

УДК 534.781:004.85

Г. О. Добрушкін, асп.;

В. Я. Данилов, д-р техн. наук, проф.

## ОСНОВНІ ПІДХОДИ ДО РОЗПІЗНАВАННЯ МОВЛЕННЕВОЇ ІНФОРМАЦІЇ (ЧАСТИНА 2)

*Проаналізовано різноманітні підходи до обробки та розпізнавання мовленнєвої інформації. Порівняно перетворення Фур'є та вейвлет-перетворення, як методи параметризації та первинної обробки сигналу, а також штучні нейронні мережі і штучні імунні системи як методи розпізнавання обробленого сигналу. Розглянуто дикторозалежну модель розпізнавання мовленнєвої інформації на основі штучних імунних систем. Обговорено перспективи створення адекватних моделей штучних імунних систем для розпізнавання мовленнєвої інформації, як найпрогресивнішої і перспективнішої з відомих на сьогодні методик.*

### Вступ

В першій частині статті було розглянуто і класифіковано популярні підходи до розпізнавання мовленнєвої інформації, означено основну мету та задачі статті. Були обговорені етапи первинної обробки мовленнєвого сигналу, які мають бути обов'язково реалізовані в якісній системі розпізнавання мовленнєвої інформації. Так, якісна первинна обробка має включати в себе: видалення шуму, сегментацію сигналу, виокремлення вокалізованих ділянок, вимір частоти основного тону та параметризацію. Було розглянуто наявні на сьогодні методи розв'язання задачі розпізнавання мовленнєвої інформації — кореляційний аналіз, фонемно-орієнтовний метод, експертні системи, приховані Марківські та змішані Гаусівські моделі.

*Метою статті*, є систематизація наявних на сьогодні підходів до первинної обробки мовленнєвого сигналу, розв'язання задачі розпізнавання мовленнєвої інформації та подання алгоритмів, що дозволяють підвищити швидкість та надійність розпізнавання мовленнєвих сигналів.

В цій частині статті наведено порівняння вейвлет-перетворення та перетворення Фур'є, з огляду на їх можливість практичної реалізації етапів якісної первинної обробки і параметризації мовленнєвого сигналу. Розглянуто новітні моделі штучних нейронних мереж, що використовуються для розв'язання задач ідентифікації диктора, розпізнавання ізольованих слів із словника і моделювання цифрових фільтрів. Описано новий апарат машинного навчання — штучні імунні системи. Наведено приклад застосування однієї з моделей штучних імунних систем: моделі негативного відбору, для розв'язання задачі ідентифікації диктору.

### Порівняння вейвлет-перетворення та перетворення Фур'є

Більшість мовленнєвих сигналів зручно аналізувати, розкладаючи їх на синусоїди (гармоніки) і цьому є декілька причин. Наприклад, саме так працює людське вухо — воно розкладає звук на окремі коливання різних частот. Крім того, синусоїди є «власними функціями» лінійних систем, тому що під час проходження через лінійні системи, вони не змінюють свою форму, а змінюють лише фазу і амплітуду.

Інтегральне перетворення і ряди Фур'є є основою гармонійного аналізу. Всі властивості і формули перетворення Фур'є отримуються за допомогою базису — комплексної синусоїдальної хвилі  $e^{it} - \cos t + i \sin t$  [1, 2].

Перетворення Фур'є є дуже зручним для обробки мовленнєвих сигналів, але разом з цим має ряд недоліків: початковий сигнал замінюється на періодичний, з періодом, рівним тривалості досліджуваного зразка; перетворення Фур'є погано працює у разі зміни параметрів процесу з часом (нестационарності), оскільки дає усереднені коефіцієнти для всього досліджуваного зразка. Наприклад, перетворення Фур'є не відрізняє сигнал, що є сумою двох синусоїд з різними частотами, від сигналу, що складається з тих же синусоїд, що включаються послідовно одна за одною [3]. Оскільки частота сигналу обернено пропорційна його тривалості, то для отримання високочастотної інформації з високою точністю важливо здобувати її із відносно малих часових інтервалів, а не зі всього сигналу; і навпаки, низькочастотну спектральну інформацію необхідно здобувати із відно-

сно широких часових інтервалів сигналу. Частина описаних труднощів знімається при використанні віконного перетворення Фур'є, однак нескінченно осцилююча базисна функція не дозволяє отримувати по-справжньому локалізовану інформацію.

Вейвлети — це сімейство математичних функцій, що дозволяють аналізувати різноманітні частотні компоненти даних. Це сімейство локальних в часі і по частоті функцій, в якому всі функції отримуються з однієї за допомогою зсуву і масштабування по осі часу (вони «йдуть один за одним»).

Термін «вейвлет» (від англ. *wavelet* — маленька хвиля) з'явився порівняно нещодавно — в 1982 році у зв'язку з аналізом властивостей сейсмічних і акустичних сигналів. В даний час сімейство аналізаторів, названих вейвлетами, починає широко застосовуватися у випадках, коли результат аналізу якогось сигналу повинен містити не тільки просте перерахування його характерних частот (масштабів), але і відомості про певні локальні координати, у яких ці частоти себе проявляють. Сюди відносяться завдання розпізнавання образів, аналіз зображень самої різної природи, вивчення властивостей турбулентних полів, стиснення великих обсягів даних і багато іншого [1, 4]. Таким чином, основним полем застосування вейвлет-аналізу є аналіз і обробка нестационарних (у часі) і неоднорідних (у просторі) сигналів різних типів.

Елементом базису вейвлет-перетворення є добре локалізована функція, що швидко наближається до нуля поза невеликим інтервалом, отже, вейвлет-перетворення володіє рухомим частотно-часовим вікном, вузьким на малих масштабах і широким на великих [1]. Тоді як базисні функції віконного перетворення Фур'є мають одну й ту саму роздільну здатність за часом і частотою, базисні функції вейвлет-перетворення мають роздільну здатність, що зменшується з масштабом а за часом і збільшується з масштабом а за частотою, що дає великі переваги під час аналізу сигналів, оскільки швидкі варіації сигналів (високочастотні характеристики) добре локалізовані, а для виявлення характеристик, що змінюються повільно, достатньо хорошої низькочастотної роздільної здатності. В [1, 4] показано, що вейвлет-перетворення, яке має рухоме частотно-часове вікно, однаково добре виявляє і низькочастотні, і високочастотні характеристики сигналів.

Розклад по вейвлетах дозволяє визначити положення особливостей функції, спостерігаючи за тими місцями, де вейвлет-коефіцієнти набувають великих значень, тоді як перетворення Фур'є не дасть ніякої інформації, скажімо, про те місце, де частота сигналу змінилася. Проте вейвлет-аналіз не завжди може замінити Фур'є-аналіз, який часто використовується для побудови ортонормованих вейвлет-базисів.

Клас вейвлет-перетворень зазвичай поділяють на дискретні вейвлет-перетворення (ДВП) і неперервні вейвлет-перетворення (НВП). ДВП використовується для кодування сигналів, тоді як НВП — для їх аналізу. В результаті, ДВП широко застосовується в інженерній справі і комп'ютерних науках, а НВП в наукових дослідженнях. Всі вейвлет-перетворення можуть розглядатися як різновид частотно-часового представлення, отже відносяться до предмету гармонічного аналізу. Дискретне вейвлет-перетворення може розглядатися як різновид фільтра скінченного імпульсного відгуку. Неперервне вейвлет-перетворення підкоряється принципу невизначеності Гейзенберга і, відповідно, базис дискретного вейвлету також може розглядатися в контексті інших форм принципу невизначеності.

Вейвлет-аналіз зберігає високу роздільну здатність на різних масштабах, тобто здатен виявити внутрішню структуру істотно неоднорідного об'єкта і вивчити його локальні властивості. Крім того, можливість вибору найпридатнішого до конкретної задачі вейвлетного базису робить вейвлет-перетворення дуже гнучким «апаратом». Оскільки мовленнєвий сигнал характеризується нелінійними флуктуаціями різних масштабів, то вейвлет-перетворення є дуже ефективним для його аналізу. На відміну від традиційного Фур'є-аналізу, вейвлет-перетворення дозволяє добре описати локальні особливості досліджуваної функції, зокрема швидкі зміни спектрального складу. Слід зазначити, що багато теорем вейвлет-аналізу доводяться за допомогою розкладання Фур'є, таким чином, ці два типи аналізу швидше доповнюють один одного, ніж змагаються в областях їх застосування [4]. До недоліків вейвлет-аналізу можна віднести той факт, що наперед неможливо визначити, на якому масштабі потрібно шукати інформацію, що нас цікавить.

### **Первинна обробка мовленнєвого сигналу за допомогою вейвлет-перетворення**

Як відомо, мовний сигнал складається з квазістационарних ділянок, що відповідають тональним, сонарним і шиплячим фонемам, між якими є ділянки з порівняно швидкими змінами спектральних характеристик сигналу (міжфонемні переходи, вибухові і глухі фонемні паузи між слова-

ми) [2, 5]. В межах стаціонарних ділянок значну роль для аналізу мовного сигналу грають спектральні особливості сигналу, що визначаються передавальною характеристикою мовного тракту, яка змінюється в процесі артикуляції. Можна сказати, що мовний сигнал характеризується нелінійними флуктуаціями різних масштабів. Тому для аналізу мовного сигналу вейвлетний кратномасштабний аналіз Делпратта і вейвлет-перетворення є ефективними.

Задля сегментації мовного сигналу необхідно розв'язати задачу локалізації міжфонемних переходів. Для фонем, що відповідають відносно протяжним квазістаціонарним ділянкам мовного сигналу вейвлет-перетворення може вирішити цю проблему [6]. На міжфонемних переходах сигнал зазнає значної зміни відразу на багатьох масштабах [2, 5], що приводить до зростання вейвлет-коефіцієнтів на багатьох рівнях деталізації, тоді як на стаціонарних ділянках фонем вейвлет-коефіцієнти виявляються згрупованими поблизу певних масштабів. Таким чином, пошук міжфонемних меж може бути зведений до пошуку моментів збільшення вейвлет-коефіцієнтів на значній кількості рівнів масштабування. У серії експериментів ці припущення підтверджувалися для випадку, коли спектральні характеристики суміжних фонем досить добре відрізняються (наприклад, поєднання «Л» — «А», «І» — «О» і т. д.). Якщо ж форма мовного тракту при переході від фонем до фонем змінюється поволі, то збільшення коефіцієнтів деталізації виявляється, як правило, тільки на одному рівні, який наперед невідомий і залежить, в першу чергу, від довжини сигналу і порядку вейвлету. Цю проблему можна вирішити вибором адекватних вейвлетних базисів для кожного класу фонем. Вибір найвідповіднішого базису в загальному випадку тривіальний і зводиться до вибору того з них, для якого кількість ненульових коефіцієнтів під час розкладання фонем даного класу мінімальна. Такий мульти-вейвлетний підхід полягає у використанні декількох вейвлетних базисів для пошуку міжфонемних переходів в кожному з них, з подальшим об'єднанням отриманих результатів. У якості можливих критеріїв для вибору конкретного вейвлету в [6] пропонується використання властивості регулярності, числа нульових моментів вейвлету, або числа вейвлет-коефіцієнтів, що перевищують деяке порогове значення. Іноді для вибору використовують функціонал інформаційної цінності (ентропії), а оптимальний базис обирають, мінімізуючи його.

Як зазначалося в першій частині, вокалізовані ділянки мовного сигналу складаються з квазіперіодичних послідовностей асиметричних коливань, що швидко згасають. З точки зору розподілу сигналу на вокалізовані та невокалізовані ділянки, вейвлети, за винятком деяких функцій, також є асиметричними і локалізованими в часі коливаннями скінченної тривалості. Природним є те, що за певного масштабу вони добре (у сенсі малої кількості відмінних від нуля коефіцієнтів) описують квазіперіодичну структуру вокалізованого мовного сигналу. Для невокалізованих ділянок мовного сигналу локальна енергія сигналу розподілена по значній кількості масштабів [2, 5]. Крім того, для вокалізованих ділянок мовного сигналу за достатньо хорошого часового масштабу перетворення, найбільші коефіцієнти вейвлет-перетворення будуть групуватися на масштабах, порівнянних з періодом основного тону (ОТ) поблизу локальних особливостей сигналу, що співпадають з квазіперіодами ОТ. Таким чином, поділ мовного сигналу на вокалізовані та невокалізовані ділянки може бути виконаний оцінюванням величини коефіцієнтів розкладання в заданому діапазоні масштабів.

З точки зору шумоочищення, вищезазначений алгоритм фільтрації дозволяє ефективно видалити високочастотний шум, що навіть перевершує за величиною досліджуваний сигнал, тоді як перетворення Фур'є втрачає інформацію про особливості низькочастотної частини сигналу, що призводить до спотворення форми корисного сигналу.

### **Параметризація сигналу за допомогою вейвлет-перетворення**

Будучи «сильним» інструментом аналізу мовних сигналів, вейвлет-перетворення також може бути використане для розпізнавання мовних сигналів. При цьому опис мовного сигналу, що використовується для розпізнавання, може включати в себе, як повний набір вейвлет-коефіцієнтів, так і їх підмножину, залежно від задачі, що вирішується. Так, під час побудови дикторонезалежних систем розпізнавання мови слід використовувати коефіцієнти великих і середніх масштабів, відкидаючи малозначні флуктуації мовного сигналу, які описуються коефіцієнтами, що відповідають малим масштабам. Перспективним є поєднання вейвлет-перетворення з нейромережевими методами розпізнавання, що дозволяє отримати вельми якісне розпізнавання.

Опис сигналу, отриманий в результаті вейвлет-перетворення, може служити основою, вхідними даними для систем розпізнавання, що ґрунтується на прихованих Марківських моделях, методах

порівняння з еталоном і т. д.

### Штучні нейронні мережі

Штучна нейронна мережа — це математична модель системи високо-паралельного обчислення, що побудована за принципом організації і функціонування біологічної нейронної мережі, складається з простих одиниць і природно може зберігати знання і використовувати їх в майбутньому. Нейронні мережі подібні до мозку людини за двома ознаками:

- 1) можливість видобутку знання з навколишнього середовища через процес навчання;
- 2) для зберігання надбаного знання використовуються сили зв'язків (ваги) між нейронами.

ШНМ розвивалися як узагальнення математичних моделей людського навчання, припускаючи, що:

1. Обробку інформації виконують нейрони;
2. Сигнали розповсюджуються за допомогою зв'язків, які мають різну силу;
3. Кожний нейрон є обробником інформації. Він застосовує активаційну функцію (як правило нелінійну) до згрупованої сукупності вхідних даних, визначаючи тим самим вихідні дані.

ШНМ, зазвичай, характеризуються трьома чіткими аспектами: архітектурою, тобто зразком зв'язності, алгоритмом навчання і функцією активації.

В 1943 році McCulloch і Pitts спроектували першу модель штучного нейрона, яка показана на рис. 1.

Штучні нейронні мережі застосовуються у задачі розпізнавання мовлення в багатьох аспектах. Вони використовуються для обробки вхідних сигналів, моделювання фільтрів, класифікації фонем, розпізнавання ізольованих слів із словника, визначення диктора в текстово-залежних задачах розпізнавання, а також як оцінювачі для інших моделей розпізнавання мовленнєвого сигналу.

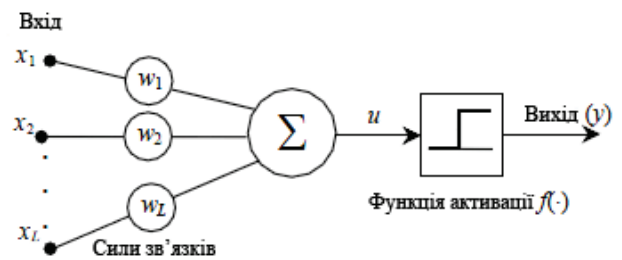


Рис. 1. Модель штучного нейрона, запропонована McCulloch і Pitts

Для розпізнавання мовленнєвої інформації використовують багаторівневі перцептрони і карти ознак Кохонена. В залежності від конкретної задачі вихідний шар перцептрону може визначати номер диктора, номер слова в словнику, або представляти згладжений вхідний сигнал.

В [7] розглядається метод застосування ШНМ для розпізнавання ізольованих слів із словника. Нейронна мережа зображена на рис. 2 і являє собою повністю зв'язний багатошаровий перцептрон.

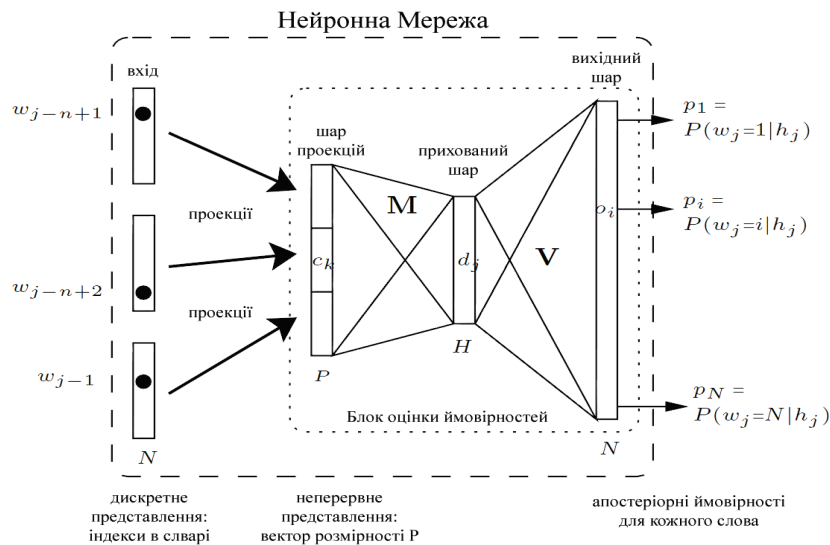


Рис. 2. Модель нейронної мережі — моделі природної мови

Вхідний шар мережі — це індекси  $(n - 1)$ -го попереднього слова зі словника, а вихідний шар — апостеріорна ймовірність появи кожного слова зі словника. Проекції використовуються для зве-

дення дискретного представлення (індекси  $(n - 1)$ -го попереднього слова зі словника) в неперервне (вектор, розмірності  $P$ ). Вихідний шар мережі відображає ймовірність того, що наступне (відносно входу мережі) слово буде деяким конкретним словом зі словника. Кількість нейронів вихідного шару дорівнює кількості слів у словнику  $N$ . Навчання мережі здійснюється методом оберненого поширення помилки з функцією перехресної ентропії в якості функції помилки. Звичайно, що при такому підході для отримання стійкого розпізнавання необхідним є великий навчальний набір. В якості навчального набору не можна використовувати ізольовані слова, адже вхід нейронної мережі залежить від попередніх слів в мовленнєвому сигналі. Тобто навчальний набір має складатися з надиктованої природної семантично вірної мовленнєвої інформації і її текстового представлення. Збір такої інформації є дуже складною задачею і окремою науковою проблематикою, якою займаються багато організації. Так, агентство міністерства оборони США DARPA (Defense Advanced Research Projects Agency — Агентство передових оборонних дослідницьких проєктів) в 2003 році створило EARS (Effective Affordable Reusable Speech-to-Text — Доступна та ефективна повторно-відтворювана технологія перетворення голосу в текст), яка фактично являє собою великий обсяг мовленнєвого сигналу від різних дикторів з синхронним текстовим відтворенням. Для проєкту EARS записувався невимушений мовленнєвий сигнал, як то розмова дикторів на радіо та телебаченні, телефонні розмови на багатьох мовах і т. д. В [7] навчальні набори EARS використовувалися для навчання нейронної мережі. Навчальний набір складався з майже 22 мільйонів слів, а тривалість мовленнєвого сигналу склала 1800 годин. Зі словником у 22 мільйони слів, а отже з вихідним шаром у 22 мільйони нейронів, нейронна мережа набула величезних розмірів, і її навчання на комп'ютері з процесором Intel Xeon з частотою 2,8 GHz зайняла 7 годин. Розпізнавання слів навченою нейронною мережею займає час, рівний 0,1RT, де RT (Real-time) — тривалість вхідного мовленнєвого сигналу [8]. Тобто, незважаючи на величезний розмір цієї ШМ, її швидкодія на етапі розпізнавання мовленнєвого сигналу є достатньою для використання її у системах, що працюють у реальному часі.

Експерименти показали, що найкращі результати нейронна мережа дає під час використання 4-грамм в якості вхідних даних. З використанням більшої кількості попередніх слів збільшується складність навчання, а якість розпізнавання не покращується. З навчальним набором у 7,2 мільйона слів (24 мільйона 4-грамм) рівень помилок під час розпізнавання склав 24,51 %. З навчальним набором у 12,3 мільйона слів (24,3 мільйона 4-грамм) — 22,19 %. З навчальним набором у 21,7 мільйона слів (28,4 мільйона 4-грамм) — 21,81 %. Для покращення якості розпізнавання запропоновано використання додаткових словників: протоколи радіо новин (240 мільйонів слів), протокол телевізійних новин CNN (80 мільйонів слів), зібрані в Інтернеті розмовні дані, що схожі на діалоги (180 мільйонів слів). Використання цих словників (загальною кількістю у 500 мільйонів слів) зменшило рівень помилок при розпізнаванні до 23,70 %, 21,77 % та 21,49 %, відповідно.

В [9] штучні нейронні мережі використовуються для визначення диктора, що вимовив наперед заданий текст. Автори використовували модель ШМ, яка показана на рис. 3. Модель побудована за архітектурою багатошарового перцептрон. Всі нейрони прихованого шару мають сигмовидну функцію активації, а нейрони вихідного шару — лінійну функцію. а вхід нейронної мережі подається деяким чином параметризований мовленнєвий сигнал. Автори пропонують використовувати Спектральну Густину Потужностей (СГП) або Кодування з Лінійним Передбаченням (КЛП) в якості методів видобутку характерних ознак промови того чи іншого диктора із мовленнєвого сигналу.

Кількість нейронів вихідного шару дорівнює кількості дикторів і в мережі, що використовувалась для експериментів, склала 5, а кількість нейронів прихованого шару для п'яти дикторів обрали рівною 30. Нейронну мережу навчали методом оберненого розповсюдження помилки, подаючи на вхідний шар мережі вектор СГП мовленнєвого сигналу. Всього мовленнєвих сигналів було сім від п'яти різних дикторів, які промовляли одну і ту ж саму фразу. Кількість елементів вектора

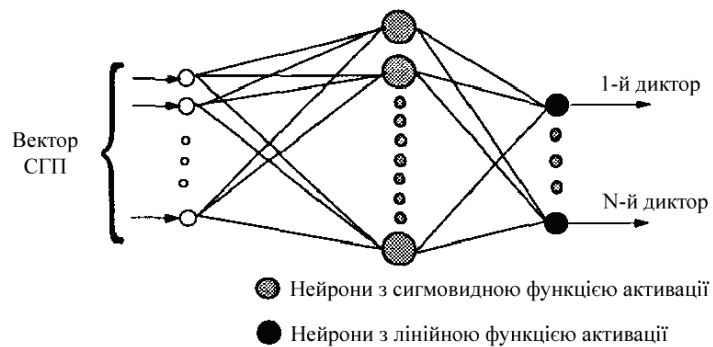


Рис. 3. Модель багатошарового перцептрон для задачі ідентифікації диктора

СГП була рівною 500. Нейронна мережа досягла прийнятної помилки (сума квадратів помилок = 0,01) за 174 ітерацію навчання. Натренована мережа показала 100 % точність розпізнавання на даних, на яких її навчали. У разі тестування мережі новими мовленнєвими сигналами від дикторів, на яких мережа була натренована, точність склала 66 %.

Також в [9] для експериментів використано нейронну мережу з архітектурою самоорганізаційної карти ознак, а за вхідні дані для неї використовували СГП (500 елементів), КЛП (20 елементів) та поєднання СГП і КЛП. Мережа навчалася значно довше, але результати були значно кращими. Так, використовуючи спектральну густину сигналу в якості вхідних даних, точність розпізнавання диктора склала 90 %, використовуючи лінійне передбачення — 98 %, їх комбінацію — 99 %. Також слід зауважити, що використання кодування з лінійним передбаченням значно спрощує складність обчислень і КЛП є кращим методом параметризації сигналу, аніж обчислення спектральних густин потужностей.

На рис. 4 показана одна з варіацій архітектури багаторівневого перцептронну, що використовується в якості фільтра зі скінченною імпульсною характеристикою (СІХ-фільтра) для обробки мовленнєвого сигналу.

На вхідний шар нейронної мережі подається деяким чином параметризований вхідний сигнал з послідовним запізненням ( $z^{-1}$ ). Активаційна функція нейронів, як правило, є сигмовидною, а вихідний шар, представлений одним нейроном, моделює фільтрований сигнал.

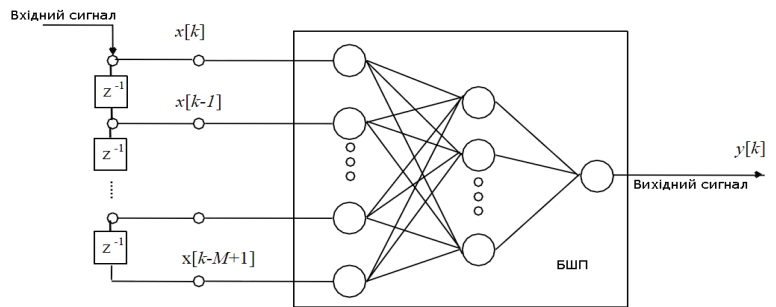


Рис. 4. Модель фільтра зі скінченною імпульсною характеристикою у вигляді багатозарового перцептронну

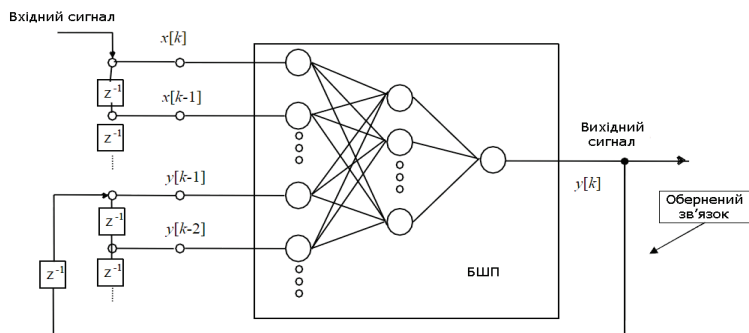


Рис. 5. Модель фільтра зі скінченною імпульсною характеристикою у вигляді багатозарового перцептронну

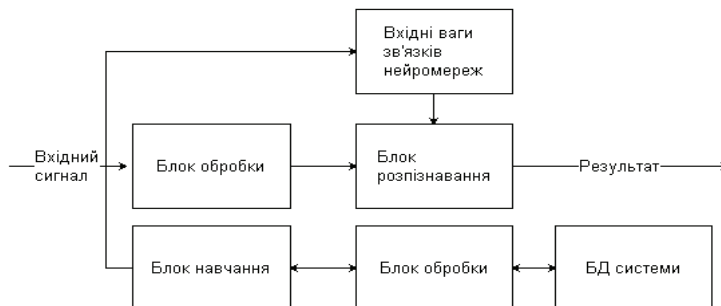


Рис. 6. Загальна схема СРМ з використанням нейромереж

На рис. 5 подана одна з варіацій архітектури багаторівневого перцептронну, що використовується в якості фільтра з нескінченною імпульсною характеристикою (НІХ-фільтра) для обробки мовленнєвого сигналу. На вхідний шар нейронної мережі подається деяким чином параметризований вхідний сигнал з послідовним запізненням ( $z^{-1}$ ) і за допомогою оберненого зв'язку — вихідний сигнал з послідовним запізненням. Активаційна функція нейронів тут також має сигмовидну форму, а вихідний шар представлений одним нейроном.

На рис. 6 наведено структуру системи розпізнавання слів природної мови з використанням нейромережі з самостійною адаптацією, розроблену в Красноярському інституті біо-

фізики Російської Академії Наук [10].

Блок розпізнавання і блок навчання містять по дві нейромережі кожний. В роботі не розглядалася можливість побудови комплексу нейронних мереж з самостійною адаптацією, що навчаються

по кінцевому результату. Нейронна мережа має топологію багат шарового перцептроні і складається з трьох рівнів: вхідний шар, прихований шар та ефекторний шар. Кожний нейрон наступного шару пов'язаний з усіма нейронами попереднього шару. Функція активації в усіх нейронах є лінійною, а у вхідному шарі моделюється конкуренція.

Навчання системи в цілому складається з трьох етапів. Спочатку системі пред'являються тільки зразки мовленнєвих сигналів, при цьому на вхідному шарі формуються нейронні ансамблі, ядрами яких є ці зразки. Потім пред'являються звуки і відповідні їм символи алфавіту, при цьому здійснюється асоціація нейронів вхідного рівня з нейронами символного рівня. На останньому етапі система розпізнає мовленнєві сигнали, а разом з тим навчається і адаптується.

Загалом, методи розпізнавання мовленнєвої інформації, що ґрунтуються на ШНМ, використовують імовірісно-мережеві моделі ухвалення рішення. Для даного підходу будь-який мовленнєвий сигнал можна представити як вектор у параметричному просторі, який потім може бути зафіксований у нейромережі. Багато нейромережних алгоритмів здійснюють паралельну обробку інформації, тобто інформацію обробляють одночасно усі нейрони, що вирішує проблему швидкості розпізнавання, адже звичайний час роботи нейромережі складає декілька ітерацій. На основі нейромереж легко будуються ієрархічні багаторівневі структури, при цьому зберігається їхня прозорість. Використання нейромереж з самостійною адаптацією дозволяє уникнути недоліків «некерованої» класифікації, що з'являються під час використання нейромережних алгоритмів з навчанням без вчителя. Проблема варіантності мови можна вирішити застосуванням нейромереж з топологією самоорганізаційних карт ознак. В таких мережах для множини вхідних сигналів формуються нейронні ансамблі, що представляють ці сигнали, а такий алгоритм має здатність до статистичного усереднення.

### Штучні імунні системи

Штучна імунна система (ШИС) — це новий алгоритм навчання, що ґрунтується на принципах функціонування природної імунної системи. Штучна імунна система об'єднує *априорне* знання з адаптивними можливостями біологічної імунної системи що створює могутню альтернативу наявним на сьогодні методам розпізнавання образів, навчання та оптимізації.

Обробка імунною системою зовнішніх сигналів здійснюється за допомогою високопаралельного процесу. Імунна система є чудовим прикладом системи, що реалізує адаптивні процеси глобального рівня на основі локальних взаємодій. Для вирішення завдань розпізнавання і класифікації імунна система використовує механізми навчання, пам'яті і асоціативного пошуку.

Природна імунна система є складною системою, що складається з декількох функціонально різних компонентів. Імунна система використовує багаторівневий захист проти зовнішніх антигенів, включаючи дію неспецифічних (природжених) і специфічних (набутих) захисних механізмів. Основна роль імунної системи полягає в розпізнаванні всіх клітин (або молекул) організму і класифікації їх на «свої» і «чужі». Чужорідні клітини піддаються подальшій класифікації з метою стимуляції захисного механізму відповідного типу. В процесі еволюції імунна система навчається розрізняти зовнішні антигени (наприклад, бактерії і віруси) і власні клітини або молекули організму.

При попаданні антигену в організм лише мала частина клітин імунної системи здатна розпізнати його пептиди. Макрофаги та інші антиген-презентуючі клітини, виставляють частини антигену (пептиди) на своїй поверхні, тим самим «привертаючи увагу» В- і Т-клітин для розпізнавання. Таке розпізнавання стимулює процеси розмноження<sup>1</sup> і диференціювання лімфоцитів, що приводить до утворення клонів ідентичних клітин (антитіл). Цей процес, що називають розмноженням клону, формує численну популяцію специфічних до антигену антитіл — продукуючих клітин. Розмноження клонів імуннокомпетентних клітин приводить до руйнування або нейтралізації антигену. Частина створених клітин зберігається як імунна пам'ять. В результаті подальша дія або вплив схожого антигену приводить до швидшої імунної реакції (вторинної відповіді).

Імунна система найцікавіша як система, що здатна ефективно обробляти значні обсяги даних.

<sup>1</sup> Процес розмноження контролюється афінністю. Афінність (спорідненість) збільшується, якщо в результаті утворення зв'язків між антитілами клону і антигеном збільшується швидкість мутації клітин. Мutowані клітини антитіл, що створюють стійкіші зв'язки з антигенами поділяться швидше.

Зокрема, вона виконує великий обсяг складних високопаралельних розподілених обчислень [11]. Поведінка імунної системи в цілому визначається всією сукупністю локальних взаємодій.

Імунна система здатна зберігати інформацію про інтенсивність попередніх контактів її клітин з антигенами і здатна давати відповідь на нові антигени, що раніше їй не зустрічалися. Опис закономірностей розвитку імунної відповіді є цікавою задачею теорії динамічних систем і її добре описано в [12].

Відносно застосування штучних імунних систем у обробленні звукових сигналів перспективними є такі властивості імунної системи:

— *Розпізнавання*. Імунна система здатна розпізнавати і класифікувати різноманітні молекулярні структури і вибірково на них реагувати. Розпізнавання відбувається в ході міжклітинних контактів, при цьому сила зв'язків, що утворюються, визначається формою молекул і величиною електростатичного заряду. Розпізнавання свого і чужого є одним з основних завдань, яке вирішує імунна система.

— *Виділення особливостей*. Антиген-презентуючі клітини (АПК) інтерпретують антигенне оточення і виділяють особливості шляхом обробки антигенів і представлення антигенних пептидів на своїй поверхні. Кожна АПК служить як «фільтр», що пригнічує молекулярний шум, і як «збільшувальне скло», що фокусує увагу лімфоцитів-рецепторів [13].

— *Різноманітність*. Імунна система використовує комбінаторний механізм (генетично-обумовлений процес) для утворення безлічі різних рецепторів лімфоцитів, з тим щоб гарантувати, що хоча б один лімфоцит зі всієї сукупності зможе вступити у взаємодію з будь-яким наперед заданим (відомим або невідомим) антигеном.

— *Навчання*. Імунна система оцінює структуру конкретного антигену, використовуючи його випадкові контакти з клітинами, що складають імунну систему. Навчання полягає в зміні концентрації лімфоцитів, яке відбувається під час первинної відповіді (в результаті першого контакту імунної системи з антигеном). Таким чином, здібність імунної системи до навчання закладена головним чином, в механізмі поповнення клонів [14], що призводить до утворення нових імуннокомпетентних клітин з урахуванням поточного стану системи (цей процес називають *розмноженням клону*).

— *Пам'ять*. Невелика частина лімфоцитів, які знаходяться в активованому стані, стають клітинами пам'яті (асоціативна пам'ять). Вважається, що час життя клітин пам'яті є динамічною величиною і визначається частотою стимуляції антигенами. Використовуючи короточасні і довгострокові механізми імунної пам'яті, імунна система підтримує ідеальний баланс між економією ресурсів і виконання функції за рахунок збереження мінімально необхідної, але достатньої пам'яті про попередні контакти з антигенами.

— *Розподілений пошук*. За своєю сутністю імунна система — це розподілена система. Клітини імунної системи, головним чином лімфоцити, безперервно циркулюють через кров, лімфу, лімфоїдні органи і решту тканин. У разі зустрічі з антигеном вони здійснюють специфічну імунну відповідь.

— *Саморегуляція*. Імунний захист має властивість саморегуляції. Центрального органу, контролюючого функції імунної системи, не існує. В залежності від способу проникнення в організм антигену та інших його властивостей, імунна відповідь може бути як локальною, так і глобальною (системною).

— *Пороговий механізм*. Імунна відповідь і розмноження імуннокомпетентних клітин відбуваються лише після подолання деякого порогу, що залежить від сили хімічних зв'язків.

— *Сумісна стимуляція*. Активація В-лімфоцитів жорстко регулюється за допомогою додаткового стимулюючого сигналу. Цей стимулюючий сигнал (від допоміжних Т-лімфоцитів) допомагає забезпечувати толерантність системи і робити відмінність між серйозними загрозами і «помилковими дзвінками» (розрізнення небезпечних і безпечних антигенів).

— *Динамічний захист*. Клональне розмноження і соматична гіпермутація дозволяють імунній системі продукувати високоафінні імуннокомпетентні клітини (цей процес називається збільшенням афінності), що створює динамічний баланс між функцією, що вивчає і захисною функцією адаптивного імунітету. Наявність динамічного захисту поступово приводить до розширення зон



спостереження, що контролюється імунною системою.

— *Імовірнісне виявлення*. Перехресні реакції в ході імунної відповіді є стохастичним процесом. До того ж виявлення антигену завжди неминуче відбувається приблизним чином, отже лімфоцит може одночасно взаємодіяти з декількома структурно схожими антигенами.

В імунній відповіді важливу роль грають і інші характеристики імунної системи, такі як адаптивність, специфічність, самотолерантність, диференціювання та інші. Всі ці чудові властивості мають відношення до обробки інформації і створюють ряд цікавих можливостей з обчислювальної точки зору.

Здібність до навчання штучної імунної системи, заснованої на принципах функціонування імунної мережі [15], досліджувалася в [16, 17]. ШІС виявилася стійкою до шумів, здібною до самонавчання і не потребувала як негативних прикладів, так і явної вказівки на те, що саме слід запам'ятовувати. У цій системі об'єднані системи класифікації з навчанням з деякими перевагами нейронних мереж, алгоритмів машинного виведення і адаптивного пошуку.

ШІС складається з керувального об'єкту (мережі клітин), а також з навчального і тестового наборів даних. Кожна клітина мережі має компонент, що дозволяє розпізнавати структури на основі порівняння. Цей компонент створюється за допомогою процесу, що нагадує генетичний механізм утворення антитіл в природній імунній системі. Так формується словник структур, і підтримується його різноманітність. Система генерує два типи відповідей: первинний і вторинний. Первинна відповідь генерується у фазі навчання, під час якої ШІС вивчає структури для формування навчального набору даних. Вторинна імунна відповідь генерується під час розпізнавання структур, коли ШІС намагається класифікувати нові дані, порівнюючи їх з раніше вивченими.

Використовуючи ШІС для конкретних цілей, необхідно спочатку провести налаштування системи на навчальній вибірці, пред'являючи її або всю цілком, або частинами (залежно від завдання). Отриману в результаті навчання інформацію можна використовувати декількома способами.

У [17] було показано, що ШІС є прикладом нелінійних адаптивних мереж з навчанням і явно адресованою пам'яттю. Метою цих досліджень було створення інструментарію для вирішення завдань машинного навчання, заснованого на принципах імунної системи.

Для вирішення завдань розпізнавання образів в роботі [18] зроблено спробу створити самоасоціативну і змістовно-адресовану пам'ять, що ґрунтується на принципах імунної мережі. На вхід системи подавалися мозаїчні чорно-білі квадратні структури розміром 64×64 пікселів, що виконували роль антигенів. Хоча авторам і не вдалося отримати стійкі рішення, запропонований підхід є вельми цікавим.

З появою методики штучних імунних систем почали з'являтися обчислювальні моделі, побудовані на цій методиці. Найпопулярнішими і застосовними моделями штучних імунних систем є: модель кісткового мозку, алгоритм негативного відбору, алгоритм клонального відбору, модель імунної мережі (aiNet).

В [19] розглянуто механізм застосування алгоритму негативного відбору для ідентифікації диктора в задачі дикторо- та текстово-залежного розпізнавання мовленнєвої інформації. Тобто спочатку система проходить фазу навчання, коли кожен диктор диктує одну і ту ж саму фразу, а потім в процесі фази ідентифікації один із дикторів вимовляє текст фрази, і система або ідентифікує диктора, або повідомляє що це новий диктор, якого в її базі ще немає.

Основна мета, що повинна бути досягнута майже в усіх програмах з використанням моделі негативного вибору, полягає у виявленні аномальних відхилень від нормальної поведінки. Алгоритм генерує детектори з деяким чином параметризованої версії сегментованих вхідних даних (набору власних клітин). Іноді для кодифікації власних елементів параметризацію не використовують, тобто кодують двійкове представлення вхідних даних [19]. Кожний елемент детектора має фіксовану довжину, яка теж може бути параметром алгоритму. Таким чином, алгоритм негативного вибору використовується для побудови набору детекторів голосового сигналу цього диктора. Згенеровані детектори потім використовуються в якості «відбитків голосу» для контролю здобутого нового голосового сигналу (фаза ідентифікації). Якщо сигнал був вимовлений тим самим диктором, форма та розподілення даних повинні бути дуже схожі на оригінал, що був використаний для побудови детекторів, тому і отримані аномалії будуть малі (в ідеальному випадку їх взагалі не буде). Відповідно до отриманого рівня аномалії, автоматична система розпізнавання диктора приймає до рішення про наявність чи відсутності нового диктора. Щоб досягти цього, використовується база даних «відбитків

голосу» різних дикторів. Відбиток голосу задається набором детекторів, отриманих в результаті застосування алгоритму негативного відбору до відповідного голосового сигналу (фаза навчання). Якщо мінімальний отриманий рівень аномалій вищий за фіксований поріг (це параметр системи, що визначає найвищий прийнятний рівень виявленої аномалії), то система вирішить, що цей голосовий сигнал не належать до будь-якого диктора у базі даних, і тому диктора не виявить. В іншому випадку, дійсним диктором визнається той, рівень аномалії для якого буде найнижчим.

На рис. 7 показано типову систему дикторозалежного розпізнавання голосу, що ґрунтується на принципах штучної імунної мережі.

В [19] штучна імунна система була навчена голосом одного диктора. Диктор промовляв один і той самий текст в різні дні і в різних ситуаціях. В фазі ідентифікації використовувалися голоси інших дикторів та голос диктора, яким навчили систему, записаним через тиждень, 2 тижні та місяць. Для виявлення особливостей сигналу використовувався набір із 100 детекторів, а сигнал параметризувався значенням тиску хвилі.

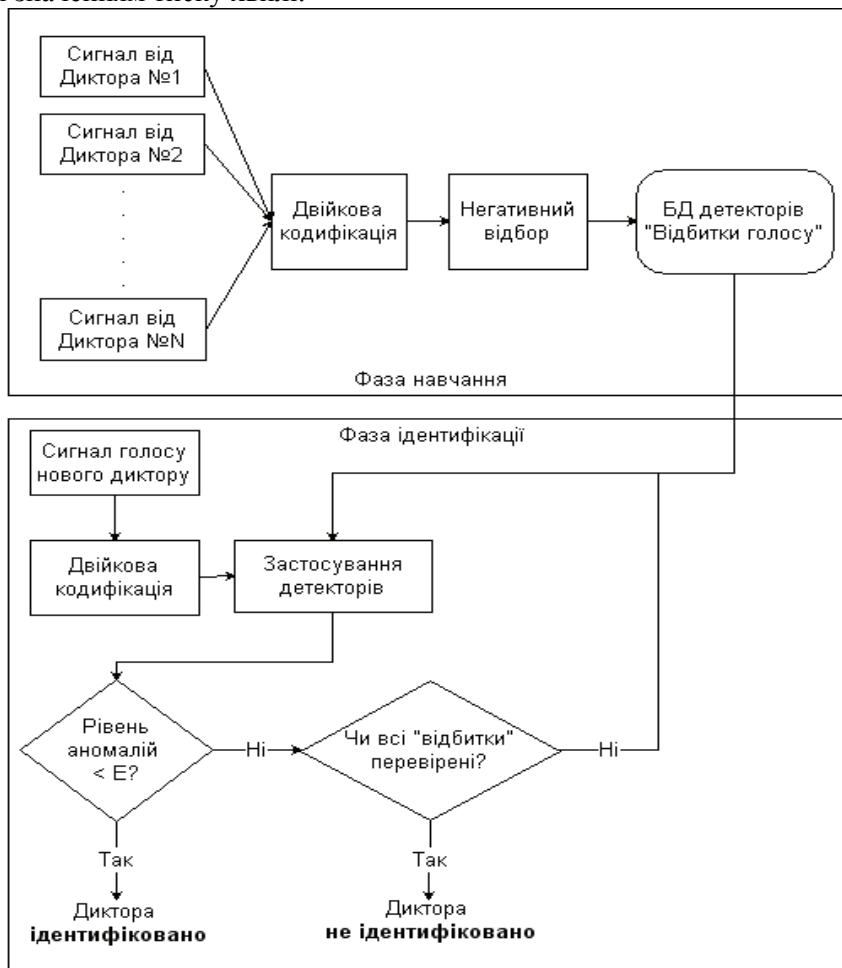


Рис. 7. Дикторозалежна система розпізнавання голосу на основі моделі негативного відбору

Диктори, що не були представлені системі на етапі навчання були ідентифіковані як такі, яких система не знає. Голос дійсного диктора, записаний через тиждень та 2 тижні системою був розпізнаний, а записаний через місяць — трохи перевищив поріг розпізнавання. Тобто, в двох з трьох випадків штучна імунна система вдало ідентифікувала диктора.

### Порівняння штучних імунних систем та штучних нейронних мереж

Вдале всебічне порівняння штучних імунних систем з штучними нейронними мережами можна здійснити, якщо в якості штучної імунної системи використати її популярну модель — штучну імунні

мережу (aiNet), яку було запропоновано імунологом Нільсом Ерном в 1974 р. [15].

З точки зору практичного застосування, функціонування штучної імунної мережі полягає у стисненні даних, використовуючи ідеї теорії імунних мереж, і принципу клонального відбору. Це стиснення приводить до створення мережі антитіл, що здатна розпізнавати антигени (вхідний набір даних) з певною (і регульованою) точністю. Зменшуючи потужність множини даних, aiNet зменшує складність проблеми, фільтруючи сторонні антитіла і використовуючи набори прототипів (антитіла або інші клітини мережі) для представлення різних кластерів даних.

З точки зору порівняння штучних імунних систем і штучних нейронних мереж, aiNet як модель ШПС обрана тому, що вона є фундаментальною моделлю ШПС зі зв'язаними клітинами. Наявність моделі зв'язності робить aiNet набагато ближчою до ШНМ, ніж будь-яка інша модель ШПС. Порівняльна характеристика aiNet та штучних нейронних мереж наведена у таблиці.

**Порівняльна характеристика aiNet та ШНМ**

Характеристика	aiNet	Штучна нейронна мережа
Основна одиниця	Клітина, що складається з низки властивостей і пов'язаних сил зв'язків	Нейрон, що складається з активаційної функції, сил зв'язків і активаційного порогу
Взаємодія з іншими одиницями	Клітина має зважені зв'язки, що розпізнають і можуть бути розпізнаними іншими клітинами. Ці ваги вказують на ступінь взаємодії з іншими клітинами	Вектори ваг можуть набувати як позитивні, так і негативні значення, вказуючи на підсилювальну чи стримувальну дію
Дія	Клітина має внутрішнє зображення навколишнього середовища, що порівнюється з інформацією, яка надходить	Нейрон оброблює інформацію, що надійшла до нього
Знання	Зберігається в силі зв'язків і низці атрибутів кожної клітини	Зберігається в вагах
Навчання	Проводиться модифікацією низки атрибутів основних одиниць і векторів сил зв'язків	Проводиться модифікацією сил зв'язків
Пам'ять	Розподілена і асоціативна	Самоасоціативна або асоціативна та розподілена
Поріг	Визначає зв'язування (розпізнавання) між клітинами і наявними стимулами	Визначає активність нейрона
Стійкість	Масштабована, гнучка, стійка до шуму та до самої себе	Високогнучка і стійка до шуму
Знаходження основних одиниць	Місцезнаходження є динамічним	Нейрони мають фіксовані позиції в мережі
Комунікація	Через клітинну спорідненість, що представлена силою зв'язків	Через сили зв'язків
Стан	Визначається концентрацією та/або спорідненістю клітин та отримуваних стимулів	Визначається активаційним рівнем вихідних нейронів
Керування	Імунні принципи визначають типи взаємодій між компонентами ШПС (керування — динамічне)	Процес корегування ваг визначається алгоритмом навчання
Характеристика	Самоорганізаційна і еволюційна	Залежить від алгоритму навчання
Структура	Шаблон зв'язності	Шаблон зв'язності
Динаміка	Зміна з часом концентрації і спорідненості клітин	Зміна з часом сили зв'язків
Метадинаміка	Неперервне продукування і смерть клітин мережі	Конструктивний алгоритм, та алгоритм навчання зі зменшення надлишкових нейронів в схованих шарах (Reed, 1993; Kwok & Young, 1997)

## Перспективи розвитку проблеми. Висновки

На акустичному рівні важливим є якісне й водночас компактне представлення звукового сигналу, що містить значиму для розпізнавання інформацію. Для створення акустичної моделі сигналу найкращі результати дає використання вейвлетного базису, що має ряд переваг перед перетворенням Фур'є: варійована роздільна здатність за часом і частотою (масштабне перетворення), рухоме частотно-часове вікно, добра обробка нестационарних процесів, локалізація в часовій і частотній області, можливість зсувів, добре відображення локальних особливостей сигналу, що особливо корисно для виявлення високочастотних характеристик сигналу. Вейвлет-перетворення може використовуватися як для акустичного представлення сигналів, так і для їхньої фільтрації.

Нейронні мережі якісно і швидко виконують задачі класифікації за рахунок паралельної обробки усіма нейронами. Нейронні мережі, подібно людському слуховому аналізатору, дозволяють отримати кореляцію інформації, що надходить на вхід мережі, чого важко досягти за допомогою прихованих Марківських моделей. До переваг нейромереж можна віднести відносно простий процес навчання, можливість самонавчання та адаптації.

Штучні імунні системи за побудовою мають великий потенціал до дикторо- і текстово- незалежного розпізнавання мовленнєвої інформації. Вони мають багато переваг перед нейронними мережами. Штучні імунні системи децентралізовані, обробляють інформацію паралельно, тобто можуть виконувати швидкий пошук у великому просторі. Кількість імунних клітин не фіксована жорстко. ШИС мають адаптивну асоціативну пам'ять, що здатна запам'ятовувати безліч мовних фонем. Штучні імунні системи мають властивість різноманітності — шляхом мутації та клонування антитіл вони мають змогу отримати подібність до будь-якої, ще небаченої до цього моменту вхідної фонемі. ШИС є високо-динамічними системами завдяки наявності динамічного балансу між вивчаючою і розпізнавальною функцією. Штучні імунні системи здатні одним антитілом реагувати на декілька антигенів, що ще більше прискорює імунну відповідь (розпізнавання мовленнєвої інформації).

Як зазначалося в першій частині роботи, ефективна система розпізнавання мови має містити в собі такі етапи обробки вхідного сигналу, як видалення шуму, сегментація, виділення вокалізованих ділянок, визначення частоти основного тону, параметризація, розпізнавання, коригування за словником з оберненим зв'язком. Зрозуміло, що не один метод не може покрити усі етапи. Ефективна система має поєднувати в собі найкращі методи виконання кожного етапу, використовуючи їх переваги. Слід зауважити, що етап параметризації тісно пов'язаний з етапом розпізнавання, оскільки для якісного розпізнавання мовленнєвого сигналу конкретним методом часто необхідно використовувати конкретний метод параметризації.

Всебічний аналіз багатьох публікацій виявляє те, що з першими трьома етапами обробки вхідного сигналу найкраще справляється вейвлет-перетворення, а власне розпізнавання найкраще виконується штучними імунними системами.

## СПИСОК ЛІТЕРАТУРИ

1. Астафьева Н. М. Вейвлет-анализ: основы теории и примеры применения / Н. М. Астафьева // Успехи физических наук. — 1996. — Т. 166, № 11. — С. 1145—1170.
2. Рабинер Л. Теория и применение цифровой обработки сигналов / Л. Рабинер, Б. Гоулд. — М. : Мир, 1978. — 848 с.
3. Walker J. Fourier Analysis and Wavelet Analysis / James S. Walker // Notices of the AMS. — 1997. — V. 44, № 6. — P. 658—670.
4. Дремин И. М. Вейвлеты и их использование / И. М. Дремин, О. В. Иванов, В. А. Нечитайло // Успехи физических наук. — 2002. — Т. 171, № 5. — С. 465—501.
5. Фант Г. Акустическая теория речеобразования / Гуннар Фант. — М. : Наука, 1964. — 284 с.
6. Ермоленко Т. В. Применение вейвлет-преобразования для обработки и распознавания речевых сигналов / Т. В. Ермоленко // Искусственный интеллект. — 2002. — № 4. — С. 200—208. — ISSN 1561-5367.
7. Schwenk H. Neural Network Language Models for Conversational Speech Recognition : праці міжн. наук. конф. International Conference on Speech and Language, 4—8 жовт. 2004, Чеджу, Корея. — С. 1215—1218.
8. Schwenk H. Training Neural Network Language Models On Very Large Corpora : праці наук. конф. Joint Conference NLT/EMNLP, 6—8 жовт. 2005, Ванкувер, Канада. — С. 201—208.
9. Moonasar V. Speaker Identification using a Combination of Different Parameters as Feature Inputs to an Artificial Neural Network Classifier : праці наук. конф. AFRICON, 28 вер. — 1 жовт. 1999, Дурбан, ПАР. — Т. 1. — С. 189—194.
10. Лалетин П. А. Использование сетей с самостоятельной адаптацией для распознавания слов человеческой речи: праці наук. конф. Всероссийская научно-техническая конференция «Нейроинформатика-2000». Т. 2. — М. : МИФИ, 2000. — С. 88—95.
11. Frank S. A. The design of natural and artificial adaptive systems / Steven A. Frank. — New York : Academic Press,

1996. — С. 451—505.

12. Perelson A. S. Immunology for physicists / A. S. Perelson, G. Weisbuch // Review of Modern Physics. — 1997. — Т. 69, № 4. — С. 1219—1267.

13. Cohen I. R. The cognitive paradigm and the immunological homunculus / I. R. Cohen // Immunol Today. — 1992. — V. 13, № 12. — P. 490—494.

14. Bersini H. The immune recruitment mechanism: A selective evolutionary strategy : праці міжн. наук. конф. Fourth international conference on genetic algorithms, 13-16 лип. 1991, San Diego. — С. 520—526.

15. Jerne N. K. Towards a network theory of the immune system / N. K. Jerne // Annales d'immunologie. — 1974. — V. 125, № 1—2. — P. 373—389.

16. Hunt J. E. An adaptive, distributed learning system, based on the immune system : праці наук. конф. IEEE international conference on systems, man and cybernetics, 22—25 жовт. 1995. — V. 3. — P. 2494—2499.

17. Hunt J. E. Learning using an artificial immune system / John E. Hunt, Denise E. Cooke // Journal of Network and Computer Applications. — 1996. — V. 19, № 2. — P. 189—212.

18. Gilbert C. J. Associative memory in an immune-based system : праці наук. конф. 12th national conference on artificial intelligence (AAAI-94), 31 лип. — 4 квіт. 1994, Seattle. — С. 852—857. — ISBN 0-262-61102-3.

19. Faraoun K. M. Artificial Immune Systems for text-dependent speaker recognition / K. M. Faraoun, A. Boukelif // INFOCOMP — Journal of Computer Science. — 2006. — V. 5, № 4. — P. 19—26.

Рекомендована кафедрою інтелектуальних систем

Надійшла до редакції 22.04.09  
Рекомендована до друку 20.05.09

*Добрушкін Григорій Олександрович* — аспірант, *Данилов Валерій Якович* — професор.

Кафедра математичних методів системного аналізу, Національний університет України «Київський політехнічний інститут»