УДК 004.934.2+621.395 **А.Ю.Небилиця**

ОПТИМІЗАЦІЯ МЕТОДУ ВИДІЛЕННЯ МІНІМАЛЬНИХ СЕГМЕНТІВ МОВНОГО ПОТОКУ

Набув подальшого розвитку метод амплітудно-інтервальної локалізації сегментів основного тону. Доведена ефективність використання оцінки амплітуд коливання за різницевою схемою із розширеним базисом в задачах попередньої обробки акустичних даних. Визначено критеріальні оцінки виділення головної пелюстки, базовий алгоритм, спосіб верифікації процесу локалізації.

Ключові слова: *мовний потік, основний тон, розпізнавання мови, амплітудно-інтервальна селекція, метод нулів сигналу, різницева схема.*

Вступ

Взаємодія з технічними пристроями і системами природнім для людини шляхом – обміном мовними повідомленнями – дозволяє досягти оперативності та гнучкості комунікації, усуває потребу постійного візуального контакту та виконання тактильних дій, надає додаткову зручність для людей з вадами зору. Такі переваги зумовили тривалі та інтенсивні пошуки методів автоматизованого розпізнавання і синтезу мови, але задача отримання повноцінного мовного інтерфейсу і на сьогодні не розв'язана [1]. Головною перешкодою залишається низька достовірність машинної ідентифікації мови, яка проявляється в сильній залежності від індивідуальних ознак, інтонації та темпу мовлення, чутливості до шуму. Стримуючим фактором у реалізації мовного інтерфейсу, особливо для вбудованих систем, є складність процесів попередньої обробки звукових даних, що зумовлює значні затрати обчислювальної потужності. Вирішення цих проблем, в межах даної роботи, вбачається в удосконаленні сегментації мовного потоку з врахуванням особливості акустичного сигналу. Найбільш вираженою ознакою мовлення є періодичність пульсації голосових зв'язок, яку прийнято називати основним тоном (ОТ). Важливість визначення ОТ засвідчує значна кількість отриманих методів, які узагальнені в [2]. Однак, всі вони не забезпечують поєднання таких показників як простота, достовірність та точність локалізації.

Виходячи із зазначеного, *метою досліджень* є оптимізація виділення сегментів ОТ за критерієм мінімізації цифрової обробки на базі «легких» машинних інструкцій. *Актуальність роботи* полягає у виявленні нових підходів попередньої обробки звукових даних для розв'язання задач ідентифікації мовних образів. *Об'єктом досліджень* є методи цифрової обробки мовних сигналів. *Предметом дослідження* є методи виділення ділянок основного тону із потоку звукових даних.

Постановка задачі

Оптимізація процесу виділення сегменту ОТ передбачає:

– досягнення мінімізації затрат обчислювальної потужності та збоїв у вигляді виділення хибних та пропуску головних пелюсток сегменту;
– здійснення верифікації процесу;
– забезпечення точності і стабільності методу шляхом адаптації до виду звукових даних та шумів, інваріантності щодо способу отримання вибірки, надійності ініціалізації.

Власне, сама задача розробки методу полягає в отриманні способу сегментації вокалізованих ділянок мовного потоку за ознаками основного тону, встановлення меж

застосування та визначенні виду налаштувань, які забезпечують відповідність зазначеним вимогам.

Теоретичні основи методу

В своїй основі метод базується на частотних характеристиках мовного потоку та параметрах його дискретизації. На даний час, більшістю розробників систем розпізнавання мови вважається доцільно обмежувати частотний спектр мовного сигналу смугою 80..6000 Гц (більш широкою, ніж в телефонії), а частоту дискретизації f_{clk} вибирати в межах 12..16 кГц [3]. Враховуючи, що нижня межа спектру основного тону f_p для чоловічого голосу становить 80 Гц, то мінімальна довжина вибірки сегменту основного тону N не буде перевищувати 200 семплів. Однак, зважаючи на вимогу здійснення швидкого спектрального перетворення, слід прийняти $N=256$. Найбільш поширеною, на сьогодні, є чотирьохформантна модель мови [4; 5]. Згідно цієї моделі умовно виділяють такі частотні діапазони, у Гц:

$$F_1 \in [470; 920],\ F_2 \in [800; 2070],\ F_3 \in [2660; 3500] \text{ та } F_4 \in [3540; 4000].$$

Найбільша щільність енергії припадає на першу форманту, тому вона є більш вираженою в мовному сигналі, що демонструє пелюстка п. 1 на рис.1. Кількість точок вибірки, які припадають на позитивний фронт пелюстки, визначаються як $n = f_{clk} /(4 \cdot F_1)$. Виходячи із встановленої частоти дискретизації $n \in [3; 8]$, такий широкий діапазон не складає перешкод у використанні, оскільки реальна ширина фронту пелюстки завжди може бути уточнена.

Прототипом способу локалізації основного тону вибрано метод Рабінера-Голда [6], в якому визначення меж сегменту проводять шляхом знаходження мінімумів та максимумів сигналу, а також враховують часові інтервали між ними. Даний метод простий у реалізації, але дуже чутливий до шумів, тому його використання можливе лише за умови проведення низькочастотної фільтрації. Згідно мети досліджень, саме усунення процесу цифрової фільтрації є первинною вимогою оптимізації.

Ідея покращення методу виділення ділянок основного тону полягає у заміні амплітудних параметрів коливань на їх оцінку, в якості якої вибрано швидкість наростання сигналу. Допустимість такої заміни ґрунтується на однозначності зв'язку між миттєвим значенням коливань $y(t) = A_m \cdot \sin(\omega \cdot t)$ та швидкістю змін сигналу $dy(t)/dt = \omega \cdot A_m \cdot \cos(\omega \cdot t)$. Такий зв'язок для одномодового випадку дозволяє визначити амплітуду коливань A_m із швидкості наростання сигналу в моменту часу, коли $y(t_i) \approx 0$. За цих умов має місце рівність $dy(t)/dt = \omega \cdot A_m = 2 \cdot \pi \cdot A_m / T_0$, де T_0 – період коливання. У випадку аналізу даних з постійним часом дискретизації $dt = 1/ f_{clk}$ похідну можливо замінити різницевою схемою $(y_i - y_{i-1}) \cdot f_{clk} = 2 \cdot \pi \cdot A_m / T_0$. Звідкіля, привівши нормалізацію часу, отримано:

$$A_m = (y_i - y_{i-1}) \cdot T_0^* /(2 \cdot \pi). \quad (1)$$

Акустичний сигнал мовного потоку полімодовий, більш того, він багатий на високочастотні складові та шум. Враховуючи дані обставини, пропонується розв'язок стосовно виділення сегменту ОТ здійснювати не шляхом співставлення абсолютних значень A_m, а їх оцінкою A_j^*, де j індекс пелюстки коливань в межах ділянки аналізу.

Вираз (1) можливо трактувати наступним чином: чисельник являє собою оцінку амплітуди, а знаменник виконує функцію корекції. Є очевидним, що еквівалентом операції множення при оцінці амплітуди може бути вираз, отриманий шляхом розширення базису $b \le 2 \cdot (n+1)$:

$$A_j^* = y_{q+m} - y_{q-l}, \tag{2}$$

$$\forall q(y_q \ge \bar{y} \wedge y_{q-1} < \bar{y}), \tag{3}$$

де $\bar{y}$ – усереднена амплітуда сигналу (загальний випадок), для центрованого сигналу $\bar{y} = 0$. Змінні m і l складають базис усереднення $b=m+l$, причому $m=n$, а $l=n+1$. Несиметричність складових базису m і l зумовлена умовою визначення позитивного фронту пелюстки коливання (3).

Візуальне представлення запропонованого способу демонструє рис.1 для випадку $n=2$. Множина точок вибірки $\boldsymbol{Y}=\{y_i\}$, де $i \in [0; N]$, на схемі відображена маркерами у формі ромба п.2. Маркерами п.3 позначені точки, які приймають участь в оцінці амплітудних значень. Положення локалізації позитивних фронтів пелюсток, які виділені програмним шляхом, відображені маркерами п.4 та п.5, відповідно для головних та поточних мікросегментів.

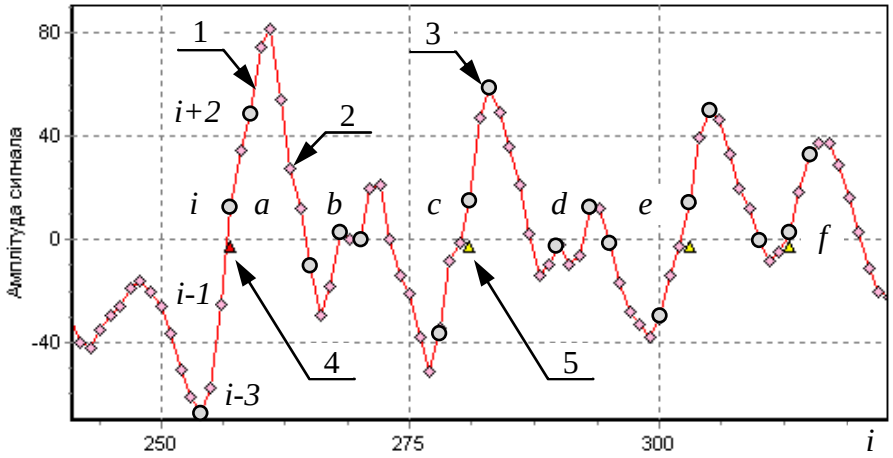

Рис. 1 Схема оцінки амплітуди хвилі

Властивість резистентності запропонованого способу до шумів демонструють фрагменти сигналу *b* і *d*. Програма локалізації основного тону вилучила їх із множини аналізу за амплітудним критерієм. Ця властивість забезпечується шляхом уточнення базису обробки сигналу за математичним очікуванням періоду «перетину нуля»:

$$n = \frac{\mu(T^*)}{4} - 1, \tag{4}$$

де $\mu(T^*)$ – математичне очікування періоду «перетину нуля», маркер '*' вказує, що період визначається в інтервалах дискретизації часу.

Оцінка амплітуди коливань забезпечує виділення мікросегментів і являє собою підготовчий етап обробки. Більш складнішою стадією є локалізація головної пелюстки коливань, яка являє собою процес сепарації множини мікросегментів ***G***, тому спосіб її проведення вартий окремого опису.

Аспекти локалізації головної пелюстки коливань

Виявлення специфіки, визначення способів мінімізації похибки та верифікації результатів процесу виділення сегмента основного тону здійснювалось на базі вибірок звукових даних, отриманих за допомогою наступних засобів: мікрофон Canyon CNR-MIC2; звукова карта Realtek HDA; стандартна утиліта «Звукозапис» Windows XP SP3 в монофонному режимі, 12 кГц, 8 біт. В якості тестових фонетичних груп обрано послідовності за схемою: приголосна + голосна → голосна → голосна + приголосна, наприклад, /па-а-ап/. Вибір такої схеми зумовлено більшою поліморфністю акустичного сигналу. Результати проведених досліджень, які застосовувались у синтезі методу виділення мінімальних сегментів мовного потоку приведені на рис. 2 та рис. 3.

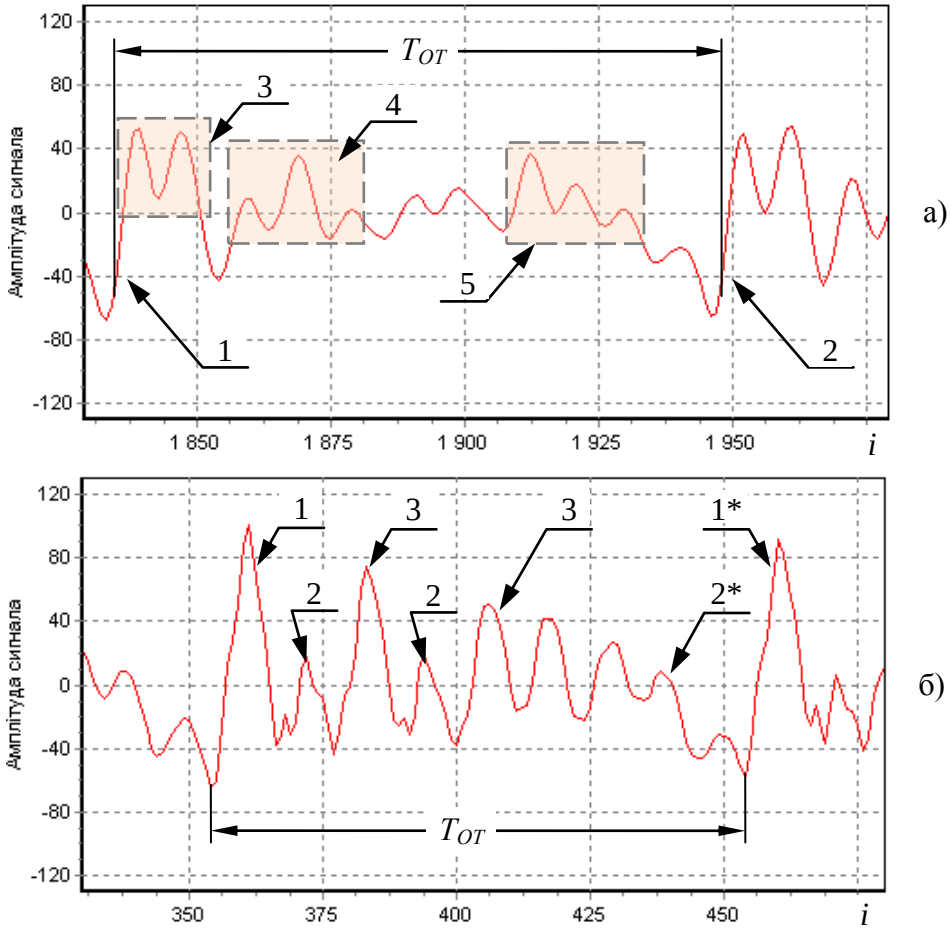

Рис. 2 Фонограми фонем /ра/ та /па/ у фазі формування приголосного звуку

У процесі розв'язування задачі виділення початку сегменту за первинні брались методи амплітудної селекції. Такий вибір обумовлений з огляду простоти реалізації. Однак, практичні досліди з сепарації пелюсток за піковим методом дали вкрай низьку достовірність виділення – $p < 0{,}4$. Основною причиною є адитивна взаємодія формант F_1, F_2 і навіть F_3, яка породжує області, позначені п. 3-5 на рис.2 *а*. Вказані області демонструють ефект суттєвого зменшення амплітуди пелюстки. У деяких випадках, амплітуда настільки зменшується, що стає співставною з амплітудами наступних коливань. Дану проблему загострює необхідність врахування можливості зменшення інтенсивності вимови фонем у фазі її спаду. Із зазначених причин головна пелюстка не

може бути ідентифікована лише шляхом зменшення критеріальної оцінки. Досягнення прийнятного рівня достовірності селекції слід пов'язувати із врахуванням часових інтервалів мікросегментів ***G***.

Проблеми локалізації сегменту ОТ створює поява пелюсток малої амплітуди. Доцільно їх розділити за характером прояву на дві групи. Перші з них гарно виражені та в міру центровані, за формою вони подібні до головних. Їх прикладом може слугувати пелюстки п.2 на рис. 2*б*. Друга група – пелюстки високочастотних мод, відмінна від перших за шириною, і як правило, розміщені асиметрично по відношенні до ізолінії, що демонструє виділення п.3 на рис. 3. Обидві групи ускладнюють процес локалізації ОТ, оскільки такий критерій, як передування головній пелюстці (п. 1* рис. 2*б*) хвилі коливання малої амплітуди (п. 2*), зумовить хибне виділення для групи пелюсток п. 3 і п. 2. Зниження ризиків такого виду можливо досягти методом амплітудної селекції за критерієм:

$$A_j^* < k_{\min} \cdot A_{\max}^*, \tag{5}$$

де $A_{\max}^*$ – поточна максимальна амплітуда, за яку приймається амплітуда попередньої головної пелюстки; $k_{\min}$ – коефіцієнт приведення до мінімального порогового значення. Для відсікання пелюсток високочастотних мод необхідно щоб $k_{\min} \in [0.125; 0.25]$. Для аналізу ділянок початку формування фонеми в якості $A_{\max}^*$ приймається довільна пелюстка, яка більша за попередні. Виключення пелюсток з множини аналізу за критерієм (5) дещо стабілізують показник достовірності локалізації на рівні 0.7, але цього не достатньо для розв'язання задач аналізу мовного потоку. Кращі результати розділення забезпечує критерій симетрії позитивного фронту $y_{q+m} - y_q \approx y_q - y_{q-l}$. Рівень симетрії для головних пелюсток найвищий, для звичайних – менший.

Складним, з точки зору виділення початку сегменту, є випадок заходження мовного сигналу в область насичення, що демонструє рис.3. Інколи, для зазначеного випадку спостерігався ефект зменшення амплітуди першої пелюстки п.1 порівняно із другою п.4, що зумовлено більшою енергетикою від'ємної півхвилі п.5. Ідентифікація головної пелюстки шляхом співставлення тривалості першого T_{F1} та останнього T_{end} періодів коливань поточного фрагменту мови, в більшості випадків, давали позитивний результат, що зумовлено затягуванням від'ємної напівхвилі останнього періоду (див. рис.2). Цей ефект для фонем чистих голосних звуків проявляється менш виражено, що демонструє рис. 3. Внаслідок значної варіабельності тривалості періодів T_{end}, використання його величини у визначенні критеріальної оцінки пошуку головної пелюстки є не доцільним.

Окремо, слід виділити результати досліджень у розв'язуванні задачі локалізації ОТ шляхом використання оператора «Teager's Energy» [7] для випадку, коли мовний сигнал заходить в область насичення. Ці результати продемонстрували, що без проведення глибокої фільтрації звукових даних, обробка за зазначеним методом зумовить появу у вихідному масиві чисельної серії високоенергетичних пульсацій. За таких обставин, процес виділення сегменту ОТ буде мати низьку ефективність.

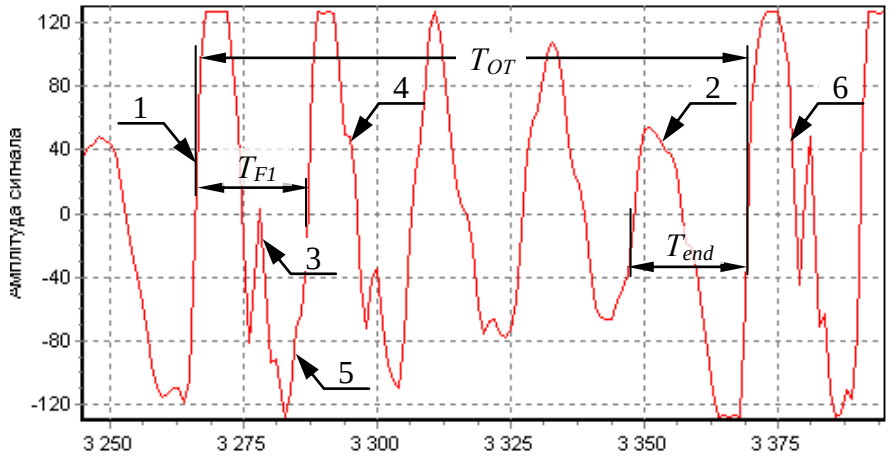

Рис. 3 Фонограма фонеми /а/ для випадку заходження сигналу в область насичення

Викладені в межах даного розділу аспекти засвідчують нетривіальність розв'язування задачі локалізації ОТ. Оптимальний її розв'язок знаходиться в самому означенні. Так, під сегментом основного тону, як складова сигналу, розуміється максимальний, квазіперіодичний фрагмент мовного потоку. Тобто кожний наступний сегмент від попереднього мало відрізняється за амплітудою, формою і часовими інтервалами пелюсток. Виходячи з такого формулювання, критеріальною оцінкою може слугувати допустима величина розбіжності за амплітудними δ_A чи часовими δ_T значеннями:

$$\frac{1}{\nu \cdot \left|A^*_{1,0} - A^*_{0,0}\right|} \cdot \sum_{j=1}^{\nu} \left|A^*_{1,j} - A^*_{0,j}\right| \le \delta_A , \tag{6}$$

$$\frac{1}{\nu} \cdot \sum_{j=0}^{\nu} \left|T^*_{1,j} - T^*_{0,j}\right| \le \delta_T , \tag{7}$$

де ν – кількість пелюсток, які знаходяться між поточною і головною пелюстками попереднього сегменту; $A^*_{1,j}$ і $A^*_{0,j}$ – оцінки амплітуд, а $T^*_{1,j}$ і $T^*_{0,j}$ – періоди пелюсток поточного, позначеного індексом 1, і попереднього сегменту ОТ, – індекс 0. В межах проведених досліджень, значення критеріїв розбіжності δ_A і δ_T приймались на рівні 5% від максимальної амплітуди та періоду ОТ, відповідно.

У випадку, коли кількість пелюсток в сегментах різна, сепарацію пелюсток за критерієм (6) проводять наступним чином. Кількість пелюсток вибирають за тим сегментом, в якому їх більше. Порівняння амплітуд для відсутньої пелюстки здійснюють за оцінкою, яка відповідає, з врахуванням зміщення, положенню перетину «нуля» поточної пелюстки базового сегменту.

Реалізація методу

З точки зору характеру отримання даних ***Y*** слід виділити два підходи: буферизації та поточний. Перший буде мати місце у процесі передачі масиву від пристрою, який перетворює акустичний сигнал в цифровий вигляд (звукова карта), методом прямого доступу до пам'яті. Другий підхід можливо реалізувати шляхом використання

мікроконтролера, в якому обробку кожного семпла можливо здійснити у процедурі обробки переривання за подією завершення перетворення АЦП. Розробка алгоритму виділення мінімальних сегментів мовного потоку здійснювалась саме для поточного виду отримання даних. Незважаючи на те, що буферизація все одно буде мати місце, такий вибір зумовлений вимогами уніфікації процесу, мінімізації обсягу пам'яті та обчислювальних витрат.

Локалізацію ОТ мовного потоку доцільно здійснювати в наступній послідовності:

1. Поточний семпл з АЦП переноситься в буфер даних сегменту. За значенням отриманого семпла перевіряється умова переходу сигналу через нуль (3). Якщо умова не виконується, то здійснюється вихід із процедури локалізації, в протилежному випадку – перехід до пункту 2.
2. Пропускаються m семплів та знаходиться оцінка амплітуди для поточної пелюстки.
3. Перевіряється подібність поточної пелюстки до головної попереднього сегменту за критеріальною оцінкою: $A^*_{1,j} \geq 0.75 \cdot A^*_{0,0}$. Якщо умова виконується, то виконується перехід до пункту 4.
4. В протилежному випадку проводиться перевірка на потужність пелюстки за критерієм (5), що дозволяє класифікувати поточну пелюстку як слабку чи як звичайну. Фіксуються параметри пелюстки та здійснюється вихід із процедури локалізації.
5. Ідентифікація поточної пелюстки як головної здійснюється за критерієм (6). За її результатами пелюстка помічається як звичайна чи головна.
6. Верифікація процесу виділення проводиться за формулою (7). У випадку, коли результати верифікації негативні, сегмент позначається як сумнівно виділений. Він приймає участь в подальшому аналізі, але знижує достовірність результатів.
7. Процес пункту 5 завершує процедуру локалізації ОТ. Змінюється вказівник на буфер даних наступного сегменту.

Застосування в критеріальних оцінках коефіцієнтів ряду 0,125, 0,25 і 0,75 забезпечує заміну використання чисел з фіксованою чи плаваючою точкою, а також операцій множення і ділення, на більш швидкі операції зсуву, віднімання та додавання. Такий підхід дозволяє у вбудованих системах розпізнавання мови застосовувати навіть такі малобюджетні мікроконтролери, як сімейство AVR.

Висновки

Підбиваючи підсумок за результатами досліджень слід виділити наступне:

1. Отриманий метод локалізації сегментів ОТ можливо віднести до групи медіанних фільтрів, якщо взяти до уваги, що ядро фільтра складають одиниці на краях і нулі в межах його тіла.
2. Виділення мінімальних сегментів мовного потоку за оцінкою амплітуди коливань можлива навіть без проведення низькочастотної чи смугової фільтрації.
3. Запізнення ідентифікації основних пелюсток становить один період основного тону.
4. Прийнятний рівень достовірності виділення головних пелюсток в мовному сигналі можливо забезпечити шляхом комбінації амплітудної та інтервальної селекції за умови врахування міри подібності групи суміжних пелюсток. В

рамках досліджень, достовірність локалізації ОТ була меншою за 0,9 лише на трьох перших сегментах ділянки наростання сигналу.

5. Кількість позитивних пелюсток у сегменті основного тону варіюється в межах 2..8.
6. Рівень стиснення звукових даних, який досягається внаслідок параметризації сигналу шляхом оцінки амплітуди коливань, не нижче 24 разів.
7. Програмну реалізації процесу локалізації сегменту ОТ можливо здійснити на базі легких операцій: порівняння, зсуву, додавання і віднімання.
8. Виділення сегментів ОТ у випадку, коли мовний сигнал заходить в область насичення, ускладнено. Зниження ризиків такого виду спотворень сигналу слід здійснювати апаратними засобами, наприклад, шляхом використання звукових компресорів типу SSM2167, при чому час затримки не повинен перевищувати 20 мс.

Зазначені властивості та показники дозволяють стверджувати, що виділення мінімальних сегментів мовного потоку шляхом оцінки амплітуд коливань надає можливість покращити процес сегментації мовного потоку та підвищить ефективність систем розпізнавання мови. Також отримані результати досліджень будуть корисними у сфері цифрової телефонії.

Література

1. Перспективы развития систем распознавания речи /Выдержка из исследования. // Закладка Работа со звуком. Исследования и прогнозы в IT. [Електронний ресурс]. – Режим доступу: http://habrahabr.ru/post/232613/
2. Голубинский А. Н. Расчёт частоты основного тона речевого сигнала на основе полигармонической математической модели // Вестник Воронежского института МВД России. – 2009. – № 1. – С. 81-89
3. Небилиця А.Ю. Часові та спектральні характеристики мовного потоку. / Актуальні проблеми природничих та гуманітарних наук у дослідженнях студентської молоді «Родзинка – 2012» / XIV Всеукраїнська студентська наукова конференція. – Черкаси: Брама-Україна, 2012. – С.422-424.
4. Физиология речи. Восприятие речи человеком. / Чистович Л. А., Венцов А. В., Гранстрем М. П. и др. // серия: «Руководство по физиологии». – Л.: Наука. – 1976. – 388 с.
5. Устойчивость оценок формантных частот / В.Н.Сорокин, А.С.Леонов, И.С.Макаров // Речевые технологии. – 2009. – №1. – С. 3-21.
6. Рабинер Л. Р. Цифровая обработка речевых сигналов / Рабинер Л. Р., Шафер Р. В.: пер. с англ. / Под ред. М. В. Назарова и Ю. Н. Прохорова. – М.: Радио и связь. – 1981. – 496 с.
7. Kaiser J. F. On a simple algorithm to calculate the ‘energy’ of a signal // in Proc. IEEE Int’l. Conf. Acoust., Speech, Signal Proc. – Albuquerque, NM. – April 3-6 1990. – pp. 381-384.

Стаття надійшла _28_._03_._2014___
Прийнято до друку _08_._04_._2014___

Аннотация

А. Ю. Небылица

Оптимизация метода выделения минимальных сегментов потока речи

Приобрел последующее развитие метод амплитудно-интервальной локализации сегментов основного тона. Доказана эффективность использования метода нулей сигнала при выделении минимальных сегментов потока речи в случае замены абсолютных значений амплитуд колебаний на их оценки, которые определены по разностной схеме с расширенным базисом. Описаны виды и специфика сепарации лепестков колебаний сегмента основного тона. В работе определенны: способ нахождения базиса, критериальные оценки выделения главного лепестка, базовый алгоритм и способ верификации результатов анализа. Приведены

рекомендации относительно аппаратной предобработки речевого сигнала и кодовой реализации оптимизированного метода сегментации потока речи.

Ключевые слова: *поток речи, основной тон, распознавание речи, амплитудно-интервальная селекция, метод нулей сигнала, разностная схема.*

Summary

A. Yu. Nebylytsia

Optimization of minimal segments localization of speech stream method

There was another development phase tracked for the amplitude-interval localization of the pitch segments method. The usage efficiency of the zero-crossing method in the allocation of the minimal segments of the speech stream in cases of replacement of the absolute values of the amplitudes with their evaluation, which were defined by the difference scheme with an extended basis, was proved. There were types and specifics of the Petals fluctuations of the pitch segment described. The article contains the following definitions of the method for finding a basis, criterial evaluation for the pitch petal selection, the basic algorithm and the method of verification of the analytical results. There were recommendations provided concerning the hardware preprocessing of the speech signal and coding implementation of the optimized method of the speech stream segmentation.

Key words: *speech stream, pitch-segment, speech recognition, peak-interval selection, zero-crossing method, deferent chart.*