УДК 681.84

АПАРАТНІ ЗАСОБИ СЕГМЕНТАЦІЇ МОВНОГО СИГНАЛУ

Шиманські З.

Постановка задачі.

Сучасний етап розвитку інформаційних технологій обробки мовних сигналів характеризується розширенням галузей застосування, в значні частині з яких вимагається опрацювання потоків даних у реальному часі на апаратних засобах, що задовольняють обмеженням щодо габаритів, енергоспоживання та вартості [1-4]. Створення таких апаратних засобів вимагає широкого використання сучасної елементної бази (напівзамовних і замовних надвеликих інтегральних схем (НВІС), однокристальних процесорів цифрової обробки сигналів), розробки апаратно-орієнтованих методів і алгоритмів. Однією із основних проблем в інформаційних технологіях обробки мовних сигналів є проблема перетворення мовних сигналів у часові області, яка використовується при розпізнаванні та синтезі мовних сигналів, верифікації та ідентифікації дикторів, в системах кодування та передавання каналами зв'язку. Перетворення мовних сигналів в часові області знаходить застосування в автоматизованих системах управління, коли необхідно змінювати темп подання мовної інформації оператору в залежності від його фізичного та емоційного стану [1,2].

При розв'язанні проблеми перетворення мовних сигналів в часові області однією із основних задач є задача автоматичного сегментування мовних сигналів, успішне розв'язання якої тісно зв'язано з наступними процедурами опрацювання мови. Аналіз процесу сегментації [2] показує, що він вимагає виконання таких етапів:

- виділення мовного сигналу на фоні шуму та пауз;
- сегментації на основі критерію правдоподібності;
- аналізу сегментів та виділення параметрів;
- класифікація сегментів.

Таким чином, актуальною задачею є розробка апаратних засобів для сегментації мовного сигналу у реальному часі з високою ефективністю використання обладнання.

Розв'язання задачі.

Алгоритми сегментації. Проаналізуємо алгоритми реалізації кожного із етапів сегментації мовного сигналу [2].

Виділення мовного сигналу на фоні шуму та пауз. Визначення точних моментів початку і завершення мовного сигналу на фоні шуму має суттєве значення для якості сегментації, наступної обробки та кількості арифметичних операцій. Для реалізації даного етапу використовуються алгоритми, які враховують тільки два параметри мовного сигналу $x(n)$ – короткочасну енергію $E(n)$ та короткочасне середнє число переходів через нуль $S(n)$. Дані параметри визначаються так:

$$E(n) = \sum_{m=-\infty}^{\infty} |x(m)| h(n-m),$$

$$S(n) = \sum_{m=-\infty}^{\infty} |sign[x(m-1)]| h(n-m),$$

де $$sign[x(m)] = \begin{cases} 1, коли & x(n) \geq 0 \\ -1, коли & x(n) < 0 \end{cases},$$

а $h(n)$ – прямокутна вагова функція тривалістю 10 мілісекунд.

Сегментація на основі критерію правдоподібності. Основною ідеєю даного методу є використання статичного критерію максимальної правдоподібності, за допомогою порівнюються декілька моделей сусідніх сегментів мовного сигналу та визначаються моменти

 ISBN 7-776-8361-7

різкої зміни параметрів моделей. Кожний однорідний сегмент мовного сигналу описується авторегресійною моделлю $M(A,\delta)$ порядку p з параметрами $A=(a_1,a_2,...,a_p)$, де a_i - коефіцієнти лінійного передбачення, δ^2 - дисперсія похибки передбачення.

Приймаються дві гіпотези-альтернативи:

H1: сигнал (y_1,...,y_n) описується моделлю $M_0(A_0,\delta_0)$;

H2: у момент часу r відбувається стрибкоподібна зміна моделі, так що сигнал (y_1,...,y_r) описується моделлю $M_1(A_1,\delta_1)$, а сигнал (y_{r+1},...,y_N) – моделлю $M_2(A_2,\delta_2)$.

Статичний критерій базується на узагальненому відношенні правдоподібності між цими двома гіпотезами:

$$D(r)=-(n-r)\ln\delta_2^2-r\ln\delta_1^2-n\ln\delta_0^2. \quad (1)$$

Рішення про наявність різкої зміни моделі приймаються якщо:

$$\min_{\delta_0 A_1 \delta_1}\ \max_{\delta_1 A_2 \delta_2}\ \max_{\gamma} D_n > D_0 .$$

Біжуче значення r визначається як аргумент виразу (1). Оскільки така багатовимірна оптимізація вимагає значних обчислювальних затрат, то на практиці використовується дещо спрощена процедура.

Аналіз сегментів та виділення параметрів. Для аналізу використовуються такі параметри:

- E –короткочасна енергія мовного сигналу, визначена на всьому діапазону частот;
- E_H/E_B – відношення енергії в діапазоні низьких частот (250-600 Гц) до енергії мовного сигналу в діапазоні високих частот(650-3000 Гц).

В системах аналізу та розпізнавання мови використовують такі параметри: тривалість виділеного сегменту мовного сигналу; середнє число переходів через нуль; частота основного тону; набір коефіцієнтів лінійного передбачення; частота формант та інші.

Класифікація сегментів. Класифікація виділених у процесі сегментації звукових сегментів мовного сигналу проводиться на основі приведених вище параметрів. На першому кроці на основі значень першого коефіцієнту авторегресії A та короткочасної енергії E всі сегменти поділяються на дві групи: вокалізовані та невокалізовані. Сегменти, для яких A_1>-0,4 класифікуються як невокалізовані, а сегменти, для яких A_1<-1,5 класифікуються як вокалізовані. Для решти сегментів обчислюється ознака O_B на основі співвідношення:

$$O_B=0{,}03E-A.$$

Рішення про вокалізованість сегменту приймається у випадку коли O_B>2,2.

Наступним кроком є розділення вокалізованих сегментів на вокалізовані голосні та вокалізовані приголосні. Основними факторами, які враховуються при цьому є:

- перевищення значення енергії E порогової величини в моменти ПГ (приголосна-голосна);
- більша концентрація енергії в низькочастотній області у голосних звуків.

Розробка структури апаратних засобів сегментації мовного сигналу. Розробку апаратних засобів сегментації мовного сигналу пропонується здійснювати на основі інтегрованого підходу, який охоплює [4]:

- сучасну елементну базу інформаційних технологій обробки мовних сигналів;
- методи та алгоритми сегментації мовного сигналу;
- методи та засоби автоматизованого проектування апаратного і програмного забезпечення;
- нові архітектурні та схемотехнічні рішення.

Апаратні засоби сегментацію мовного сигналу повинні забезпечувати високу ефективність використання обладнання та роботу в реальному часі. Проведений аналіз алгори-

тмів сегментації мовних сигналів показав, що дані алгоритми є складними з великою кількістю логічних операцій. Ефективно реалізувати такі алгоритми можна шляхом одночасного використання універсальних і спеціальних підходів, апаратних і програмних засобів. В основу структурної організації таких апаратних засобів пропонується покласти принципи, які дозволять підвищити ефективність використання обладнання, зменшити вартість і терміни розробки. Аналіз показує, що забезпечити такі вимоги можна при використанні наступних принципів: змінного складу обладнання, модульності, узгодженості та відкритості програмного забезпечення.

Елементна база. Основною елементною базою, яка використовується для реалізації апаратних засобів сегментації мовного сигналу є однокристальні програмовані мікропроцесори і мікро-ЕОМ, архітектура яких орієнтована на розв’язання задач цифрової обробки сигналів (ЦОС) та спеціалізовані НВІС [5,6].

В мікропроцесорах і мікро-ЕОМ ЦОС за рахунок вдалих архітектурних і технологічних рішень вдалось поєднати високу швидкодію виконання базових операцій ЦОС з ефективною реалізацією алгоритмів управління та прийняття рішень. Особливостями архітектури мікропроцесорів і мікро-ЕОМ ЦОС є:

- розвинута багатошинна організація, яка дозволяє виконувати пересилання даних між різними функціональними вузлами;
- одночасне транспортування всіх операторів до операційного пристрою;
- використання апаратних засобів для одночасного виконання операції множення з підсумовуванням;
- виконання паралельно з основними операціями функцій адресації, переадресації, розгалуження програм;
- застосування апаратно вбудованих засобів організації циклічних обчислень, що переважають в алгоритмах ЦОС;
- використання модифікованої гарвардської архітектури;
- набір команд для виконання з підвищ енною точністю арифметичних операцій;
- коротший командний цикл і виконання більшості команд за один цикл;
- здатність підтримувати конвеєрну обробку;
- зменшення довжини арифметичного конвеєра за рахунок використання багатопортової пам’яті;
- використання для обміну швидкодіючих каналів прямого доступа до пам’яті і механізму підтримки багатопроцесорності.

Спеціалізовані НВІС за способом проектування і виготовленням, тобто налаштуванням на виконання конкретного алгоритму діляться на два класи: замовні і напівзамовні [5].

Замовні НВІС - це мікросхеми, розроблені на основі стандартних або спеціально створених елементах і вузлах за схемою замовника. Особливістю замовних НВІС є оптимізація елементів і зв'язків, що дозволяє досягнути граничних значень параметрів для кожного рівня технології.

Напівзамовні НВІС - це мікросхеми, що складаються з двох частин: наперед спроектованої постійної та змінної - замовної, структура якої визначається замовником. До напівзамовних НВІС відносяться мікросхеми на основі базових матричних кристалів (БМК) та програмовані користувачем логічні інтегральні схеми (ПЛІС). Проектування пристроїв на базі ПЛІС здійснюється методом завантаження необхідної конфігурації в елементи "тіньової" пам'яті. При використанні БМК спеціалізація НВІС здійснюється за рахунок нанесення відповідних шарів з'єднань. Основними елементами БМК є базові комірки, що складаються з набору незкомутованих елементів-транзисторів і резисторів. На базі таких елементів реалізуються функціонально завершені вузли, які виконують елементарні функції типу І-НЕ, АБО-НЕ та інші.

 ISBN 7-776-8361-7

Структура апаратних засобів сегментації мовного сигналу. Структура апаратних засобів сегментації мовного сигналу наведена на рис.1, де АЦП – аналого-цифровий перетворювач, БП – багатопортова пам'ять, ПЦОС – процесор ЦОС, АР – апаратний розширювач.

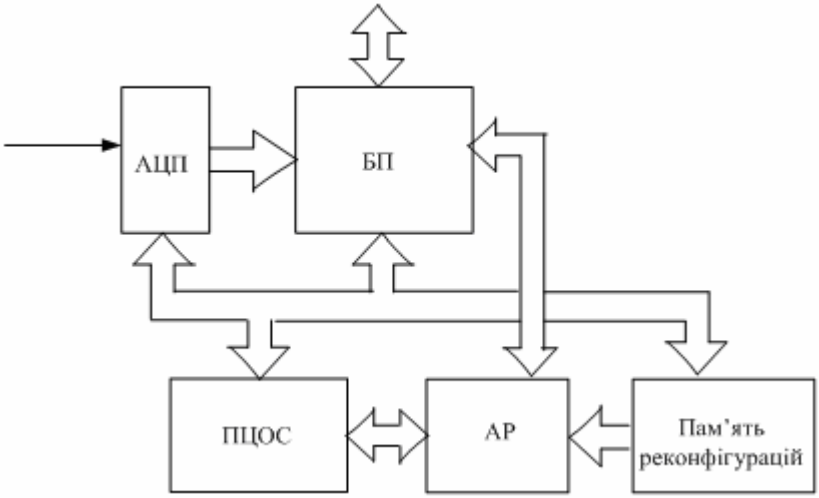

Рис.1 Структура апаратних засобів сегментації мовного сигналу

Основним елементом апаратних засобів сегментації мовного сигналу є ПЦОС, основні характеристики якого визначаються особливостями архітектури і технічними характеристиками мікропроцесора ЦОС. До числа таких характеристик відносяться: довжина інформаційного слова, час реалізації основних команд, обсяг внутрішньокристальної пам'яті даних і програм. Порівняння характеристик мікропроцесорів ЦОС різних фірм показує, що відмінності між мікропроцесорами одного покоління є несуттєвими. Це пояснюється близькістю архітектури і використанням для їх реалізації приблизно однакових за рівнем технологій. Тому структури апаратних засобів сегментації мовного сигналу на основі мікропроцесорів ЦОС різних фірм одного покоління не будуть мати істотних відмінностей. За повнотою родини, за існуючими технологічними інструментальними засобами та за кількістю розробленого програмного забезпечення мікропроцесори ЦОС фірми Texas Instruments TMS 320 переважають МП інших фірм. Тому апаратні засоби сегментації мовного сигналу доцільно реалізовувати на мікропроцесорі ЦОС серії TMS 320C50. Самостійно мікропроцесор ЦОС не відповідає вимогам, які ставляться до ПЦОС за ємністю пам'яті, завадостійкістю та навантажувальною здатністю інтерфейсу. Задовольнити такі вимоги можна шляхом доповнення мікропроцесора ЦОС зовнішньою оперативною пам'яттю та розширювачем інтерфейсу. Особливістю структурної організації ПЦОС є відділення мікропроцесора від зовнішніх пристроїв, що дозволило підвищити завадостійкість і сумістити у часі роботу мікропроцесора та зовнішніх пристроїв.

У апаратних засобах сегментації мовного сигналу для взаємодії між ПЦОС і зовнішніми пристроями використовується БП, яка забезпечує зменшення часу обміну та розв'язує проблеми пов'язані з синхронізацією роботи всіх пристроїв, що до неї підключаються [7]. Для забезпечення режиму реального часу найскладніші алгоритми з великою кількістю обчислювальних операцій реалізуються апаратним шляхом за допомогою АР. Архітектура АР відображає структуру алгоритму розв'язання задачі та дозволяє повною мірою використовувати можливості НВІС-технології, враховувати вартість площі кристала, а також кількість вхідних і вихідних виводів. АР реалізується на ПЛІС з динамічним репрограмуванням, що забезпечує оперативне переналаштування на реалізацію конкретного алгоритму. Узгодження обчислювальної здатності апаратних засобів сегментації мовного сигналу з інтенсивність надходження даних забезпечує високу ефективність використання обладнання. Для оцінки ефективності використання обладнання E доцільно використовувати критерій, який зв'язує продуктивність системи з витратами обладнання та

дає оцінку її елементам (вентилям) за продуктивністю [4]. Кількісна величина ефективності використання обладнання для такого компоненту визначається наступним чином:

$$E = \frac{R}{W_c T_p}$$

де R – складність алгоритму розв'язання задачі у кількості елементарних операцій; W_c - витрати обладнання на реалізацію системи обробки мовних сигналів в вентилях; T_p - час розв'язання задачі.

Висновки:

1. Запропоновано розробку апаратних засобів сегментації мовного сигналу в реальному часі з високою ефективністю використання обладнання здійснювати на основі інтегрованого підходу, який охоплює сучасну елементну базу, методи та алгоритми сегментації мовних сигналів, нові, орієнтовані на НВІС-реалізації, алгоритмічні, архітектурні та схемотехнічні рішення.
2. Показано, що для ефективної сегментації мовного сигналу в реальному часі необхідно використовувати універсальні та спеціальні підходи, апаратні і програмні засоби.
3. Для вибору і оцінки апаратних засобів сегментації мовних сигналів у реальному часі запропоновано використовувати критерій ефективності використання обладнання, який зв'язує продуктивність з витратами обладнання та дає оцінку елементам системи за продуктивністю.
4. При побудові апаратних засобів сегментації мовних сигналів універсалізація розглядається як засіб скорочення номенклатури та зменшення затрат, а спеціалізація – як засіб підвищення швидкодії.

Methods and algorithms for segmentation of speech signals are analyzed, proposed the structure of hardware for segmentation of speech signal in real time and analytical expression for estimation effectiveness of use equipment are developed.

1. Вінцюк Т.К. Інтелектуальні усно мовні інформаційні технології та системи// Праці третьої всеукраїнської конференції "Укробраз-96".- К.: ІК АН України, 1996.- С117-120.
2. Рашкевич Ю.М. Перетворення часового масштабу мовних сигналів. Львів. Академічний Експрес, 1997. – 143с.
3. З. Шиманські, Р. Фігура, Р. Марцишин. Особливості перетворення часової структури дифтонгів польської мови. Вісник Національного університету "Львівська політехніка". №468, 2002. – С.170-174.
4. Цмоць І.Г. Інформаційні технології та спеціалізовані засоби обробки сигналів і зображень у реальному часі. – Львів: УАД, 2005.- 227с.
5. Грушицкий Р.И., Мурсаев А.Х., Угрюмов Е.П. Проектирование систем на микросхемах программируемой логики. – СПб.: БХВ-Петербург, 2002. – 608с.
6. Кунг С. Матричные процессоры на СБИС. – М.- Мир.- 1991.- 672с.
7. Пат. №23358А Україна, МПК G11 C11/00. Багатопортова пам'ять / Демида Б.А, Рашкевич Ю.М, Цмоць І.Г. Бюл. №4, 1998.

 ISBN 7-776-8361-7