

БАГАТОШВИДКІСНА ПОЛІКАНАЛЬНА ОБРОБКА СИГНАЛУ МОВНОГО ПОТОКУ**А. Ю. Небилиця**

Черкаський національний університет імені Богдана Хмельницького

бульв. Т. Шевченка, 81, м. Черкаси, 18031, Україна. E-mail: anton.nebylytsia@gmail.com

Обґрунтовано можливість та доцільність використання багатошвидкісної обробки сигналу для поліпшення експлуатаційної ефективності систем розпізнавання природної мови. Запропоновано метод передискретизації сигналу за трьохканальною схемою та адаптивного управління параметрами функціонування. Встановлено, що роздільна обробка за високочастотним, вокалізованим і інфразвуковим каналами підвищує гнучкість та ефективність застосування методів цифрової обробки задля виділення інформаційних ознак мови. Описані принципи динамічного управління частотою дискретизації за деривацією частоти основного тону мовного потоку. Визначено критерії зниження валідності проміжних результатів та механізм корекції параметрів цифрової обробки сигналу за даними процесу ідентифікації фону. Наведено результати порівняльного аналізу ефективності багатошвидкісної обробки та типових методів обробки мовного потоку.

Ключові слова: мовний людино-машинний інтерфейс, розпізнавання мови, мовний потік, сегмент основного тону, багатошвидкісна обробка сигналу, передискретизація.

МНОГОСКОРОСТНАЯ ПОЛИКАНАЛЬНАЯ ОБРАБОТКА СИГНАЛА РЕЧЕВОГО ПОТОКА**А. Ю. Небылиця**

Черкасский национальный университет имени Богдана Хмельницкого

бульв. Т. Шевченко, 81, г. Черкассы, 18031, Украина. E-mail: anton.nebylytsia@gmail.com

Обоснована возможность и целесообразность использования многоскоростной обработки сигнала для улучшения эксплуатационной эффективности систем распознавания речи. Предложен метод передискретизации сигнала по трехканальной схеме и адаптивного управления параметрами функционирования. Установлено, что раздельная обработка по высокочастотным, вокализованным и инфразвуковым каналам повышает удобство, гибкость и эффективность применения методов цифровой обработки для выделения информационных признаков языка. Описаны принципы динамического управления частотой дискретизации исходя из деривации частоты основного тона речевого потока. Определены критерии снижения валидности промежуточных результатов и механизм коррекции параметров цифровой обработки сигнала по данным процесса идентификации фону. Приведены результаты сравнительного анализа эффективности многоскоростной обработки и типовых методов обработки речевого потока.

Ключевые слова: речевой человеко-машинный интерфейс, распознавание речи, речевая последовательность, сегмент основного тона, многоскоростная обработка сигнала, передискретизация.

АКТУАЛЬНІСТЬ РОБОТИ. Використання засобів обчислювальної техніки, з часів її появи і по сьогодні, проводилось шляхом тактильного управління та візуального контакту з монітором. Такого роду людино-машинний інтерфейс (НМІ) є класичним і задовольняє більшість вимог користувачів. Однак, мають місце випадки, які обмежують зручність використання класичного інтерфейсу взаємодії людини з комп'ютеризованою технікою чи системами. До таких випадків слід віднести експлуатацію утилітарних роботів, систем «розумний дім», бортових систем управління і т.п. Причина полягає у неможливості забезпечення постійного контакту із засобами введення та виведення інформації класичного виду. Альтернативу класичному НМІ може скласти природний мовний інтерфейс. На сьогодні більшість операційних систем вже мають вбудований підсистеми розпізнавання та синтезу голосу, які дозволяють реалізовувати утиліти голосового набору чи озвучування тексту, а також голосового управління або синтезу інформаційних повідомлень природною мовою. На жаль, забезпечувана ними достовірність розпізнавання голосу суттєво поступається можливостям людини і є недостатньою для ефективного використання [1]. Крім того, програмні засоби мовного НМІ часто недостатньо мобільні, оскільки орієнтовані на використанні віддалених

серверів, занадто громіздкі і потребують значних обчислювальних ресурсів, а значить зумовлюють підвищений рівень споживання живлення. За вказаних обставин, підвищення ефективності використання природомовного НМІ є актуальною задачею.

Мета роботи – комплексне вирішення проблем підвищення достовірності розпізнавання мови та зниження обчислювальних затрат у мовному людино-машинному інтерфейсі.

Для досягнення даної мети були вирішені наступні задачі:

- аналіз характеру мовного потоку та процесу розпізнавання мови;
- визначення параметрів складових мовного сигналу;
- вибір первинної частоти дискретизації мовного потоку;
- встановлення способу декомпозицій сигналу природної мови;
- розробка способу реалізації багатошвидкісної обробки мовного сигналу;
- встановлення обсягу обчислювальних ресурсів для проведення багатошвидкісної обробки звукових даних;
- визначення впливу багатошвидкісної обробки на показники ефективності виділення інформаційних ознак мовного потоку.

Базова теорія і методи багатошвидкісної обробки добре розроблені [2–4], але їх використання обмежене сферою передачі даних. З цієї причини результати проведених досліджень будуть подаватись із більш розлогим поданням специфіки процесу розпізнавання мови.

МАТЕРІАЛ І РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ. Мовний потік являє собою чергування двох складових – мовних послідовностей та пауз. З фізичної точки зору, мовна послідовність проявляється у вигляді полігармонічних акустичних коливань, амплітуда яких перевищує рівень шумів понад 10 дБ. З фонетичної – вона є послідовністю фонем. Паузи в потоці мовлення, як правило, виступають роздільниками між окремими лексичними одиницями, хоча можуть мати місце і в межах останніх.

За такої структури мовного потоку процес розпізнавання мови проводиться в наступній послідовності стадій: виділення із акустичного сигналу первинних інформаційних ознак, сегментації потоку, виділення вторинних інформаційних ознак, локалізації та ідентифікації фонем, розпізнавання мовних образів. До групи первинних ознак доцільно віднести: признаки початку та кінця лексичних одиниць, фонем, сегментів основного тону (ОТ); статистичні параметри мовлення. Група вторинних ознак включає модельне або параметричне представлення форми сигналу чи його спектру. Як первинні, так і вторинні інформаційними ознаки можуть бути отримані в часовій або спектральній областях сигналу. Однак, вони відрізняються за функціональним навантаженням, частотою та характером прояву, що відображено в табл. 1.

Таблиця 1 – Відмінності між підгрупами інформаційних ознак

Назва груп інформаційних ознак	Функціональне навантаження	Частотний діапазон, Гц	Характер прояву
первинні	локалізація, адаптація	0,1 .. 320	динамічний
вторинні	нормалізація, ідентифікація	80 .. 8000	статичний

Доцільність застосування такої класифікації стосовно розгляду процесу розпізнавання мовних образів зумовлена можливістю часового рознесення операцій виділення первинних та вторинних інформаційних ознак мовного потоку. Перекриття частотних діапазонів в межах 80 .. 320 Гц не створює перепон, оскільки ця смуга охоплює всі можливі частоти ОТ, яка на кожний момент часу має конкретне значення і може визначатись за методикою [5]. Саме роздільність операцій є першим прецедентом, оскільки надає можливість проведення багатошвидкісної обробки сигналу природної мови.

Теорія обробки сигналів за шириною частотного діапазону поділяє сигнали на вузько-, широко- та надширокопasmові. Такий поділ обумовлено доцільністю використання уніфікованих підходів щодо вибору виду процесів, методів і параметрів проведення маніпуляцій над сигналом. За критерій поділу вибирають відносний показник ширини смуги μ , який обчислюють виходячи з величин мінімальної f_{\min} та максимальної f_{\max} частот сигналу за разом $\mu = 2 \cdot \Delta F / (f_{\min} + f_{\max})$, де величина ширини смуги $\Delta F = f_{\max} - f_{\min}$. Для частотного діапазону мовного потоку від 0,1 Гц до 8000 Гц (див. табл. 1) показник $\mu = 0,5$, тому сигнал природної мови слід відносити до широкопasmових сигналів. Дана властивість акустичного сигналу є другим прецедентом застосування багатошвидкісної обробки внаслідок гарантованого підвищення ефективності обробки і аналізу таких сигналів за рахунок розбиття значної за шириною смуги на ряд частотних піддіапазонів.

Третій прецедент, який впливає на вибір методів обробки сигналу, є величина селективності за часо-

тою $\Delta f = f_{clk} / N$, де f_{clk} – частота дискретизації, а N – ширина віка аналізу сигналу. Змінюючи f_{clk} або N , можливо управляти значенням показника Δf , що дозволяє досягти оптимізації обчислень і спростити проведення аналізу даних на стадіях ідентифікації та розпізнавання мовних образів.

Аналогічні наслідки застосування багатошвидкісної обробки зумовлює четвертий прецедент, який являє собою необхідність синхронізації та узгодження параметрів обробки чи аналізу за моментами виявлення у мовному потоці його характерних фрагментів, в першу чергу, сегментів ОТ. Переваги обробки сигналів на інтервалі ОТ обґрунтовано А.Н. Огородніком [6], а необхідність узгодження параметрів обробки шляхом передискретизації – у роботах [7–8].

Узагальнюючи результати аналізу характеру мовного потоку та процесу його обробки, слід відзначити наступне. Підґрунтям багатошвидкісної обробки стали такі прецеденти: можливість цілковитого розділення виконання базових операцій, широкопasmовий вид мовного сигналу, необхідність оптимізації обчислень та мінімізація спотворень інформаційних ознак на всіх етапах обробки мовного потоку.

Теоретичні основи методу. Основна енергія мовного сигналу знаходиться в діапазоні частот від 80 Гц до 8 кГц, але людина здатна розпізнавати мову з прийнятною достовірністю у значно вузькому діапазоні частот – від 300 Гц до 3,5 кГц. Таку властивість сприйняття мови людиною широко використовують, наприклад, для звуження каналу передачі даних у телефонії. Ще за більшого гіршого рівня розбірливості частотний поріг може бути обмежений частотою 2,4 кГц. Однак машинний спосіб розпізнавання мови, в силу своєї недосконалості, низь-

кого рівня адаптивності, відсутності можливості забезпечення просторової селективності та обмеженого рівня когнітивної логіки, потребує залучення до аналізу всього діапазону частот мовного сигналу. За таких обставин достовірність машинного розпізнавання мови в реальних умовах комунікації не перевищує 80 %, що є недостатньо для більшості сфер використання [1].

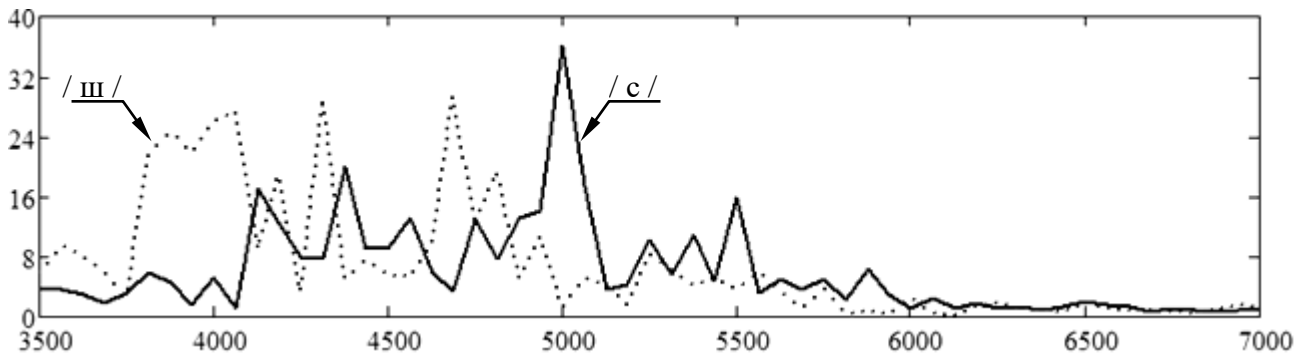


Рисунок 1 – Амплітудно-частотні характеристики звуків /с/ та /ш/

Як засвідчує АЧХ, найбільша потужність сигналу шиплячих і свистячих звуків припадає на діапазон 3..6 кГц. Враховуючи таку властивість, за граничну частоту сигналу мовного потоку можливо прийнятий поріг у 6 кГц. За такого вибору мінімальна частота дискретизації $f_{clk0} = 12$ кГц, і саме вона визначає первинну швидкість обробки акустичних даних.

Сегментація мовного потоку проводиться у смузі починаючи з частоти слідування фонем, а це 0,1 Гц, і аж до f_p – частоти слідування сегментів основного тону (pitch frequency). За таких обставин доцільно за частоту дискретизації на етапі сегментації мовного потоку вибрати $f_{clk1} = f_p$.

Частотна смуга сигналу вокалізованих фонем, які містять голосні звуки, не може бути однозначно визначена, оскільки величина як нижньої, так і верхньої меж залежить від параметрів тракту диктора та підходів модельного представлення мови. Так, діапазон нижньої межі визначається частотою ОТ і має деривацію в чотири октави, а це від 80 Гц до 150 Гц для чоловічого голосу та від 150 Гц до 320 Гц для дитячих і жіночих голів. А ось варіативність верхньої межі смуги вокалізованих фонем зумовлена вибором моделі мови, причому величину визначають за більш високочастотним жіночим голосом. За трьохформантною моделлю мови верхня межа становить 3000 Гц, а за чотирьохформантною – 3800 Гц. Найбільшу відповідність вимові вокалізованих звуків, на думку автора, відповідає трьохформантна модель мови. Підґрунтям такої точки зору є зменшення до 8 дБ амплітуд четвертої форманти відносно першої, що зумовлено суттєвим затуханням високочастотних компонент голосу внаслідок дисипації акустичної енергії в голосовому тракті та випромінювання в оточуючий простір

У відповідності до теореми Котельнікова про вибірку частота дискретизації, f_{clk} повинна, як мінімум, у два рази перевищувати максимальну частоту сигналу f_{max} . Для мовного сигналу високочастотна ділянка спектру більшою мірою притаманна свистячим та шиплячим звукам, АЧХ деяких з них приведено на рис. 1, на якому вісь абсцис – частота, Гц.

через відкриту ротову порожнину. Значне зменшення амплітуди четвертої форманти фактично робить її співставною з рівнем шумів мовного сигналу.

Наведений опис частотного характеру сигналу вокалізованих фонем спонукає до трьох важливих висновків. По-перше, обробка вокалізованої частини сигналу може проводитись із пониженням частоти дискретизації у два рази. По-друге, значний рівень кореляції між частотами формант звуків та частотою ОТ вказує на доцільність проведення над сигналом вокалізованих фонем операції частотної нормалізації, причому таку операцію задля мінімізації спектральних спотворень доцільно виконувати у часовій області сигналу шляхом передискретизації [7–8]. По-третє, здійснення дробової передискретизації потребує обробки та аналізу сигналу мовного потоку в смузі частот від 80 Гц до 3000 Гц, тому проводиться окремо від іншої частини мовного сигналу.

Реалізація методу. Призначенням процесу багатшвидкісної цифрової обробки мовного сигналу є здійснення ефективної трансформації вхідного потоку даних $y(t)$ в множину значень достовірності ідентифікації фонем $\{p_F\}$ та множину параметрів опису сегментів $\{p_S\}$, які комплексно забезпечують стабільну роботу підсистеми розпізнавання мовних образів. Синтез процесу проводився за критерієм підвищення достовірності ідентифікації фонем та виду сегментів мовного потоку. Інші показники якості роботи системи розпізнавання образів вважалися вторинними.

Задля наглядності процес багатшвидкісної обробки мовного сигналу приведений у вигляді схеми на рис. 2, з якого можливо виділити три канали обробки вхідних даних: високочастотний, вокалізованої смуги частот, інфразвуковий. Опис схемних позначень процесів та їх функціонального призначення приведено нижче за текстом.

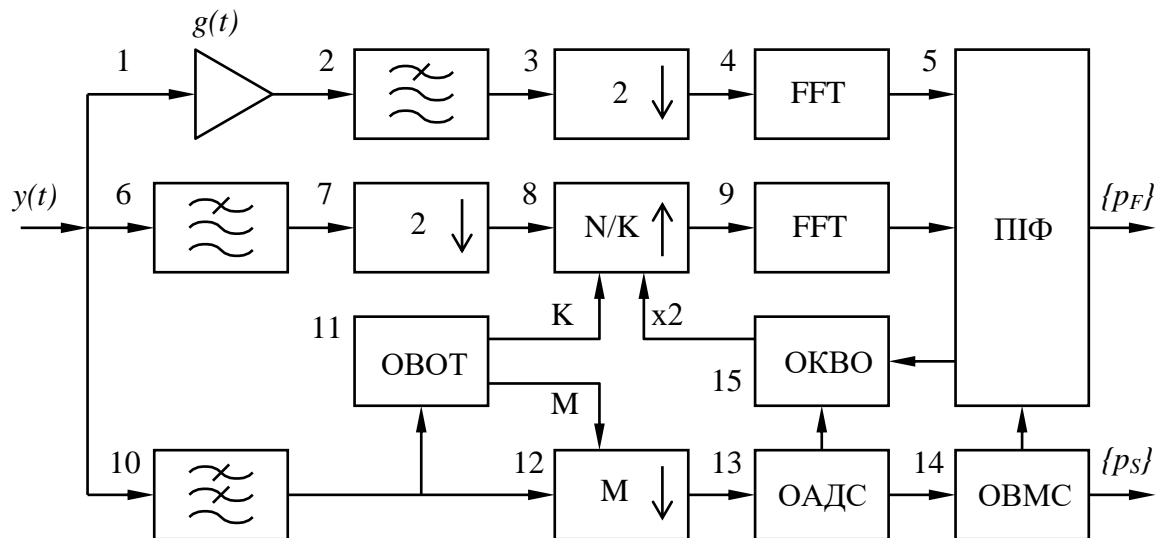


Рисунок 2 – Схема процесу багатошвидкісної обробки сигналу мовного потоку

Високочастотний канал обробки утворений послідовністю операцій п. 1-4 і призначений для виділення та перетворення частини сигналу, яка відповідає акустичному сигналу шиплячих та свистячих звуків. Операції п. 1 і п. 2 забезпечують гетеродинне перенесення спектру сигналу [9] на $0,5 \cdot f_{clk0}$ шляхом перемноження $y(t)$ на $g(t) = \cos(\pi \cdot f_{clk0} \cdot t)$. Враховуючи дискретність часової осі, косинус набуде значень $g_i = \{1, -1, 1, -1, \dots\}$, що, в силу тривіальності операцій множення на 1 або -1, еквівалентно операції інвертування знаку кожного другого значення потоку. Операція п. 2 являє собою низькочастотний фільтр із частотою зрізу $0,5 \cdot f_{clk0}$. В результаті зазначених маніпуляцій над сигналом можливо понизити частоту дискретизації у два рази, що дозволить тримати необхідну роздільну здатність спектра високочастотної компоненти мовного сигналу за умов урізаної ширини вікна аналізу [10].

Канал обробки вокалізованих звуків утворений послідовністю операцій п. 6-9. Особливість цього каналу полягає у проведенні передискретизації сигналу в дві стадії, що зумовлено потребою мінімізації обчислень і досягається в наступний спосіб. Операція п. 8 являє собою частотне еспандування сигналу з раціональним коефіцієнтом інтерполяції, яка реалізується шляхом використання поліноміального фільтру третього порядку. Такі фільтри на одну ітерацію потребують проведення трьох операцій множення та додавання, а також три множення і вісім додавань на кожну із $K-2$ точок вхідних даних. Отже, зменшення довжини вибірки в два рази зменшить витрати обчислювальних ресурсів у 3..4 рази.

Функціональне навантаження інфразвукового каналу обробки сигналу полягає у локалізації сегментів ОТ та макросегментів мовного потоку. Перші виділяються в процесі п. 11, яким позначена на схемі операція виділення основного тону (ОВОТ), яка

крім здійснення локалізації забезпечує визначення тривалості та енергію ОТ. Отримані параметри слугують для знаходження коефіцієнтів інтерполяції K та децимації частоти M . У подальшому частота дискретизації на етапі виконання операцій аналізу динаміки сигналу (ОАДС) та операції виділення макросегментів (ОВМС) плаваюча і становить величину $f_{clk1} = f_p$.

Результати обробки сигналу мовного потоку всіх трьох каналів складають основу проведення процесу ідентифікації фонем (ПФ). У випадку, коли рівень ідентифікації поточної фонемою незадовільний, то ця подія активізує операцію корекції визначення інформаційних ознак (ОКВО). Критерієм настання події є виділення із загального списку більше ніж трьох фонем із достовірністю $p_{F_i} \geq 0,2$. Суть корекції полягає у підвищенні частотної селективності за рахунок розширення вікна аналізу.

Результати, отримані в процесі багатошвидкісної обробки, $\{p_F\}$ та $\{p_s\}$, є вхідними даними процесу розпізнавання мовних образів (РМО). У загальному випадку РМО також може потребувати корекції результатів отриманих на попередньому етапі. Проте розгляд такого виду зворотних зв'язків виходить за межі даних досліджень, тому з метою спрощення схеми рис. 2 вони не приводились.

Ефективність методу. З огляду на схему рис. 2 домінує первинна думка, що запропонований метод занадто громіздкий і нездатен мінімізувати обчислення, тому доречними будуть такі доводи. Складність етапу виділення інформаційних ознак більшою мірою визначається витратами обчислювальної потужності на проведення спектрального перетворення, обчислювальна складність якого $O(N \cdot \log_2 N)$. Для класичного методу, тобто без синхронізації вибірки із періодом ОТ, довжина вікна аналізу значна $N \geq 512$ відліків, що зумовлено вимогами перекриття вікон і накладання зважувальної функції [11-12]. Натомість,

в запропонованому методі довжина вікна перемінна і може становити 64 відліків для слідуючого режиму та 128 відліків у випадку переходу в режим корекції. Нагадуємо, що зменшення розміру вікна аналізу пов'язано з прив'язкою до тривалості мінімального сегменту мови – ОТ. Частотний діапазон високочастотного сигналу обмежений, а спектр зміщений (рис.1), тому необхідна вибірка для проведення FFT – мінімальна і становить 64 відліків, що відповідає селективності $\Delta f = 3000/32 = 93,75$ Гц. За таких обставин класичний метод потребує виконання 4608 умовних операцій типу «метелик», натомість запропонований метод для обчислення спектрів по вокалізованому та високочастотному каналам потребує всього $128 \cdot 7 + 64 \cdot 6 = 1280$ операцій, тобто в 3,6 рази менше.

ВИСНОВКИ. Підбиваючи підсумки за результатами досліджень слід, виділити так:

1. Характер мовного потоку дозволяє проводити його багатошвидкісну обробку.

2. Оптимальна первинна частота дискретизації 12 кГц.

3. Доцільно багатошвидкісну обробку проводити у трьохканальній реалізації за частотними смугами – високочастотною, вокалізованою та інфразвуковою.

4. Зміна частоти дискретизації у відповідності до деривації частоти ОТ дозволяє мінімізувати спотворення інформаційних ознак.

5. Багатошвидкісна обробка забезпечує гнучкість управління якістю виділення інформаційних ознак мовного потоку.

Зазначені властивості дають підстави стверджувати, що використання багатошвидкісної обробки мовного сигналу у відповідності до запропонованого методу дозволяє покращити процес сегментації мовного потоку, достовірність ідентифікації фонем та підвищити експлуатаційну ефективність систем розпізнавання мови. Альтернативою основному призначенню метода є його використання у сфері цифрової телефонії, наприклад, задля параметризації мовного сигналу.

ЛІТЕРАТУРА

1. Семенова Я.С. Автоматизированная система распознавания интонации, подсчета слов и устойчивых словосочетаний в речи человека. / CLAIM – научно-образовательный кластер. [Електронний ресурс]. – Режим доступу: http://it-claim.ru/Persons/Semenova_Yana/speech.htm (25.05.16).

2. Fredric J. Harris Multirate Signal Processing for Communication Systems. Prentice Hall PTR. – 2004. – 478 p.

3. Bojan Vrcelj Multirate Signal Processing Concepts in Digital Communications. California Institute of Technology. Pasadena, California. – 2003. – 113 p.

4. Multirate DSP and its technique to reduce the cost of the analog signal conditioning filters. Richa Maheshwar, Manisha Bharadia, Meenakshi Gupta. International Journal of Computer Applications. –Vol. 4, No. 10. – 2010. – P. 27–34.

5. Небилиця А.Ю. Оптимізація методу виділення мінімальних сегментів мовного потоку – Вісник Черкаського університету: Серія «Прикладна математика. Інформатика», 2014. – № 18 (311). – С. 59–67.

6. Огородников А.Н. Выбор интервалов анализа сигнала при распознавании речи. Вестник Томского государственного университета, 2003. – № 280. – С. 285–304.

7. Азаров И.С. Изменение основного тона речевого сигнала в реальном масштабе времени // Международная научно-техническая конференция, приуроченная к 50-летию МРТИ-БГУИР (Минск, 18-19 марта 2014 г.): Материалы конф. в 2 ч. – Ч. 1. – Минск. – 2014. – С. 274–275.

8. Вашкевич М.И. Передискретизация речевого сигнала, согласованная с частотой основного тона // Международная научно-техническая конференция, приуроченная к 50-летию МРТИ-БГУИР (Минск, 18-19 марта 2014 г.): Материалы конф. в 2 ч. Ч. 1. – Минск. – 2014. – С. 308–309.

9. Ричард Лайонс Цифровая обработка сигналов: Второе издание. Пер. с англ. – М.: ООО «Бином-Пресс». – 2006. – 656 с.

10. High resolution speech feature parametrization for monophone-based stressed speech recognition. R. Sarikaya, J. H. L. Hansen. IEEE Signal Processing Letters, Vol. 7, Iss.: 7. – 2000. – P. 182–185.

11. The use of variable frame rate analysis in speech recognition. K.M. Ponting, S.M. Peeling. Computer Speech&Language, Vol. 5, Iss. 2. – 1991. – P. 169–179.

12. Entropy-based variable frame rate analysis of speech signals and its application to asr H. You, Q. Zhu, A. Alwan. IEEE International Conference. Acoustics, Speech, and Signal Processing, Vol. 1. – 2004. – P. 549–552.

MULTIRATE AND MULTICHANNEL SIGNAL PROCESSING OF SPEECH STREAM

A. Nebylytsia

The Bohdan Khmelnytsky National University of Cherkasy

aven. Shevchenko, 81, Cherkasy, 18031, Ukraine. E-mail: anton.nebylytsia@gmail.com

Purpose. The research is intended to contribute to the challenges concerning the increasing of confidence in speech recognition and reducing the computational effort in voice human machine interface (HMI). **Methodology.** The method of three-channel scheme signal resampling and adaptive management of functioning parameters is proposed. **Results.** The paper substantiates the possibility and feasibility of using multirate signal processing to improve operational efficiency of the speech recognition systems. It was determined, that the separate processing of high-frequency, vocalized and infrasound channels increases convenience, flexibility and efficiency of using digital processing techniques to extract information language signs. The results of comparative study of the multirate processing effectiveness compared

to the usual methods of processing the speech stream are presented. **Originality.** The principles of dynamic control of the discretization frequency taking into account the derivation frequency of the main speech stream have been described. It has been established the criteria for validity reducing of the intermediate results and the parameters adjustment mechanism of digital signal processing according to the phoneme identification process. **Practical value.** The method of multirate processing in three-channel and two-contour adaptive implementation, which was claimed in this research, ensures easiness and increases the reliability of the speech recognition subsystems work, provides high operational efficiency of voice HMI. References 12, table 1, figures 2.

Key words: voice human machine interface, voice recognition system, speech recognition, speech series, segment pitch, multi-rate signal processing, resampling.

REFERENCES

1. Semenova, Ya.S., (2015), "Automatic tone recognition system, counting words and phrases in the human speech ", *CLAIM – nauchno-obrazovatelny klaster*, available at: http://it-claim.ru/Persons/Semenova_Yana /speech.htm (accessed April 15, 2016).
2. Harris, F.J. (2004), *Multirate Signal Processing for Communication Systems*, Prentice Hall PTR, Upper Saddle River, USA.
3. Vrcelj, B, (2003), *Multirate Signal Processing Concepts in Digital Communications*, California Institute of Technology. Pasadena, California, USA.
4. Maheshwar, R., Bharadia, M., Gupta, M. (2010), "Multirate DSP and its technique to reduce the cost of the analog signal conditioning filters", *International Journal of Computer Applications*, vol. 4, no.10, pp. 27–34.
5. Nebilitsya, A.Yu. (2014), "Optimization method of allocation of minimum segment the speech stream", *Visnik Cherkaskogo universitetu*, no. 18 (311), pp. 59–67.
6. Ogorodnikov, A.N. (2003), "Selection intervals signal analysis for speech recognition", *Vestnik Tomskogo gosudarstvennogo universiteta*, no. 280, pp. 285–304.
7. Azarov, I.S. (2014), "Changing the pitch of the speech signal in real time", *Proceeding of the International scientific and technical conference dedicated to the 50th anniversary of the MRTI-BGUIR*, V.1, pp. 274–275.
8. Vashkevich, M.I. (2014), "Oversampling of the speech signal, consistent with the fundamental frequency", *Proceeding of the International scientific and technical conference dedicated to the 50th anniversary of the MRTI-BGUIR*, Vol. 1, pp. 308–309.
9. Lyons, R.G. (2004), *Tsifrovaya obrabotka signalov: Vtoroe izdanie* [Understanding digital signal processing. 2nd Edition], OOO «Binom-Press», Moscow, Russia.
10. Sarikaya, R., Hansen, J.H.L. (2000), "High resolution speech feature parametrization for monophone-based stressed speech recognition", *IEEE Signal Processing Letters*, vol. 7, iss. 7, pp. 182–185.
11. Ponting, K.M., Peeling, S.M. (1991), "The use of variable frame rate analysis in speech recognition", *Computer Speech & Language*, vol. 5, iss. 2, pp. 169–179.
12. You, H., Zhu, Q., Alwan, A. (2004), "Entropy-based variable frame rate analysis of speech signals and its application to asr", *IEEE International Conference. Acoustics, Speech, and Signal Processing*, vol. 1, pp. 549–552.

Стаття надійшла 28.05.2016