

Національне агентство з питань інформатизації при
Президентові України

Державний науково-дослідний інститут інформаційної
інфраструктури

РАШКЕВИЧ Юрій Михайлович

УДК 681.84.087.4

**МОДЕЛІ, МЕТОДИ ТА ЗАСОБИ
ПЕРЕТВОРЕНЬ МОВНИХ СИГНАЛІВ В
ЧАСОВОМУ ПРОСТОРИ**

05.13.06 - *Автоматизовані системи управління та
прогресивні інформаційні технології*

Автореферат

дисертації на здобуття наукового ступеня
доктора технічних наук

Львів - 1997



Дисертацією є рукопис.

Робота виконана в Державному університеті "Львівська політехніка" Міністерства освіти України

Науковий консультант: член-кореспондент НАН України, доктор технічних наук, професор **ГРИЦИК Володимир Володимирович**, Державний науково-дослідний інститут інформаційної інфраструктури, директор

Офіційні опоненти: доктор технічних наук, професор **ВІНЦЮК Тарас Климович**, Міжнародний науково-навчальний центр інформаційних технологій та систем НАН України та Міністерства освіти України, завідувач відділом

доктор фізико-математичних наук, професор **ДРАГАН Ярослав Петрович**, Фізико-механічний інститут НАН України, пров. наук. спів.

доктор технічних наук, професор **ЛУЦКІВ Микола Михайлович**, Національна академія друкарства, завідувач кафедру.

Провідна установа: Харківський державний технічний університет радіоелектроніки (кафедра програмного забезпечення електронних обчислювальних машин) Міністерства освіти України, Харків

Захист відбудеться «26» XII 1997 р. о 13 годині на засіданні спеціалізованої вченої ради Д 35.813.01 при Державному науково-дослідному інституті інформаційної інфраструктури (290053, м. Львів, вул. Наукова, 5а).

З дисертацією можна ознайомитися в бібліотеці інституту (290053, Львів, вул. Наукова, 5а)

Автореферат розісланий «25» XI 1997 р.

Вчений секретар
спеціалізованої вченої ради
докт. техн. наук

Бунь Р.А.

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Звукова мова є не тільки найважливішим і найприроднішим засобом комунікації між людьми, але й однією із основних форм представлення інформації в людино-машинних системах оброблення інформації та управління.

Серед важливих і складних проблем оброблення мовних сигналів особливе місце займає в силу надзвичайно широкого спектру застосування проблема розроблення технологій перетворення структури мовних сигналів в часовій області, яке є необхідним при розв'язуванні задач розпізнавання та синтезу мовних сигналів (нормалізація за темпом на етапі попереднього оброблення, забезпечення необхідної швидкості відтворення мови в синтезаторах), верифікації та ідентифікації дикторів, в системах кодування та передавання мовної інформації каналами зв'язку (мовна пошта) а також в багатьох інших задачах аналізу, перетворення і синтезу інформації у вигляді мовного повідомлення.

Предметом роботи є розроблення та застосування ефективних методів часової трансформації мовних сигналів для задач зміни темпу мовної інформації - регулювання темпу мови. Виникнення даної проблеми пояснюється тим, що поряд з очевидними перевагами звукової мови як засобу обміну інформацією, фізіологічні особливості процесу мовотворення (механічні властивості артикуляційних органів людини) не дозволяють в загальному випадку підвищити темп мови до рівня характеристик продуктивності слухового каналу, який здатен суттєво швидше обробляти інформацію, що поступає на його вхід.

Базові ідеї розроблення засобів прискореного прослуховування наперед записаної на магнітофонну плівку мовної інформації виникли наприкінці 40-х років і стосувалися забезпечення можливості незрячим людям засвоювати мовну інформацію з магнітофону в темпі, швидшому від темпу запису, а відповідно - в короткий час.

Запропонований Герві та Феірбенксом підхід базувався на властивості значної надлишковості мовного сигналу, яка дозволяє вилучати або вводити в мовний сигнал короткі звукові відрізки і тим змінювати час звучання (темп мови) при збереженні розбірливості.

В зв'язку з бурхливим розвитком методів цифрової обробки сигналів протягом 70-80-х років на Заході були розроблені доско-

наліші інформаційні технології модифікації мовних сигналів в часовій області, в основі яких в тій чи іншій формі використовується ідея фільтрації початкового сигналу набором фільтрів з наступною модифікацією часового масштабу кожного із проміжних сигналів і синтезом вихідного сигналу зміненої тривалості. Найбільш відомими серед цих методів є: алгоритм гармонічного часового масштабування (Малах), аналіз-синтез за допомогою короткочасного перетворення Фур'є (КПФ) (Портнофф), оцінка мовного сигналу на основі модифікованого КПФ (Гріффін, Лім), метод підсумовування з перекриванням (Роукос, Вілгус), синхронне з основним тоном підсумовування з перекриванням (Моулінез, Шарпентьєр), аналіз-синтез мови на основі синусоїдального представлення (Кватієрі, МакАулай).

В Україні роботи, пов'язані з проблемами розтягування та стискування мовних сигналів, проводилися науковими колективами під керівництвом Т.Вінцюка, М.Бондаренка, В.Грицика, М.Деркача і стосувалися питань розпізнавання та синтезу мови, вокодерної телефонії, регулювання темпу мови.

На сьогодні, незважаючи на існування цілого ряду методів та засобів перетворення часової структури мовних сигналів, ще не вдалося досягти можливості відтворення та надійного розпізнавання мови в темпі, наближеному до граничного темпу сприйняття інформації на слух, який приблизно в чотири рази перевищує темп мовотворення. Причиною цього є відсутність глибоких теоретичних та експериментальних досліджень процесів зміни темпу мови людиною, як наслідок того, що існуючі технології використовують в основному лінійні підходи до перетворення структури мовного сигналу, які не в повній мірі відповідають реальним процесам мовотворення, а тому не забезпечують вирішення поставленої задачі.

Таким чином, дослідження і створення методів, алгоритмів та засобів часової трансформації мовних сигналів з метою забезпечення можливості ефективного прискорення та сповільнення темпу мовної інформації в широкому діапазоні надалі являє собою актуальну задачу та має теоретичний і практичний інтерес.

Зв'язок роботи з науковими програмами, планами та темами. Робота виконувалася в рамках: наукової програми 05.08 "Системи зв'язку" Державної науково-технічної програми ДКНТПП України; держбюджетних тем Міністерства освіти України "Розробка математичних моделей засобів часового нормування мовних сигналів" (1991-93 рр.), "Шифрування мовної інформації в телефонно-

му каналі зв'язку" (1993-95 рр.), "Захист мовної інформації в телефонному каналі зв'язку" (1997 р.); науково-дослідної теми "Дослідження і розробка методів та засобів збору і оброблення інформації в задачах автоматизації наукових досліджень", затвердженої Постановою Президії АН УРСР № 535 від 25.11.1983р; госпдоговірної теми "Підготовка фонограм до семантичного аналізу мовної інформації" з ЛВ Київського НДІ судової експертизи (1993 р.) та ряду інших тем.

Мета і задачі дослідження. Метою роботи є розроблення теоретичних основ перетворення часової структури мовних сигналів в різних темпах мовлення та створення нових інформаційних технологій і технічних засобів регулювання темпу мови. Досягнення мети здійснюється шляхом розв'язання таких задач:

- розроблення нової моделі мовного сигналу та нових технологій перетворення часової структури мовного сигналу для задач регулювання темпу мови;
- розроблення нових адаптивних методів та алгоритмів перетворення часової структури мовного сигналу, які забезпечують якісно вищий рівень розбірливості та натуральності мови;
- створення алгоритмів сегментації та маркірування сигналів для адаптивних методів регулювання темпу мови;
- створення нових алгоритмів та структур технічних засобів в рамках існуючих підходів до регулювання темпу мови;
- вивчення закономірностей зміни структури мовного сигналу в різних темпах мовлення;
- проведення тестових експериментів, прикладне і практичне використання отриманих результатів в наукових дослідженнях та інженерних роботах.

Наукова новизна одержаних результатів. На основі виконаних теоретичних та експериментальних досліджень отримано такі нові результати:

- вперше розроблено новий напрямок досліджень - нелінійні перетворення часової структури мовного сигналу - і запропоновано ряд нових методів та алгоритмів регулювання темпу мови, які дозволяють значно розширити діапазон регулювання та підвищити натуральність мовного сигналу;
- на основі вивчення закономірностей зміни структури мовного сигналу в різних темпах мовлення побудовано нову темпоральну модель мовного сигналу та розроблено моделі перетворення

часової структури мови;

- розроблено новий підхід до регулювання темпу мови на основі запропонованих функцій темпоральних перетворень, який дозволяє здійснювати перетворення структури мовного сигналу максимально наближено до природніх процесів при зміні диктором темпу мовлення;

- дістали подальший розвиток методи лінійних перетворень часової структури мовного сигналу: розроблено ефективні алгоритми та структури технічних засобів регулювання темпу мови на основі синхронізації процедур перетворення структури сигналу з основним тоном та підсумовування попередньо зважених відрізків вихідного сигналу;

- запропоновано алгоритм сегментації та маркірування мовних сигналів для адаптивних методів регулювання темпу мови.

Практичне значення одержаних результатів. Розроблені методи та алгоритми перетворення часової структури мовних сигналів дозволяють суттєво збільшити робочий діапазон зміни коефіцієнта регулювання темпу мови та розширити спектр практичних застосувань. Адаптивні технології регулювання темпу мови не лише дають можливість забезпечити трикратне прискорення та сповільнення темпу мовної інформації (для традиційних методів та алгоритмів граничною є зміна темпу в 2,5 разів), але й істотно покращити якість перетвореного мовного сигналу.

На основі отриманих наукових основ та одержаних експериментальних результатів розроблено нові типи технічних засобів - регулятори темпу мови (РТМ), які дозволяють автоматизувати процедури тренування та тестування операторів та диспетчерів, лікування логопедичних хворих, фоноскопічних експертиз, вивчення іноземних мов. Створені технології перетворення часової структури мовного сигналу підвищують швидкість передавання мовного сигналу каналами зв'язку.

Результати впроваджено в Львівському науково-дослідному радіотехнічному інституті при розробленні методів та спеціалізованих технічних засобів захисту від несанкціонованого доступу та засобів передавання мовної інформації каналами зв'язку; Державній санітарно-епідеміологічній службі України при проведенні психофізіологічного тестування операторів та диспетчерів на Львівській залізниці; Державному університеті "Львівська політехніка" в науково-дослідних роботах по шифруванню мовної інформації в телефонних каналах зв'язку, а також в навчальному процесі.

Особистий внесок здобувача. В роботах, написаних в співавторстві, авторові дисертації належить: в [7, 11] - технології регулювання темпу мови; в [2, 5, 6, 19] - методики нелінійних перетворень часової структури мовних сигналів; в [3, 4, 8, 10, 15, 18, 20, 21] участь співавторів була рівнотворчою.

Апробація результатів дисертації. Основні результати роботи викладено та обговорено на 12 Міжнародних, 6 Всесоюзних і 3 Республіканських науково-технічних конференціях (НТК) та семінарах, а саме: 2-й Республіканській конференції "Автоматизація наукових досліджень" (Київ, 1981); 12-й, 13-й та 14 Всесоюзних школах-семінарах "Автоматичне розпізнавання слухових образів" (Київ-Одеса, 1982; Новосибірськ, 1984; Каунас, 1986); 4-й, 5-й та 6-й Всесоюзних школах семінарах "Розпаралелювання обробки інформації" (Львів, 1983, 1985, 1987); 11-му Міжнародному конгресі фонетичних наук (Таллінн, 1987); 20-му Міжнародному колоквіумі з інформаційної техніки (Дрезден, 1987); Міжнародній НТК по обробці сигналів (Рига, 1990); 40-му Міжнародному науковому колоквіумі (Ільменау, ФРН, 1995); Міжнародних конференціях "Методології проектування в обробці сигналів" (Краків, 1994; Закопане, 1996); 1-й та 3-й Всеукраїнській НТК "Оброблення сигналів і зображень та розпізнавання образів" (Київ, 1994, 1996); Всеукраїнській НТК "Комп'ютерні технології друкарства: алгоритми, сигнали та системи" (Львів, 1996); 3-й республіканській НТК "Автоматика-96" (Севастопіль, 1996); 10-й польській національній конференції "Застосування мікропроцесорів в автоматичному контролі та вимірюваннях" (Варшава, 1996); Міжнародній НТК "Сучасні проблеми автоматизованої розробки і виробництва радіоелектронних засобів, застосування засобів зв'язку та підготовки інженерних кадрів" (Львів, 1996); 4-й Польській НТК "Комп'ютерне забезпечення наукових досліджень" (Вроцлав, 1997); Літній школі з використання нейронних мереж для задач оброблення сигналів (Ченстохова, Польща, 1997), а також на щорічних конференціях Державного університету "Львівська політехніка".

Публікації. По темі дисертаційної роботи опубліковано 52 наукових праці, в тому числі одна монографія, 15 статей в наукових журналах та збірниках наукових праць, 31 публікація в збірниках матеріалів та тез конференцій, отримано 5 авторських свідоцтв.

Структура та обсяг роботи. Дисертаційна робота складається із вступу, шести розділів, списку використаної літератури та до-

датків і викладена на 289 сторінках. Число ілюстрацій 61, таблиць 23, додатків 3. Список використаних літературних джерел включає 250 найменувань.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовується актуальність проблеми дослідження і розроблення методів, алгоритмів та засобів перетворення часової структури мовних сигналів з метою забезпечення можливості ефективного прискорення та сповільнення темпу мовної інформації в широкому діапазоні, формулюються мета та основні задачі досліджень, подається анотація основних положень роботи.

В першому розділі аналізуються анатомо-фізіологічні передумови творення та слухового сприйняття мовних звуків, моделі мовотворення та існуючі методи перетворення часової структури мовного сигналу, розроблена структура системи та введено поняття регулювання темпу мови в реальному часі.

Характеристики продуктивності процесів мовотворення в основному визначаються обмеженими можливостями механічних артикуляторів. В повному стилі мовлення середній темп мови складає порядку 70 слів за хвилину, максимальний - 90-100 слів/хв. Експериментальні дослідження слухового сприйняття скомпресованої мови показують, що людина здатна сприймати до 30 фонем за секунду, що відповідає приблизно 250 словам за хвилину. Ця диспропорція є аргументом для розробки систем регулювання темпу мови, в яких зміна часу звучання досягається відповідною зміною швидкості відтворення попередньо записаної мовної фрази, а частотні спотворення, що виникають при цьому, компенсуються регулятором темпу мови.

Аналіз цифрових моделей мовних сигналів показує, що для розроблення методів лінійного розтягування та стискування на основі систем аналізу-синтезу мови найбільш придатними є синусоїдальна модель та модель мовного сигналу на базі короткочасного перетворення Фур'є.

В першій моделі на етапі аналізу мовний сигнал подається у вигляді суми хвильових компонент з виділеними для кожної із хвиль значеннями амплітуди та фази. Часова модифікація проводиться в два етапи: на першому на основі КПФ визначаються значення частот, амплітуди та фази синусоїдальних складових, на

другому - проводиться розділення компонент впливу сигналу збудження та голосового тракту. Для кожної складової параметри збудження та тракту модифікуються так, щоб розширити або скоротити частотні траєкторії, не змінюючи частоту основного тону. При цьому модифікація параметрів тракту відображає процеси прис-

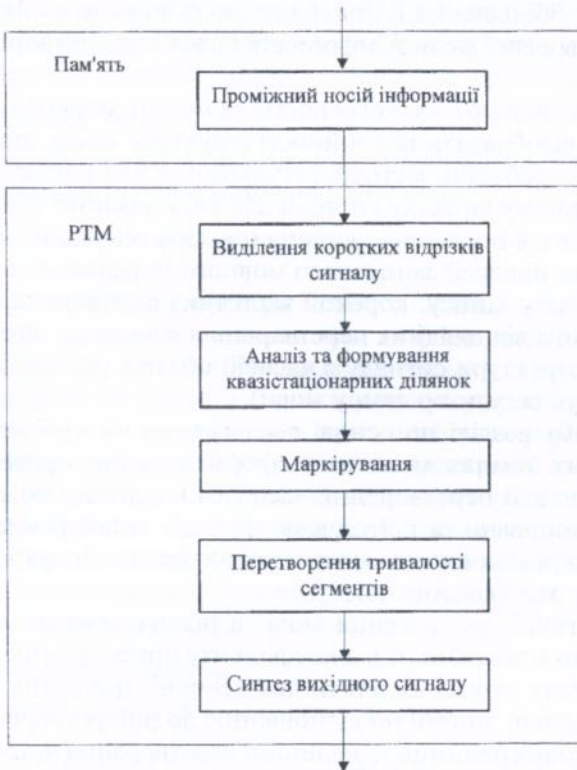


Рис.1. Блок-схема системи для регулювання темпу мови

корення або сповільнення роботи артикуляторів, модифікація параметрів збудження - зміну лінійного масштабу частотних траєкторій.

В моделі на основі КПФ мовний сигнал подається у вигляді лінійної комбінації гармонічно зв'язаних комплексних експонент, де параметри голосового тракту відображаються в повільно змінній в часі амплітуді експоненти, а параметри сигналу збудження

- в швидкозмінній фазовій компоненті. Модифікація КПФ для синтезу на зміненому інтервалі часу полягає в лінійній інтерполяції значень модуля КПФ відповідно до заданого коефіцієнту регулювання темпу мови β та нелінійній модифікації фази.

Обидві моделі забезпечують можливість високоякісного перетворення часового масштабу мовного сигналу в границях зміни β від 0,5 до 2. Збільшення β приводить до різкого падіння розбірливості перетвореної мовної інформації внаслідок спотворення коротких звуків.

Перспективною є кусково-постійна модель мовного сигналу, яка дозволяє відобразити нелінійності структури мови, що є необхідним при розробленні методів регулювання для високих β .

В загальному випадку система для регулювання темпу мови (рис.1) забезпечує послідовне виконання двох операцій: зміни часу відтворення наперед записаного мовного повідомлення по відношенню до часу запису; корекції частотних спотворень мовного сигналу шляхом відповідних перетворень в частотній області, або модифікації структури сигналу в часовій області (власне цю функцію і виконує регулятор темпу мови).

В другому розділі на основі дослідження часової структури мови в різних темпах мовлення сформульовано принципи та розроблено моделі перетворення часового масштабу мовних сигналів, запропоновано та побудовано функції зміни темпоральної структури для різних класів звуків, проаналізовано існуючі методи та алгоритми регулювання темпу мови.

Статистичні дослідження мови в різних темпах мовлення показують, що в залежності від коефіцієнту прискорення або сповільнення темпу можна виділити два основні принципи регулювання темпу мови: лінійні по відношенню до β перетворення тривалостей звукових одиниць та нелінійні перетворення часової структури мовного сигналу. В рамках даного підходу можна сформулювати три основні моделі перетворення часового масштабу мови:

1. Модель на основі пауз.

Мовний сигнал X подається у вигляді послідовного з'єднання звукових ділянок та пауз між словами:

$$X = (X_{s1}, X_{n1}, X_{s2}, X_{n2}, \dots, X_{sn}, X_{nn}, \dots, X_{sn}),$$

де X_{si} - тривалість i -ї звукової ділянки, X_{ni} - тривалість паузи між i -

м та $(i + 1)$ -м словами. Зміна часу відтворення сигналу X досягається шляхом зменшення або збільшення тривалостей пауз. Максималь-

ний коефіцієнт прискорення $\beta_{\max} = \left(\sum_{i=1}^n X_{zi} + \sum_{i=1}^{n-1} X_{ni} \right) / \sum_{i=1}^n X_{zi}$ як

правило не перевищує значення 1,3. Модель забезпечує повне збереження якості та розбірливості мовної інформації, оскільки безпосередньо над звуками ніяких операцій не проводиться.

2. Лінійна модель.

Мовний сигнал X зображений у вигляді послідовного з'єднання коротких ділянок однакової довжини:

$$X = (X_1, X_2, \dots, X_i, \dots, X_m).$$

В процесі зміни структури мовного сигналу тривалість кожної із виділених ділянок перетворюється за формулою: $t_{in} = t_i/\beta$ (де t_i та t_{in} - тривалості ділянки X_i відповідно на вході і на виході РТМ) шляхом вилучення частини ділянки у випадку прискорення темпу, або повторення частини ділянки при сповільненні. Дана модель є найпростішою для реалізації в технічних пристроях, але можливість суттєвого збільшення коефіцієнту прискорення та сповільнення обмежується різким зменшенням розбірливості мовного сигналу за рахунок спотворення та повної редукції окремих коротких звуків та складів. Рекомендований максимальний коефіцієнт регулювання темпу мови для даної моделі складає 2.

3. Адаптивна модель.

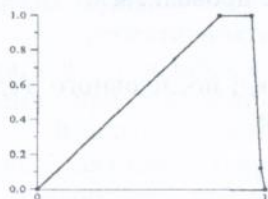
Мовний сигнал X подається у вигляді:

$$X = (X_1(k_1, t_1), X_2(k_2, t_2), \dots, X_i(k_i, t_i), \dots, X_n(k_n, t_n)),$$

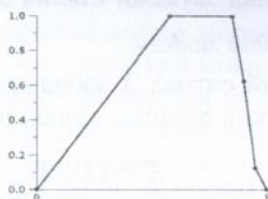
де $X_i(k_i, t_i)$ - позначає виділену в процесі сегментації ділянку мовного сигналу з номером i , тривалість якої t_i , а на основі результатів сегментації та маркірування даних ділянки присвоєний клас k_i . Зміна тривалостей t_i відбувається на основі функцій перетворень $F(k_i, t_i)$, які враховують як клас ділянки, так і її початкову тривалість. Ця модель найбільш повно відповідає реальним процесам трансформації структури мовного сигналу при зміні диктором темпу мовлення і тому забезпечує як найбільший ступінь прискорення/сповільнення, так і натуральність перетвореного сигналу. Рекомен-

дований максимальний коефіцієнт регулювання темпу мови для даної моделі складає 3-3,5 в залежності від початкового темпу.

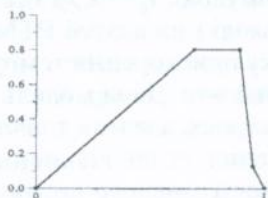
Аналіз особливостей перетворення часової структури мовних сигналів в різних темпах мовлення показує, що в залежності від масштабу зміни тривалості звуку в цілому, перетворення окремих ділянок (стаціонарних та перехідних) кожної фонемі зокрема всі звуки української мови можна розділити на 5 класів: наголошені



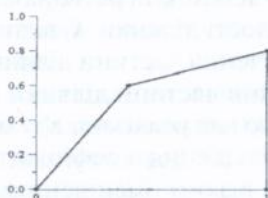
а)



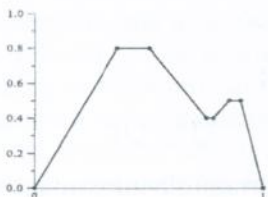
б)



в)



г)



д)

Рис.2. Функції зміни темпоральної структури звуків

- а) наголошеного голосного звуку
- б) ненаголошеного голосного звуку
- в) вокалізованого приголосного звуку
- г) невокалізованого приголосного звуку
- д) вибухового звуку

голосні, ненаголошені голосні, вокалізовані приголосні, невокалізовані приголосні та вибухові (паузи аналізуються окремо). В межах кожного класу при зміні темпу перетворення як тривалості звуку в цілому, так і його окремих ділянок відбуваються подібним чином. На основі обширного статистичного матеріалу для кожного

класу побудовані функції зміни темпоральної структури звуків (рис. 2), які в подальшому використовуються при виконанні перетворень часової структури мовних сигналів.

Існуючі на сьогодні методи регулювання темпу мови (рис. 3) можна класифікувати за двома ознаками: область виконання основних перетворень - часова або частотна - та підхід до процесу трансформації структури - лінійні методи та адаптивні до мовного сигналу (в основному - нелінійні) перетворення. Аналіз-порівняння показує, що лінійні методи та побудовані на їх основі регулятори темпу мови можуть успішно використовуватися в ряді прикладних задач при коефіцієнтах прискорення, що не перевищують величини $\beta = 2$. Розширення діапазону регулювання та підвищення якості звучання досягається шляхом використання адаптивних методів, які виконують нелінійні перетворення часової структури мови на основі поділу сигналу на квазістаціонарні ділянки та диференційованого перетворення їх тривалостей.

В третьому розділі розроблено алгоритм регулювання темпу мови, який дозволяє зменшити частотні спотворення, що виникають при використанні відомого методу вибіркової сегментації. Запропоновано новий метод стискування мовного сигналу за рахунок підсумовування попередньо зважених трапецієдальною вагою функцією сусідніх сегментів початкового сигналу, розроблено методику регулювання темпу мови на основі побудови систем аналізу-синтезу з проміжним параметричним представленням мовного сигналу у вигляді короткочасного перетворення Фур'є.

При використанні відомого методу вибіркової сегментації сигналів внаслідок амплітудних та фазових перепадів, що виникають на границях сегментів в процесі формування вихідного сигналу, виникають спотворення частотного спектру, які погіршують натуральність та знижують розбірливість мови. Розроблено алгоритм перетворення часової структури мовного сигналу шляхом визначення в мовному сигналі опорних точок, віддалі між якими відповідають біжучому періоду основного тону для вокалізованого сигналу та є постійними для невокалізованого, знаходження відповідних їм опорних точок вихідного сигналу та формування прискореної або сповільненої мови на виході шляхом переставлення відрізків вхідного сигналу. Алгоритм працює як при постійному, так і при залежному від часу коефіцієнті регулювання темпу мови β .

Принципово іншим підходом до скорочення тривалості мов-

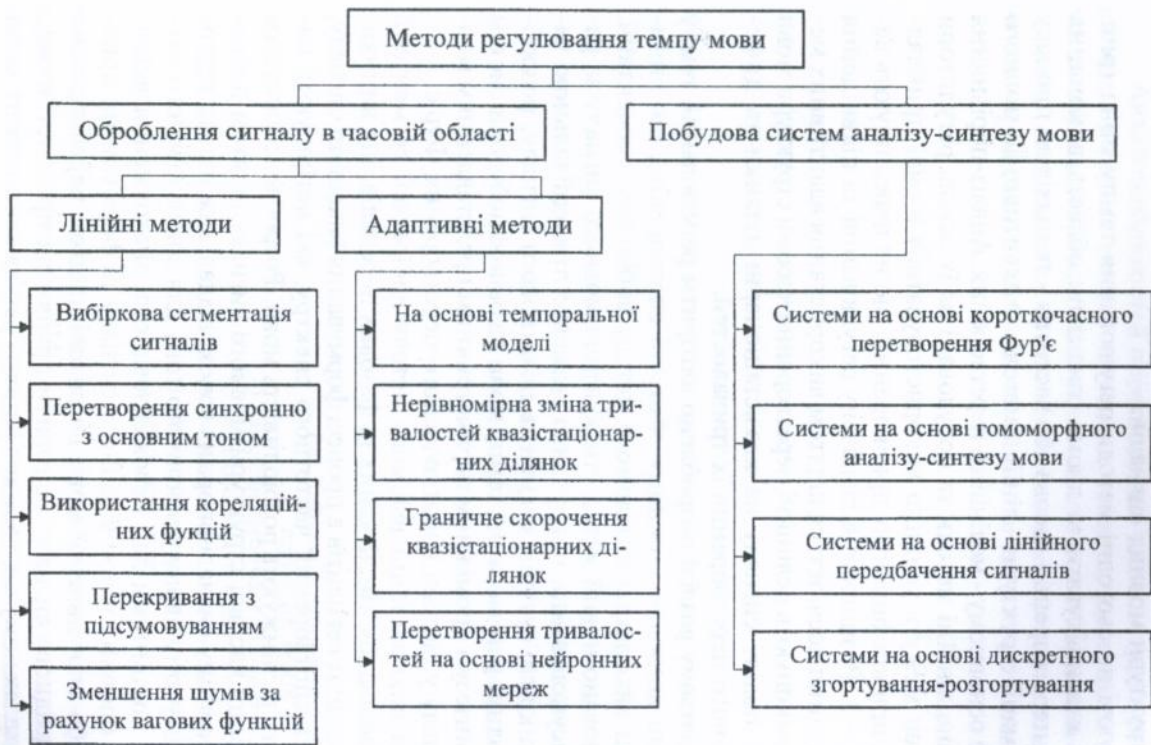


Рис.3. Класифікація методів регулювання темпу мови

ного сигналу є розроблений метод регулювання темпу мови на основі підсумовування попередньо зважених трапеційдальною ваговою функцією сусідніх сегментів початкового сигналу. Даний метод не передбачає вилучення в процесі перетворення фрагментів вхідного мовного сигналу. Скорочення тривалості досягається за рахунок перекривання з підсумовуванням сусідніх відрізків мовного сигналу, спектральний склад яких є близьким. Це дозволяє зберегти всі структурні одиниці початкового сигналу і уникнути зниження розбірливості мови за рахунок значного спотворення або повного вилучення коротких звуків, яке має місце в відомих алгоритмах.

Мовний сигнал ділиться на однакові відрізки тривалістю t (рекомендується тривалість $t = 80$ мілісекунд). Сигнал на кожному із відрізків перемножується на рівнобедренну трапеційдальну вагову функцію, в якій ширина бічної грані дорівнює $t/(1-1/\beta)$, а ширини нижньої та верхньої граней відповідно рівні t та $t(2/\beta - 1)$. В процесі формування вихідного сигналу ділянки, які відповідають верхнім граням трапецій, залишаються без змін, а ділянки сусідніх відрізків початкового сигналу, що відповідають бічним граням, додаються. Метод використовується для значень коефіцієнта регулювання темпу мови β від 1 до 2. У випадку $\beta = 2$ трапеційдальна вагова функція переходить в трикутну.

В загальному, проведені дослідження показують, що лінійні методи перетворення часової структури мовного сигналу, при яких ступінь перетворення тривалості кожного із виділених відрізків мовного сигналу є однаковим, можуть бути успішно використані для розв'язування задач регулювання темпу мови в основному лише в діапазоні зміни коефіцієнту регулювання 0,5-2, що достатньо для задач логопедії, фоноскопічних експертиз, вивчення іноземних мов.

Четвертий розділ присвячено розробці алгоритмів сегментації мовних сигналів в задачах регулювання темпу мови. Розроблено алгоритм автоматичної сегментації та маркування мови (рис. 4), який дозволяє розділити мовний сигнал на відрізки, що відповідають кожному із виділених у другому розділі класів звуків та паузам. Аналіз та інтерпретація мовного сигналу здійснюється на основі 4 параметрів: першого коефіцієнту фільтру авторегресії A_1 , загальної енергії E , відношення E_n/E_n енергії в області низьких частот (250-600 Гц) до енергії на високих частотах (650-3000 Гц) та тривалості сегменту.

Після виділення мовного сигналу на фоні шуму та пауз, по-

ділу його на квазістаціонарні сегменти на основі критерію правдоподібності та виділення параметрів (інформативних ознак) здійснюється класифікація сегментів на 4 групи: голосні, вокалізовані приголосні, невокалізовані приголосні та паузи. Класифікація здійснюється послідовно в два етапи.

На першому етапі відбувається розділення всіх звукових оди-

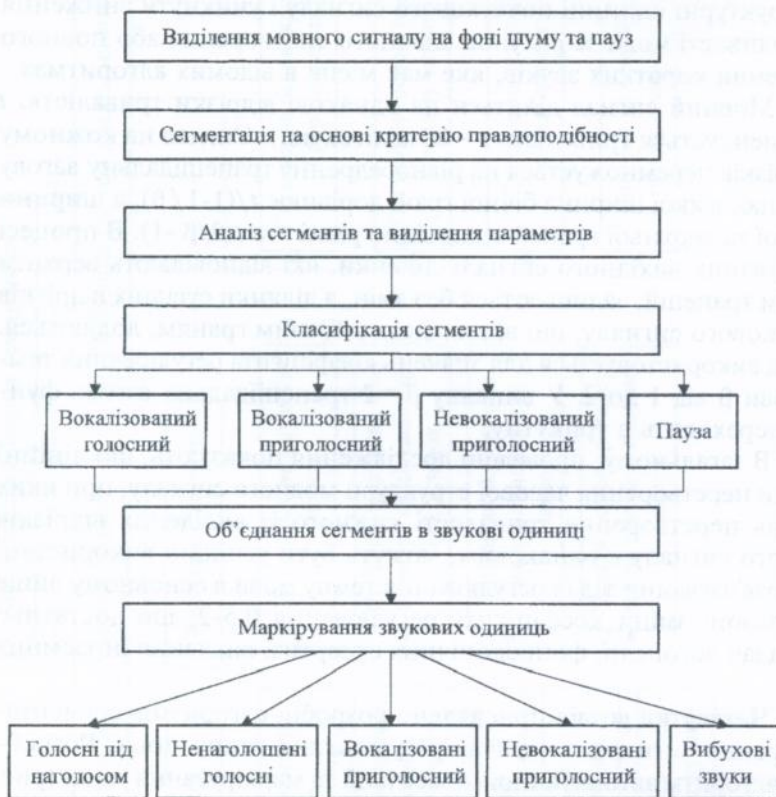


Рис.4. Блок-схема алгоритму автоматичної сегментації та маркірування

ниць початкового сигналу на вокалізовані та невокалізовані звуки, основними ознаками при цьому є значення першого коефіцієнту фільтру моделі авторегресії A_1 та короточасної енергії E . Сегменти, для яких $A_1 > -0,8$, класифікуються як невокалізовані; сегменти, для яких $A_1 < -1,2$, класифікуються як вокалізовані; для решти

сегментів обчислюється ознака вокалізованості $O_v = 0,03E - A_1$, рішення про вокалізованість сегменту приймається у випадку, якщо $O_v > 2,2$.

На другому етапі відбувається розділення вокалізованих сегментів на вокалізовані голосні та вокалізовані приголосні. Основними факторами, які враховуються при цьому є: перевищення значенням короткочасної енергії порогової величини в моменти переходу приголосна-голосна, більша концентрація енергії в низькочастотній області у приголосних звуків в порівнянні з голосними. В результаті сегменти, для яких $A_1 < -1,6$, класифікуються як вокалізовані голосні; серед ділянок, що залишилися, ті, для яких $E > 66$, також класифікуються як вокалізовані голосні; для решти ділянок на основі значень A_1 , E та E_n/E_v обчислюється ознака голосності O_r за експериментально отриманою формулою:

$$O_r = (17E - 70 E_n/E_v)/1000 + 1,1|A_1|,$$

сегменти, для яких $O_r > 2,2$, класифікуються як вокалізовані голосні.

Таким чином, в результаті виконання процедури класифікації, всі виділені звукові сегменти поділені на 4 класи: вокалізовані голосні, вокалізовані приголосні, невокалізовані приголосні та попередньо виділені паузи.

Після цього на основі класифікаційних ознак з урахуванням ряду експериментально отриманих евристичних правил відбувається об'єднання однойменних сегментів у звукові одиниці.

Заключною процедурою є маркірування, яке полягає в присвоєнні одній або декільком звуковим одиницям відповідного класу звуку. На даному етапі нетривіальними процедурами є формування вибухових звуків та виділення голосних під наголосом. Необхідною умовою наявності вибухового звуку є пауза перед звуковою частиною. Додатковою особливістю, яка враховується в процесі маркірування, є короткотривалість звукової частини. Таким чином, виділення вибухових звуків проводиться на основі двох правил:

1. Наявність паузи перед звуковою частиною. В проведених експериментах середня тривалість паузи у вибухових звуках складала 80 мс, середня тривалість звукової частини - 35 мс. При маркіруванні у випадку, якщо слово починалося із вибухового звуку, тривалість паузи приймалася втричі більшою від тривалості звукової частини, решта паузи класифікувалася як пауза між словами.

2. Тривалість звукової частини не повинна перевищувати 40 мілісекунд.

Виділення голосних під наголосом проводиться за наступним правилом: наголошеним вважається голосний звук із найбільшою тривалістю на виділеному фрагменті мови.

Експериментальні випробування підтверджують надійність даного алгоритму та можливість його використання на етапі попереднього оброблення в методах нелінійних перетворень часової структури мови.

В п'ятому розділі запропоновано нову темпоральну модель мовного сигналу, особливістю якої є задання у вигляді функцій темпоральних перетворень алгоритму трансформації часової структури мовного сигналу шляхом, максимально наближеним до природних процесів. На основі цієї моделі розроблено алгоритм адаптивного перетворення мовного сигналу, який дозволяє в 3-3,5 рази прискорювати темп мовної інформації. Описано також два методи регулювання темпу мови - шляхом нерівномірної зміни тривалостей квазістаціонарних ділянок та метод граничного скорочення тривалостей квазістаціонарних ділянок. Поставлено та аналітично розв'язано для обох методів задачу забезпечення необхідної тривалості вихідного мовного сигналу в умовах детермінованості алгоритму перетворення та стохастичності результатів попередньої сегментації. Розроблено методіку використання штучних нейронних мереж для нелінійних перетворень часової структури мовного сигналу в задачах регулювання темпу мови.

У відповідності із розробленим в попередньому розділі алгоритмом автоматичного визначення границь сегментів на основі комплексного статистичного опису сигналу фрагмент мовного сигналу X подається у вигляді послідовного з'єднання ділянок:

$$X = (X_1, X_2, \dots, X_p, \dots, X_n).$$

Кожна ділянка X_i являє собою окремий звук тривалістю I_p , віднесений до певного класу (наголошений голосний, вибуховий, невокалізований приголосний і т.п.), який в моделі зображується відповідним номером класу k_i .

Вводиться поняття функції темпоральних перетворень (ФТП) звуку $w(k_i)$. Ця функція є кусково-лінійною апроксимацією відповідної функції зміни темпоральної структури звуків, введеної в другому розділі. Основне завдання функції темпоральних перетворень - задання алгоритму перетворення тривалості даного звуку, який буде залежати не тільки від типу звуку, його тривалості та коефіцієнту регулювання темпу мови, але й від результатів

сегментації оброблюваного в даний момент фрагменту мовного сигналу. Останнє характеризує важливу особливість регулювання темпу мови в реальному часі, коли тривалість звучання вихідного сигналу з одного боку є детермінованою тривалістю вхідного сигналу та коефіцієнтом регулювання темпу мови, а з другого боку вона визначається стохастичними по своїй природі результатами сегментації.

На рис. 5 зображені приклади функції зміни темпоральної структури звуку (ФЗТС) (рис. 5а) та побудованої відповідно до неї функції темпоральних перетворень (рис. 5б) для ділянки сигналу тривалістю 60 мс при використанні кроку апроксимації 5 мс.

Побудована таким чином ФТП не тільки визначає послідовність вилучення 5-мілісекундних сегментів для даного звуку (для наведеного прикладу спочатку вилучаються сегменти з номерами від 4 до 10, в другу чергу - 3 та 11, пізніше - 2, і насамкінець - 1 та 12), але й дозволяє розставити пріоритети вилучення сегментів між окремими звуками, оскільки обробляються фрагменти мовного сигналу в цілому (кожен фрагмент включає декілька звуків).

Узагальнюючи вищеописане, темпоральна транскрипція фрагменту мовного сигналу тепер має вигляд послідовності тріад $(l, k, w(k))$:

$$\tau = ((l_1, k_1, w(k_1)), (l_2, k_2, w(k_2)), \dots, (l_n, k_n, w(k_n))).$$

Оскільки клас звуку використовується лише для вибору виду функції темпоральних перетворень, то попередній вираз можна дещо спростити:

$$\tau = ((l_1, w(k_1)), (l_2, w(k_2)), \dots, (l_n, w(k_n))).$$

На основі темпоральної моделі розроблено метод нелінійного перетворення часової структури мовного сигналу, характерними особливостями якого є:

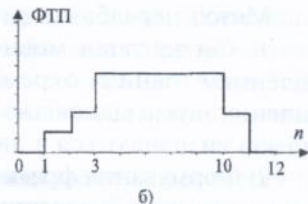
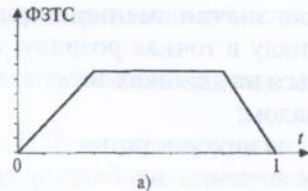


Рис.5. Побудова функції темпоральних перетворень на основі функції зміни темпоральної структури звуку

- максимальне забезпечення відповідності проведених перетворень часової структури мовного сигналу до змін, які відбуваються в структурі мови при переході диктора від одного темпу мовлення до іншого;

- використання ФТП для розширення/стиснення звуків дозволяє значно зменшити рівень шумів при "склеюванні" мовного сигналу в точках розриву, оскільки, як правило, ці точки знаходяться на ділянках звуку з максимально однорідним спектральним складом;

- використання ФТП дозволяє значно спростити проблему забезпечення необхідної тривалості вихідного сигналу в умовах початкової невизначеності результатів сегментації.

Метод передбачає виконання таких операцій:

1. Сегментація мовного сигналу на стаціонарні відрізки з виділенням границь окремих звуків та маркуванням кожного із виділених звуків відповідно до його класу. Тривалості стаціонарних відрізків визначаються з дискретністю 5 мс.

2. Формування фрагментів сигналу, що обробляються як єдине ціле та містять в середньому 4-6 звуків.

3. Відповідно до тривалості t_{ϕ} виділеного фрагменту та коефіцієнту регулювання темпу мови β обчислюється тривалість вихідного фрагменту $t_{\text{вих}} = t_{\phi} / \beta$ та отримане значення заокруглюється до цілого числа з дискретністю 5 мс. Обидва значення представляються у вигляді цілого числа 5-мілісекундних елементарних сегментів (відповідно l_{ϕ} та $l_{\text{вих}}$).

4. Для кожного виділеного звуку відповідно до його класу та тривалості будується своя функція темпоральних перетворень.

5. Проводиться перетворення тривалості відрізків шляхом зміни кількості елементарних сегментів відповідно до заданої ФТП.

У дисертації розроблено алгоритм проведення перетворень для випадку прискорення темпу мови (основне застосування РТМ).

Алгоритм забезпечує збалансоване по всіх відрізках послідовне один за одним вилучення елементарних сегментів (ЕС) тривалістю 5 мілісекунд і включає в себе наступні операції:

- на виділеному фрагменті знаходиться максимальне значення ФТП f_{max} , і відрізки, ФТП яких мають максимум f_{max} , включаються в процес вилучення ЕС;

- виконується циклічна процедура вилучення по одному ЕС на кожному із виділених відрізків, при цьому після кожного вилучення перевіряється, чи загальне число вилучень є рівним $l_{\phi} - l_{\text{вих}}$,

у випадку досягнення мети - процес припиняється;

- якщо після вилучення всіх ЕС для заданого рівня ФТП необхідне скорочення тривалості фрагменту ще не є забезпеченим, то знаходяться наступні максимальні значення f та новий перелік відрізків, які включаються в процес вилучення; для зменшення кількості розривів в мовному сигналі першочергово вилучаються ЕС, які прилягають до вже вилучених;

- на кожному відрізу проводиться "склеювання" частин, що залишилися після вилучення.

6. Отриманий в результаті проведених перетворень фрагмент мовного сигналу зі зміненою часовою структурою зчитується на вихід, і починається обробка нового фрагменту.

Цей метод забезпечує можливість прискорення та сповільнення темпу мовної інформації при коефіцієнтах регулювання темпу мови порядку 2,5 з надійністю розпізнавання не менше 95%, що значно перевищує можливості лінійних методів регулювання темпу мови.

В рамках розглянутої вище адаптивної моделі перетворення часового масштабу мовного сигналу в дисертації запропоновані два нові алгоритми регулювання темпу мови, які не передбачають виділення на етапі сегментації окремих класів звуків.

Для їх опису адаптивна модель дещо спрощується: $X_i(k_p, t_i)$ тепер є функцією лише одного аргументу t_p , тобто характеризує квазістаціонарну ділянку мови. Сегментація початкового сигналу X на послідовність квазістаціонарних ділянок здійснюється за допомогою розробленого в розділі 4 алгоритму блокової сегментації на основі похибки лінійного передбачення. Якщо ділянку $X_i(k_p, t_i)$ зобразити через кількість блоків s_p , що входять в неї, тобто $X_i(s_p)$, то сигнал X має вигляд:

$$X = (X_1(s_1), X_2(s_2), \dots, X_i(s_i), \dots, X_n(s_n)).$$

Перетворення тривалості квазістаціонарної ділянки здійснюється за формулою: $X_i(s_p)_n = F(s_p)$, тобто довжина вихідної ділянки залежить лише від її початкової тривалості.

Алгоритм 1.

Функція перетворення тривалості ділянки має вигляд: $F(s_i) = s_i/k$, де k - експериментально визначений дільник для забезпечення заданого коефіцієнту регулювання темпу мови β (в таблиці 1 подані значення k , які забезпечують можливість зміни β в діапазоні значень

від 1 до 3). При цьому, якщо s_i/k є дробовим числом, то у вихідну ділянку $X_i(s_i)_n$ включаються s_i блоків початкової ділянки, де s_i' - найближче ціле число, більше від s_i/k .

Таблиця 1
Залежність коефіцієнту регулювання темпу мови β від дільника k .

β	1,0	1,35	1,6	2,0	2,35	2,7	2,85	3
k	1,0	1,2	1,5	2,0	2,5	3,5	4,5	10

Особливостями перетворень даного типу є:

- збереження принаймі одного блоку з кожної квазістаціонарної ділянки, що дозволяє в максимальній мірі зберегти всі структурні одиниці початкового сигналу і уникнути втрати розбірливості мови за рахунок вилучення коротких звуків, яка має місце в методах, основаних на вибірковій сегментації мовного сигналу;

- скорочення/збільшення тривалості перетвореного мовного сигналу в основному за рахунок довгих квазістаціонарних ділянок, питома (на один блок) інформативність яких є меншою від питомої інформативності коротких квазістаціонарних, внаслідок чого вилучення частини вхідного сигналу не приводить до суттєвої втрати розбірливості мови;

- оскільки в процесі перетворення мовного сигналу скорочення/збільшення тривалості досягається шляхом вилучення або введення цілого числа блоків фіксованої тривалості (порядку 15-20 мілісекунд), то даний метод може забезпечити можливість зміни коефіцієнту регулювання темпу мови з дискретністю не менше 3-5 відсотків.

Алгоритм 2.

Функція перетворення тривалості ділянки передбачає вилучення з ділянки $X_i(s_i)$ $s_i - 1$ блоків. Очевидно, що в даному випадку тривалість вихідного сигналу визначається числом квазістаціонарних ділянок, на які поділено вхідний фрагмент, тобто - порогом сегментації. Аудиторні випробування якості перетворення мовного сигналу показують подібні результати для обох методів - прискорення/сповільнення в 2,5 рази з надійністю порядку 93%.

Аналогічний підхід до перетворення часової структури мовного сигналу використовується в задачах передавання мовного сигналу каналами зв'язку та в комп'ютерних синтезаторах. Але задача

регулювання темпу мови має ту особливість, що в процесі роботи в реальному часі, коли час звучання вихідного сигналу є детермінованим заданими часом звучання вхідного сигналу та коефіцієнтом регулювання темпу мови, виникає проблема забезпечення потрібної тривалості вихідного сигналу. Виникнення такої проблеми пояснюється тим, що тривалість перетвореного відрізка мовного сигналу на виході РТМ визначається стохастичними по своїй суті результатами сегментації, тобто тривалість $F(s_i)$, отримана в результаті обчислень, є випадковою величиною, яка в загальному випадку не є рівною потрібній тривалості s_i/β . Відтворення на виході РТМ такої ділянки протягом невідповідного їй проміжку часу приведе до виникнення частотних спотворень у вихідному сигналі. Тому в роботі поставлена і для обидвох алгоритмів аналітично розв'язана задача забезпечення необхідної тривалості вихідного мовного сигналу в умовах детермінованості алгоритму перетворення та стохастичності результатів попередньої сегментації.

Розглянуто можливість та запропоновано методику виконання нелінійних перетворень часової структури мови на основі прогнозування за допомогою нейронних мереж. Рекомендованою структурою є 5-шарова мережа з п'ятьма входами та п'ятьма виходами. Опис мовного сигналу задається у вигляді тривалостей біжучої, двох попередніх та двох наступних квазістаціонарних ділянок. Для покращення результатів прогнозування значення тривалостей на вході мережі домножуються на симетричну вагову функцію з коефіцієнтами 1, 0,8 та 0,6.

У шостому розділі розглянуто питання реалізації пристроїв для регулювання темпу мови, які дозволяють підвищити розбірливість та натуральність перетвореної мови в порівнянні з відомими пристроями на основі вибіркової сегментації сигналів.

Розроблені структури регуляторів темпу мови на базі оперативного запам'ятовувального пристрою із зменшенням шумів шляхом використання трикутної вагової функції та РТМ на основі підсумовування попередньо зважених трапецієдальною ваговою функцією сусідніх сегментів. Дані РТМ реалізовані у вигляді лабораторних зразків та успішно апробовані для розв'язування задач фоноскопічних експертиз, логопедії та передавання мовного сигналу каналами зв'язку. Аналіз-порівняння обидвох структур дозволив запропонувати узагальнену базову структуру РТМ на основі вибіркової сегментації сигналів (рис.6). Основними елементами структури є оперативний запам'ятовувальний пристрій (ОЗП), спе-

ціалізований контролер (СК), що має шинну організацію і складається із програмованого інтервального таймера та лічильника адресів, аналого-цифровий (АЦП) та цифроаналоговий (ЦАП) перетворювачі, вхідний (Φ_1) та вихідний (Φ_2) фільтри. В процесі розроблення базової структури основна увага була зосереджена на реалізацію окремих вузлів у вигляді спеціалізованих надвеликих інтегральних схем. Оскільки сучасні інтегральні технології доз-

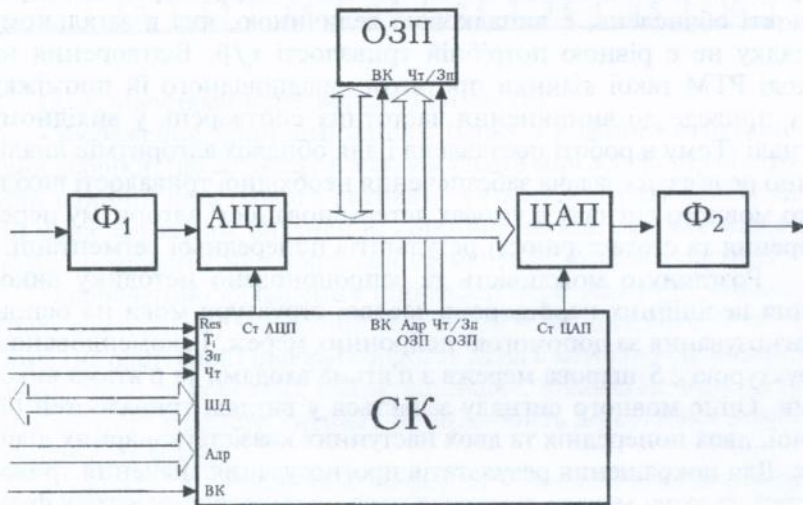


Рис.6. Базова структура регулятора темпу мови на базі ОЗП та спеціалізованого контролера

воляють на одному кристалі реалізувати як цифрові, так і аналогові елементи, то перспективною є реалізація СК з аналоговими входом та виходом. Для реалізації такої інтегральної схеми необхідно фільтри АЦП та ЦАП розмістити на кристалі разом із схемою СК. Це дозволить значно зменшити габарити, потужність споживання і ціну РТМ в цілому.

Розроблена структура дозволяє реалізувати РТМ у вигляді однієї надвеликої інтегральної схеми та використовувати його в комплекті із засобами відтворення мовної інформації, що мають регульовану швидкість.

Для обидвох алгоритмів регулювання темпу мови проведені аудиторні випробування, які показали, що пристрої забезпечують можливість прискорення та сповільнення темпу мовної інформації

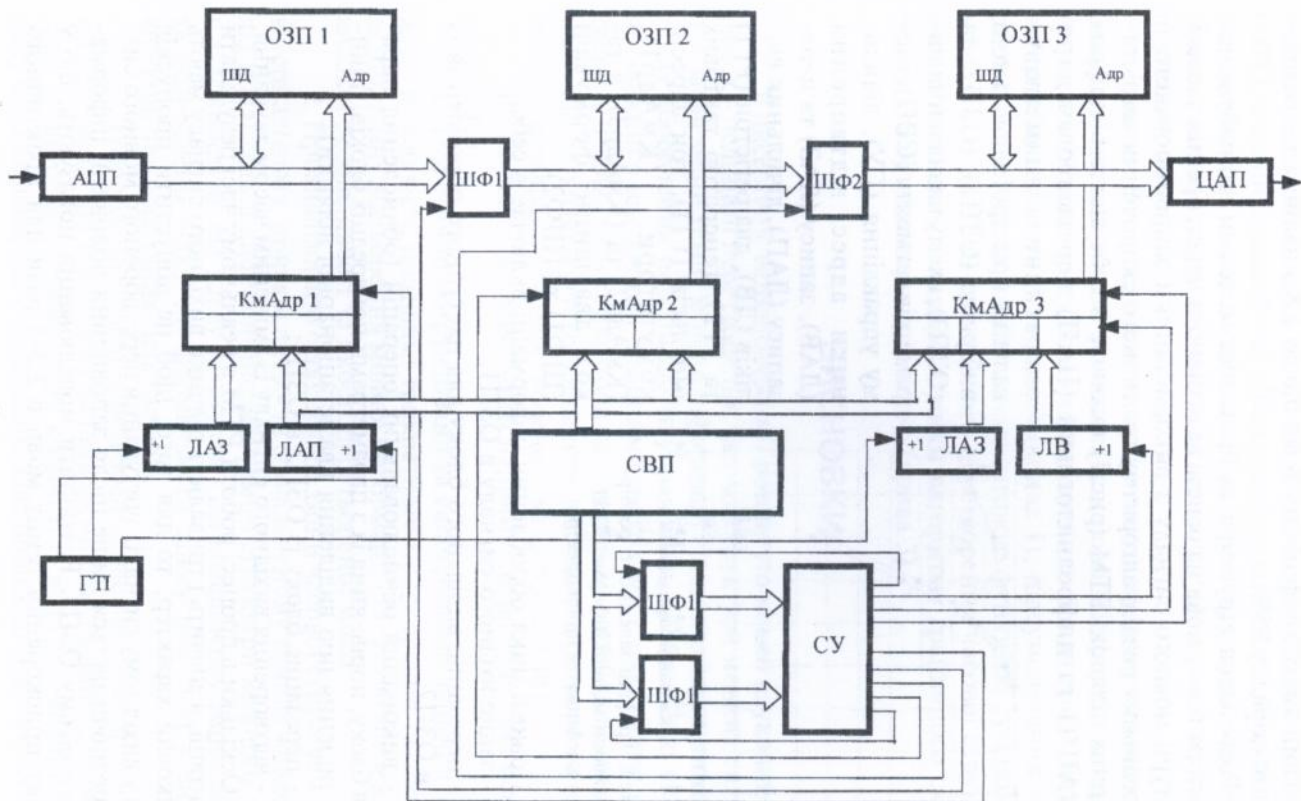


Рис.7. Регулятор темпу мови на основі адаптивного перетворення часової структури мовного сигналу

в діапазоні зміни коефіцієнту β від 0,6 до 1,8 з надійністю розпізнавання порядку 95%.

Розроблена структура та проведені модельні випробування регулятора темпу мови на основі адаптивного перетворення часової структури мовного сигналу з використанням запропонованого в попередньому розділі алгоритму граничного скорочення квазістаціонарних ділянок. РТМ (рис.7) включає в себе аналого-цифровий (АЦП) та цифроаналоговий (ЦАП) перетворювачі, три

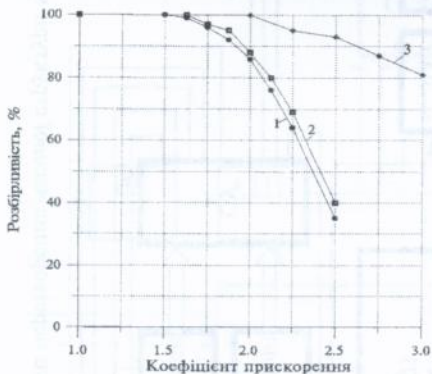


Рис.8. Графіки залежності розбірливості мовної інформації від коефіцієнту прискорення

оперативних запам'ятовувальних пристроїв з довільною вибіркою (ОЗП1), (ОЗП2) та (ОЗП3), схему визначення параметрів сегменту (СВП), схему управління (СУ), лічильники адрес відтворення (ЛАВ), запису (ЛАЗ) та перезапису (ЛАП), лічильник відліків (ЛВ), два регістри (РГ1 та РГ2), генератор тактових імпульсів (ГП), три адресні комутатори (КмАдр1), (КмАдр2) та (КмАдр3), а також два шинні формувачі (ШФ1) та (ШФ2).

Повний цикл оброблення інформації включає в себе:

- запис вхідного сигналу в ОЗП1;
- перезапис виділеного блоку для наступного оброблення із ОЗП1 в ОЗП2;
- виконання обчислювальних операцій (обчислення параметрів блоку, порівняння їх з параметрами попередніх блоків, прийняття рішення про виділення квазістаціонарної ділянки);
- перезапис блоку із ОЗП2 в ОЗП3;
- відтворення вихідного сигналу із змінним часом звучання.

Оскільки в процесі роботи РТМ в реальному часі результати сегментації, а значить і тривалість ділянки вихідного сигналу мають випадковий характер, то для того, щоб не допустити пропусків даних у вихідному сигналі, зчитування перетвореного мовного сигналу розпочинається лише після заповнення половини інформаційного об'єму ОЗП3. Результати моделювання показують, що у випадку прискорення темпу мови в 2,5-3 рази для практичного

уникнення (ймовірність менше 0,01) можливості переповнення ОЗПЗ, його інформаційний об'єм повинен дозволяти зберігати порядку 1 секунди мовного сигналу. Для даної структури розроблені схемні рішення для реалізації адаптивного РТМ на основі багатопортової пам'яті.

На рис.8 приведені отримані шляхом аудиторних випробувань графіки залежності розбірливості мовної інформації від значення коефіцієнту регулювання темпу мови β для трьох методів: вибіркової сегментації сигналів (крива 1), підсумовування попередньо зважених трапецієдальною ваговою функцією сусідніх сегментів (крива 2) та адаптивного перетворення часової структури мовного сигналу з використанням алгоритму граничного скорочення квазістаціонарних ділянок (крива 3).

ВИСНОВКИ

В дисертації запропоновано і розвинуто новий науковий напрямок - нелінійні перетворення часової структури мовних сигналів для задач зміни темпу мовної інформації в людино-машинних системах оброблення інформації та управління, особливістю якого в порівнянні з існуючими лінійними підходами є виконання перетворень часової структури мови адаптивно до оброблюваного сигналу, що дозволяє суттєво збільшити діапазон перетворень та підвищити розбірливість перетвореного мовного сигналу.

В рамках цього наукового напрямку отримано такі основні результати:

1. Теоретично обґрунтовано та експериментально підтверджено можливість принаймі трикратного прискорення та сповільнення темпу відтворення мовної інформації при умові збереження високої розбірливості та якості перетвореної мови.

2. Проаналізовано особливості зміни часової структури мови в нормальному, прискореному та сповільненому темпах мовлення, які полягають в диференційованій зміні структури кожного звуку в залежності як від класу звуку, так і від коефіцієнту зміни темпу, та розроблено підхід до формального опису цих змін у вигляді функцій темпоральних перетворень.

3. Розроблено моделі перетворення часового масштабу мовних сигналів, які є основою для створення як лінійних, коли ступінь розширення/стискування кожного із відрізків мовного сигналу є однаковим, так і нелінійних, адаптивних до мовного сигналу,

методів та алгоритмів регулювання темпу мови.

4. На основі концепції кусково-постійної структури мовного сигналу розроблено нову темпоральну модель мови, яка дозволяє узагальнити та уніфікувати процедури перетворень часової структури в процесі зміни темпу мовлення.

5. Розроблено новий метод нелінійних перетворень часової структури мовного сигналу на основі використання функцій темпоральних перетворень, який базується на виділенні в процесі аналізу окремих класів звуків. Метод дозволяє прискорювати швидкість відтворення мовної інформації в 2,5 рази при збереженні розбірливості 96 %.

6. Запропоновано алгоритми нерівномірної зміни тривалостей квазістаціонарних ділянок та граничного скорочення квазістаціонарних ділянок, які використовують розділення початкового сигналу на ділянки з однорідним спектральним складом. Для кожного із алгоритмів поставлено та теоретично розв'язано задачу забезпечення необхідної тривалості вихідного мовного сигналу в умовах детермінованості алгоритму перетворення та стохастичності результатів попередньої сегментації. Обидва алгоритми дозволяють прискорювати швидкість відтворення мовної інформації в 2,5 рази при збереженні розбірливості 91 %.

7. Розроблено методику використання нейронних мереж для визначення алгоритму перетворення тривалостей структурних одиниць мовного сигналу, яка дозволяє використати в процесі перетворення часової структури інформацію про кореляційні залежності між сусідніми відрізками вихідного сигналу.

8. Розроблено нові лінійні алгоритми регулювання темпу мови з використанням засобів зменшення частотних спотворень сигналу за рахунок:

- синхронізації процедур перетворення структури сигналу з основним тоном;

- підсумовування попередньо зважених трапеційдальною вагою функцією сусідніх сегментів початкового сигналу.

9. Розроблено алгоритм сегментації мовного сигналу на квазістаціонарні ділянки, що відповідають перехідним та стаціонарним ділянкам звуків та паузам.

10. Розроблено методику маркірування структурних одиниць мовного сигналу у відповідності із запропонованою класифікацією звуків української мови для задач регулювання темпу мови.

11. Розроблено та реалізовано у вигляді спеціалізованих при-

строїв нові структури регуляторів темпу мови на базі запам'ятовувальних пристроїв з довільною вибіркою. Аналіз та порівняння структур РТМ дозволили запропонувати узагальнену структуру РТМ на основі вибіркової сегментації сигналів, орієнтовану на реалізацію у вигляді надвеликої інтегральної схеми.

12. Розроблено нову структуру нелінійного регулятора темпу мови та експериментально показано його переваги перед існуючими РТМ за рахунок підвищення якості та розбірливості перетвореного мовного сигналу: у випадку прискорення темпу мови в 2 рази нелінійний РТМ забезпечує збереження розбірливості мовної інформації на рівні 100%, в той час, як РТМ на базі вибіркової сегментації сигналів в аналогічних умовах дозволяє забезпечити розбірливість 88%.

13. Результати розробок та досліджень впроваджено на практиці. До найважливіших практичних результатів можна віднести:

- використання регуляторів темпу мови для тренування операторів в системах диспетчеризації;
- використання методів стиснення інформації в каналах передавання мовного сигналу;
- використання технологій перетворення часової структури мовного сигналу в задачах шифрування мовної інформації;
- використання регуляторів темпу мови для проведення фоноскопічних експертиз в судово-криміналістичних дослідженнях та корекції дефектів мови в логопедії.

СПИСОК ОПУБЛІКОВАНИХ АВТОРОМ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Рашкевич Ю.М. Перетворення часового масштабу мовних сигналів. - Львів: Академічний експрес, 1997. - 140 с.

2. Гепперт Х., Ребайн Ф., Рашкевич Ю.М. Особенности построения нелинейного РТР//Контрольно-измерительная техника. - Львов: Вища школа, 1987. - № 42. - С. 90-93.

3. Балицкий С.В., Гнатив Я.Н., Рашкевич Ю.М. Формирование речевых сообщений вычислительной системы на проблемно-ориентированном языке//Вестник Львовского политехнического института. - Львов: Вища школа, 1987. - № 219. - С. 65 - 68.

4. Рашкевич Ю.М., Марцишин Р.С. Система обробки мовних сигналів на базі мікро-ЕОМ "Искра-226"//Контрольно-вимірваль-

на техніка. - Львів: Світ. - 1992. - № 49. - С. 88 - 91.

5. Шпак З.Я., Рашкевич Ю.М. Регулирование темпа подачи речевой информации//Распознавание и синтез звуковых сигналов. - Киев: ИК АН УССР, 1987. - С. 70-77.

6. Гнатив Я.Н., Рашкевич Ю.М., Шпак З.Я. Методы нелинейных преобразований временной оси речевых сигналов//Речевая информатика. - Киев: Ин-т кибернетики им. В.М.Глушкова, 1989. - С. 105 - 109.

7. Марчук Д.С., Мелень Л.А., Гнатив Я.Н., Рашкевич Ю.М., Шпак З.Я. Метод распознавания слов, использующий пороговые принципы анализа и структурное описание сигнала. Алгоритм ритмизации речи//Автоматическое распознавание и синтез речевых сигналов. - Киев: Ин-т кибернетики им. В.М.Глушкова, 1989. - С. 105 - 108.

8. Марцишин Р.С., Рашкевич Ю.М. Статистичне дослідження темпоральних характеристик мовного сигналу//Вісник Львівського політехнічного інституту. - Львів: Вища школа, 1992. - № 267. - С. 73-76.

9. Ю.Рашкевич. Аналіз методів часового масштабування мовних сигналів//Праці 3-ї Всеукраїнської міжнародної конференції "Оброблення сигналів і зображень та розпізнавання образів". -Київ: Ін-т кібернетики ім. В.М.Глушкова НАН України. - 1996. - С. 129 - 131.

10. Demyda B., Rashkevich Y., Tsmots I. Multi-processor system based on multi port memory for digital signal processing//Proc. of the X Polish national conf. "Application of microprocessors in automatic control and measurement". - Warsaw (Poland). - 1996. -Vol. 2. - P. 9 - 15.

11. Gnativ Y.N., Rashkevich Y.M., Shpak Z.Y. A Device for Correction of Rhythmical Disorders of Speech Functions//Proc. of XIth ICPhS. - Tallinn (USSR). - 1987. - Vol.1. - P. 206 - 209.

12. Rashkevich Y. Non-Linear Time-Scale Modification of Speech by Neural Networks//Proc. of the Summer School on Neural Network Application to Signal Processing. - Czestochowa (Poland). - 1997.- P. 393-395.

13. Rashkevich Y. A Time-Domain Method for Time-Scale Modifications of Speech//Internationales Wissenschaftliches Kolloquium. - Ilmenau (Germany). - 1995. - Band 1. - P. 530 - 532.

14. Rashkevich Y. Nonlinear Transforms of Time Structure of Speech Signals//Latvian Signal Processing Int. Conf. - Riga (USSR). - 1990. - Vol.1. - P. 291-295.

15. Рашкевич Ю., Гнатів Я., Марцишин Р., Шпак З. Регулювання темпу мовної інформації//Праці першої Всеукраїнської конф. "Обробка сигналів і зображень та розпізнавання образів". - Київ: Ін-т кібернетики ім. В.М.Глушкова НАН України. - 1992. - С. 139-140.

16. Rashkevich Y. Speech Time-Scale Modifications: Methods and Trends of Investigation//IV Krajowa Konferencja "Komputerowe wspomaganie badan naukowych". - Wroclaw (Poland), 1997 - P. 517-521.

17. Rashkevich Y. The Methods for Time-Scaling of Speech Signals// Proc. of the Workshop on "Design methodologies for signal processing". - Zakopane (Poland). - 1996. - P. 34 - 37.

18. Демида Б.А., Рашкевич Ю.М., Цмоць І.Г. Мультипроцесорні системи цифрової обробки сигналів на базі багатопортової пам'яті//Матеріали міжнар. конф. "Сучасні проблеми автоматизованої розробки і виробництва радіоелектронних засобів, застосування засобів зв'язку та підготовки інженерних кадрів". - Ч. 2. - Львів. - 1996. - С. 191 - 192.

19. А.с. 1406636 СССР, МКИ G 10 L 3/02. Способ ускоренного воспроизведения речевых сообщений и устройство для его осуществления/ С.В. Балицкий, Я.Н. Гнатив, Ю.М. Рашкевич (СССР). - № 3833250/24-10; Заявлено 02.01.85; Оpubл. 30.06.88, Бюл. №24. - 3с.

20. А.с. 1352522 СССР, МКИ G 10 L 3/02. Устройство для изменения темпа речевой информации/ Я.Н. Гнатив, Ю.М. Рашкевич, З.Я. Шпак (СССР). - № 4032819/24-10; Заявлено 05.03.86; Оpubл.15.11.87, Бюл. № 42. - 3 с.

21. 1633449 (СССР), МКИ G 10 L 1/00. Устройство для изменения темпа речи/Я.Н. Гнатив, Ю.М. Рашкевич, Л.М. Сергийчук (СССР). - № 4685021/10; Заявлено 26.04.89; Оpubл. 07.03.91, Бюл. № 9. - 3 с.

Рашкевич Ю.М. Моделі, методи та засоби перетворень мовних сигналів в часовому просторі. - Рукопис.

Дисертація на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.06 - автоматизовані системи управління

та прогресивні інформаційні технології. - Державний науково-дослідний інститут інформаційної інфраструктури Національного агентства з питань інформатизації при Президентові України, Львів, 1997.

Дисертацію присвячено питанням перетворення часової структури мовних сигналів в задачах регулювання темпу мовної інформації. В дисертації розроблено новий науковий напрямок - нелінійні перетворення часової структури мовних сигналів, який ґрунтується на адаптивному використанні інформації про структуру перетворюваної мови. Запропоновано нові методи, алгоритми та структури технічних засобів регулювання темпу мови, ефективність яких обґрунтована теоретично і підтверджена практично. Основні результати праці знайшли практичне впровадження в ряді технологій оброблення мовної інформації.

Ключові слова: мовний сигнал, перетворення часової структури, регулятор темпу мови.

Рашкевич Ю.М. *Модели, методы и средства преобразования речевых сигналов во временном пространстве.* - Рукопись.

Диссертация на соискание ученой степени доктора технических наук по специальности 05.13.06 - автоматизированные системы управления и прогрессивные информационные технологии. - Государственный научно-исследовательский институт информационной инфраструктуры Национального агентства по вопросам информатизации при Президенте Украины, Львов, 1997.

Диссертация посвящена вопросам преобразования временной структуры речевых сигналов в задачах регулирования темпа речевой информации. В работе развивается новое научное направление - нелинейные преобразования временной структуры речевых сигналов, который основан на адаптивном использовании информации о структуре преобразуемой речи. Предложены новые методы, алгоритмы и структуры технических средств регулирования темпа речи, эффективность которых обоснована теоретически и подтверждена практически. Основные результаты работы нашли практическое применение в некоторых технологиях преобразования речевой информации.

Ключевые слова: речевой сигнал, преобразование временной структуры, регулятор темпа речи.

Rashkevich Y.M. *Models, methods and devices for time-scale modification of speech signals.* - Manuscript.

Thesis for a doctor's degree by speciality 05.13.06 - automatic control systems and progressive information technologies. - The State Institute of Informational Infrastructure of the President's of Ukraine National Agency on Informatization, Lviv, 1997.

The dissertation is devoted to design frontier topics that are connected with the development of the new methods and technical devices for transformation of speech signals time structure for the time-scale modification of speech. A new scientific direction in speech time-scale modification is elaborated which is based on the nonlinear transformation of the time structure of the speech signals by using the information about the current structure of the processed speech. New methods, algorithms and devices are developed, the high effectiveness of them is stated by using of the theoretical investigations and reinforced by illustrative examples. The results of the work have found a practical implementation in some speech signals processing technologies.

Key words: speech signal, time-scale modification, variable speech control system.

A handwritten signature in black ink, consisting of a large, stylized loop followed by a smaller loop and a long horizontal stroke extending to the right.

Підп. до друку 21.11.97р. Формат 60×90/16.
Тираж 100 примірників. Друк різнограф.

ТЗОВ "Простір М"

113/117

AB 39.114

AB 39.114