

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
"КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ"**

**ЛАДОШКО ОЛЬГА МИКОЛАЇВНА**



УДК 534.78, 004.934

**ПІДВИЩЕННЯ РОБАСТНОСТІ СИСТЕМ АВТОМАТИЧНОГО  
РОЗПІЗНАВАННЯ МОВЛЕННЯ МЕТОДАМИ ОБРОБКИ СИГНАЛІВ**

Спеціальність: 05.09.08 – прикладна акустика та звукотехніка

**Автореферат**  
дисертації на здобуття наукового ступеня  
кандидата технічних наук

Київ – 2016

**Дисертацією є рукопис.**

Робота виконана в Національному технічному університеті України «Київський політехнічний інститут» Міністерства освіти і науки України на кафедрі акустики та акустoeлектроніки.

**Науковий керівник**

доктор технічних наук, професор  
**Продеус Аркадій Миколайович,**  
Національний технічний університет України  
«Київський політехнічний інститут»,  
професор кафедри акустики та акустoeлектроніки.

**Офіційні опоненти:**

доктор фізико-математичних наук, професор  
**Калюжний Олександр Якович,**  
ТОВ «НВП Омега»,  
провідний науковий співробітник-консультант

доктор технічних наук, професор  
**Розорінов Георгій Миколайович,**  
Державний університет телекомунікацій  
заступник директора навчально-наукового  
інституту захисту інформації,  
завідувач кафедри систем захисту інформації

Захист відбудеться «\_\_»\_\_\_\_\_2016 р. о\_\_\_\_\_ на засіданні спеціалізованої вченої ради Д 26.002.19 в НТУУ «КПІ» за адресою 03056, м. Київ-56, просп. Перемоги, 37, корпус 12, аудиторія 412.

З дисертацією можна ознайомитись у бібліотеці НТУУ «КПІ» за адресою: 03056, м. Київ-56, просп. Перемоги, 37.

Автореферат розісланий «\_\_»\_\_\_\_\_2016 р.

Учений секретар  
спеціалізованої вченої ради



В. Б. Швайченко

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** Стрімкий розвиток сучасних досягнень в області цифрової обробки сигналів сприяє широкому поширенню апаратно-програмних систем автоматичного стенографування, портативних пристроїв, керованих голосом, таких як мобільні телефони та планшети, систем автоматичного розпізнавання мовлення в телефонних та мережевих лініях зв'язку. Нестійка і ненадійна робота систем автоматичного розпізнавання мовлення (АРМ) в цих системах та пристроях в умовах шуму і реверберації призводить до вимушеного обмеження їх використання.

Незважаючи на багаторічну історію і значний внесок у досягнення в області розпізнавання мовлення здійснених такими українськими вченими як Т. К. Вінцюк, М. М. Сажок, Т. В. Людовик, В. В. Пилипенко, О. М. Карпов, Т. В. Єрмоленко, а також рядом зарубіжних вчених Л. Рабінер, С. Янг, Р. Стерн, Д. Еліс, Г. Хемански та багатьма іншими, вирішення завдання автоматичного розпізнавання мовлення в умовах завад залишається актуальним до теперішнього часу.

Для підвищення точності розпізнавання мовлення в умовах шуму і реверберації необхідно проводити належне коригування спотворених сигналів. Вирішення більшості задач корекції мовленнєвих сигналів для підвищення точності розпізнавання систем АРМ базується на методах попередньої обробки сигналів та методах параметричного подання мовленнєвих сигналів в системі АРМ.

Центральне місце серед методів попередньої корекції мовленнєвих сигналів займають спектральні методи ослаблення шумової завади та пізньої реверберації. Найбільш популярними на даний час є такі методи ослаблення шумової та ревербераційної завади як спектральне віднімання, вінерівська фільтрація, а також запропоновані Я. Ефραίмом та Д. Малахом метод мінімізації середньоквадратичної похибки оцінки короткочасного амплітудного спектру (MMSE) та метод мінімізації середньоквадратичної похибки оцінки логарифму короткочасного амплітудного спектру (logMMSE).

Між тим, слід зазначити, що задача оптимізації такого ослаблення в системах АРМ недостатньо розглянута в літературі. Зокрема, не досліджена в повній мірі оцінка ступеня зниження точності розпізнавання мовлення при ослабленні шуму та пізньої реверберації, а також недосить дослідженим є питання узгодженості точності розпізнавання із критеріями якості мовленнєвих сигналів.

Вирішення більшості задач параметричного подання мовленнєвих сигналів в системах АРМ базується на обмеженому використанні класичних моделей слухового сприйняття, як це викладено в працях С. Снефа, О. Гітза, Р. Лайона, що зумовлено бажанням зменшити обчислювальну складність систем АРМ. Крім того, існуючі методи параметричного подання мовного сигналу, такі як мел-кепстральні частотні коефіцієнти (MFCC), розроблені П. Мермелстейн, С. Девіс та перцепційні коефіцієнти лінійного прогнозування (PLP), запропоновані Г. Хемански, наразі є морально застарілими, оскільки конструювалися без огляду на необхідність

забезпечення робастності систем АРМ в складних заводських умовах. Такий стан справ призвів до необхідності розробки більш складних методів параметричного подання мовленнєвих сигналів, одним з яких є метод нормалізованих за потужністю кепстральних коефіцієнтів (PNCC). При обчисленні таких коефіцієнтів намагаються видалити повільно змінювану шумову складову сигналу, а також врахувати часове маскування мовленнєвого сигналу шумом. На жаль, слід зазначити неефективність застосування цього методу в умовах нестаціонарного шуму.

Таким чином, розробка нових, більш дієвих в системах автоматичного розпізнавання мовлення, методів обробки мовленнєвих сигналів є **актуальною** науково-технічною задачею корекції мовленнєвих сигналів, що має важливе прикладне значення.

### **Зв'язок роботи з науковими програмами, планами, темами**

Дисертаційну роботу було виконано на кафедрі акустики та акустoeлектроніки Національного технічного університету України «Київський політехнічний інститут». Дослідження, результати яких викладені в дисертації, отримані при виконанні держбюджетної теми "Розробка нового покоління медичних приладів - ширококугових ультразвукових фізіотерапевтичних випромінювачів із можливістю одночасного іонофорезу" (№ ДР 0114U002485) у вигляді математичного забезпечення корекції сигналів. Отримані в дисертаційній роботі результати впроваджено у практику створення цифрової гідролокаційної системи виявлення та класифікації акустичних сигналів в рамках дослідно-конструкторської роботи «Зірниця-58250» № 0107U000073Т у вигляді математичного забезпечення систем виявлення та класифікації сигналів.

Впровадження результатів дисертаційної роботи підтверджено відповідними актами.

**Мета і задачі дослідження.** Метою роботи є розробка нових методів забезпечення завадостійкості систем АРМ.

Для досягнення поставленої мети необхідним було вирішення наступних задач:

1. Підвищення точності АРМ при ослабленні пізньої реверберації методом попередньої корекції сигналу та при сліпому оцінюванні часу реверберації.
2. Визначення доцільності використання параметризації мовленнєвих сигналів в просторі ознак PNCC.
3. Обґрунтування структури детектора голосової активності, що забезпечує робастність системи АРМ при використанні PNCC ознак.
4. Виявлення можливості підвищення робастності системи АРМ за рахунок введення нової ознаки «траєкторія частоти основного тону» в детекторі голосової активності.

*Об'єктом дослідження* є процеси обробки акустичних мовленнєвих сигналів, спрямовані на підвищення робастності систем АРМ.

*Предметом дослідження* є методи обробки спотворених сигналів в системі АРМ.

**Методи дослідження.** У роботі використовувалися методи частотної корекції з використанням алгоритму  $\log\text{MMSE}$  для обчислення коефіцієнта передачі коригуючого фільтра для ослаблення пізньої реверберації. Для сліпого оцінювання часу реверберації використовувався метод максимальної правдоподібності. Методи спектрального аналізу мовленнєвих сигналів, положення фізіологічної акустики на основі психоакустичного згладжування та стиснення спектральної характеристики мовленнєвого сигналу, методи теорії фільтрації використовувалися для визначення доцільності використання ознак PNCC. Моделювання систем АРМ виконувалося з використанням апарату прихованих марківських моделей для побудови акустичних моделей мовлення та методів статистичного моделювання мовлення. Обґрунтування структури детектора голосової активності виконувалося методами побудови нейронних мереж та цифрової обробки сигналів. Оцінювання нової ознаки «траєкторія частоти основного тону» в детекторі голосової активності виконувалося методами автокореляційної обробки сигналів та методом динамічного програмування для отримання найвірогіднішої траєкторії частоти основного тону. Для експериментальної перевірки розроблених алгоритмів використані реальні та штучні мовленнєві сигнали.

#### **Наукова новизна одержаних результатів:**

1. Вперше вирішено задачу побудови нейромережевого детектора голосової активності системи АРМ із використанням ознак нормалізовані за потужністю кепстральні коефіцієнти PNCC для експлуатації в умовах нестационарних завад.
2. Вперше запропоновано метод підвищення завадостійкості детектора голосової активності за рахунок оцінювання ознаки «траєкторія основного тону», включеної до переліку класифікаційних ознак нейромережевого детектора голосової активності системи автоматичного розпізнавання мовлення.
3. Удосконалено метод навчання нейромережевого детектора голосової активності на основі адаптивної корекції параметрів, що дозволяє прискорити процедуру навчання.
4. Удосконалено метод ослаблення пізньої реверберації, що дозволяє підвищити точність систем автоматичного розпізнавання мовлення навіть в умовах недостатності апріорної інформації про параметри реверберації.

#### **Практичне значення отриманих результатів полягає в наступному:**

1. Запропонований спосіб визначення тонових, шумових та паузних ділянок мовленнєвого сигналу, на якому базується розроблений нейромережевий детектор мовленнєвої активності, дозволяє підвищити точність систем автоматичного розпізнавання мовлення.
2. Запропоноване використання додаткової класифікаційної ознаки «траєкторія частоти основного тону», а також адаптивна корекція вагових коефіцієнтів у розробленому нейромережевому детекторі мовленнєвої активності дозволяють зменшити обчислювальні витрати й досягти бажаної точності визначення тонових, шумових та паузних ділянок мовленнєвого сигналу в умовах дії шумових завад.

3. Удосконалення методу ослаблення пізньої реверберації дозволяє суттєво підвищити точність систем автоматичного розпізнавання мовлення навіть в умовах недостатності апріорних даних про час реверберації та його залежність від частоти.

Отримані й описані в даній роботі результати комп'ютерного моделювання, експериментальних досліджень та розрахунків знайшли відображення в патенті України на корисну модель, що охоплює трансформацію зразків або операції, які спрямовані на підвищення стійкості системи АРМ до дії шумів у каналі або до зміни робочих умов.

Отримані результати також можуть використовуватися у навчальному процесі вищих навчальних закладів України, у т. ч. при підготовці інженерів-акустиків в курсах лекцій кафедри акустики та акустoeлектроніки НТУУ «КПІ», а саме: у курсі лекцій з дисципліни «Пристрої реєстрації та відображення інформації», у розділі «Кодування акустичних сигналів», а також у курсі лекцій з дисципліни «Комп'ютерні акустичні системи», у розділі «Корекція мовленнєвих сигналів».

**Особистий внесок здобувача** відображений публікаціями [10-13, 15]. У наукових працях, виконаних у співавторстві здобувачу належить: в роботі [1] моделювання системи автоматичного розпізнавання мовлення, проведення експериментальних досліджень та аналіз отриманих результатів, в роботах [2–3] – розробка комп'ютерних засобів аналізу та проведення статистичного аналізу отриманих результатів, створення розмітки спонтанного українського мовлення, в роботі [4] – розробка алгоритму для комп'ютерного моделювання траєкторії основного тону, проведення експериментальних досліджень та аналіз отриманих результатів; в роботі [5] здобувач сформулював задачу, дослідив вплив адаптивної корекції вагових коефіцієнтів нейромережі, провів необхідні експериментальні дослідження та проаналізував отримані результати; в роботі [6] особистий внесок здобувача полягає в дослідженні та моделюванні методів ослаблення ревербераційної завади, проведенні експериментальної перевірки правильності теоретичних розрахунків, в роботі [7] – в аналізі отриманих результатів за критерієм точності розпізнавання мовлення.

В роботі [8] здобувач розробив спосіб визначення тонових, шумових та паузних ділянок мовного сигналу, провів моделювання детектора виділення голосової активності на основі запропонованого способу, провів експериментальну перевірку запропонованої корисної моделі. В роботі [9] здобувач провів експериментальні дослідження системи автоматизованого стенографування та виконав статистичний аналіз отриманих результатів, в роботі [14] – здобувач розробив спосіб виділення тональних, шумових і паузних ділянок усного мовлення, в [16] – провів моделювання методів ослаблення ревербераційної завади та перевіряв експериментально правильність теоретичних розрахунків.

**Апробація результатів дисертації.** Основні положення і результати дисертації обговорювалися на 8-ми науково-технічних конференціях та 2-ох школах-семінарах МННЦТiС НАН і МОН України: Міжнародна науково-технічна конференція «Штучний інтелект. Інтелектуальні системи ШІ», АР Крим, Кацивелі,

2010, 2011; Міжнародна науково-технічна конференція Акустичний симпозиум «Консонанс», м. Київ 2011, 2013; Міжнародна науково-технічна конференція «Моделювання і комп'ютерна графіка», м. Донецьк, 2011; III Міжнародна науково-технічна конференція студентів, аспірантів та молодих вчених. "Інформаційні управляючі системи та комп'ютерний моніторинг", м. Донецьк, 2011; Одинадцята всеукраїнська міжнародна конференція «Оброблення сигналів і зображень та розпізнавання образів», м. Київ, 2012; Тридцять четверта міжнародна наукова конференція IEEE «Електроніка та нанотехнології», Київ, 2014. Школа-семінар Української асоціації з оброблення інформації та розпізнавання образів спільно з МННЦТiС НАН і МОН України «Усномовні технології та проблеми створення корпусів», м. Київ, 2010; «Розпізнавання та синтез спонтанного мовлення», м. Київ, 2011.

**Публікації.** За результатами досліджень опубліковано 16 наукових праць, у тому числі 7 статей у наукових фахових виданнях України, з них 1 стаття у виданнях України, які включено до міжнародних наукометричних баз даних, 1 патент на корисну модель, 8 тез доповідей в збірниках матеріалів конференцій.

**Структура та об'єм дисертаційної роботи.** Дисертаційна робота складається із вступу, чотирьох розділів, висновків, списку використаних літературних джерел, додатків. Загальний об'єм становить 185 сторінок, в тому числі 150 сторінок основного тексту. Робота містить 89 рисунків та 38 таблиць, 8 додатків та список використаних джерел з 122 найменувань.

### ОСНОВНИЙ ЗМІСТ РОБОТИ

**Вступ** містить обґрунтування актуальності дисертаційного дослідження, мету і задачі дослідження, новизну отриманих результатів та їх практичне значення.

В **першому розділі** приділяється увага проблемам підвищення робастності (стійкості до дії завад) систем АРМ в умовах впливу шуму та реверберації на систему АРМ (рис. 1). Для цього проаналізовані можливі методи підвищення робастності систем АРМ.

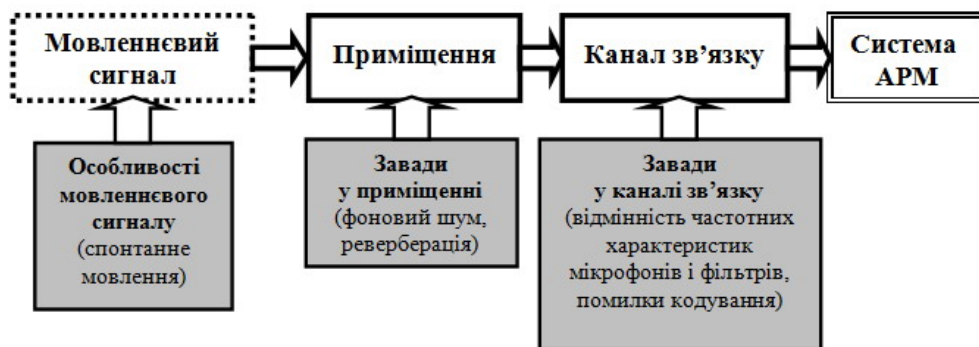


Рис. 1. Вплив завад на робастність системи АРМ

Аналітична модель спотворення мовленнєвого сигналу шумом та реверберацією визначається виразом:

$$y(t) = x(t) \otimes h(t) + n(t), \quad (1)$$

де  $x(t)$  - неспотворений мовленнєвий сигнал;  $h(t)$  - імпульсна характеристика (ІХ) приміщення;  $n(t)$  - випадковий шумовий процес.

В загальному випадку мета системи АРМ полягає в отриманні оптимальної, в сенсі критерію максимальної правдоподібності, послідовності слів  $W$ , розпізнаних у спотвореному мовленнєвому сигналі  $Y$ :

$$\hat{W} = \arg \max_W P_\Gamma(W) \cdot \sum_{\theta} \prod_{t=1}^T p_\Lambda(\bar{x}_t(Y) | \theta_t) \cdot P_\Lambda(\theta_t | \theta_{t-1}), \quad (2)$$

де  $P_\Gamma(W)$  - ймовірнісна модель відповідної мови;  $\bar{x}_t(Y)$  - вектор ознак, отриманий з відкоригованого сигналу для методів попередньої корекції або робастний вектор ознак, отриманий зі спотвореного сигналу  $Y$  для методів параметричного подання сигналу в системах АРМ;  $\{x_t\}$  - сукупність векторів, що описується прихованими Марківськими моделями з послідовністю усіх можливих станів  $\theta_t$  для транскрипції  $W$ ;  $t$  - момент спостереження.  $\Lambda$  та  $\Gamma$  - параметри відповідно акустичної та мовленнєвої моделі.

Точність розпізнавання слів на виході системи АРМ визначається формулою:

$$Acc\% = \frac{N - D - S - I}{N}, \quad (3)$$

де  $N$  - загальна кількість розпізнаваних одиниць мовного сигналу;  $D$  - кількість помилково вилучених одиниць мовного сигналу;  $S$  - кількість замієнених одиниць мовного сигналу;  $I$  - кількість помилково вставлених одиниць мовного сигналу.

Застосування методів попередньої корекції або робастного параметричного подання сигналу у системі АРМ дозволяє виключити необхідність адаптувати параметри акустичної моделі  $\Lambda$  до спотвореного сигналу  $Y$  системи АРМ, не ускладнювати обчислення та не змінювати структуру і параметри вже існуючих систем АРМ, що притаманно усім методам, заснованим на адаптації моделей. Тому в даній дисертації такому напрямку, як корекція мовних сигналів шляхом їх попередньої обробки, вирішено приділити значну увагу.

Детальний аналіз аналітичної моделі спотворень сигналу під дією пізньої реверберації (ПР)  $y(t) = h_i(t) \otimes x(t) + r(t)$  показав, що дія ПР визначається адитивним та нестационарним компонентом  $r(t) = h_l(t) \otimes x(t - T_l)$ , де  $h_i(t)$  та  $h_l(t)$  - відповідно імпульсна характеристика ранніх та пізніх віддзеркалень,  $T_l$  - момент початку ПР.

Найбільшого поширення набув метод частотної корекції, який є найбільш простим, в плані технічної реалізації, та дозволяє отримати оцінку амплітудного спектру відновленого сигналу:

$$\hat{\lambda}_x^{1/2}(l, k) = G(l, k) \lambda_y^{1/2}(l, k). \quad (4)$$

Оцінювання амплітудної частотної характеристики коригувального фільтра  $G(l, k)$  потребує попереднього оцінювання спектру потужності пізньої реверберації  $\lambda_r(l, k) = e^{-2\delta(k)T_l} \cdot \lambda_y(l - N_l, k)$ , де  $N_l$  - зсув фреймів;  $\delta(k) = 2\ln 10 / T_{60}(k)$ ;  $T_{60}(k)$  - час реверберації. В роботах К. Леберта та Е. Хабетса визначено, що оцінювання спектру потужності  $\lambda_y(l, k)$  сигналу  $y(t)$  потребує усереднення у часі:

$$\hat{\lambda}_y(l, k) = \eta_z(k) \hat{\lambda}_y(l - 1, k) + (1 - \eta_z(k)) |Y(l, k)|^2, \quad (5)$$

де  $Y(l, k)$  - дискретне перетворення Фур'є  $l$ -го сегмента сигналу  $y(t)$ ,

$$\eta_z(k) = \begin{cases} \eta_z^d(k), & |Y(l, k)|^2 \leq \hat{\lambda}_y(l-1, k); \\ \eta_z^a(k), & \text{в інших випадках.} \end{cases}$$

- параметр, який регулює ступінь усереднення.

На жаль, викладені в зазначених роботах рекомендації з вибору параметрів  $\eta_z(k)$  та  $T_l$  є недостатньо аргументованими. Як наслідок, це не гарантує досягнення максимальної якості відновленого сигналу  $\bar{x}(t)$  та точності його розпізнавання.

Слід зауважити, що попереднє ослаблення завад супроводжується спотворенням мовленнєвих сигналів, що негативно позначається на точності АРМ. Тому, оскільки системи АРМ використовують інформацію у вигляді спектрального подання мовленнєвого сигналу на виході гребінки фільтрів

$$MF[m, l] = \sum_{k=0}^{(K/2)-1} |X[m, e^{j\omega_k}] H_l[e^{j\omega_k}]|^2, \quad (6)$$

де  $X[\cdot]$  - короткочасний спектр сигналу;  $H_l[\cdot]$  - зважувальна функція;  $\omega_k = 2\pi k / F_s$ ,  $F_s$  - частота дискретизації;  $m$  та  $l$  - індекс фрейму та частотного каналу аналізу, можна припустити, що включення компенсації залишкового шуму, що має місце у відновленому сигналі  $\bar{x}(t)$ , у процедуру (6), дозволить уникнути надмірного спотворення сигналу й таким чином підвищити точність АРМ.

Найбільш поширеним спектральним поданням мовленнєвого сигналу у системах АРМ є мел-частотні кепстральні коефіцієнти та перцепційні коефіцієнти лінійного прогнозування, основні етапи обчислення яких зображені на рис. 2.

Оскільки методи MFCC та PLP дозволяють отримувати лише усереднені значення спектральних характеристик мовленнєвого сигналу на виході гребінки фільтрів (6), дослідниками Х. Занг та Л. Кані була розроблена більш складна «фізіологічна» модель для систем АРМ, що дозволила описати відгук активності слухового нерву на мовлення, масковане шумом. Однак об'єм обчислень, що потребувала їх модель, у 250 разів перевищував аналогічний для моделі MFCC. Ця обставина стала причиною розробки більш ефективного подання мовленнєвого сигналу на основі нормалізованих за потужністю кепстральних коефіцієнтів PNCC.

Компенсація шуму, за методом PNCC, проводиться за оцінками середньої за часом потужності, яку отримують шляхом усереднення впродовж кількох фреймів короткочасної оцінки спектру потужності на виході гребінки фільтрів (6).

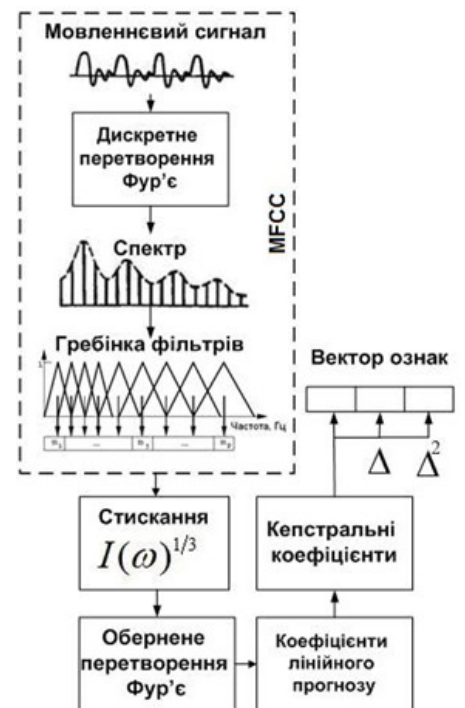


Рис.2. Стадії обчислення MFCC та PLP коефіцієнтів

Оцінюючи змінюваний у часі поріг шуму та віднімаючи його від короточасної оцінки спектру потужності, одержують відкориговану мовленнєву складову сигналу:

$$Q_{\text{вих}}[m, l] = \begin{cases} \lambda_a Q_{\text{вих}}[m-1, l] + (1 - \lambda_a) Q_{\text{вх}}[m, l], \\ \text{if } Q_{\text{вх}}[m, l] \geq Q_{\text{вих}}[m-1, l] \\ \lambda_b Q_{\text{вих}}[m-1, l] + (1 - \lambda_b) Q_{\text{вх}}[m, l], \\ \text{if } Q_{\text{вх}}[m, l] < Q_{\text{вих}}[m-1, l] \end{cases}, \quad (7)$$

де  $Q_{\text{вх}}[m, l]$  - результат ковзаючого усереднення  $MF[m, l]$  впродовж  $M$  фреймів;  $Q_{\text{вих}}[m, l]$  - усереднена за часом потужність, при коефіцієнтах забування  $1 > \lambda_a > \lambda_b > 0$ .

Проблема обчислення коефіцієнтів PNCC полягає у недосконалості процедури виявлення мовленнєвого сигналу на тлі шуму. Так, після одержання відкоригованого мовленнєвого сигналу до нього застосовують ідеальний лінійний напівперіодний випрямляч. Після цього мовленнєвий сигнал повинен оброблюватися у першому каналі, шляхом повторного застосування фільтру (7) для визначення порогового рівня потужності лише для фреймів, де відсутнє мовлення. В той самий час мовленнєвий сигнал повинен оброблюватися у другому каналі шляхом застосування процедури часового маскування

$$Q_m[m, l] = \begin{cases} Q_0[m, l], & Q_0[m, l] \geq \lambda_t \cdot Q_p[m-1, l] \\ \mu_t \cdot Q_p[m-1, l], & Q_0[m, l] < \lambda_t \cdot Q_p[m-1, l] \end{cases}, \quad (8)$$

де  $Q_0[m, l]$  - сигнал на виході напівперіодного випрямляча;  $\lambda_t$  - коефіцієнт забування;  $Q_p[m, l] = \max(\lambda_t \cdot Q_p[m-1, l], Q_0[m, l])$ , для кожного каналу  $l$  та фрейму  $m$ .

Таким чином формування вектору коефіцієнтів PNCC повністю залежить від правильного визначення мовленнєвого сигналу, що в даному випадку визначається на основі порогового енергетичного детектора мовленнєвої активності, обираючи значення 1-го (співвідношення (7)) або 2-го (співвідношення (8)) каналів обробки у такий спосіб:

$$PNCC_n[m, l] = \begin{cases} \max(Q_m[m, l], Q_f[m, l]), & \text{мовлення} \\ Q_f[m, l], & \text{мовлення відсутнє} \end{cases}, \quad (9)$$

де  $n$  - кількість отриманих коефіцієнтів;  $Q_m[m, l]$  - сигнал на виході процедури часового маскування;  $Q_f[m, l]$  - значення нижньої огибаючої, отримане за допомогою (7) на виході напівперіодного випрямляча.

Визначений істотний недолік в обчисленні PNCC коефіцієнтів може призвести до некоректної обробки мовленнєвого сигналу у випадку нестационарного шуму оточення. Вказані вище недоліки обумовили постановку задач, що розв'язуються у даній дисертації.

У другому розділі розглянуто завдання оптимізації оцінювання спектра пізньої реверберації за критеріями точності автоматичного розпізнавання (3) та за показниками якості мовленнєвих сигналів, зокрема такими як перцептуальний показник якості мовлення PESQ (Perceptual Evaluation of Speech Quality). Крім того, перевірено можливість забезпечення високої точності розпізнавання мовлення при сліпому оцінюванні часу реверберації, що має місце в умовах відсутності інформації про імпульсну характеристику приміщення.

Попередня обробка спотворених мовленнєвих сигналів має ту перевагу, що дозволяє зберегти незмінною структуру і параметри системи АРМ (рис. 3).

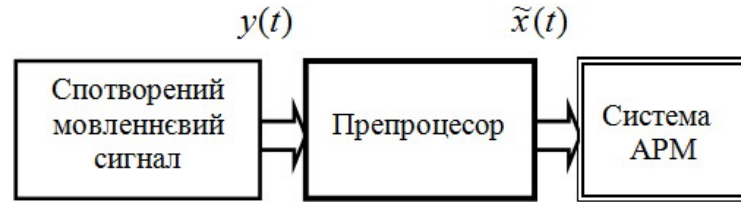


Рис. 3. Попередня обробка реверберованого сигналу

Найбільш простим, в плані технічної реалізації, методом корекції мовленнєвих сигналів є метод частотної корекції (4). Для оцінки коефіцієнта передачі коригуючого фільтра  $G(l, k)$  на сьогоднішній день одним з найбільш ефективних є алгоритм logMMSE (Logarithmic Minimum Mean-Square Estimation), згідно з яким коефіцієнт передачі коригуючого фільтра  $G(l, k)$  обчислюють відповідно до співвідношення:

$$G(l, k) = \frac{\xi(l, k)}{1 + \xi(l, k)} \exp\left(\frac{1}{2} \int_{v(l, k)}^{\infty} \frac{e^{-t}}{t} dt\right), \quad v(l, k) = \frac{\xi(l, k)}{1 + \xi(l, k)} \gamma(l, k),$$

де  $\xi(l, k) = \lambda_x(l, k) / \lambda_n(l, k)$  і  $\gamma(l, k) = \lambda_y(l, k) / \lambda_n(l, k)$  - апіорне і апостеріорне відношення сигнал-завада, відповідно;  $\lambda_n(l, k)$  - спектр потужності  $l$ -го сегмента завади;  $\lambda_x(l, k)$  і  $\lambda_y(l, k)$  - спектри потужності  $l$ -го сегмента сигналів  $x(t)$  та  $y(t)$ , відповідно, на частоті  $f_k = kF_s / N_{fft}$ .

Результати виявлення оптимального, в сенсі показника  $Acc\%$ , значення параметра  $T_l$  в задачі ослаблення пізньої реверберації наведено на рис. 4,а, де наведені графіки залежності  $Acc\%(T_l)$  для різних значень параметра усереднення  $\eta_z^d$ , не залежного від змінної  $k$  (при цьому  $\eta_z^a = 0,5\eta_z^d$ ).

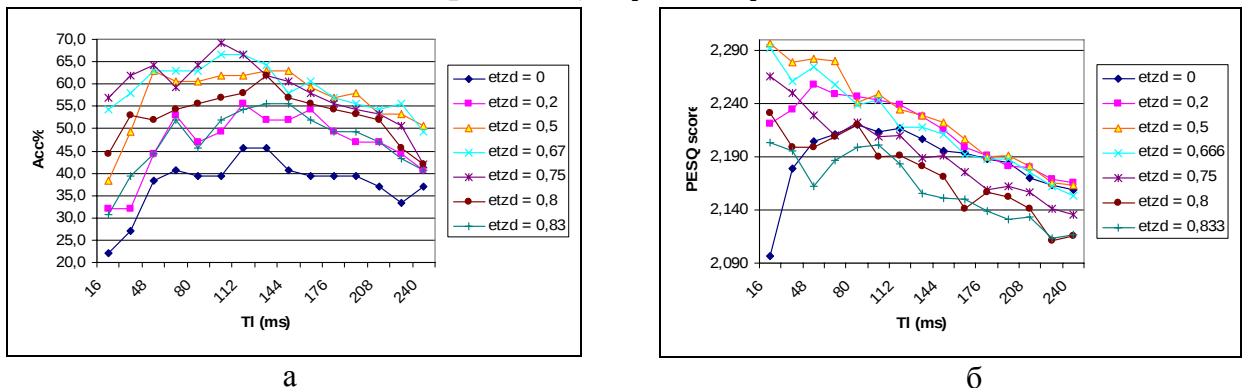


Рис. 4. Показники  $Acc\%$  (а) та PESQ як функції параметрів  $T_l$  та  $\eta_z^d$

Визначено (рис. 4,а), що максимальне значення  $Acc\%(T_l) \approx 67...69\%$  досягається при  $\eta_z^d \approx 0,67...0,75$  для  $T_l \approx 100$  мс. Також визначено (рис. 4,б), що показник  $PESQ \approx 2,3$  за умови  $\eta_z^d \approx 0,5...0,66$ . Максимальна якість мовлення  $PESQ \approx 2,22-2,25$  досягається при наявності екстремумів лише для значень  $\eta_z^d \approx 0...0,2$ , що відповідає  $T_l \approx 50...80$  мс, однак при цьому точність  $Acc\% = 40-45\%$  є незадовільною для практичного використання. Проведені в дисертаційній роботі дослідження показали, що параметр усереднення  $\eta_z(k)$  у співвідношенні (5) не повинен залежати від частоти  $f_k = kF_s / N_{ff}$ , що дозволяє значно спростити обчислення.

Нарешті, в дисертаційній роботі одержано оцінку точності розпізнавання мовлення при ослабленні пізньої реверберації шляхом сліпого оцінювання часу реверберації  $T_{60}$ . Таке оцінювання є можливим, оскільки в паузах мовленнєвого сигналу дія реверберації описується співвідношенням:  $y(n) = \xi(n)a(n)$ ,  $a(n) = e^{-\delta n/F_s}$  де  $\xi(n)$ ,  $n \geq 0$ ;  $n = tF_s$  - дискретний гаусовий білий шум з параметрами  $[0, \sigma]$ .

Представляючи  $y(n)$  як  $N$ -мірний вектор у з багатовимірною щільністю ймовірностей:

$$L(\mathbf{y}; \mathbf{a}, \sigma) = \frac{1}{a(0) \cdots a(N-1)} \left( \frac{1}{2\pi\sigma^2} \right)^{N/2} \exp \left( - \frac{\sum_{n=0}^{N-1} (y(n)/a(n))^2}{2\sigma^2} \right), \quad (10)$$

оптимальні значення параметрів  $a$  та  $\sigma$  знаходять ітераційним методом як такі, що відповідають максимуму функції максимальної правдоподібності (10), а оцінку часу реверберації визначають за співвідношеннями:  $T_{60} = 6,91/\delta$ ,  $\delta = -F_s \ln a$ .

Результати досліджень наведені на рис. 5, показали, що використання оцінки  $T_{60}$ , виміряної сліпим методом (спосіб 3), дозволило на 20-30% підвищити показник  $Acc\%$ , хоча при цьому значення  $Acc\%$  в середньому на 20% поступаються таким при прямому оцінюванні параметра  $T_{60}$  (спосіб 2).

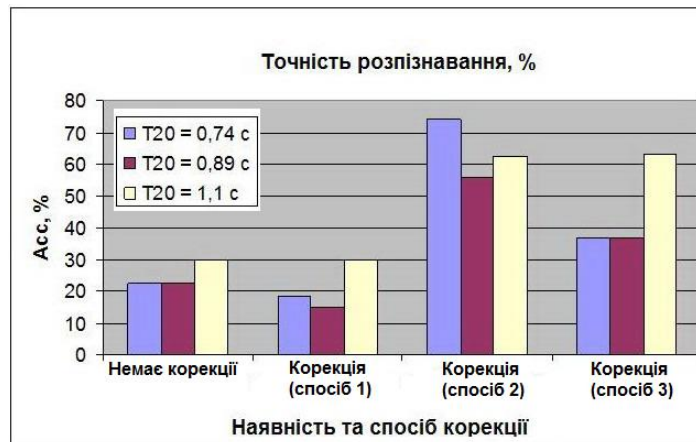


Рис. 5. Точність розпізнавання за відсутності або за наявності корекції

В дисертаційній роботі розглянута можливість подальшого підвищення точності  $Acc\%$  при сліпому оцінюванні  $T_{60}$ , за рахунок наближеного оцінювання

залежності часу реверберації від частоти  $T_{60}(f)$ . Таке оцінювання здійснювалося по сигналам на виходах гребінки з 2-х, 3-х і 7-ми октавних фільтрів (рис. 6).

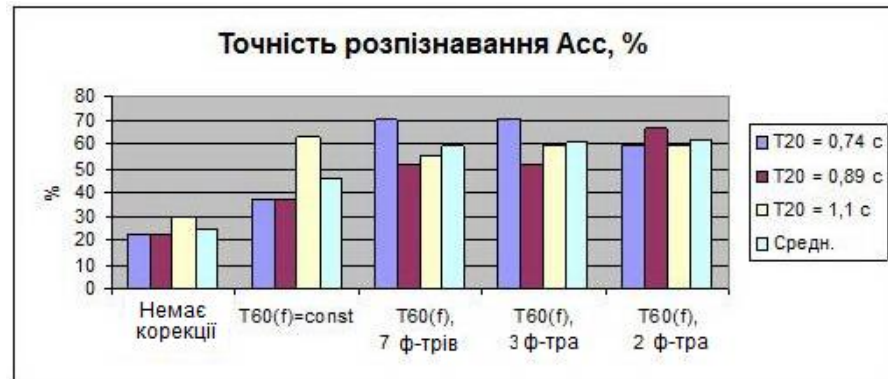


Рис. 6. Точність розпізнавання при врахуванні залежності  $T_{60}(f)$

Визначено (рис. 6), що гребінка всього з 2-ох октавних фільтрів забезпечує прийнятну якість сліпого оцінювання залежності  $T_{60}(f)$ , при цьому можна досягти точності  $A_{cc}\%$ , що є лише на 3-5% нижчою за таку для вимірювання часу реверберації прямим методом.

У третьому розділі наведені результати комп'ютерного моделювання двох сучасних систем АРМ, а саме, НТК та CMU Sphinx. Схему такого моделювання показано на рис. 7. При цьому експериментально перевірено доцільність використання подання мовленнєвих сигналів в просторі ознак PNCC, у порівнянні з традиційним методом параметричного подання мовленнєвого сигналу у просторах ознак MFCC та PLP. Дослідження виконано на реальних сигналах, спотворених двома телефонними каналами зв'язку 1 (NTIMIT) та 2 (STC-TIMIT) типу. В якості критерію оптимальності отриманих результатів була використана точність  $A_{cc}\%$ . Обчислювалися 12 кепстральних коефіцієнтів.

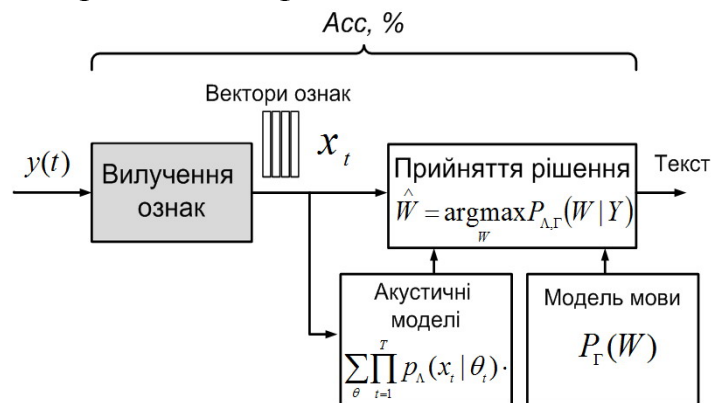


Рис. 7. Схема моделювання системи АРМ

На рис. 8 наведено результати отриманої експериментальним шляхом залежності точності розпізнавання  $A_{cc}\%$  від виду параметричного подання мовленнєвого сигналу (рис. 8,а – для MFCC, рис.8,б - для PLP) та видів спотворених сигналів.

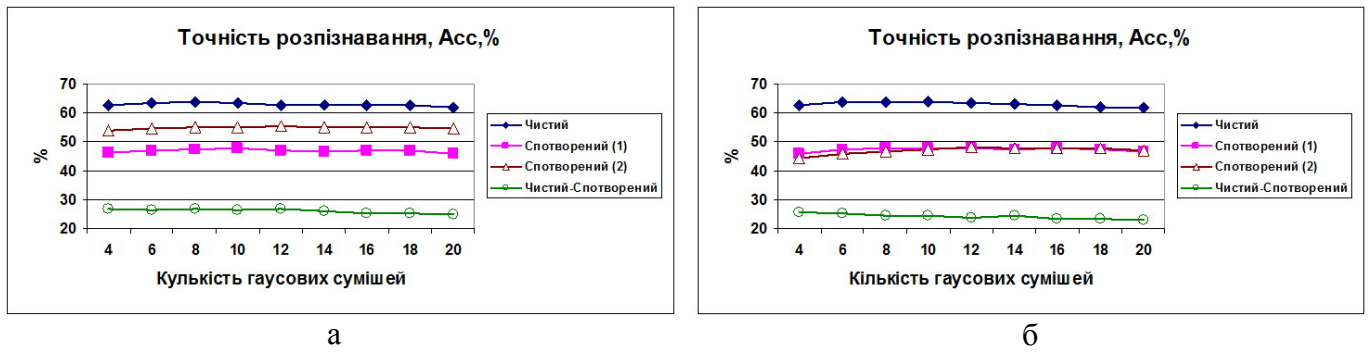


Рис. 8. Показники  $Ass\%$  для різних умов спотворення мовленнєвих сигналів

Як впливає з рис.8, процедура параметризації сигналів PLP дозволила підвищити точність  $Ass\%$  для випадку спотвореного сигналу 2 (STC-TIMIT) в порівнянні з MFCC. Однак у всіх інших випадках суттєвої різниці між MFCC та PLP поданням при розпізнаванні спотвореного телефонними каналами зв'язку не виявлено. При цьому, вкрай неприйнятним виявилася  $Ass\%$  для обох методів в різних умовах навчання та тестування акустичної моделі (чистий-спотворений). У цьому випадку поліпшення при зміні методу параметризації MFCC на PLP склало 2%, при точності MFCC подання  $Ass\% = 24,3$  та  $Ass\% = 26,2$  для PLP, що є вкрай неприйнятним для ефективної роботи систем АРМ. Звідси впливає висновок про доцільність дослідження нового методу подання мовленнєвого сигналу у вигляді PNCC коефіцієнтів в системах АРМ.

Для виявлення доцільності розпізнавання з використанням PNCC подання, в дисертації експериментальним шляхом визначено оптимальні, в сенсі максимуму  $Ass\%$ , значення параметрів процедури фільтрації (7) та часового маскування (8), які відповідно склали  $\lambda_b = 0,25$  и  $0,45$ ,  $\lambda_a = 0,9$ , коефіцієнт забування  $\lambda_t = 0,85$  при  $\mu_t = 0,1$ . Була створена статистична мовна модель  $P_T(W)$  з використанням словника фонетичних транскрипцій з 6070 слів, що складався з 40 фонем. При цьому використовувалося 8 гаусових сумішей, 40 каналів гребінки фільтрів (для отримання сорока PNCC коефіцієнтів).

На рис. 9 показані графіки залежності точності розпізнавання  $Ass\%$  для MFCC та PNCC параметричного подання мовленнєвого сигналу. Очевидно, що параметричне подання PNCC призвело до підвищення точності  $Ass\%$  у всіх випадках рис. 9 на 1,2-1,9%.

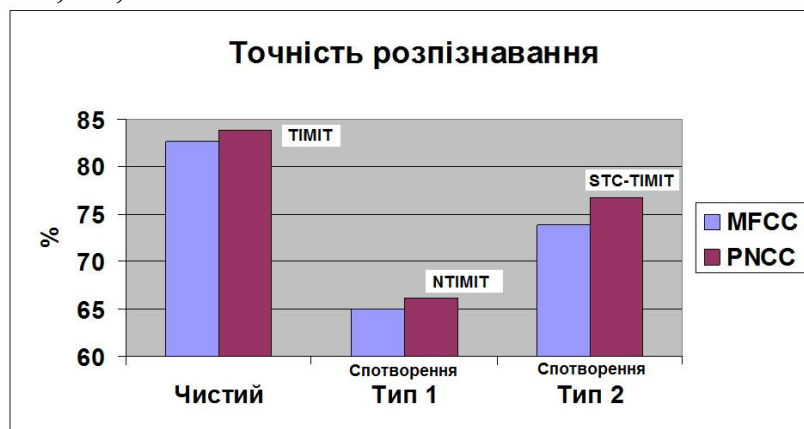


Рис. 9. Точність розпізнавання для MFCC та PNCC подання

За отриманими результатами експериментів зроблено висновок про необхідність удосконалення методу PNCC шляхом заміни процедури роздільної обробки голосової активності на основі енергетичного підходу на більш стійкі щодо дії нестационарних шумів методи.

**Четвертий розділ** присвячений удосконаленню процедури роздільної обробки мовленнєвого сигналу при визначенні голосової активності та при визначенні параметрів шуму методом PNCC. Результати моделювання перевірені експериментальним шляхом на штучних та реальних сигналах. Для цього було розроблено програмну модель (мова програмування C++) нейромережевого детектору голосової активності (рис. 10) системи APM.

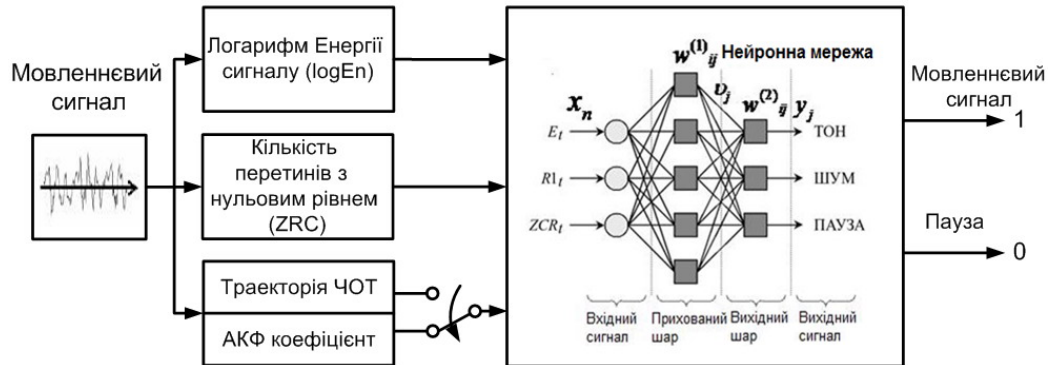


Рис. 10. Нейромережевий детектор голосової активності (VAD)

В такому детекторі, котрий в подальшому позначатимемо аббревіатурою VAD (Voice Activity Detector), використовувалися 3 класифікаційні ознаки: логарифм короточасної енергії, кількість перетинів амплітуди сигналу з нульовим рівнем та автокореляційний коефіцієнт з одиничною затримкою між сусідніми відліками сигналу. Обґрунтування структури детектора голосової активності, що забезпечує робастність системи APM при використанні PNCC ознак, обумовлене основними вимогами: проста система ознак, яка може бути обчислена в реальному часі; ефективний алгоритм прийняття рішення; робота на інтервалах стаціонарності (фреймах) системи APM.

У якості класифікатору в просторі ознак обрано класичну стаціонарну нейронну мережу у вигляді багатошарового перцептрону, котрий в подальшому позначатимемо аббревіатурою MLP (Multilayer Perceptron). Однією з важливих властивостей MLP мережі є те, що складність операцій в ній можна визначити наперед, шляхом зміни структури зв'язків між її шарами у процесі її проектування. Окрім того, запропонована MLP мережа дозволяє розширювати вибір або замінювати класифікаційні ознаки, не змінюючи методи визначення структури зв'язків між її шарами.

У першому підрозділі четвертого розділу запропоновано розширити простір ознак розпізнавання класів «голосний-приголосний-пауза» за рахунок введення ознаки «траєкторія частоти основного тону». Метою такого розширення є підвищення завадостійкості запропонованого детектору.

Траєкторія оцінок частоти основного тону (ЧОТ)  $F_0$  визначається для голосних звуків як неперервна лінія, що поєднує оцінки, отримані для групи фреймів. ЧОТ  $F_0 = 1/T_0$  пов'язана з періодом основного тону (ОТ)  $T_0 = p_{F_0}/F_d$ ,  $F_d$  -

частота дискретизації сигналу,  $p_{F0}$  - період ОТ у відліках сигналу. Оцінки ОТ отримують шляхом пошуку таких періодів ОТ  $p_m$ , де  $m$  - кількість відібраних оцінок, що відповідають  $p_m = \arg \max R(p)$ , де  $R(p)$  - функція нормованої автокореляції (ФНАК) для зсувів у обмеженому діапазоні  $p \in [p_{\min}, p_{\max}]$ .

Традиційно при автокореляційному аналізі шукається максимальне значення автокореляційної функції (АКФ) при ненульовому зміщенні, вважаючи що АКФ має піки на зсувах, кратних періоду ОТ. Однак для мовленнєвих сигналів такий підхід може призвести до помилкового вибору максимумів, що відповідають частотам, кратним  $F0$ . Для врахування можливості вибору помилкових максимумів АКФ пропонується проводити сукупний аналіз отриманих оцінок. На рис. 11 схематично зображено сітку пошуку найбільш ймовірної траєкторії ОТ.

Ймовірність вибору оцінки  $p_m$  в траєкторію ЧОТ пропорційна величині  $q_N(m_k) = R_k(p(m_k))$ . Для зменшення помилкового вибору максимумів ФНАК, що обумовлені дією шумової завади, пропонується обмежувати границі пошуку траєкторії ОТ у суміжних фреймах  $k$  та  $k_{k+1}$  у такий спосіб:  $q_T(m_k, m_{k+1}) = 0$ , якщо  $|p(m_k) - p(m_{k+1})| \leq \alpha \cdot p(m_k)$ , де  $\alpha$  - обмежує можливі відхилення траєкторії ОТ для суміжних фреймів.

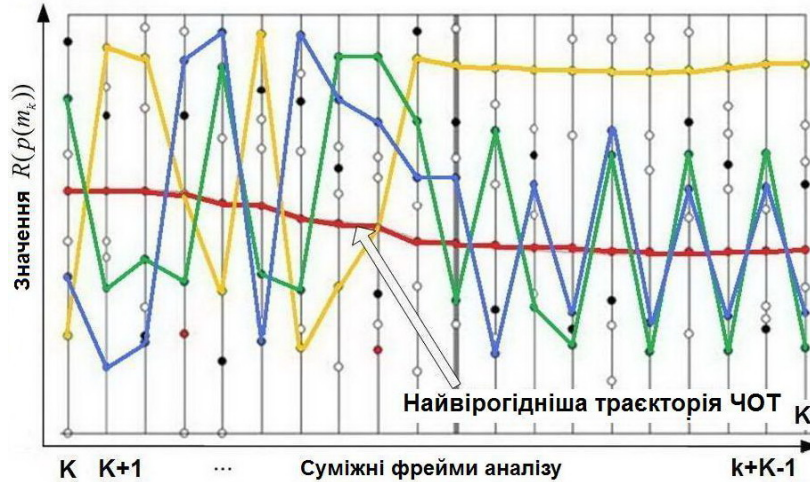


Рис. 11. Сітка пошуку найбільш ймовірної траєкторії ОТ

Поеднання отриманих оцінок у суміжних фреймах у траєкторію ОТ виконується для групи суміжних фреймів тільки за умови  $q_T(m_k, m_{k+1}) = 0$ , підсумовуванням  $q_N(m_k)$  для групи кадрів  $K$  та обираючи ту траєкторію, сума ймовірностей вибору оцінок якої максимальна.

Розроблений в дисертаційній роботі алгоритм для отримання «траєкторії частоти основного тону» був експериментально перевірений на еталонних сигналах ларингограм та на тестових сигналах. Процент грубих помилок (GPE) отриманої траєкторії ОТ, тобто відхилення еталонної та тестової траєкторії, для жіночих голосів склав  $GPE = 0,8\%$ , а для чоловічих  $GPE = 1\%$ . Для оцінки завадостійкості отриманої траєкторії ОТ проведено експерименти для чистих та спотворених білим шумом (з відношенням сигнал-шум  $SNR = 0\text{дБ}$ ) сигналів, при цьому похибка склала всього  $GPE = 1,4\%$ .

У другому підрозділі четвертого розділу представлено результати удосконалення методу навчання нейромережевого детектора голосової активності на основі адаптивної корекції параметрів.

Зазвичай для налаштування зв'язків MLP мережі використовують класичний метод найшвидшого спуску для оптимізації параметрів  $w$ , зі сталим коефіцієнтом швидкості навчання  $\eta$ :  $w(n+1) = w(n) + \Delta w$ , де  $\Delta w(n) = -\eta \cdot \nabla E(w)$ ;  $\nabla E(w)$  – вектор градієнту щодо параметрів  $w$  усіх шарів мережі;  $E(w) = \frac{1}{2} \sum_{k=1}^M (y_k - d_k)^2$  – глобальна функція помилок для вихідного шару MLP мережі, де  $y_k$  – вихідні сигнали та  $d_k$  – очікувані сигнали на виході останнього  $k$ -го шару;  $M$  – кількість нейронів у вихідному шарі.

В даній дисертації запропоновано алгоритм, що дозволяє застосувати метод інкрементної зміни дельт ( $\delta(t)$  – помилки на виході окремого нейрону) Incremental Delta-Bar-Delta (IDBD), розроблений Р. С. Саттоном, для обраного типу стаціонарної нелінійної MLP мережі у такий спосіб:

$$w_i(t+1) = w_i(t) + \alpha_i(t+1) \cdot \delta(t) \cdot x_i(t) \quad (11)$$

де  $\delta(t) = d(t) - y(t)$  – різниця між очікуваним та фактичним виходом;  $x_i(t)$  – вхідні сигнали у нейрон,  $i = \overline{1, n}$ ;  $\alpha_i(t) = e^{\beta_i(t)}$  – коефіцієнт швидкості навчання, де  $\beta_i(t+1) = \beta_i(t) + \theta \cdot \delta(t) \cdot x_i(t) \cdot h_i(t)$  – параметр адаптації,  $\theta$  – коефіцієнт мета-швидкості навчання,  $h_i(t+1) = h_i(t) \cdot [1 - \alpha_i(t+1) \cdot x_i^2(t)]^+ + \alpha_i(t+1) \cdot \delta(t) \cdot x_i(t)$  – параметр пам'яті для кожного вхідного сигналу.

На рис. 12 наведені результати експериментального порівняння розглянутих в дисертації методів коригування параметрів  $w$  для різної архітектури (кількості нейронів у прихованому шарі) MLP мережі, а саме: для чистого (рис. 12,а) та спотвореного телефонним каналом (рис. 12,б) мовлення.

З рис. 12 випливає, що коригування параметрів  $w$  за методом IDBD дозволяє істотно скоротити кількість циклів навчання MLP мережі для більшості запропонованих архітектур з 20, 40, 60, 80 та 100 нейронами у прихованому шарі.

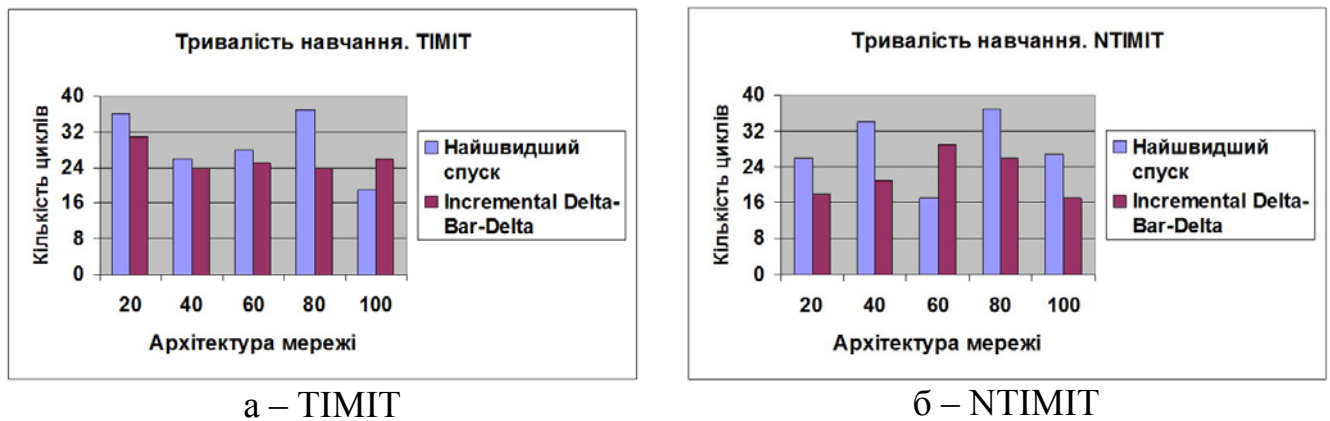


Рис. 12. Порівняння тривалості навчання

Зокрема для обраної моделі з 40 нейронів у прихованому шарі для сигналу, спотвореному телефонним каналом зв'язку (рис. 12,б), кількість циклів вдається зменшити з 34 до 21 циклів, тобто, в 1,6 разів.

У третьому підрозділі **четвертого розділу** працездатність та ефективність запропонованого детектору голосової активності була експериментально підтверджена шляхом тестування на стандартних сигналах, спотворених білим та рожевим шумами із співвідношеннями сигнал-шум від -12 дБ до 20дБ та на реальних сигналах, отриманих з телефонного каналу зв'язку NTIMIT. Результати порівняння запропонованого детектору MLP-IDBD з алгоритмами Д. Їнг, Д. Согн та алгоритмами міжнародних стандартів ETSI AMR та ITU G.729 приведені на рис. 13.

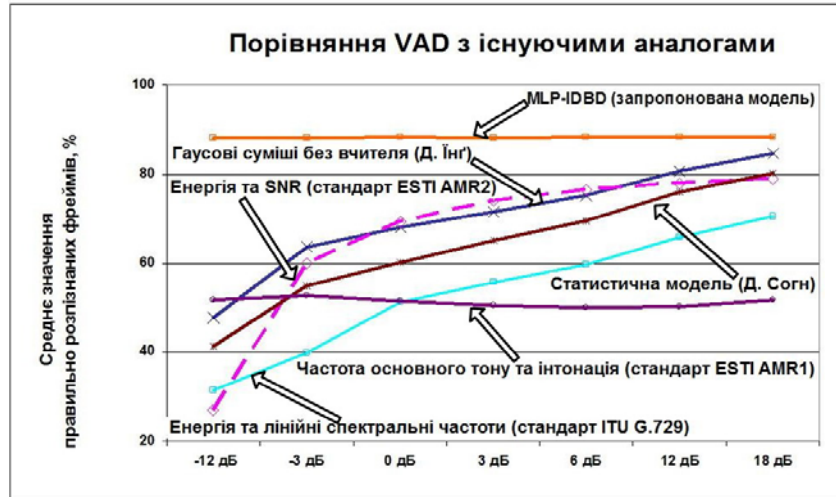


Рис. 13. Порівняння VAD алгоритмів

З рис. 13 видно, що запропонований в даній дисертації детектор MLP-IDBD має перевагу над конкурентними аналогами за середнім критерієм правильно розпізнаних фреймів. Разом із тим, вагомою перевагою запропонованого детектору є можливість його одночасного використання як для підвищення якості роботи PNCC подання в умовах нестаціонарного шуму, так і для зменшення кількості помилкових спрацювань системи АРМ на ділянках, що не містять мовлення. Це може бути досягнуто шляхом налаштування детектору на виключення з обробки фреймів сигналу, що не містять мовленнєвий сигнал.

## ВИСНОВКИ

З метою вирішення актуального науково-технічного завдання розробки нових методів забезпечення завадостійкості систем АРМ методами обробки сигналів одержано наступні результати.

1. Вперше вирішено задачу побудови нейромережевого детектора голосової активності для системи автоматичного розпізнавання мовлення, що є підґрунтям для використання таких ознак як нормалізовані за потужністю кепстральні коефіцієнти PNCC при роботі з нестаціонарними шумами телефонного каналу зв'язку у діапазоні співвідношень сигнал-шум від -12 до +18 дБ.

2. Обґрунтовано вибір структури детектора голосової активності, що забезпечує робастність системи АРМ при використанні PNCC ознак на основі вибору алгоритму корекції параметрів нейронної мережі, пошуку оптимальних, за критерієм похибки класифікації розпізнаних мовленнєвих фреймів, значень його параметрів та вибору раціональної архітектури нейронної мережі, що дозволяє

обмежити складність запропонованої моделі та використовувати запропонований детектор в режимі реального часу роботи системи АРМ.

3. Запропоновано алгоритм адаптивної корекції параметрів стаціонарної нелінійної MLP мережі, що дозволило удосконалити нейромережевий детектор голосової активності і прискорити процедуру навчання, скоротивши число циклів навчання з 34 до 21 циклів для сигналу, спотвореного телефонним каналом зв'язку.

4. Вперше запропоновано метод підвищення завадостійкості детектора голосової активності за рахунок оцінювання ознаки «траєкторія основного тону», що дало можливість підвищити робастність його роботи, знизивши процент грубих помилок до 1,4% при співвідношенні сигнал-шум 0 дБ.

5. Удосконалено метод ослаблення пізньої реверберації шляхом оптимізації значень його параметрів за критерієм точності автоматичного розпізнавання мовлення, при цьому визначено, що оптимальним значенням моменту початку пізньої реверберації є 100 мс, а параметр згладжування періодограм при оцінюванні спектру пізньої реверберації не повинен залежати від частоти.

6. Виявлено, що при сліпому оцінюванні часу реверберації методом максимальної правдоподібності можна досягти точності автоматичного розпізнавання мовлення, яка на 3-5% поступається такій для вимірювання часу реверберації за наявною імпульсною характеристикою приміщення.

## СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Ладошко О. М. Оптимізація алгоритмів системи розпізнавання мови з використанням інструментарію НТК / О. М. Ладошко, А. М. Продеус // Електроніка та зв'язок. – 2007. – № 4 (39). – С. 53-60.

2. Ладошко О. М. Анотація та врахування мовленнєвих збоїв в задачі автоматичного розпізнавання спонтанного українського мовлення / О. М. Ладошко, В. В. Пилипенко // Штучний інтелект. – 2010. – № 3. – С. 238-248.

3. Ладошко О. М. Розмітка спонтанної української мовлення / О. М. Ладошко, А. М. Продеус // Електроніка та зв'язок. Тематичний випуск «Електроніка і нанотехнології». – 2011. – С. 97-103.

4. Ладошко О. М. Оцінка надійності виділення частоти основного тону для акустичного аналізу мови / О. М. Ладошко, А. М. Продеус // Інформатика, кібернетика та обчислювальна техніка. – 2012. – № 15. – С. 162-169.

5. Ладошко О. М. Нейромережевий алгоритм виділення тональних, шумових і паузних ділянок мовлення / О. М. Ладошко, І. Ю. Бондаренко // Електроніка та зв'язок. – 2012. – №6 (71). – С. 19-25.

6. Ладошко О. М. Залежність показників систем ослаблення ревербераційної завади від ступеня спотворення сигналу / О. М. Ладошко, А. М. Продеус // Стандартизація, сертифікація, якість. – 2014. – №3(88). – С. 45-49.

7. Оцінка ефективності захисних конструкцій за критерієм розбірливості мови / В. С. Дідковський, А. М. Продеус, О. М. Ладошко, Н. О. Самійленко // Вісник вищих навчальних закладів. Радіоелектроніка. – 2014, Том 57. – № 2 (620). – С. 55-60. (Включено до наукометричної бази SCOPUS).

8. Пат. на кор. мод. 150553 Україна МПК (2015.01) G10L 17/20. Спосіб визначення тонових, шумових та паузних ділянок мовного сигналу. / Ладощко О. М., Дідковський В. С, Продеус А. М. – № u201511068; заявл. 12. 11. 2015.
9. Ладощко О. М. Анотація та облік мовленнєвих збоїв в задачі автоматичного розпізнавання спонтанної української мови / О. М. Ладощко, В. В. Пилипенко // Міжнародна науково-технічна конференція «Штучний інтелект. Інтелектуальні системи ШІ-2010», 20-24 вересня 2010 р.: тези доп., Т. 1. – Донецьк, 2010. – С. 223-227.
10. Ладощко О. М. Дослідження акустичних особливостей вокалізованих пауз спонтанної мови / О. М. Ладощко // Міжнародна науково-технічна конференція «Штучний інтелект. Інтелектуальні системи ШІ-2011», 19-23 вересня 2011 р.: тези доп., Т. 3. – Донецьк, 2011. – С. 82-86.
11. Ладощко О. М. Дослідження характеристик вокалізованих пауз спонтанної мови / О. М. Ладощко // Акустичний симпозіум «Консонанс-2011», 27-29 вересня 2011 р.: тези доп. – К. – 2011. – С.188-193.
12. Ладощко О. М. Моделювання виділювача частоти основного тону для досліджень спонтанного мовлення / О. М. Ладощко // Міжнародна науково-технічна конференція "Моделювання і комп'ютерна графіка", 5-8 жовтня 2011 р.: тези доп. – Донецьк, 2011. – С. 304-308.
13. Ладощко О. М. Дослідження впливу характеристик телефонного каналу зв'язку на надійність розпізнавання фонем / О. М. Ладощко // III Міжнародна науково-технічна конференція студентів, аспірантів та молодих вчених. Інформаційні управляючі системи та комп'ютерний моніторинг (ІУС і КМ-2012), 16-18 квітня, 2012р.: тези доп. – Донецьк. – 2012. – С. 143-148.
14. Ладощко О. М. Нейромережевий алгоритм виділення тональних, шумових і паузних ділянок усного мовлення /О. М. Ладощко, І. Ю. Бондаренко // Одинадцята всеукраїнська міжнародна конференція «Оброблення сигналів і зображень та розпізнавання образів» (УкрОбраз'2012), 15-19 жовтня 2012 р.: тези доп. – Київ: УАсОІРО. – 2012. – С. 55-58.
15. Ладощко О. М. Дослідження впливу параметризації мовленнєвого сигналу і характеристик каналів зв'язку на надійність розпізнавання фонем / О. М. Ладощко // Акустичний симпозіум «Консонанс-2013», 1-2 жовтня 2013.: тези доп. – К. – 2013. – С. 169-174.
16. Ladoshko O. On existence of optimal boundary value between early reflections and late reverberation /O. Ladoshko, A. Prodeus // Proc. of IEEE 34th International Scientific Conference «Electronic and Nanotechnology», 15-18 April 2014: Proceedings. – Kyiv, 2014. – P. 442-446.

## АНОТАЦІЯ

**Ладошко О. М. Підвищення робастності систем автоматичного розпізнавання мовлення методами обробки сигналів.** – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.09.08 – Прикладна акустика та звукотехніка. – Національний технічний університет України "Київський політехнічний інститут" МОН України. – Київ, 2016.

Дисертацію присвячено вирішенню актуальної задачі підвищення робастності систем автоматичного розпізнавання мовлення шляхом розробки нових методів обробки мовленнєвих сигналів.

Удосконалено метод ослаблення пізньої реверберації, що дозволяє підвищити точність систем автоматичного розпізнавання мовлення навіть в умовах недостатності апіорної інформації про параметри реверберації.

Проведена експериментальна перевірка доцільності використання подання мовленнєвих сигналів в просторі ознак PNCC разом з використанням детектора голосової активності, що дозволяє забезпечити робастність системи автоматичного розпізнавання мовлення при використанні PNCC ознак в умовах нестаціонарного шуму.

За отриманими результатами зроблено висновок про необхідність вдосконалення методу PNCC шляхом заміни процедури роздільної обробки голосової активності на основі енергетичного підходу на більш стійкі щодо дії нестаціонарних шумів методи.

Розроблено нейромережевий детектор голосової активності системи автоматичного розпізнавання мовлення, що дало можливість використовувати такі ознаки як нормалізовані за потужністю кепстральні коефіцієнти при роботі з нестаціонарними шумами.

Розширено перелік ознак запропонованого нейромережевого детектору голосової активності за рахунок введення ознаки «траєкторія основного тону», що дозволило підвищити завадостійкість його роботи.

Удосконалено метод навчання нейромережевого детектора голосової активності. Для цього запропоновано алгоритм адаптивної корекції параметрів стаціонарної нелінійної MLP мережі, що дозволило прискорити процедуру навчання такого детектора.

Працездатність та ефективність запропонованого детектору голосової активності була експериментально підтверджена шляхом тестування на стандартних сигналах, спотворених білим та рожевим шумами та на реальних сигналах, отриманих з телефонного каналу зв'язку NTIMIT.

Результати порівняння запропонованого детектору MLP-IDBD з алгоритмами Д. Їнг, Д. Соґн та алгоритмами міжнародних стандартів ETSI AMR та ITU G.729 показали, що запропонований в даній дисертації детектор MLP-IDBD має перевагу над конкурентними аналогами за критерієм проценту правильно розпізнаних фреймів.

Ключові слова: детектор голосової активності, корекція мовних сигналів, пізня реверберація, робастність систем автоматичного розпізнавання мовлення, частота основного тону, штучні нейронні мережі.

## АННОТАЦИЯ

**Ладошко О. Н. Повышение робастности систем автоматического распознавания речи методами обработки сигналов.** – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.09.08 – Прикладная акустика и звукотехника. – Национальный технический университет Украины «Киевский политехнический институт». – Киев, 2016.

Диссертация посвящена решению актуальной задачи повышения робастности систем автоматического распознавания речи путем использования новых и усовершенствованных методов обработки речевых сигналов.

Рассмотрены и проанализированы современные методы повышения робастности систем АРМ.

Указано, что использование методов предварительной коррекции или робастного параметрического представления сигнала в системе АРМ позволяет исключить необходимость адаптировать параметры системы АРМ к искаженному сигналу и, таким образом, избежать усложнений, обусловленных необходимостью изменять структуру и параметры уже существующих систем АРМ, что характерно для методов, основанных на адаптации моделей.

Поэтому в данной диссертации таким направлениям, как коррекция речевых сигналов путем их предварительной обработки и робастное параметрическое представление сигнала в системе АРМ, уделено первостепенное внимание.

Поскольку, в отличие от вопроса подавления шумовой помехи, задача подавления поздней реверберации исследована в значительно меньшей степени, главное внимание в диссертации уделено оптимизации структуры и параметров алгоритма подавления поздней реверберации.

Усовершенствован метод подавления поздней реверберации, что позволило повысить точность систем автоматического распознавания речи путем оптимизации структуры и параметров оценки спектра поздней реверберации даже в условиях недостаточности априорной информации о параметрах реверберации.

Проведена экспериментальная проверка целесообразности использования представления речевых сигналов в пространстве признаков PNCC вместе с использованием детектора голосовой активности, что обеспечивает робастность системы автоматического распознавания речи при использовании PNCC признаков в условиях нестационарного шума.

Полученные результаты позволили сделать вывод о необходимости усовершенствования метода PNCC путём замены процедуры отдельной обработки голосовой активности на основе энергетического подхода на более устойчивые к воздействию нестационарных шумов методы.

Для усовершенствования процедуры отдельной обработки речевого сигнала в системе автоматического распознавания речи разработана программная модель нейросетевого детектора голосовой активности с использованием трех классификационных признаков в виде логарифма кратковременной энергии, количества нуль-пересечений амплитуды сигнала и одиночного значения автокорреляционной функции.

Обоснован выбор структуры детектора голосовой активности, который обеспечивает робастность системы автоматического распознавания речи при использовании PNCC признаков. Экспериментально обоснован выбор алгоритма коррекции параметров нейронной сети, определение его оптимальных значений и поиск оптимальной архитектуры нейронной сети, что позволило ограничить сложность разработанной модели и использовать предложенный детектор в режиме реального времени работы системы АРМ.

Предложено расширить пространство признаков, используемых в детекторе голосовой активности, для распознавания классов «гласный-согласный-пауза» за счет введения признака «траектория частоты основного тона», что позволило повысить помехоустойчивость предложенного детектора. Разработанный алгоритм формирования «траектории частоты основного тона» экспериментально проверен на эталонных сигналах. При этом для искаженных белым шумом, с соотношением сигнал-шум 0 дБ, сигналов процент грубых ошибок составил всего 1.4%.

Усовершенствован метод обучения нейросетевого детектора голосовой активности. Для этого предложен алгоритм адаптивной коррекции параметров стационарной нелинейной MLP сети, что позволило ускорить процедуру обучения такого детектора.

Работоспособность и эффективность предложенного детектора голосовой активности была экспериментально подтверждена путем тестирования на стандартных сигналах, искаженных белым и розовым шумами, а также на реальных сигналах речевого корпуса NTIMIT для телефонных каналов связи. Результаты сравнения предложенного детектора MLP-IDBD с алгоритмами Д. Инг, Д. Соги и алгоритмами международных стандартов ETSI AMR и ITU G.729 показали, что предложенный в данной диссертации детектор MLP-IDBD имеет преимущество перед конкурентными аналогами по критерию процента правильно распознанных фреймов.

Ключевые слова: детектор голосовой активности, искусственные нейронные сети, коррекция речевых сигналов, поздняя реверберация, робастность систем автоматического распознавания речи, частота основного тона.

## ABSTRACT

### **Ladoshko O. Robustness enhancement of automatic speech recognition systems by signal processing techniques. – Manuscript.**

The thesis for a candidate of technical sciences degree in the specialty 05.09.08 – Applied Acoustics and Audio Engineering. – National Technical University of Ukraine "Kyiv Polytechnic Institute" Ministry of Education and Science of Ukraine. – Kyiv, 2016.

The goal of thesis: development of speech enhancement and robust feature extraction methods for robust automatic speech recognition.

The analysis of modern methods to improve the robustness of automatic speech recognition system is presented. It is indicated that the use of pre-correction enhancement techniques or robust feature-processing methods in automatic speech recognition system allows excluding the need to adapt the parameters of automatic speech recognition system to the distorted signal and thus avoiding complications arising from the need to change the structure and parameters of the existing automatic speech recognition systems.

Therefore, in this thesis such directions as the correction (enhancement) of speech signals by their preprocessing and robust parametric signal representation in the automatic speech recognition system, it decided to devote considerable attention.

Modification of existing logMMSE methods have proposed. It is used instead of noise spectrum estimator for enhancement of speech distorted by reverberation.

Neural network based voice activity detector for automatic speech recognition system have proposed. It is enabling the use of robust features power normalized cepstral coefficients with non-stationary noise.

It had been proposed to include as a classification feature trajectory of the pitch. For this purpose, it had been proposed the use of pitch tracking algorithm in noisy speech.

Adaptive correction parameters algorithm of neural network based voice activity detector had been proposed to accelerate the learning process. Systematic evaluations shows that the proposed neural network based voice activity detector are robust to different noise conditions. The proposed approach also outperforms other state-of-the-art voice activity detection algorithms.

Keywords: late reverberation suppression, neural network, pitch tracking, robust speech recognition, speech enhancement, voice activity detection.