УДК 621.39

О. М. Ткаченко, Н. О. Біліченко, О. В. Дзісь
Вінницький національний технічний університет
Хмельницьке шосе, 95, 21021 Вінниця, Україна

Сегментація мовленнєвих сигналів на основі алгоритму Вітербі

Розглянуто питання сегментації мовлення при створенні баз мовленнєвих даних. Запропоновано використовувати для попередньої сегментації алгоритм Вітербі у поєднанні з методами розпізнавання. Для підвищення точності визначення границь сегментів запропоновано використовувати статистичну інформацію про тривалість фонем. Отримані результати може бути використано при розробці україномовної бази мовленнєвих даних.

Ключові слова: *сегментація, алгоритм Вітербі, розпізнавання мовлення, мовленнєві дані, розмітка.*

Вступ

Дослідження в таких галузях як розпізнавання, ущільнення та синтез мовлення потребують накопичування великої кількості мовленнєвих фрагментів, що супроводжується описом відповідних деталей цих фрагментів (розміткою). Створення та розмітка достатньо повних мовленнєвих баз даних є однією з головних передумов успішного розвитку сучасних мовленнєвих технологій.

Відсутність розміченої фонетичної бази даних великого обсягу для українського мовлення зумовлює труднощі в процесі навчання та оцінювання якості систем автоматичного розпізнавання та синтезу мовлення, фонетичних вокодерів тощо. Більшість експериментів у цих дослідженнях виконується на іншомовному матеріалі — найчастіше це англомовні фонетичні бази, зокрема TIMIT [1]. Проте такий підхід не дозволяє врахувати особливості, притаманні українському усному мовленню. Тому існує необхідність створення аналогічної бази даних для української мови. Для цього потрібно розв’язати дві задачі:

1) записати мовленнєвий матеріал у достатній кількості та якості;

2) розмітити цей матеріал на окремі фонетичні елементи (фонеми).

Очевидно, що перша задача носить суто технічний характер, і для її вирішення потрібно лише час та ресурси. Що ж стосується другої, то вона потребує застосування досить складних підходів. По суті виділення у мовленні окремих елементів (фонем, слів, фраз тощо) є задачею сегментації. Підходів до її вирішення існує

достатньо багато. Наприклад, у [2] запропоновано використовувати метод виділення монологічних складових з метою сегментації мовлення на ділянки, вимовлені різними дикторами. В основі методу лежить пошук відмінностей між сусідніми ділянками. У роботі [3] пропонується алгоритм сегментації мовленнєвого сигналу на групи вокалізованих і невокалізованих звуків.

Взагалі найпоширенішою класифікацією алгоритмів сегментації є їхній поділ на два типи. До першого типу відносяться алгоритми, що працюють за умови відомої послідовності фонем у фразі. Алгоритми другого типу не використовують апріорну інформацію про фразу [4].

У роботі [5] охарактеризовано такі методи сегментації як:

— сегментація на основі обчислення значення енергії в заданому часовому вікні;

— сегментація голосних звуків на основі спектрального аналізу;

— сегментація на основі кореляції між спектрами фрагментів сигналу однакової тривалості;

— сегментація з використанням алгоритмів швидкого вейвлет-перетворення;

— сегментація на основі використання штучних нейронних мереж.

Ці алгоритми не використовують відомості про послідовність фонем у фразі, оскільки орієнтовані на сегментацію невідомого потоку мовлення.

Взагалі вибір алгоритму сегментації зумовлений особливостями та специфікою вирішуваної задачі, а саме типом сегментів, що потрібно виділити, наявною інформацією про сигнал і фразу, лімітом ресурсів і часу.

Вважається, що найкращий результат сегментації можна отримати лише за допомогою кваліфікованих спеціалістів з фонетики. Та не варто забувати, що ручна розмітка — це досить складна та трудомістка процедура, яка вимагає багато часу й ресурсів. Ситуація ускладнюється тим, що одній людині з великим обсягом записаних аудіоданих явно не впоратися, тому для отримання результату потрібна праця багатьох спеціалістів. Зрозуміло, що і в такому випадку процес буде довготривалим і високовартісним.

На рис. 1 показано фрагмент розміченої вручну фрази (слово «sun») з БД ТІМТ.

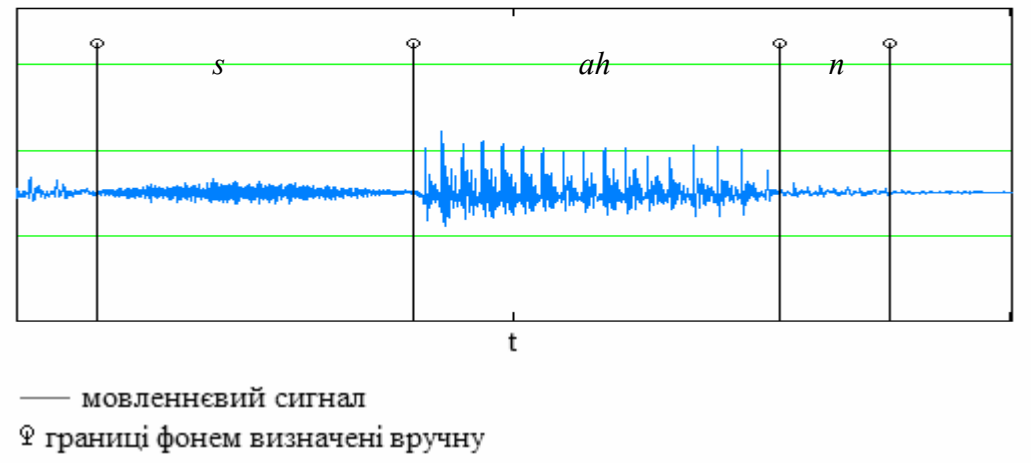

Рис. 1. Фрагмент з ручною розміткою

Наведений рисунок наглядно ілюструє суть сегментації. Варто відзначити, що для визначення границі, спеціалісту потрібно кілька разів прослуховувати частину фрази, поступово уточнюючи її межі, порівнювати її з сусідніми фразами і т.д. У результаті, на визначення однієї границі може піти приблизно 0,5–3 хвилини. У фразі в середньому буває 30–40 фонем, на розмітку яких загалом треба в середньому 15–40 хвилин. А таких фраз у БД може налічуватися від кількох тисяч до кількох десятків тисяч [6].

З іншого боку, існують автоматичні системи сегментації, які здатні виконати цю роботу практично без людини. Але це зазвичай призводить до значних неточностей і відхилень, оскільки на характер вимови впливає величезна кількість факторів, таких як тембр голосу, настрій, оточуюче середовище, фізичні особливості диктора і навіть погода. Врахувати певною мірою це все може лише досвідчений лінгвіст.

Тому для розв'язання задачі сегментації доцільно знайти компромісний варіант, який би дозволив поєднати інтелект і досвід спеціаліста та потужність обчислювальної системи. Таким варіантом є створення автоматизованої системи сегментації, котра б виконувала попередню розмітку, яка згодом без особливих зусиль могла бути уточнена людиною.

Метою роботи є підвищення точності встановлення границь між фонемами при сегментації мовлення.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

— проаналізувати існуючі підходи до сегментації мовлення;

— удосконалити метод сегментації мовлення при відомій фразі;

— увести показник для оцінювання результатів сегментації;

— розробити програмне забезпечення для реалізації запропонованого методу та провести аналіз отриманих результатів.

Особливості сегментації при відомій фразі

Як було зазначено вище, існують методи сегментації мовлення при невідомій та відомій фразі. Перший випадок має місце під час оброблення мовлення у реальному масштабі часу, коли дізнатись фразу заздалегідь неможливо. Безперечно, у такому разі задача сильно ускладнюється. Не знаючи транскрипцію фрази, не можна жорстко обмежити кількість встановлених границь фонем, що спричиняє велику кількість помилок. Головною вадою сегментації без відомої фрази є саме встановлення зайвих границь, що викликано здебільшого особливостями зміни звучання переходів між фонемами, які складно класифікувати. У другому випадку при сегментації з метою розмітки бази мовленнєвих даних, як правило, фрази, що вимовляються дикторами, є заздалегідь відомими, і це дозволяє суттєво підвищити точність розмітки [6].

Відомо [7], що на часових інтервалах тривалістю 10–30 мс мовленнєвий сигнал можна розглядати як стаціонарний випадковий процес. Тому доцільно поділити цей сигнал на окремі кадри (фрейми), та аналізувати не безпосередньо відліки сигналу, а окремі фрейми. При цифровому обробленні мовленнєвих сигналів відліки одного фрейму прийнято описувати деякою параметричною моделлю. Звичайно, при цьому точність встановлення границь буде обмежена щонайменше

тривалістю одного фрейму, але, враховуючи, що ця розмітка буде згодом уточнюватись людиною, такий підхід є допустимим.

Найпростішим шляхом пошуку границь фонем буде порівняння сусідніх фреймів та визначення тих пар, що найбільше відрізняються між собою, тобто:

$$\begin{aligned}(d(f_i, f_{i+1}) > \varepsilon) \Rightarrow f_i \in F_j, f_{i+1} \in F_{j+1},\\ (d(f_i, f_{i+1}) < \varepsilon) \Rightarrow f_i \in F_j, f_{i+1} \in F_j,\end{aligned} \quad (1)$$

де f_i — i-й фрейм; ε — деяке порогове значення; $d(f_i, f_{i+1})$ — функція відстані між сусідніми фреймами, $i \in 1 \ldots K$; K — кількість фреймів; F_j — j-та фонема, $j \in 1 \ldots N$; N — кількість фонем у фразі.

Зрозуміло, що такий підхід є занадто примітивним, і до того ж отримані границі не обов'язково будуть прив'язані до конкретних фонем.

При відомій кількості фреймів і фонем у фразі варіанти розміщення границь обмежені числом

$$N_{вар} = C_K^N = \frac{K!}{N!(K-N)!}. \quad (2)$$

Маючи усі можливі варіанти, можна оцінити кожен з них за деяким критерієм і обрати кращий. Проте кількість цих варіантів частіше за все занадто велика, у чому нескладно переконатися, підставивши у формулу (2) реальні значення — наприклад, при K = 300, N = 30 отримаємо приблизно $1,7 \cdot 10^{41}$ варіантів. Тому повний перебір варіантів є неефективним шляхом розв'язання задачі. Але оскільки відома ще й послідовність фонем, то цю кількість можна суттєво зменшити. Враховуючи, що кожен наступний фрейм може відноситися до попередньої або до наступної фонеми (тобто перебувати в одному з цих двох станів), то для скорочення кількості варіантів встановлення границь доцільно застосувати алгоритм Вітербі [8].

Застосування алгоритму Вітербі для сегментації мовлення

Нехай відліки кожного фрейму представлено у вигляді вектора $\bar{\boldsymbol{Y}}$ коефіцієнтів деякої параметричної моделі представлення мовленнєвих даних (це можуть бути, наприклад, коефіцієнти лінійного прогнозування або кепстральні коефіцієнти). Тобто:

$$\bar{\boldsymbol{Y}}_i = \{y_{i,1}, y_{i,2}, \ldots, y_{i,M}\}, \quad (3)$$

де M — порядок моделі; i — номер фрейму, з якого отримано вектор коефіцієнтів, $i \in 1 \ldots K$.

Нехай функція $J(Y_i, Y_{i+1})$ буде оцінювати можливість того, що вектори відносяться до різних фонем, а функція $S(Y_i, Y_{i+1})$ — можливість того, що вектори від-

носяться до однієї фонеми. Вибір виду цих функцій є окремим питанням, яке буде розглянуте далі.

Мережа Вітербі у нашому випадку буде мати K шарів по N вузлів, відповідно до кількості фреймів і фонем. Кожен шар відповідає своєму фрейму, а кожен вузол у шарі — своїй фонемі. Значення оцінки зміни та збереження стану є вагами ребер, що з'єднують відповідні вузли. Зрозуміло, що у першому шарі значення має лише один вузол, оскільки фрейм, що йому відповідає, може відноситися лише до однієї фонеми — першої. І оскільки від j-ї фонеми можливий перехід тільки до $(j+1)$-ї фонеми, то у перших та останніх $(N-1)$ шарах, буде потрібно менше, ніж N вузлів. Для n-го вузла у k-му шарі оцінки зміни та збереження стану будуть визначатися таким чином:

$$\begin{gathered} Jump_{k,n} = \max\{Stay_{k-1,n}, Jump_{k-1,n-1}\} + J(\boldsymbol{Y_n}, \boldsymbol{Y_{n+1}}), \\ Stay_{k,n} = \max\{Stay_{k-1,n}, Jump_{k-1,n-1}\} + S(\boldsymbol{Y_n}, \boldsymbol{Y_{n+1}}), \\ Jump_{1,1} = J(\boldsymbol{Y}_1, \boldsymbol{Y}_2), \\ Stay_{1,1} = S(\boldsymbol{Y}_1, \boldsymbol{Y}_2), \\ J, S \geq 0. \end{gathered} \tag{4}$$

Як видно з (4), загальна оцінка поступово накопичується, але при цьому рішення про розміщення границь не приймається, поки не буде отримано останні дві оцінки. Коли їх визначено, пошук оптимального шляху буде проходити у зворотному напрямі, і полягатиме у порівнянні попередніх оцінок. Результат заноситься у вектор $\boldsymbol{R}$ за таким правилом:

$$\begin{gathered} (Jump_{i-1,d-1} > Stay_{i-1,d}) \Rightarrow \boldsymbol{R_i} \leftarrow 1, d \leftarrow d-1, \\ (Jump_{i-1,d-1} \leq Stay_{i-1,d}) \Rightarrow \boldsymbol{R}_i \leftarrow 0. \\ i = N \ldots 1, \end{gathered} \tag{5}$$

де початково $d = N$.

Знаючи тривалість одного фрейму, по вектору $\boldsymbol{R}$ нескладно визначити тривалість кожної фонеми. Роботу алгоритму для випадку, коли $N = 10$, $K = 4$, проілюстровано прикладом, зображеним на рис. 2.

За рахунок такого підходу кількість варіантів розміщення границь значно зменшилась. Але відкритим питанням залишається вибір функцій $J(\boldsymbol{Y}_i, \boldsymbol{Y}_{i+1})$ та $S(\boldsymbol{Y}_i, \boldsymbol{Y}_{i+1})$, від адекватності яких і залежить точність сегментації. При використанні як аргументів цих функцій лише векторів коефіцієнтів не враховується наявна інформація про конкретну фонему, що знаходиться між визначеними границями. Таким чином, отримані сегменти не прив'язані до фонем, а є лише найбільш несхожими ділянками. Через це похибка сегментації може бути досить суттєва. Зменшити її можна за рахунок використання інформації про параметри фонеми, що містяться у словнику ознак. Для цього доцільно використати методи розпізнавання.

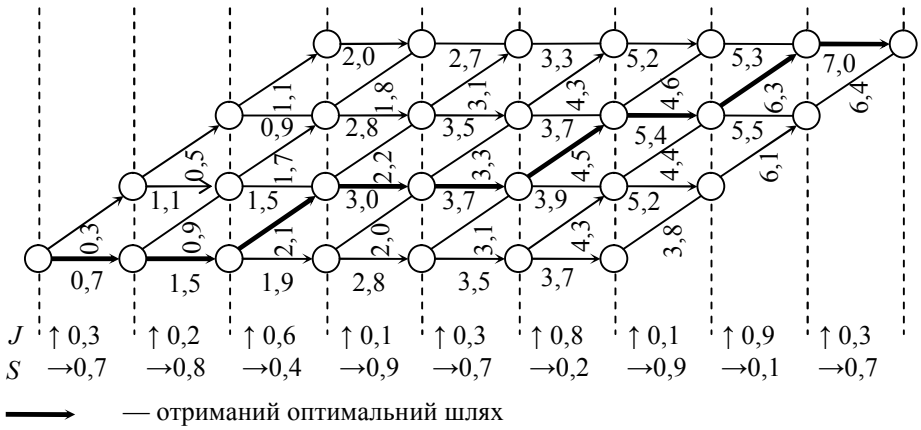

Рис. 2. Приклад роботи алгоритму Вітербі

Використання методів розпізнавання при сегментації мовлення

У випадку, коли послідовність фонем відома, задача сегментації полягає у встановленні відповідності кожній фонемі деякої послідовності фреймів. Задача ж розпізнавання полягає у встановленні відповідності деякій ділянці мовлення (тобто послідовності фреймів) одній з фонем із словника. Як бачимо задачі дуже подібні, і їх можна деякою мірою поєднати.

З огляду на сказане, ідею сегментації можна сформулювати так: розпізнати кожен окремий фрейм і визначити послідовність фреймів, що знаходяться поруч та відносяться до тої ж самої фонеми. Відповідно, коли починається послідовність фреймів, що належить іншій фонемі, слід встановити границю. Виглядає все просто, але ефективно працює такий підхід лише тоді, коли якість розпізнавання майже стовідсоткова. Проте досягти такого рівня доволі складно. І пов'язано це здебільшого з тим, що зазвичай фонем у словнику достатньо багато для того, щоб система розпізнавання помилялась.

Застосування алгоритму Вітербі, дозволяє зменшити кількість помилок при сегментації мовлення. Як показано раніше, для кожного фрейму може бути лише два випадки — або він відноситься до тієї ж фонеми що і попередній фрейм, або він є початком наступної фонеми:

$$(\boldsymbol{Y}_i \in F_j) \oplus (\boldsymbol{Y}_i \in F_{j+1}). \qquad (6)$$

Таким чином, обирати треба лише з двох варіантів, хоча у словнику знаходиться вся множина фонем.

Як показано у [9], доцільно представити мовленнєві дані у вигляді кепстральних коефіцієнтів, і для визначення відстаней у просторі ознак скористатися зваженою евклідовою метрикою. Для підвищення фонетичної коректності варто застосувати методи кластеризації при формуванні словника ознак. З огляду на це, функції оцінювання J та S будуть мати вигляд:

$$J(\boldsymbol{Y}_i) = \frac{1}{D(\boldsymbol{C}_{j+1}, \boldsymbol{Y}_i)},$$
$$S(\boldsymbol{Y}_i) = \frac{1}{D(\boldsymbol{C}_j, \boldsymbol{Y}_i)}. \tag{7}$$

де C_j — центроїд, який характеризує фонему, що відповідає j-му вузлу; $D(C_j, Y_i)$ — відстань за зваженою евклідовою метрикою між j-м центроїдом та i-м фреймом, яка тим більша, чим менше вектор, схожий на центроїд.

Аналіз результатів сегментації

Для проведення дослідження було використано загальнодоступну частину англомовної бази даних TIMIT. Ця база даних широко використовуються для тестування та налаштування систем розпізнавання мовлення. Також вона містить достатньо великий набір різноманітних фраз і дикторів. Основною перевагою TIMIT є те, що матеріал цієї бази розмічений на фонеми. Таким чином, доцільно провести апробацію методу сегментації на матеріалі бази даних TIMIT, використовуючи цю розмітку як еталонну, після чого застосувати запропонований метод для розмітки бази акустичних фрагментів українського мовлення.

Звукові файли TIMIT було переконвертовано з формату (PCM Raw data, 16 кГц, 16 біт, моно) у формат (Windows PCM (wav), 16 кГц, 16 біт, моно). Для отримання коефіцієнтів вхідний сигнал розбивався на окремі кадри (фрейми) довжиною 10 мс. При частоті дискретизації 16 кГц, кожен такий фрейм містив 160 відліків вхідного сигналу. З одного фрейму отримувався вектор кепстральних коефіцієнтів розмірністю $M = 10$.

Для застосування методів розпізнавання було утворено словник центроїдів, при формуванні якого було використано 7 дикторів. Кількість представлених у словнику фонем — 53. Для підвищення фонетичної коректності розпізнавання застосовано методи кластеризації, як це запропоновано у роботі [9]. Відстань у просторі ознак обчислювалася за зваженою евклідовою метрикою.

Що стосується оцінювання результатів, то для задачі сегментації це досить складне і неоднозначне питання. Загалом немає гарантії, що навіть два досвідчених фахівця виконають розмітку фрази на фонеми абсолютно однаково. Проте ручна розмітка, виконана кваліфікованим спеціалістом наразі вважається найточнішою. Тому доцільно порівнювати отримані результати саме з ручною розміткою.

Для оцінювання точності сегментації введемо показник

$$CS = \left(1 - \frac{\sum_{i=1}^{N-1} |b_i - a_i| + |b_{i+1} - a_{i+1}|}{\sum_{i=1}^{N-1} b_{i+1} - b_i}\right) \cdot 100\ \%. \tag{8}$$

де b_i — значення *i*-ї границі ручної розмітки; a_i — значення *i*-ї границі визначеної розмітки; N — кількість фонем у фразі. Схожий показник використовувався для оцінювання результатів у роботі [3].

Запропонований показник характеризує тривалість «вірно» розміченої частини фрази. Це зручно для загального оцінювання результату. Проте, якщо розглядати окремі фонеми, то можливий випадок, коли відносно усієї фрази відхилення встановленої границі незначне, а порівняно з тривалістю однієї фонеми є досить суттєвим. Тому варто додатково враховувати кількість таких помилок. Будемо вважати, що при сегментації допущено помилку, якщо виконується нерівність:

$$\left(\left(\frac{|b_i - a_i|}{b_{i+1} - b_i}\right) > \delta\right) \vee \left(\left(\frac{|b_{i+1} - a_{i+1}|}{b_{i+1} - b_i}\right) > \delta\right), \tag{9}$$

де $i \in 1 \ldots N-1$, δ — порогове значення; для даного дослідження обрано $\delta = 0,3$.

Отже, для аналізу результатів введемо ще один показник, який буде характеризувати кількість правильно встановлених границь:

$$BC = \frac{N_r}{N} \cdot 100\ \%, \tag{10}$$

де N_r — кількість правильно встановлених границь.

Для практичної перевірки запропонованих підходів було розроблено програмне забезпечення, що реалізує метод сегментації мовлення при відомій фразі з використанням алгоритму Вітербі та засобів розпізнавання мовлення.

Фрагмент фрази, розміченої таким шляхом, показано на рис. 3. Як видно, визначені автоматично границі знаходяться досить близько до встановлених при ручній розмітці, хоча і наявні деякі відхилення.

Оцінки результатів сегментації запропонованим методом наведено в табл. 1.

Таблиця 1. Оцінки результатів сегментації

Параметр	**Значення**
Кількість фраз	18
Загальна кількість фонем	652
CS, %	87,5
BC, %	77,3

Сумарна тривалість правильно розмічених фреймів складає 87,5 %. Що стосується окремих фонем, то для 77,3 % з них встановлені границі знаходилися в допустимих межах. Проте використання статистичної інформації про тривалість фонем, що міститься у словнику ознак, дає можливість покращити отримані показники.

Однією із задач, що розв'язуються за допомогою алгоритму Вітербі, є згладжування брязкоту, який виникає при розпізнаванні на акустичному рівні. Досяга-

ється це за рахунок того, що алгоритм Вітербі значно скорочує кількість варіантів фонем, до яких може відноситися поточний фрейм. Але навіть при цьому може виникнути така ситуація, коли однозначно визначити потрібну фонему важко і програма «коливається» між декількома фонемами, по черзі віддаючи перевагу кожній з них. Звичайно, таке явище негативно впливає на результати сегментації — наслідком є неправильно встановлені границі.

Щоб уникнути помилок такого характеру, доцільно враховувати статистичну інформацію, яку можна отримати під час формування словника ознак. В процесі розстановки границь між фонемами по суті відбувається визначення тривалості кожної фонеми. Зрозуміло, що одні фонеми (наприклад голосні) мають більшу тривалість, інші навпаки меншу (короткі вибухові приголосні). Знаючи середню максимальну і середню мінімальну тривалість кожної фонеми, можна корегувати результати сегментації. Це неважко реалізувати програмно, оскільки за умовами послідовність фонем у фразі відома заздалегідь.

Таке уточнення доцільно реалізувати на основі розглянутого алгоритму Вітербі. Знаючи максимальну та мінімальну тривалість фонеми, не можна жорстко обмежувати ними реальний звук, оскільки необов'язково його тривалість має бути максимальною чи мінімальною. Але якщо ці межі порушуються, то доцільно їх скорегувати. Границі фонем визначаються по графу Вітербі при зворотному його обході, і під час цієї процедури нескладно визначати тривалість поточної фонеми. Для цього достатньо змінити формулу (5) таким чином:

$$\begin{gathered}
(Jump_{i-1,d-1} > Stay_{i-1,d}) \wedge (Dr_c \geq DrMin_d) \Rightarrow \boldsymbol{R_i} \leftarrow 1, d \leftarrow d-1, Dr_c \leftarrow 0, \\
(Jump_{i-1,d-1} > Stay_{i-1,d}) \wedge (Dr_c < DrMin_d) \Rightarrow \boldsymbol{R_i} \leftarrow 0, \\
(Jump_{i-1,d-1} \leq Stay_{i-1,d}) \wedge (Dr_c \leq DrMax_d) \Rightarrow \boldsymbol{R}_i \leftarrow 0, \\
(Jump_{i-1,d-1} \leq Stay_{i-1,d}) \wedge (Dr_c > DrMax_d) \Rightarrow \boldsymbol{R}_i \leftarrow 1, d \leftarrow d-1, Dr_c \leftarrow 0, \\
i = N \ldots 1.
\end{gathered} \tag{11}$$

де Dr_c — тривалість поточної ділянки (дорівнює кількості фреймів); $DrMax_d$, $DrMin_d$ — відповідно середня максимальна і середня мінімальна тривалість d-ї фонеми.

Загалом при цьому відбувається деяке згладжування розмітки — нехарактерні тривалості фонем уточнюються згідно статистичної інформації, що додатково зберігається у словнику ознак.

На рис. 3 і 4 наведено фрагмент фрази, розмічений відповідно без і з урахуванням тривалості фонем. В обох випадках для сегментації використовувалися алгоритм Вітербі та моделі фонем із словника ознак.

З рисунків видно, що порівняно з попередніми результатами, отримано суттєве уточнення для багатьох границь. Водночас, деякі границі залишились такими ж, що свідчить про те, що вони знаходились у допустимих межах.

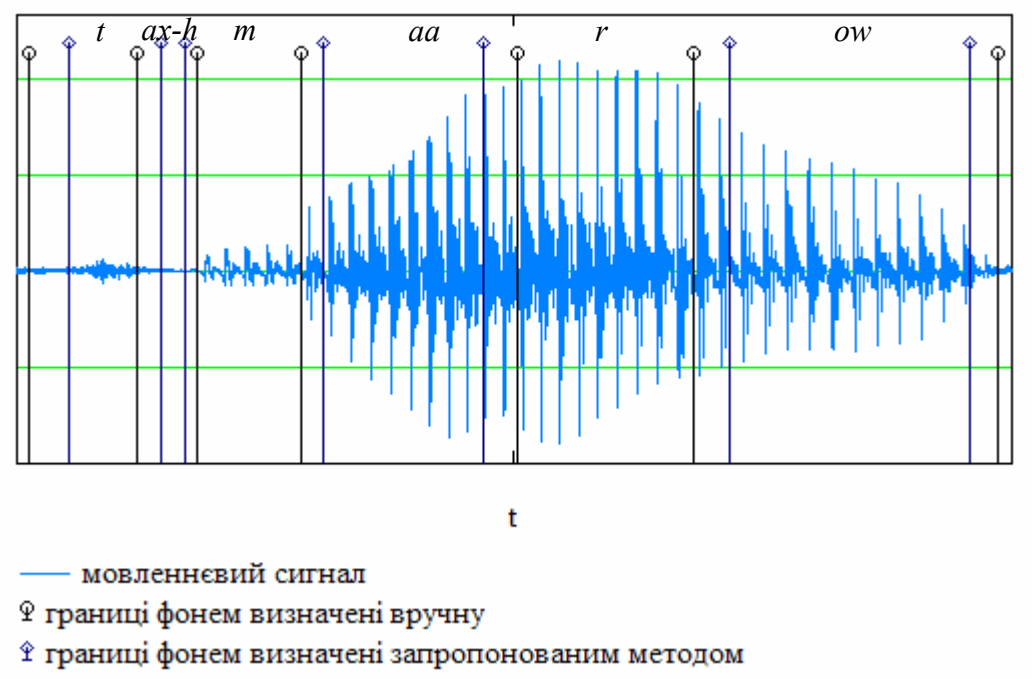

Рис. 3. Фрагмент фрази, розміченої без врахування інформації про тривалість фонем

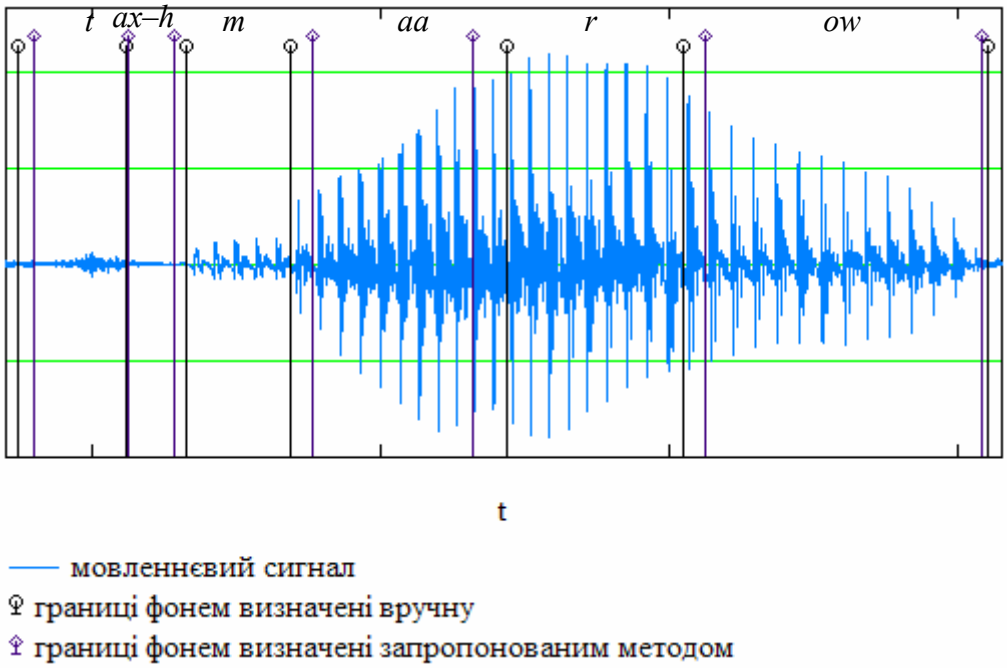

Рис. 4. Фрагмент фрази, розміченої з урахуванням інформації про тривалість фонем

Чисельні оцінки результатів, отриманих із таким уточненням, наведено у табл. 2.

Таблиця 2. Оцінки результатів сегментації після уточнення

Параметр	**Значення**
Кількість фраз	18
Загальна кількість фонем	652
CS, %	90
BC, %	83,8

Як бачимо, при врахуванні інформації про тривалість фонем спостерігаються значні покращення. Показник CS збільшився на 2,5 %, що свідчить про те, що загалом сегментація стала точнішою. Кількість вірно встановлених границь стала вище на 6,5 %, про що свідчить показник BC. Як видно з рис. 3 та табл. 2, за рахунок застосованого корегування тривалості фонем, значна частина встановлених границь стала ближче до тих, які були визначено вручну.

Висновки

У роботі проаналізовано підходи до розв'язання задачі сегментації мовлення при відомій фразі. Показано, що застосування методів розпізнавання у поєднанні з алгоритмом Вітербі дає змогу досягти точності сегментації на рівні 87,5 % та 77,3 % правильно встановлених границь. Удосконалено метод сегментації мовлення при відомій фразі. Запропоновано при зворотному пошуку шляху по дереву Вітербі враховувати статистичну інформацію про тривалість фонем, що дозволило підвищити точність сегментації до 90 % і збільшити відсоток правильно встановлених границь до 83,3 %. Отримані результати можуть застосовуватися для попередньої сегментації мовленнєвих даних в автоматизованій системі сегментації з метою створення україномовної бази мовленнєвих даних.

1. *Carson-Berndsen J.* Framework for Cross-Language Automatic Phonetic Segmentation / Kalu U. Ogbureke, Julie Carson-Berndsen // ICASSP 2010: IEEE International Conference on Acoustics, Speech and Signal Processing. — Dallas, Texas (USA). — 2010.

2. *Григорян Р.Л.* Метод выделения монологических составляющих с использованием идентификации дикторов на основе векторного квантования / Р.Л. Григорян, С.А. Репалов, С.С. Коршунов // «Штучний інтелект» 3'2006. — Донецьк, 2006.

3. *Жуйков В.Я.* Алгоритм автоматической классификации сегментов речи на основе автокорреляционных и энергетических характеристик / В.Я. Жуйков, Н.Н. Кузнецов, А.Н. Харченко // Электроника и связь. — НТУУ «КПИ». — 2010. — № 5(58).

4. *Сорокин В.Н.* Сегментация и распознавание гласных / В.Н. Сорокин, А.И. Цыплихин // Информационные процессы. — 2004. — Т. 4, № 2. — С. 202–220.

5. *Каркульовський В.І.* Особливості методів сегментації мовленнєвих сигналів / В.І. Каркульовський, В.С. Ткаченко // Комп'ютерні системи проектування. Теорія і практика. Вісник НУ «Львівська політехніка» № 651. — Львів: Видавництво НУ «Львівська політехніка», 2009. — С. 144–148.

6. *Богданов Д.С.* База речевых фрагментов русского языка «ISABASE» / Д.С. Богданов, О.Ф. Кривнова, А.Я. Подрабинович, В.В. Фарсобина // Интеллектуальные технологии ввода и обработки информации. — М., 1998.

7. *Рабинер Л.Р.* Цифровая обработка речевых сигналов / Л.Р. Рабинер, Р.В. Шафер. — М.: Радио и связь, 1981. — 496 с.

8. *Биков М.М.* Моделювання процесу аналізу і класифікації голосових команд / М.М. Биков, Т.В. Грищук // Монографія. — Вінниця: ВНТУ, 2009. — 129 с.

9. *Ткаченко О.М.* Аналіз підходів до розпізнавання мовлення при побудові фонемних вокодерів / О.М. Ткаченко, О.В. Дзісь // Інформаційні технології та комп'ютерна інженерія. — 2009. — № 3. — С. 93–101.

Надійшла до редакції 26.10.2010