів, яких не позбавлений кожний мовний сигнал.

На рис. 3, *а* наведено осцилограму слова “євро”, а на рис. 3, *б* частину цієї осцилограми, в якій акустично чути “клацання” (відліки 240–260 на рис. 3, *б*).

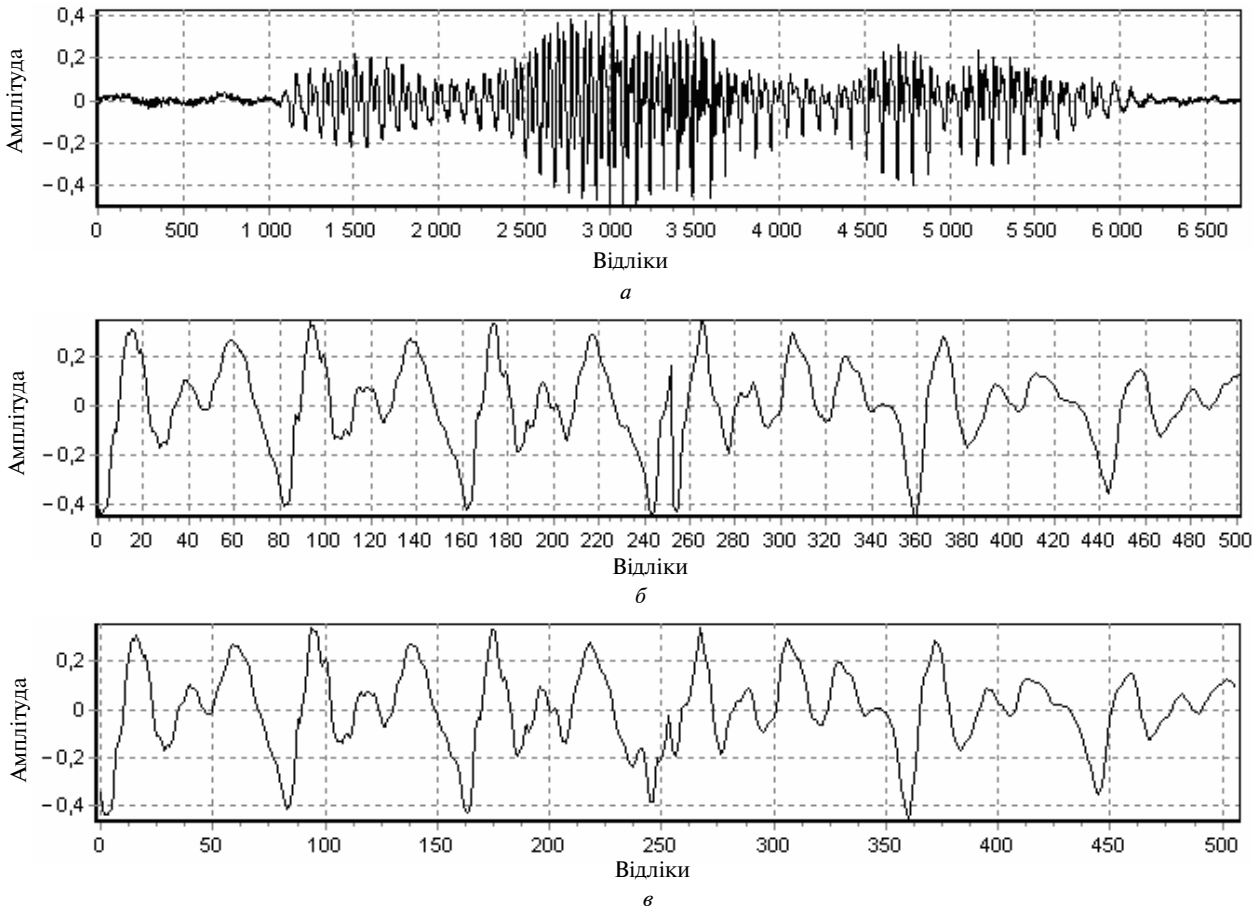

Рис. 3. Осцилограма слова “євро”: *а* – повна осцилограма; *б* – частина осцилограми, спотворена неякісним АЦП; *в* – осцилограма спотвореної частини після процедури відновлення

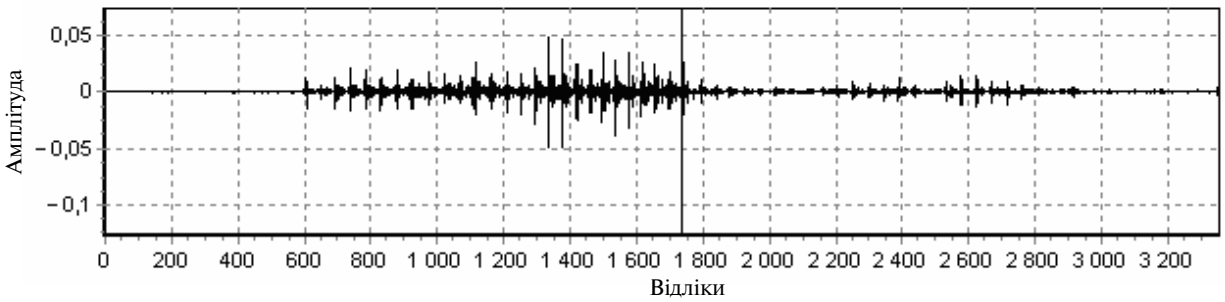

Рис. 4. Коефіцієнти 1-го рівня вейвлет-перетворення, спотвореного неякісним АЦП слова “євро”

Незважаючи на те, що на осцилограмі неозброєним оком можна побачити місце “спотворення” сигналу, автоматизувати процес пошуку місця спотворення в рамках традиційних підходів неможливо. Спектрограма також не може дати інформацію про місце спотворення, адже перетворення Фур’є оперує лише амплітудно-частотними характеристиками і повністю втрачає інформацію про часову складову сигналу. Розв’язання цієї задачі полягає в застосуванні вейвлет-перетворення, оскільки воно зберігає як частотну, так і часову інформацію про вхідний сигнал. Розклавши сигнал по вейвлетах Добеші 10-го порядку і поглянувши на коефіцієнти деталізації 1-го рівня розкладу (рис. 4), неозброєним оком можна побачити місце “клацання”. Автоматизувати алгоритм локалізації таких сплесків також нескладно. Сплески вейвлет-коефіцієнтів ми видаляли окремо для кожного з перших чотирьох рівнів деталізації. На подальших рівнях ці спотворення не проявлялися, адже вони високочастотні, а із збільшен-

ням рівня розкладу вейвлет-перетворення "акцентує" свою увагу на все нижчих частотах.

Після очищення вейвлет-коефіцієнтів від миттєвих сплесків і синтезування сигналу за допомогою оберненого вейвлет-перетворення акустичні "клацання" зникли, хоч візуально сигнал не відновився на сто відсотків (рис. 3, *в*, місце спотворення – відліки 240–260).

Сегментація мовного сигналу за допомогою вейвлет-перетворення

Задача сегментації мовного сигналу на фонеми досить складна. Вейвлет-перетворення значно допомагає в її розв'язанні, але повної автоматизації процесу досягти досить складно. Як зазначалося на початку статті, з погляду на сегментацію мовного сигналу на фонеми задача полягає в пошуку міжфонемних переходів. Мовний сигнал на стику фонем, що за артикуляцією дуже різняться, зазнає неабияких змін, що характеризується значним підвищенням вейвлет-коефіцієнтів на багатьох масштабах дослідження. В той же час, зміни коефіцієнтів вейвлет-перетворення на стаціонарних ділянках мовного сигналу згруповані поблизу невеликої кількості масштабів.

Основним питанням є те, на якому саме рівні розкладу шукати зміни вейвлетних коефіцієнтів? Універсального розв'язання цієї задачі досягти не вдалося.

Перший крок сегментації полягає *у виборі* такого *вейвлетного базису*, розклад по якому найкраще описує мовний сигнал (кількість ненульових елементів розкладу мінімальна). Вибір адекватного базису не є тривіальним, адже для кожного класу фонем цей базис різний і без наявності апріорного знання про сигнал, що розпізнається, вибрати базис досить складно. Вирішення цієї проблеми, на нашу думку, полягає в переборі можливих вейвлетних базисів і підрахунку або кількості ненульових коефіцієнтів деталізації, або рівня інформаційної ентропії сигналу відносно до базису:

$$\exp\left(-\sum_{j,k}\gamma_j(k)^2\log(\gamma_j(k)^2)\right).$$

Чим менша ентропія, тим краще сигнал описується вейвлетним базисом [8].

Якщо сегментацію проводити на окремому рівні деталізації, то ***другим кроком*** буде *вибір* конкретного *рівня деталізації*.

Третім кроком є підрахунок і *згладжування* центрованим ковзним середнім *квадратів коефіцієнтів деталізації* $\{\gamma_j(k)^2\}_k$ вибраного рівня. Для осцилограми слова "вода", використаної як тестовий приклад, згладжені квадрати коефіцієнтів деталізації другого рівня розкладу зображено на рис. 5, *а*.

Четвертим кроком є *прийняття рішення* про те, в якому місці фонеми розпочинаються, закінчуються чи змінюють одна одну. Одним із варіантів виконання цього кроку є детектування місця, в якому виконується умова

$$\left(\frac{\tilde{\gamma}_j(i)}{\max\limits_k\{\tilde{\gamma}_j(k)\}}-\alpha\right)\left(\frac{\tilde{\gamma}_j(i+1)}{\max\limits_k\{\tilde{\gamma}_j(k)\}}-\alpha\right)<0, \quad (3)$$

де α – поріг детектування, який дорівнював 0,01; $\tilde{\gamma}_j(k)$ – згладжені центрованим ковзним середнім квадрати коефіцієнтів деталізації вейвлет-перетворення на *j*-му рівні дослідження:

$$\tilde{\gamma}_j(k) = \frac{1}{n_j}\sum_{i=-n_j/2}^{n_j/2-1}\gamma_j(k+i)^2,$$

$$k=\frac{n_j}{2},\ldots,N_j-\frac{n_j}{2}+1,$$

де N_j – загальна кількість коефіцієнтів деталізації на *j*-му рівні масштабування; n_j – розмір вікна ковзання на *j*-му рівні масштабування.

У наших експериментах n_j розраховувався ітераційно за такою схемою:

$$n_0=100 \quad \forall j\ge 1: n_j=\left[\frac{n_{j-1}}{2}\right].$$

Виконання умови (3) свідчить про те, що у вхідному сигналі міжфонемний перехід знаходиться в моменті часу, який відповідає відліку i.

Під час прийняття рішення також враховувався той факт, що фонеми мають мінімальну тривалість промовляння, яку ми прийняли в 50 мс, тобто, якщо знайдені межі фонем містяться на відстані менш ніж 50 мс одна від одної, то вони об'єднувались в одну межу. Результат сегментації слова "вода" за цим критерієм зображено на рис. 5, *б*. Слово "вода" складається з двох складів – "во" (відліки 1500–4000 на осцилограмі) і "да" (відліки 4800–7200).

Для розпізнавання мовлення часто достатньо знайти лише межі фонем і виділити їх з неперервного мовлення. Не потрібно намагатися сегментувати самі фонеми. Але, тим не менш, склад "во" можна чітко розділити на фонему "в" (відліки 1500–3000) і вибухове "о"

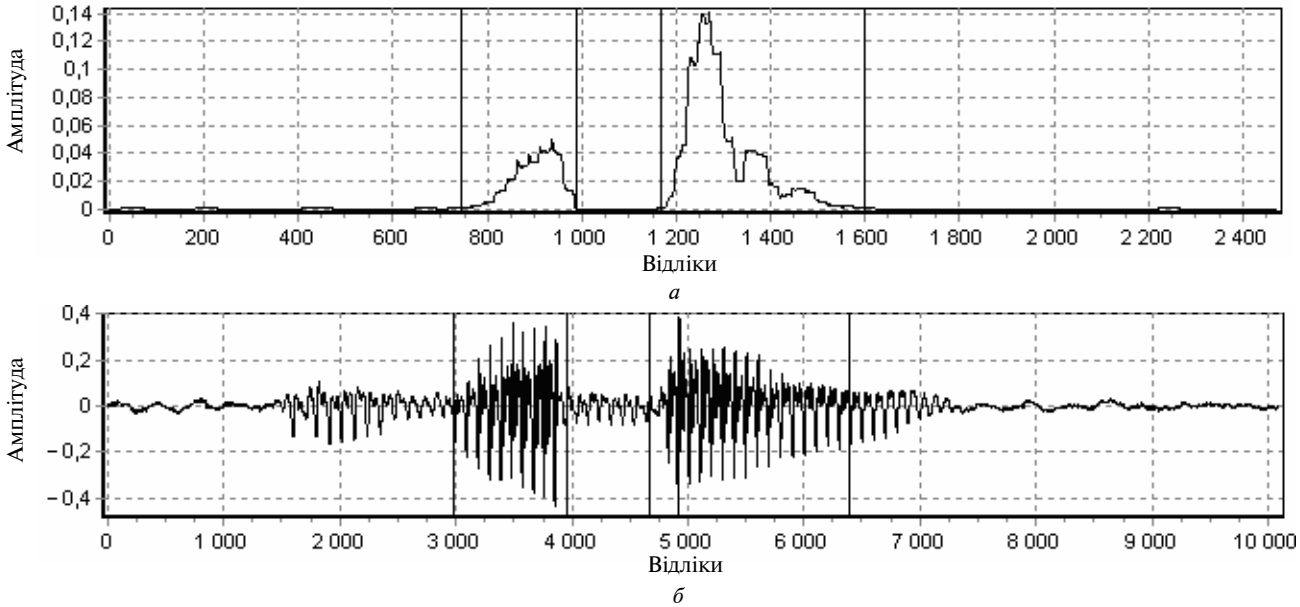

Рис. 5. Згладжені квадрати коефіцієнтів деталізації 2-го рівня розкладу осцилограми слова “вода” (*а*) і сегментація осцилограми за цими коефіцієнтами (*б*)

(відліки 3000–4000). Склад “да” сегментувати на фонеми не так легко, але завдяки дзвінкій вибуховій приголосній “д” його неважко виділити з неперервного мовлення.

Отже, інформація з 2-го рівня розкладу надала нам можливість сегментувати ослилограму слова “вода” на “в”, “о”, “да”, “а”. На цьому прикладі одразу проявляються “недоліки”: склад “во” було сегментовано, але алгоритм не надав ніякої інформації про фонему “в”, а фонема “а” із складу “да” була сегментована на дві частини.

Експеримент показав, що фонема “в” у складі “во”, за наведеним вище критерієм, “проявляється” лише на 5-му рівні розкладу. Слід очікувати, що всі губні приголосні будуть проявляти себе лише на високих рівнях деталізації, адже саме ці рівні відображають смугу низьких частот. Сегментацію по 5-му рівню розкладу подано на рис. 6, *б*, а згладжені квадрати коефіцієнтів деталізації – на рис. 6, *а*. Напівжирними вертикальними лініями помічено міжфонемні переходи, які були “склеєні” через їх близькість, як було сказано вище.

Як бачимо, основною проблемою такого методу сегментування є те, що апріорно невідомо, на якому рівні масштабування необхідно шукати важливу, з погляду на сегментування, інформацію.

Четвертий крок алгоритму (*прийняття рішення*) є дуже важливий. Інколи для сегментування мовлення краще працює логарифмічний, а не ковзний середній критерій знаходження міжфонемних переходів. У таких випадках для прийняття рішення використовуються такі параметри:

$$d_i = 10\log\left(\sum_{k=0}^{n-1}\gamma_j(i+k)^2\right),$$

де n – розмір вікна; $\gamma_j(k)$ – коефіцієнти деталізації вейвлет-перетворення на j-му рівні дослідження.

А міжфонемні переходи знаходяться за умовою

$$|d_i - d_{i-1}| < \beta,$$

де β – рівень детектування, який підбирається експериментально і є індивідуальним для кожної задачі. Виконання цієї умови свідчить про те, що у вхідному сигналі міжфонемний перехід знаходиться в моменті часу, який відповідає відліку i, тобто фактично прийняття рішення про наявність у конкретному місці міжфонемного переходу ухвалюється за наявності перепаду квадратів поряд стоячих вейвлетних коефіцієнтів у $10^{(\beta/10)}$ разів.

На рис. 7, *б* зображено результат сегментування слова “вода” по наведеному вище логарифмічному критерію за коефіцієнтами 2-го рівня деталізації вейвлет-перетворення, а на рис. 7, *а* наведено послідовність $\{d_i\}$. Параметри такі: $\beta = 1{,}5$; $n = 50$.

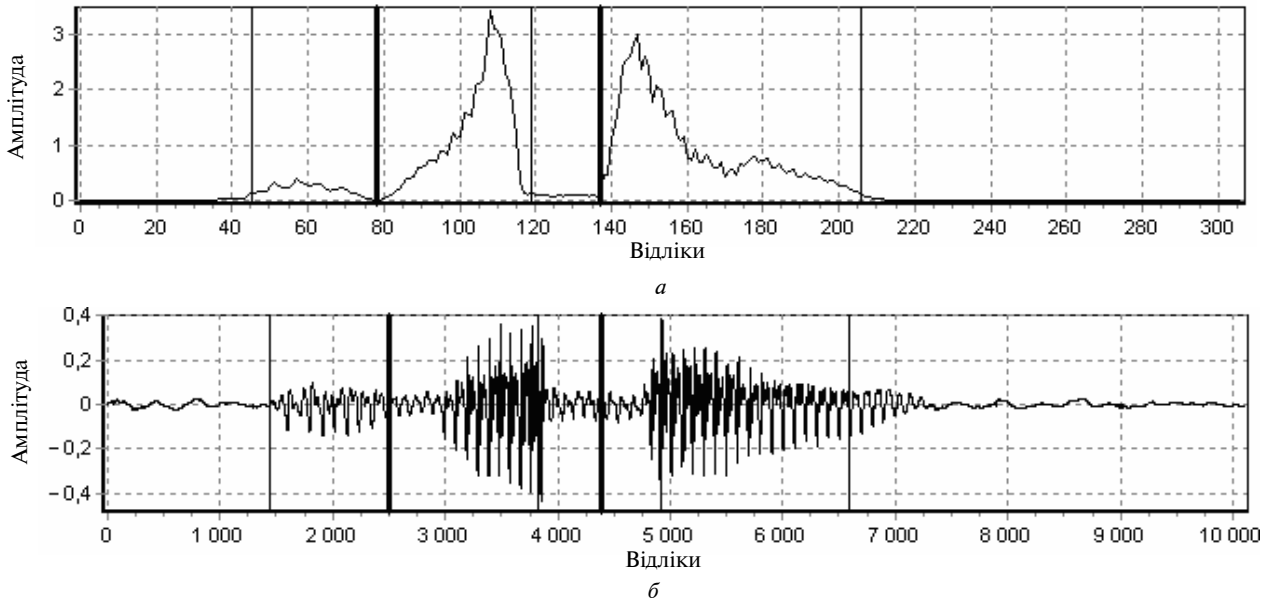

Рис. 6. Згладжені квадрати коефіцієнтів деталізації 5-го рівня розкладу осцилограми слова “вода” (*а*) і сегментація осцилограми за цими коефіцієнтами (*б*)

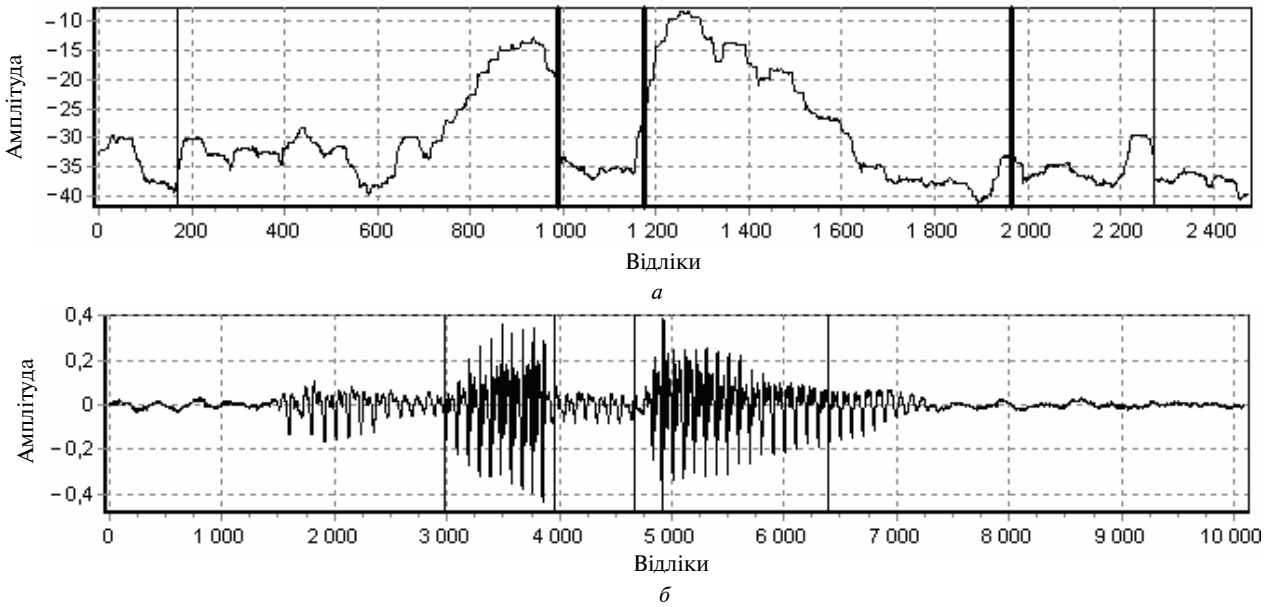

Рис. 7. Логарифми згладжених квадратів коефіцієнтів деталізації 2-го рівня розкладу осцилограми слова “вода” (*а*) і сегментація осцилограми за цими логарифмічними коефіцієнтами (*б*)

Висновки

Методи очищення мовних сигналів від шумів і артефактів АЦП розв’язують поставлену задачу і вже використовуються нами в реальних системах. За допомогою наведених вейвлетних методів у мовному сигналі без його спотворення вдається трохи збільшити відношення сигнал/шум, чим полегшується подальша задача розпізнавання мовлення. Це збільшення не таке істотне (лише до 35–40 дБ), але слід враховувати, що експерименти проводилися не в спеціальній кімнаті із спеціальною музичною апаратурою, а у звичайному приміщенні із звичайним конденсаторним мікрофоном.

Запропоновані логарифмічний і ковзносередній критерії допомагають у прийнятті рішення про сегментацію мовного сигналу і дають можливість розв’язати задачу лише за умови наявності апріорного знання про інформа-

тивність рівнів вейвлетного розкладу сигналу. При відсутності апріорної інформації про те, який рівень вейвлетного розкладу за конкретним вейвлетним базисом буде найбільш інформативним з точки зору пошуку конкретного міжфонемного переходу залежно від класу і природи фонеми, автоматизація методу здається важкою. Подолання цього обмеження, на наш погляд, полягає в реалізації паралельного пошуку за багатьма класами фонем на всіх рівнях вейвлетного розкладу, однак, з іншого боку, це може зменшити швидкість обробки, яка є важливим фактором у розв'язанні задач цього типу.

Г.А. Добрушкин, В.Я. Данилов

ПРИМЕНЕНИЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ ДЛЯ СЕГМЕНТАЦИИ И УДАЛЕНИЯ ШУМА ИЗ РЕЧЕВЫХ СИГНАЛОВ

На основе элементов теории вейвлет-преобразования представлен автоматизированный метод очистки речевого сигнала от некоторых шумов, помех и артефактов, которые появляются в результате использования АЦП. Предложен новый метод нахождения межфонемных переходов в непрерывном речевом сигнале, а также логарифмический и скользяще-средний критерии, которые помогают принимать решения о сегментации речевого сигнала. Проанализированы преимущества и недостатки предложенных критериев, показана сложность автоматизации решения задачи сегментации речевого сигнала при отсутствии априорного знания о характере фонем, переход между которыми необходимо детектировать.

G.O. Dobrushkin, V.Ya. Danilov

USE OF DISCRETE WAVELET-TRANSFORM FOR SEGMENTATION AND SPEECH SIGNALS DENOISING

Based on the components of the wavelet-transform theory, we present an automated method of speech signals denoising, interference cleaning and ADC artifacts refining. We propose a new method of detecting interphoneme transitions in the continuous speech signals as well as logarithmic and moving average criteria that help make a decision on the speech signals segmentation. In addition, we analyze advantages and disadvantages of the proposed criteria. We also show the complexity of automation of the problem solving of speech signal segmentation under the conditions of absence of a priori knowledge on phonemes natures and transition between them.

1. *Daubechies I.* Ten Lectures on Wavelets // CBMS-NSF Series in Applied Mathematics, T. 61. – Philadelphia: SIAM Publications, 1992. – 357 p.
2. *Zheng F., Zhang G., Zhanjiang Song J.* Comparison of Different Implementations of MFCC // Computer Science & Technology. – 2001. – **16**, N 6. – P. 582–589.
3. *Ganchev T.* Comparative evaluation of various MFCC implementations on the speaker verification task // 10th International Conference on Speech and Computer SPECOM'2005, 17–19 жовтня 2005. – Patras, Greece, 2005. – **1**. – P. 191–194.
4. *Бойков Ф.Г.* Применение вейвлет-анализа сигнала в системе распознавания речи // Матер. XI Всерос. конф. "Математические методы распознавания образов", 23–29 ноября 2003. – Пущино, 2003. – С. 247–250.
5. *Сорокин В.Н., Цыплихин А.И.* Сегментация и распознавание гласных // Информ. процессы. – 2004. – **4**, № 2. – С. 202–220.
6. *Фант Г.* Акустическая теория речеобразования. – М.: Наука, 1964. – 284 с.
7. *Рабинер Л., Гоулд Б.* Теория и применение цифровой обработки сигналов. – М.: Мир, 1978. – 848 с.
8. *Дремин И.М., Иванов О.В., Нечитайло В.А.* Вейвлеты и их использование // Усп. физ. наук. – 2002. – **171**, № 5. – С. 465–501.

Рекомендована Радою
Навчально-наукового комплексу
"Інститут прикладного системного
аналізу" НТУУ "КПІ"

Надійшла до редакції
8 липня 2009 року