

Образний комп'ютер: концепції, методологія, підходи

Тарас Вінцюк

Міжнародний науково-навчальний центр ЮНЕСКО інформаційних технологій та систем
40 проспект Академіка Глушкова, Київ 03680, Україна
Тел.: +380 44 266-4356 Факс: +380 44 266-1570
vintsiuk@uasoiro.freenet.kiev.ua

ABSTRACT

Taras K. Vintsiuk. Pattern Computer: Concepts, Methodology, Approaches. Problems of designing, creation, developing, production and dissemination for principally new so called pattern computers and respective intellectual information technologies and systems are discussed. Research strategy and technical policy for Ukraine on 2000-2010 years period as well as on nearest 2000-2003 years period are considered too.

Pattern computer (PatCom) is such intellectual cybernetic system that provides a functional simulation of intelligent, mainly subconscious activity of all living and a human being particularly. In PatCom this activity is related to image, sound and other patterns perception, scene analysis, action and movement planning, generalisation of observations, discovering of regularities, prediction, decision making etc. Pattern computer operates with patterns and other complex notions. It actualises both pattern and logical reasoning.

Pattern computer is a parallel system. It has several information perception channels (acoustic, visual, scential) that is multimodal perception, pattern operation system, improved human-machine interface. Unlike usual computer, which is based on a rapid arithmetic-logical processor and a large RAM space, PatCom is grounded on external world models including physical, geometry, acustical, language, linguistic, semantical, canonical forms etc models.

Two ways how to create a pattern computer are debated. The first approach is based on so-called generative model of pattern recognition and understanding. A multimodal information synthesis is used as a feedback in the pattern analysis and understanding. The second way to create PatCom is so-called multilevel multidecision model. As example, ways how to create a dictation and spoken translation machine are discussed.

ВСТУП

Сьогодні науково-технічний прогрес визначається значними проривами екстенсивного характеру в галузі розроблення та впровадження інформаційних систем, засобів телекомунікації та комп'ютеризації. Досягнені фантастичні швидкодії та обсяги пам'яті у комп'ютерів, глобалізується Інтернет, надається широкий спектр інформаційних послуг обчислювального та пошукового характеру.

Спостерігаються й чіткі негативні тенденції, пов'язані з світовою монополізацією розроблення та виробництва комп'ютерів та засобів комунікації, зі штучним стримуванням розвитку передових інформаційних технологій та систем, в яких має нагальну потребу людство.

А людство потребує створення принципово нових комп'ютерів, які здатні сприймати та розуміти звуки, зображення, людську мову, рукописні тексти, крес-

лення, просторові та звукові сцени, інші образи, описувати та озвучувати зображення, перекладати з однієї мови на іншу тощо.

Розроблення таких комп'ютерів, які виконують не тільки обчислення, але й моделюють образне сприйняття світу та образне прийняття рішень відносять до проривних напрямів у науково-технологічному поступі.

Кабінет Міністрів України своєю Постановою від 8.11.2000 № 1652 схвалив Державну науково-технічну програму "Образний комп'ютер" та встановив два періоди її розвитку: 2000-2003 та 2004-2010 роки.

НАЗВА ТА МЕТА ПРОГРАМИ

"Розроблення нових, так званих образних комп'ютерів (ОК)". Скорочена назва програми — "Образний комп'ютер".

Програма має на меті створення принципово нових інформаційних технологій та систем — образних комп'ютерів, входження України в міжнародний науково-технологічний розподіл праці, гарантування ринків збуту вітчизняної наукомісткої продукції.

ЩО ТАКЕ ОБРАЗНИЙ КОМП'ЮТЕР?

Образний комп'ютер — це така кібернетична система, в якій виконується функційне моделювання інтелектуальної, головню підсвідомої, діяльності людини та всього живого, що пов'язана зі сприйняттям зорових, слухових та інших образів, аналізом сцен та складних ситуацій, плануванням дій та рухів, узагальненням спостережень, встановленням закономірностей, прогнозуванням, прийняттям рішень тощо. ОК оперує образами та іншими складними поняттями, реалізує як образне, так і логічне мислення.

Образний комп'ютер є паралельною мультимодальною системою, яка має у своєму складі декілька каналів сприйняття інформації (слухової, зорової, текстової, смакової, нюхової тощо), образну операційну систему, моделі зовнішнього світу (в тому числі акустичну, оптичну, геометричну, лінгвістичну, семантичну, канонічних форм тощо), розвинений інтерфейс з людиною, засоби взаємодії з існуючими комп'ютерними та телекомунікаційними мережами.

Образна операційна система "синхронізує" оброблення інформації, що надходить різними каналами її сприйняття, та, оперуючи моделями зовнішнього світу, виконує комплексну семантичну інтерпретацію всієї отриманої інформації.

На відміну від звичайного комп'ютера, основу якого складають арифметико-логічний процесор та оперативна пам'ять, у центрі образного комп'ютера знаходяться моделі зовнішнього світу з усіма його об'єктами, явищами та їх проявами.

Розкриємо поняття образного комп'ютера переліком деяких базових його властивостей, які розглядатимемо як вимоги до ОК.

Сприймає та аналізує звукові сигнали, визначає характер звуків, їх джерело, напрямок та відстань до джерела, автоматично розпізнає тип джерела.

Сприймає людську мову, автоматично розпізнає та розуміє її, озвучує довільні тексти. Автоматично редагує та друкує тексти під диктування.

Сприймає та розуміє надруковані та написані від руки тексти. Вміє перекладати з однієї мови на іншу.

Робить усний переклад з однієї мови на іншу.

Аналізує та розуміє зображення (фото, написи, креслення, графіки тощо). Аналізує та розуміє сцени, обстежує об'єкти, вимірює відстань до об'єктів, вказує їх місцезнаходження.

Орієнтується в часі та просторі. Аналізує причини та наслідки.

Верифікує та ідентифікує людину за її голосом.

Сприймає запахи та тактильну інформацію.

Стежить за здоров'ям користувача.

Здатен комплексно сприймати інформацію про зовнішній світ. Здатен до узагальнення спостережень, встановлення закономірностей, прогнозування, цілеспрямованих дій та рухів.

Дає зорову інтерпретацію усних повідомлень, описує текстами та усною мовою ті зображення, які спостерігає.

Сприймає, інтерпретує та описує аудіовідеоінформацію.

Виконує автоматичний пошук аудіовідео- та текстової інформації за заданою темою. Автоматично реферує тексти та аудіовідеоінформацію.

КОНЦЕПЦІЯ ПРОГРАМИ

Ряд ОК. Створюється ряд моделей образних комп'ютерів – від супер-ОК до персонального ОК масового використання.

Єдність ближніх і дальніх цілей. Створення образних комп'ютерів у повному обсязі потребує не менш ніж 10-річного періоду науково-технологічного поступу. Перші прототипи ОК можуть бути розроблені за дво(чотири)річний період. Окремі компоненти образних комп'ютерів можуть бути підготовлені для впровадження в найближчий рік-два. Ці компоненти вбудовуватимуться в існуючі ЕОМ, суттєво покращуючи споживчі властивості останніх.

Створення прототипів ОК кожного поточного року та коригування науково-технічної програми є необхідною умовою робіт.

Основні напрями робіт. Розроблення образного комп'ютера виконується за такими напрямками:

1. Методологія, концепція та архітектура ОК.
2. Образні операційні системи та системи програмування ОК.
3. Автоматичне розпізнавання, розуміння та синтез мовних сигналів для усного спілкування з ЕОМ природною мовою, створення диктувальних машин та машин усного перекладу. ОК "Усна Мова".
4. Сприйняття, автоматичне розпізнавання, розуміння та генерація зображень (просторових сцен, плоских зображень, текстової та креслярської інформації, зокрема друкованої та рукописної). ОК "Око".
5. Автоматичне розуміння, інтерпретація та генерація текстової інформації та її переклад з однієї мови на іншу. ОК "Перекладач".
6. Зорова інтерпретація усномовних повідомлень, усномовна інтерпретація зображень, комплексне сприйняття та генерація аудіовідеозображень. ОК "Інтерпретатор".
7. Самопрограмування цілеспрямованих дій комп'ютерних систем в реальному світі. ОК "Робот".
8. Автоматичне реферування та пошук текстової та аудіовідеоінформації за заданою темою. ОК "Авто-референт".

Дослідження за пп. 1-2 є фундаментальними, за пп. 3-8 — фундаментальними та науково-дослідними і дослідно-конструкторськими.

Перелік пілотних проектів 2000-2001 років:

1. Пакет програмних засобів для озвучення текстів українською та іншими мовами в ПК.
 2. Портативний текстовий словник-перекладач з української та на українську з озвученням українською мовою.
 3. Усний словник-перекладач (англійсько-український та українсько-англійський) для персонального комп'ютера.
 4. Програмно-апаратний комплекс для оброблення та змістовної інтерпретації креслярсько-графічних зображень (схем, карт, планів місцевості, креслень споруд тощо).
 5. Система телевізійного сприйняття навколишнього світу.
 6. Засоби усномовного введення-виведення даних для ЕОМ з технологіями мультимедія.
 7. Голосові замки.
 8. Комп'ютерна фоноскопія.
 9. Самопрограмування цілеспрямованих рухів маніпуляційних роботів для обминання перепон (наприклад, для збирання уламків радіоактивного палива).
 10. Тривимірний дисплей з контролером для нього.
- Першочергові завдання і проекти програми** зведені у перелік, що наводиться нижче.

1. Методологія, концепція та архітектура ОК

- 1.1. Методологія, концепція та архітектура образних комп'ютерів
- 1.2. Розроблення моделей та формальних мов для опису зовнішнього світу та подій у ньому
- 1.3. Автоматизація процесів формування понять про об'єкти, явища та закономірності за зразками, подібностями та аналогіями
- 1.4. Навчання, самонавчання та самоорганізація моделей

2. Образні операційні системи та системи програмування ОК

- 2.1. Ієрархічна структуризація знань і даних про зовнішній світ через мультимодальне сприйняття та організація управління пошуком інформації й прийняттям рішень
- 2.2. Принципи організації та управління в образних операційних системах
- 2.3. Системи програмування образних комп'ютерів

3. Автоматичне розпізнавання, розуміння та синтез мовних сигналів для усного спілкування з ЕОМ природною мовою, в тому числі створення диктувальних машин та машин усного перекладу. ОК "Усна Мова"

- 3.1. Набір засобів для озвучення текстів українською та іншими мовами в ПК
- 3.2. Портативний текстовий словник-перекладач з української та на українську з озвученням українською мовою
- 3.3. Усний словник-перекладач (англійсько-український та українсько-англійський) для персонального комп'ютера
- 3.4. Засоби усного діалогу для персонального комп'ютера
- 3.5. Диктувальна машина (машина для автоматичного друкування та редагування текстів під послівне диктування)
- 3.6. Диктувальна машина під зв'язну мову
- 3.7. Машина усного перекладу для англійської та української мов при послівному вимовлянні
- 3.8. Машина усного перекладу для англійської та української мов при злитому вимовлянні
- 3.9. Засоби стискання, компресованої передачі та відтворення усномовної інформації та їх інтеграція з Інтернет для організації комп'ютерної телефонії
- 3.10. Автоматична ідентифікація та верифікація особи за її голосом, голосові замки
- 3.11. Комп'ютерна фоноскопія
- 3.12. Генерація усної мови з моделюванням індивідуальних особливостей голосу людини

4. Сприйняття, автоматичне розпізнавання, розуміння та генерація зображень (просторових сцен, плоских зображень, текстової та креслярської інформації, зокрема друкованої та рукописної). ОК "Око"

- 4.1. Оброблення та змістовна інтерпретація креслярсько-графічних зображень (схем, карт, планів місцевості, креслень споруд тощо)
- 4.2. Програмно-апаратні засоби сприйняття рукописної інформації
- 4.3. Автоматичне сприйняття текстової інформації з книжок, газет, друкованих рукописів, її редагування та переформатування
- 4.4. Сприйняття, аналіз та розуміння просторових сцен
- 4.5. Автоматична ідентифікація та верифікація підписів
- 4.6. Генерація зображень та просторових сцен
- 4.7. Тривимірний дисплей з засобами його підтримки для персонального комп'ютера

5. Автоматичне розуміння, інтерпретація та генерація текстової інформації та її переклад з однієї мови на іншу. ОК "Перекладач"

- 5.1. Автоматичне розуміння та інтерпретація текстової інформації
- 5.2. Автоматичний синтез текстової інформації за заданим смислом
- 5.3. Автоматизація перекладу текстової інформації з однієї мови на іншу — лексично-граматичний переклад
- 5.4. Автоматичний переклад текстової інформації з однієї мови на іншу, що ґрунтується на розумінні текстів — семантичний переклад
- 5.5. Створення базового набору комп'ютерних мультимедійних словників української мови
- 5.6. Створення комп'ютерних версій українсько-іншомовних словників

6. Зорова інтерпретація усномовних повідомлень, усномовна інтерпретація зображень, комплексне сприйняття та генерація аудіовідеозображень. ОК "Інтерпретатор"

- 6.1. Зорова інтерпретація текстової інформації — перетворювач "текст—зображення"
- 6.2. Сміслова, текстова та усномовна інтерпретація зображень — перетворювач "зображення—смісл—текст—усна мова"
- 6.3. Комплексне сприйняття та опис аудіовідеоінформації — перетворювач "аудіовідео—смісл-текст"
- 6.4. Комплексна генерація аудіовідеоінформації за заданим "смісл-текстом"

7. Самопрограмування цілеспрямованих дій комп'ютерних систем в реальному світі. ОК "Робот".

- 7.1. Система телевізійного сприйняття навколишнього світу
- 7.2. Самопрограмування цілеспрямованих рухів маніпуляційних роботів для обминання перепон (наприклад, для збирання уламків радіоактивного палива).

8. Автоматичне реферування та пошук текстової та аудіовідеоінформації за заданою темою. ОК "Автореферент"

- 8.1. Автоматичне реферування текстів
- 8.2. Пошук текстової інформації за темою реферату
- 8.3. Автоматичне реферування фонограм
- 8.4. Автоматичний пошук фонограм на задану тему
- 8.5. Автоматичне реферування та пошук аудіовідеоінформації

МЕТОДОЛОГІЯ СТВОРЕННЯ ОК

Першу в континентальній Європі електронну обчислювальну машину, що мала назву "МЭСМ", було створено в Києві ще на початку 50-х років. Уже тоді її творці усвідомлювали, що машину треба навчити не тільки робити обчислення, а й мислити, сприймати й розуміти людську мову, а також звуки та зображення.

З того часу комп'ютери досягли гігантської швидкодії та обсягів пам'яті, значних успіхів у виконанні обчислень, пошуку інформації, формальної логіки. Але дуже скромні результати мають місце в моделюванні підсвідомої діяльності людини та всього живого. По правді кажучи, комп'ютери все ще залишаються "сліпими" та "глухими" в широкому розумінні цих слів.

Знаменитий Білл Гейтс, керівник фірми Microsoft, вважає, що майбутнє за інтелектуальними інформаційними технологіями та системами, які сприймають і розуміють людську мову, друковані та рукописні

тексти, обробляють і розуміють різноманітні зображення, креслення тощо, сприймають просторові та звукові сцени. Розробкою цих технологій Гейтс збирається заволодіти світовим інформаційним простором та отримати надприбутки.

З часу створення першої ЕОМ "МЭСМ" в Україні сформувалась та утвердилась наукова школа світового рівня з проблем автоматичного розпізнавання сигналів і зображень та створення інтелектуальних інформаційних технологій і систем на їх основі.

Складність проблеми створення ОК зумовлена багатьма чинниками. Чи не найголовнішим з-поміж них є той, що розумова діяльність людини, підсвідомі дії всього живого, що пов'язані з розпізнаванням образів, є "за сімома замками", тобто прихованими від нас Природою. Часто цю діяльність та дії називають підсвідомими або інтуїтивними, а, отже, такими, що погано піддаються формалізації, дослідженням точними науками.

Українські науковці пішли шляхом функційного моделювання підсвідомої діяльності людини та всього живого, ґрунтуючись на загальновизнаних принципах оброблення інформації в живій природі, техніці та суспільстві, таких як: аналіз через синтез, індуктивне та дедуктивне виведення, генерація та направлений перебір варіантів, зворотній зв'язок, використання апріорної інформації, адаптація, навчання та самонавчання, розпізнавання образів, побудова моделей зовнішнього світу та розумової діяльності. Проблема полягає в тому, як конструктивно об'єднати та реалізувати ці принципи, щоб створити інтелектуальні інформаційні технології та системи.

Такими конструктивними моделями стали названі нами генеративні моделі розпізнавання сигналів і зображень. Відомо, що кожне вимовлення одного й того ж слова чи написання однієї й тієї ж літери навіть однією й тією ж людиною, скажімо, з інтервалом в одну секунду часу, завжди передаються різними ("двічі в одну й ту ж воду ввійти неможливо"), але чимось схожими сигналами або зображеннями.

Отже, універсальним алгоритмом розпізнавання міг би бути такий. Спочатку запам'ятати всі можливі сигнали чи зображення, які передають один і той самий образ (клас, слово, букву тощо), а потім при автоматичному розпізнаванні порівнювати пред'явлений для оброблення сигнал чи зображення з усіма раніше запам'ятованими сигналами чи зображеннями й віднести розпізнаваний сигнал (зображення) до того образу (класу), з чим раніше запам'ятованим сигналом (зображенням) він збігся. На жаль, такий алгоритм розпізнавання хоч і є "сильним", але він не є конструктивним: не знайдеться такий звичайний комп'ютер, ні тепер, ні в майбутньому, який був би в змозі запам'ятати всі можливі, розмаїті сигнали (зображення) та виконувати в реальному часі необхідні порівняння.

Труднощі, що виникають при реалізації цього універсального алгоритму розпізнавання можна

подолати, якщо спробувати навчитись запам'ятовувати, структурувати всі можливі сигнали (зображення) якимсь економним способом (адже вони пов'язані сильними залежностями, бо є схожими!), а порівняння сигналів (зображень) виконувати направленим перебором варіантів.

Відомо, що за допомогою диференційних рівнянь або рівнянь з частинними похідними можна генерувати величезну кількість близьких, "схожих" розв'язків. Для економного ж опису сигналів і зображень з подальшим їх розпізнаванням були розроблені ієрархічно організовані одно- та двовимірні стохастичні породжувальні граматики (генеративні моделі), які використані в якості зворотного зв'язку для направленої перебору та порівняння сигналів і зображень, що стало ефективною конструктивною реалізацією ідеї аналізу сигналів (зображень) через їх синтез.

Українськими науковцями розроблені принципово нові, інтелектуальні, системи, теорія навчання та самонавчання розпізнаванню образів, принципи розумної поведінки кібернетичних пристроїв, принципи планування цілеспрямованої поведінки роботів у складних умовах, теорія і методи автоматичного розпізнавання, розуміння та синтезу мовних сигналів, математична теорія розпізнавання зображень та теорія комп'ютерного стереозору, методи розпізнавання об'ємних тіл та просторових сцен і побудови моделі навколишнього середовища, теорія кластеризації спостережень та індуктивного виведення, принципи самоорганізації нейронних мереж, принципи побудови людино-машинних інформаційних технологій оброблення зображень.

На базі отриманих фундаментальних результатів побудовані та впроваджені інтелектуальні інформаційні технології та системи для оброблення інформації в таких ділянках: розпізнавання, аналіз, синтез та розуміння усної мови; розпізнавання зображень, зокрема креслень, картографічних, текстурних, стерео-скопичних зображень; аерофотограмметричний аналіз земної поверхні для відновлення рельєфу місцевості за стереопарами аерофотознімків; аналіз та розуміння сцен для керування автономними роботами з автоматичним плануванням дій та прийняттям рішень; адаптивні системи прогнозування складних процесів в природі та суспільстві.

В Україні були створені перші системи автоматичного розпізнавання окремо вимовлюваних слів, системи типу "око-рука", системи автоматичного розпізнавання, розуміння та синтезу усної мови та системи усного діалогу, системи автоматичного розпізнавання креслярсько-графічних зображень, системи технічного зору для контролю якості виробів та керування виконавчими механізмами, перші в світі експериментальна система автоматичного розпізнавання зв'язного мовлення та багатомовна система усного діалогу для ЮНЕСКО, що оперувала сімома мовами та виконувала усний переклад з однієї мови на іншу. За даними японської фірми WACOM система оброблення креслярсько-гра-

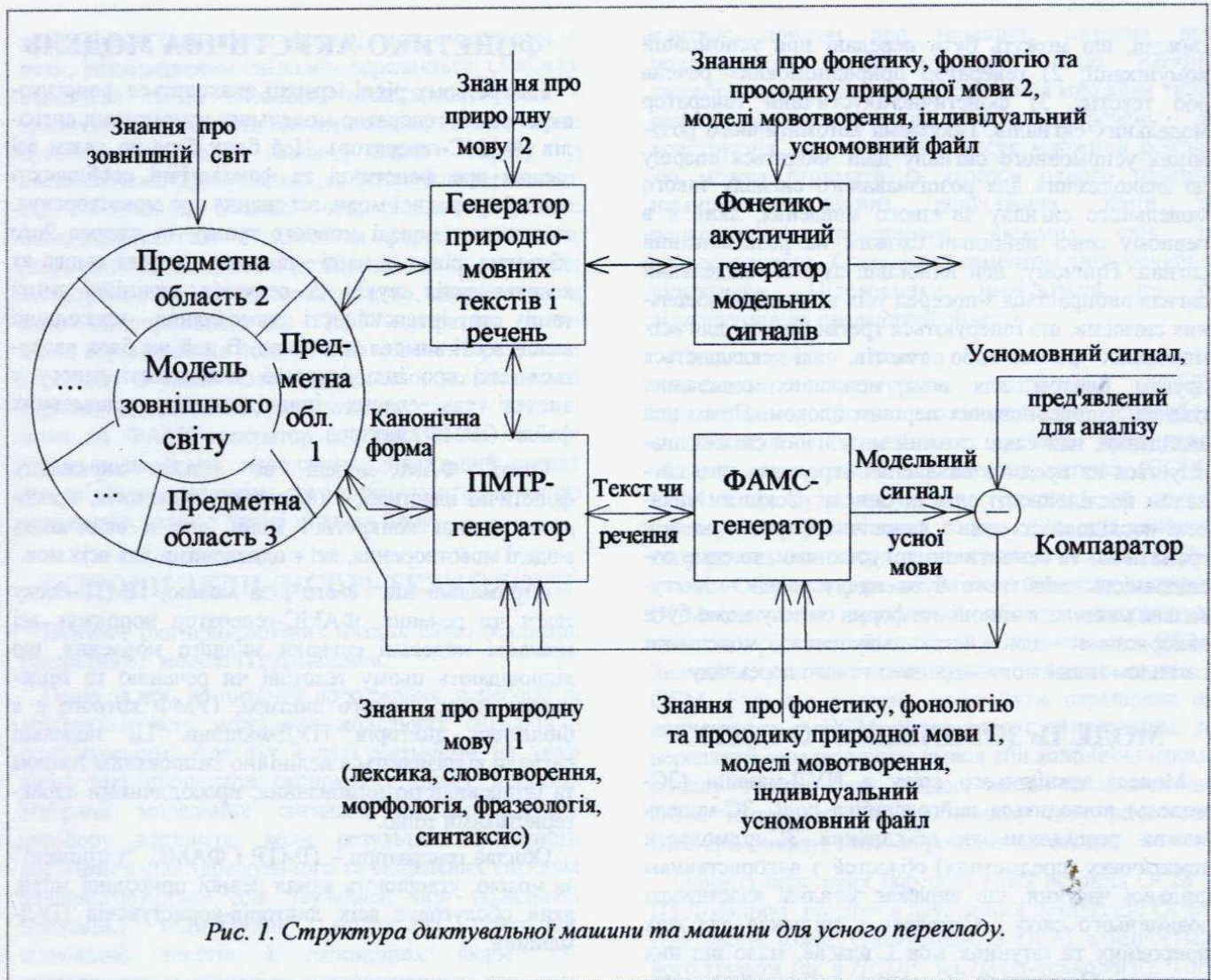


Рис. 1. Структура диктувальної машини та машини для усного перекладу.

фічних документів є найефективнішою з-посеред систем такого ж призначення. Розроблено програмне забезпечення першої вітчизняної, конкурентноздатної на зовнішньому ринку, цифрової фотограмметричної станції "Дельта".

В Україні сформувалася визнана у світі наукова школа, що обіймає широке коло наукових та прикладних проблем створення інтелектуальних інформаційних технологій та систем. Ці роботи є загально-визнаним у світі внеском у становлення та розвиток автоматичного розпізнавання образів (сигналів і зображень) як розділу науки. Українські роботи широко цитуються, визнані піонерними та такими, що мають значний вплив на світовий рівень досліджень у цій галузі і зміцнюють авторитет української науки.

Науковці України беруть участь у виконанні міжнародних контрактів з автоматичного розпізнавання сигналів і зображень, запрошуються до читання лекцій в західних університетах. На жаль, частина вчених виїхала на наукову роботу за кордон.

До сьогодні розробки з автоматичного розпізнавання, розуміння та синтезу образів велись розрізнено: кожна група науковців зосереджувалась на окремих вузьких проблемах. Тепер же на часі є об'єднання науковців навколо розроблення методології, концеп-

ції та архітектури ОК, навколо розроблення образних операційних систем та систем програмування ОК, окремих підсистем ОК, створення прототипів ОК.

ПІДХОДИ ДО СТВОРЕННЯ ОК

Розглянемо два підходи до створення образного комп'ютера на прикладі Диктувальної машини та машини Усного Перекладу (ПУД-машини). А потім покажемо, як моделі ПУД-машини можуть бути доповнені каналами сприйняття іншої, неусномовної, інформації та перетворені в прототипи ОК.

Перший підхід ґрунтується на так званій генеративній моделі автоматичного розпізнавання, розуміння та синтезу мовних сигналів.

Другий підхід використовує багатозначну багаторівневу модель.

ГЕНЕРАТИВНА МОДЕЛЬ ПУД-МАШИНИ

Узагальнена структура ПУД-машини, що ґрунтується на генеративній моделі, подана на Рис. 1.

Трьома найголовнішими блоками ПУД-машини є: 1) модель зовнішнього світу, яка описує всі можливі

сенси, що можуть бути передані при усномовній комунікації; 2) генератор природномовних речень або текстів; 3) фонетично-акустичний генератор модельних сигналів. Проблема автоматичного розуміння усномовного сигналу далі зводиться спершу до знаходження для розпізнаваного сигналу такого модельного сигналу зв'язного мовлення, який є в певному сенсі найбільш схожим на розпізнаваний сигнал. Причому, цей найбільш схожий модельний сигнал вибирається з-посеред усіх можливих модельних сигналів, що генеруються третім блоком для всіх можливих речень або текстів, які складаються другим блоком для всіх можливих смислових гіпотез, запропонованих першим блоком. Потім цей знайдений, найбільш схожий модельний сигнал аналізується на предмет виявлення переданих цим сигналом послідовності слів та смислу. Оскільки виявлені послідовність слів і канонічна форма смислу є граматично та семантично допустимими, то сама послідовність слів може бути надрукована – диктувальна машина, а канонічна форма смислу може бути використана для його озвучення модельним сигналом іншої мови – машина усного перекладу.

МОДЕЛЬ ЗОВНІШНЬОГО СВІТУ

Моделі зовнішнього світу в ПУД-машині (ЗС-модель) відводиться найголовніша роль. ЗС-модель можна розглядати як об'єднання ЗС-підмоделей тематичних (предметних) областей з використанням спільної частини, що виражає загальні властивості зовнішнього світу. ЗС-модель є спільною для всіх природних та штучних мов і, власне, мало від них залежить. Описується ЗС-модель спеціальною математичною мовою, наприклад мовою канонічних форм. Опускаючи тут подробиці задання ЗС-моделі, підкреслимо лише, що ЗС-модель генерує канонічні форми смислів, які можуть передаватись в процесі усномовної комунікації. Очевидно, що для кожної предметної області може бути вказана скінченна множина канонічних форм, можливих при діалозі. Наприклад, якщо йдеться про усномовний калькулятор на чотири арифметичні дії, то в цьому випадку допустимі канонічні форми мають просту структуру: операція та перший і другий операнди.

МОДЕЛЬ ПРИРОДНОЇ МОВИ

Наступною, другою в ієрархії частиною ПУД-машини є генератор природномовних речень або текстів (ПМТР-генератор). Цей блок ґрунтується на знаннях про лексику, словотворення, морфологію, фразеологію та синтаксис кожної конкретної природної мови. Отримуючи від ЗС-моделі канонічну форму передаваного смислу, ПМТР-генератор породжує всі допустимі для даної мови речення і тексти, котрі виражають один і той же смисл, що визначений прийнятою канонічною формою. Можна сказати, що ПМТР-генератор є семантико-граматичною текстовою моделлю конкретної природної мови.

ФОНЕТИКО-АКУСТИЧНА МОДЕЛЬ

На третьому рівні ієрархії знаходиться фонетико-акустичний генератор модельних усномовних сигналів (ФАМС-генератор). Цей блок бере до уваги всі знання про фонетичні та фонологічні особливості певної природної мови, всі знання про мовотворення, включаючи моделі мовного тракту та джерел його збурення, рівно ж наші уявлення про такі явища як коартикуляція звуків, їх редукція, нелінійні зміни темпу та інтенсивності вимовляння, просодичні властивості вимовляння тощо. В цей же блок вводяться дані про індивідуальні особливості голосу у вигляді так званих індивідуальних усномовних файлів (ІУМФ) диктора.

Отже, ФАМС-моделі не тільки виражають фонетичні властивості (фонетика, фонологія, просодика) певної конкретної мови, але й включають моделі мовотворення, які є однаковими для всіх мов.

Отримавши від "свого", за мовою, ПМТР-блоку текст чи речення, ФАМС-генератор породжує всі можливі модельні сигнали зв'язного мовлення, що відповідають цьому текстові чи реченню та моделюють голос певного диктора, ІУМФ котрого є в бібліотеці дикторів ПУД-машини. Ці модельні сигнали відрізняються нелінійно змінюваним темпом та інтенсивністю вимовляння, просодичними характеристиками тощо.

Обидва генератори – ПМТР і ФАМС, "узгоджені" за мовою, утворюють канал певної природної мови, який обслуговує всіх дикторів-користувачів ПУД-машини.

ІДЕЯ ВИРІШЕННЯ ПРОБЛЕМИ

Далі в рамках генеративної моделі автоматичного розпізнавання, розуміння та синтезу мовлення проблема автоматичного усного перекладу з однієї мови на іншу або проблема автоматичного редагування та друкування під диктування формулюється та вирішується так.

Одночасно з подачею пред'явленого до аналізу усномовного сигналу за допомогою генеративної моделі породжуються всі можливі модельні сигнали, які порівнюються з розпізнаваним сигналом. Ці модельні сигнали, при всьому їх розмаїтті, відповідають певним текстам та реченням, які, в свою чергу, як це впливає з опису ПУД-машини, виражають певний смисл, що заданий канонічною формою. Перебираючи всі можливі сигнали зв'язного мовлення, а, отже, всі допустимі тексти чи речення та всі можливі канонічні форми передаваних смислів, знаходимо такий модельний сигнал, який з-посеред усіх можливих модельних сигналів є найбільш схожим, у певному розумінні, на розпізнаваний сигнал.

Далі, аналізуючи цей найбільш схожий модельний сигнал, визначаємо, якій послідовності слів, тобто якому реченню або тексту, він відповідає, а заодно,

який смисл (канонічна форма) цим модельним і, отже, розпізнаваним сигналом передається. Оскільки визначена таким способом послідовність слів є граматично та семантично коректною, то вона може бути надрукована, якщо маємо справу з автоматичним друкуванням та редагуванням текстів під диктування. Більше того, оскільки знайдені послідовність слів і канонічна форма виражають результат автоматичного розпізнавання та смислової інтерпретації зв'язного мовлення, то, звертаючись з цією інтерпретацією до ПМТР-генератора іншої мови, отримаємо на його виході текстовий результат перекладу з усної мови. Аналогічно, подавши цей отриманий текстовий результат перекладу з усної мови на ФАМС-генератор для тієї мови, на яку робиться переклад, породжуємо модельний сигнал для зв'язного мовлення, що і є результатом повного усного перекладу з однієї мови на іншу.

ОСНОВИ ІКДП (НСДР)-ТЕХНОЛОГІЇ

Виникає ряд принципових питань щодо реалізації генеративної моделі ПУД-машини.

Перш за все, це питання породження, перебору та порівняння всіх можливих модельних сигналів з розпізнаванням. Але тут і далі матимемо на увазі лише такі процедури економного опису процесів генерації модельних сигналів та направленою перебору варіантів, коли результати поточного порівняння спостережуваного та модельних сигналів використовуються для звуження, при подальшій генерації, підмножини модельних сигналів, відповідно текстів і канонічних форм, що претендують у відповідь розпізнавання, так, щоб гарантувати не втратити оптимальне рішення. Саме ці згадані дві суперечливі вимоги: економна специфікація породжувальних процедур та направлений перебір варіантів, – задовольняються ІКДП(НСДР)-технологією автоматичного розпізнавання, розуміння та синтезу усномовних сигналів [1-6]. Перша вимога задовольняється використанням ієрархічно (I (H – *Hierarchy*)) структурованих автоматних породжувальних граматики, які синтезують (композиція (K (C – *Composition*))) модельні сигнали, а друга – заснована на динамічному програмуванні (ДП (DP – *Dynamic Programming*)).

На Рис. 1 зображена не тільки ієрархія ПМТР- і ФАМС-генераторів, але й ідея направленою перебору варіантів – зворотніми зв'язками, що йдуть від каналного компаратора до всіх рівнів (блоків) ПУД-машини.

ПРИКЛАД РЕАЛІЗАЦІЇ

Розіб'ємо всі речення предметної області за передаваним смислом на категорії або типи смислу. Так, для інформаційно-довідникової служби аеропорту можна виділити такі категорії смислу: питання, які стосуються прибуття літаків; питання стосовно відльоту літаків; довідки про наявність

квитків; довідки про маршрут; питання про розташування служб аеропорту тощо. Кожній категорії смислу відповідає скінченна множина типів речень, які передають цей смисл. Тип речень – це конструкція, що економно описує множини речень, які можна отримати із якогось одного речення шляхом незалежних допустимих замінів та допустимих перестановок окремих слів та словосполучень. Основним елементом типу речень є підсловник. Підсловники іменуються по їх відношенню до предметної області.

Типи речень можна економно задавати списковими структурами, наприклад, *LISP*-структурами. Іншим еквівалентним способом специфікації всіх можливих речень предметної області є орієнтована семантична мережа (ОСМ). ОСМ може бути побудована, зокрема, на основі категорій смислу та типів речень. Так, для кожного типу речень спочатку будується своя часткова ОСМ, а потім ОСМ для окремих категорій смислу та повна ОСМ для всієї предметної області утворюється шляхом простого об'єднання всіх часткових ОСМ. Звичайно, важливо мінімізувати кількість станів в ОСМ. Будь-яке речення може бути перевірене на допустимість в ОСМ. Якщо воно є допустимим, то переданий цим реченням смисл або канонічна форма смислу можуть бути знайдені також за допомогою ОСМ, наприклад, шляхом аналізу типу речень та категорії смислу, яким це речення належить.

Один із можливих способів використання ОСМ в ПУД-машині полягає в акустичній деталізації ОСМ. Для цього кожне слово-місце в ОСМ заміщується його графом (як в *НСДР*- чи *НММ*-технології). Тоді цей граф генеруватиме всі можливі сигнали цього слова в контексті злитого мовлення, а сама ОСМ стає усномовною, тобто такою, яка генерує модельні усномовні сигнали для вибраної предметної області. Тепер вирішення проблеми ПУД-машини для цієї предметної області полягає у знаходженні найбільш схожого модельного сигналу та у його аналізі на предмет виявлення послідовності слів та канонічної форми смислу, переданого пред'явленим усномовним сигналом [1-3].

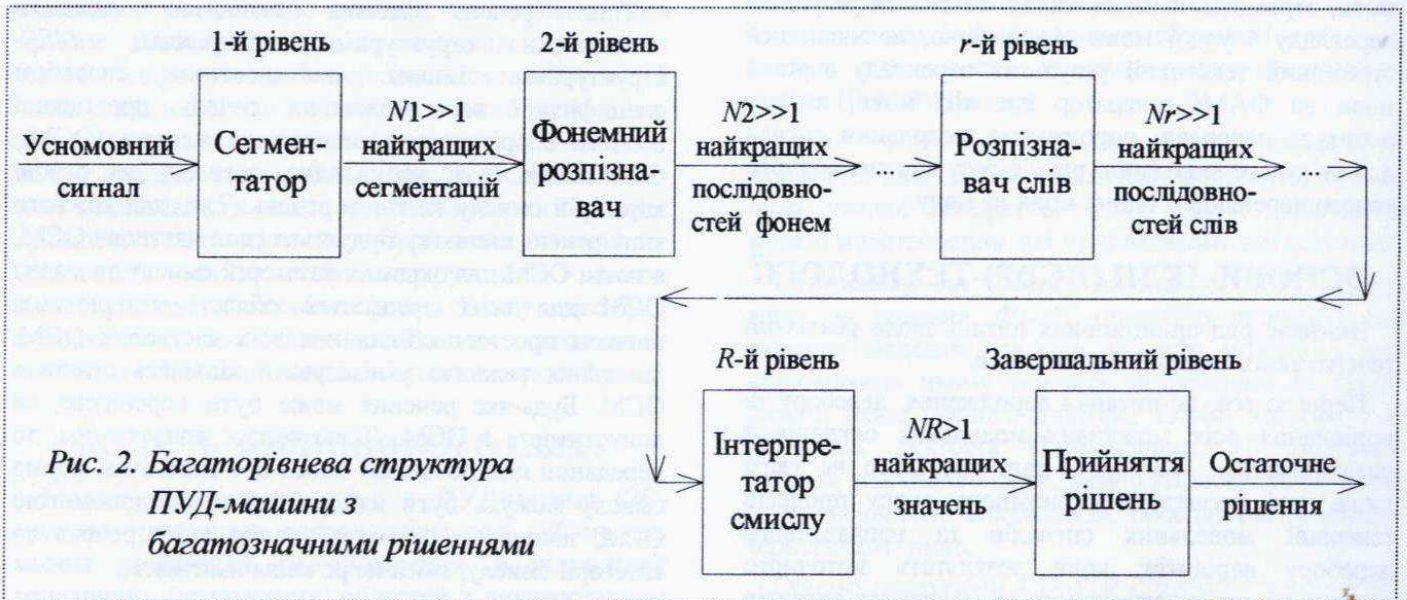
ГЕНЕРАТИВНА МОДЕЛЬ ОК

Доповнивши генеративну модель ПУД-машини (див. Рис. 1), зокрема, приєднавши до моделі зовнішнього світу канал генерації зображень та просторових сцен, що складається з композитора зображень і сцен та генератора модельних зображень і сцен, отримаємо модель образного комп'ютера, який не тільки є диктувальною машиною та машиною усного перекладу, але й сприймає зображення та аналізує просторові сцени, дає їм усномовну інтерпретацію та робить їх текстовий опис, будує зображення за усномовним повідомленням.

БАГАТОРІВНЕВА БАГАТОЗНАЧНА МОДЕЛЬ

Другий підхід до створення ПУД-машини – це багаторівнева багатозначна модель. Вона також є ієрархічно організованою. Тут використовуються багатозначні рішення на всіх рівнях оброблення усномовного сигналу. Наприклад, на першому, нижньому, рівні знаходять $N1 \gg 1$ найкращих розбивок усномовного сигналу на сегменти, що

елементам. Далі, на другому рівні, ця багатозначна сегментація відображається в $N2 \gg 1$ найкращих композицій з більш крупних елементів (літер, окремих деталей, дрібних предметів, елементів фону тощо). На третьому рівні ієрархії знаходить $N3 \gg 1$ найкращих сегментацій зображення, що відповідають можливим композиціям з більш крупних елементів (надписів, предметів, людей). На вищих рівнях ієрархії виконується опис та смисловий аналіз зображення і сцени в цілому.



відповідають фонемам. Далі, на другому рівні, ця багатозначна сегментація в умовах, наприклад, вільного порядку слідування фонем, відображається в $N2 \gg 1$ найкращих послідовностей фонем. Отже, на другому рівні ієрархії маємо справу з багатозначним фонемним розпізнавачем. На третьому рівні ієрархії розташований розпізнавач слів, який в умовах вільного порядку слідування слів знаходить $N3 \gg 1$ найкращих послідовностей слів. Це – так званий узагальнений розпізнавач послідовностей слів. І так далі. На наступних рівнях ієрархії виконується багатозначна синтаксична та семантична фільтрація послідовностей слів. Нарешті, на найвищому рівні ієрархії знаходять $Nr \gg 1$ найкращих результатів смислової інтерпретації усномовного сигналу та приймають остаточне рішення.

В багаторівневій багатозначній моделі зворотний зв'язок явно не показаний. В ній замість повного перебору варіантів використовується “обережна” стратегія прийняття багатозначних рішень, яка має на меті не “загубити” і врешті-решт віднайти найкраще рішення.

Структура багаторівневої багатозначної ПУД-машини показана на Рис. 2.

Стосовно багаторівневого багатозначного оброблення зображень і сцен – то діємо аналогічно. Наприклад, на першому, нижньому, рівні знаходимо $N1 \gg 1$ найкращих розбивок зображення на сегменти, що відповідають дугам, відрізкам, дрібним

ПРИКІНЦЕВІ ПОЛОЖЕННЯ

Щоб взяти участь в розробках за Державною науково-технічною програмою “Образний комп'ютер”, необхідно виграти тендерні змагання (конкурс запитів та пропозицій). Ці змагання проводитиме Міжнародний науково-навчальний центр ЮНЕСКО інформаційних технологій та систем, на який покладено функції органу управління програмою.

ЛІТЕРАТУРА

- [1] Т.К. Vintsiuk, “Element-Wise Recognition of Continuous Speech Composed of Words from Specified Vocabulary”, *Kibernetika*, 1971, No. 2, pp 133-143.
- [2] Т.К. Vintsiuk, “Speech Recognition and Understanding”, *Kibernetika*, 1982, No 5, pp 101-111.
- [3] Т.К. Винцюк. “Анализ, распознавание и интерпретация речевых сигналов”, Киев: “Наукова думка”, 1987, 264 с.
- [4] Т.К. Vintsiuk, “HCDP-Technique for Automatic Analysis, Recognition and Understanding of Speech Signals”, *Proc. First Intern. Conf. on Information Technology for Image Analysis and Pattern Recognition*, L'viv, 1990, Vol 1, pp 108 - 112.
- [5] Taras K. Vintsiuk, “Two Approaches to Create a Dictation/ Translation Machine”, *Proc. Second Intern. Workshop “Speech and Computer”*, Cluj-Napoca, 1997, pp 1-6.
- [6] Тарас Винцюк, “Образний комп'ютер”, Зб. наук. праць “Сучасні проблеми в комп'ютерних науках”, Вид-во Нац. ун-ту “Львівська політехніка”, Львів, 2000, сс 5-14