

ГЕНЕРАТИВНА МОДЕЛЬ ОБРАЗНОГО КОМП'ЮТЕРА¹

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 44 266-4356
vintsiuk@uasoiro.org.ua

ABSTRACT

Taras K. Vintsiuk. Generative Model for Pattern Computer. 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.

Here a so-called generative model for how to create a pattern computer is debated. This model is based on both the generative grammar hierarchy for multimodal prototype scene composition and the comparison of them with a scene to be perceived. So a multimodal information synthesis is used as a feedback in the pattern analysis and understanding. Examples are given in reference to dictation and spoken translation machine creating and scene analysis.

ВСТУП

Людство потребує створення принципово нових комп'ютерів, які здатні сприймати та розуміти звуки, зображення, людську мову, рукописні тексти, креслення, 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".

Програма має на меті створення принципово нових інформаційних технологій та систем — образних комп'ютерів.

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.

Отже, змістовно визначені два базових поняття: образний комп'ютер (ОК) та образна операційна система (ООС).

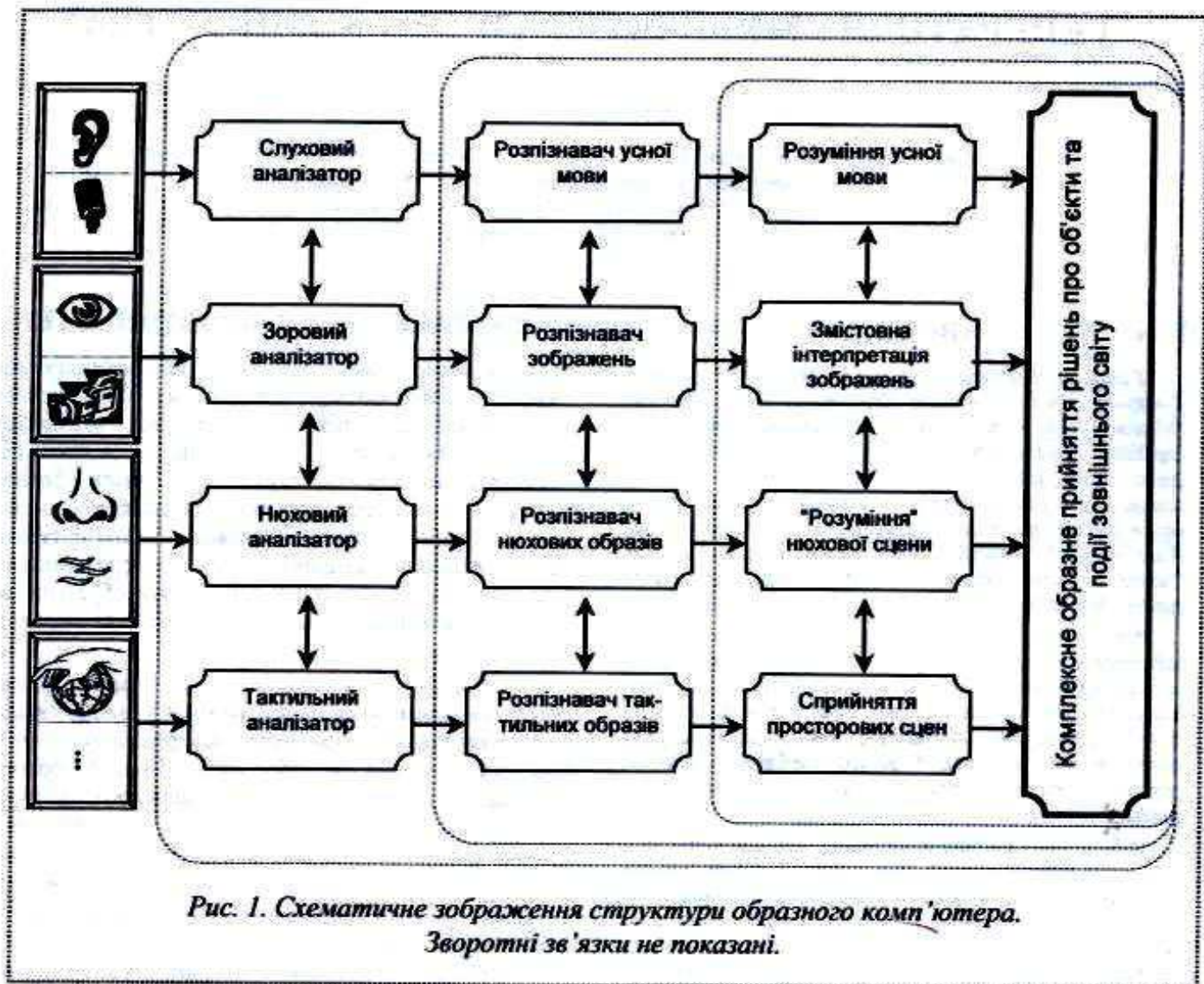
Підкреслимо, що на відміну від звичайного комп'ютера у центрі образного комп'ютера знаходяться моделі зовнішнього світу з усіма його об'єктами, явищами та їх проявами.

СТРУКТУРА ОБРАЗНОГО КОМП'ЮТЕРА

На рис. 1 подано схематичне зображення ОК. Наведені чотири канали сприйняття інформації: слуховий, зоровий, нюховий та тактильний. В кожному каналі виділено персептор, аналізатор, розпізнавач та інтерпретатор.

Пунктирними лініями зображається ієрархія рівнів оброблення інформації. Найнижчий рівень

¹ Робота виконана в рамках Контракту № ОК_2002_1_МПШ_ЦЕНТР за ДНТП "Образний комп'ютер"



стосується аналізу, найвищий — змістовної інтерпретації та розуміння.

Образна операційна система виконує комплексну, узгоджену за всіма каналами, інтерпретацію та образне прийняття рішень про об'єкти та події зовнішнього світу. Ієрархії в ООС відображено вкладеннями пунктирними фігурами.

Взаємодії між каналами показані вертикальними стрілками. Зворотні ж зв'язки в каналах та між каналами не відмічені.

КОНЦЕПЦІЯ ГЕНЕРАТИВНОЇ МОДЕЛІ

Генеративна модель образного комп'ютера ґрунтується на конструктивній реалізації загальноновизнаних принципів 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.

Відомо, що кожне вимовляння 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 образів 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, різноманітні сигнали (зображення) та виконувати в реальному часі необхідні порівняння.

Труднощі, що виникають при реалізації цього універсального алгоритму розпізнавання в

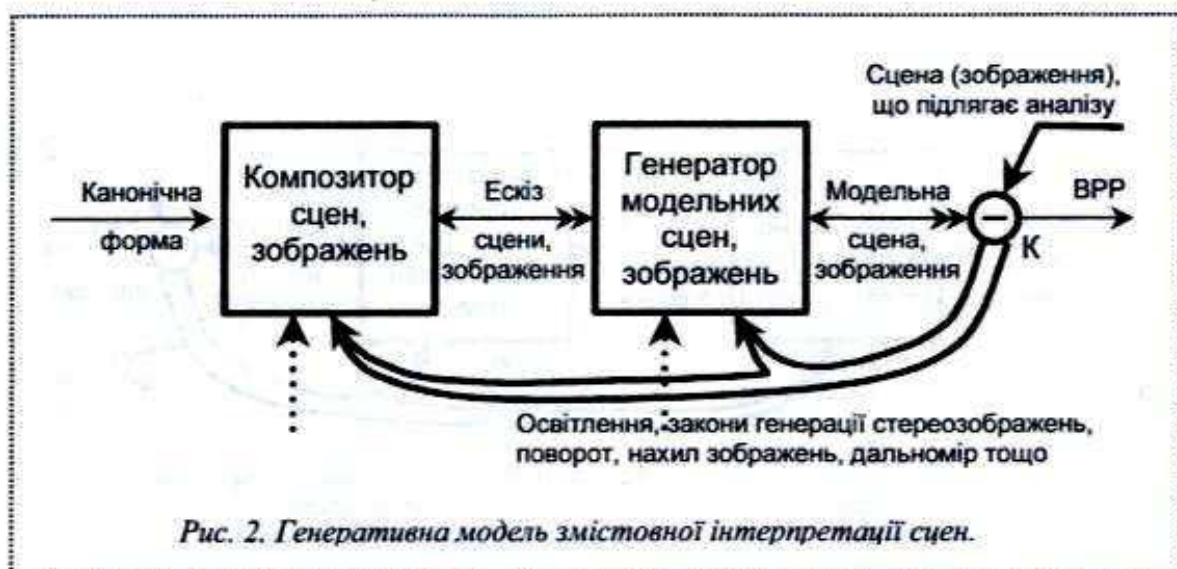


Рис. 2. Генеративна модель змістовної інтерпретації сцен.

генеративній моделі можуть долатися таким чином: 1) треба навчитися запам'ятовувати, структурувати всі можливі сигнали (зображення) якимсь економним способом (адже вони пов'язані сильними залежностями, бо є схожими!); 2) а порівняння сигналів (зображень) треба виконувати направленим перебором варіантів.

СПЕЦИФІКАЦІЯ ОБРАЗІВ

Відомо, що за допомогою диференціальних чи різницевих рівнянь або рівнянь з частинними похідними чи приростами можна генерувати величезну кількість близьких, "схожих" розв'язків. Для економного ж опису сигналів і зображень з подальшим їх розпізнаванням були запропоновані 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 [1—5].

ГЕНЕРАТИВНІ МОДЕЛІ ДЛЯ ОКРЕМИХ КАНАЛІВ ОК

На рис. 2 схематично подана трирівнева генеративна модель розуміння зорових сцен (зображень). Модель зовнішнього світу (ЗС-модель, див. далі) породжує (генерує) канонічну форму сцени. "Композитор" (генератор) сцен (ГС) за канонічною формою синтезує ескіз сцени. Генератор модельних сцен і зображень (ГМСЗ) за ескізом сцени, моделюючи освітлення, розміри об'єктів, їх взаємне розташування, повороти, нахил, проектування на площину тощо, породжує модельну сцену або зображення.

Компаратор (К) порівнює модельну сцену з пред'явленою для розпізнавання. Результат порівняння використовується як зворотній зв'язок,

під впливом якого направленим перебором відсікаються неперспективні варіанти та відшукується найкраща модельна сцена, яка й аналізується і на підставі якої формується результат розуміння сцени.

На рис. 3 схематично подана трирівнева генеративна модель розуміння мовного сигналу. За канонічною формою передаваного смислу, що надходить з ЗС-моделі, генератор семантично еквівалентних речень (ГСЕР) породжує всі можливі речення, що передають один й той самий смисл, визначений вхідною канонічною формою. Генератор модельних сигналів зв'язного мовлення (ГМСЗМ) ставить у відповідність прийнятому орфографічному текстові всі можливі модельні сигнали злиглого мовлення, які відображають розмаїті мовні сигнали, що відрізняються темпом, інтенсивністю, інтонацією, моделюють індивідуальні особливості мовлення тощо.

В компараторі (К) модельні сигнали порівнюються з мовним сигналом, що пред'являється для аналізу. Результат поточного порівняння використовується у зворотньому зв'язку для направленої відбору та пошуку найкращого модельного сигналу. Останній аналізується, тобто вказується, якому реченню (орфографічному текстові) він відповідає (текст можна надрукувати — автоматична машинка, що друкує під диктування) і/або отримати відповідну цьому текстові канонічну форму, що передається (автоматичне розуміння мовного сигналу — цей результат можна використати для виконання дії, щоб задовольнити інтереси людини, яка говорить).

Рис. 4 ілюструє функціонування генеративних моделей ОК, наведених на рис. 2 та рис. 3.

Відображено, що кожній предметній області відповідає скінченна множина канонічних форм смислів або сцен, що передаються; що кожній канонічній формі відповідає своя скінченна множина всіх можливих речень або ескізів сцен; що кожному реченню або ескізові сцени відповідає своя скінченна



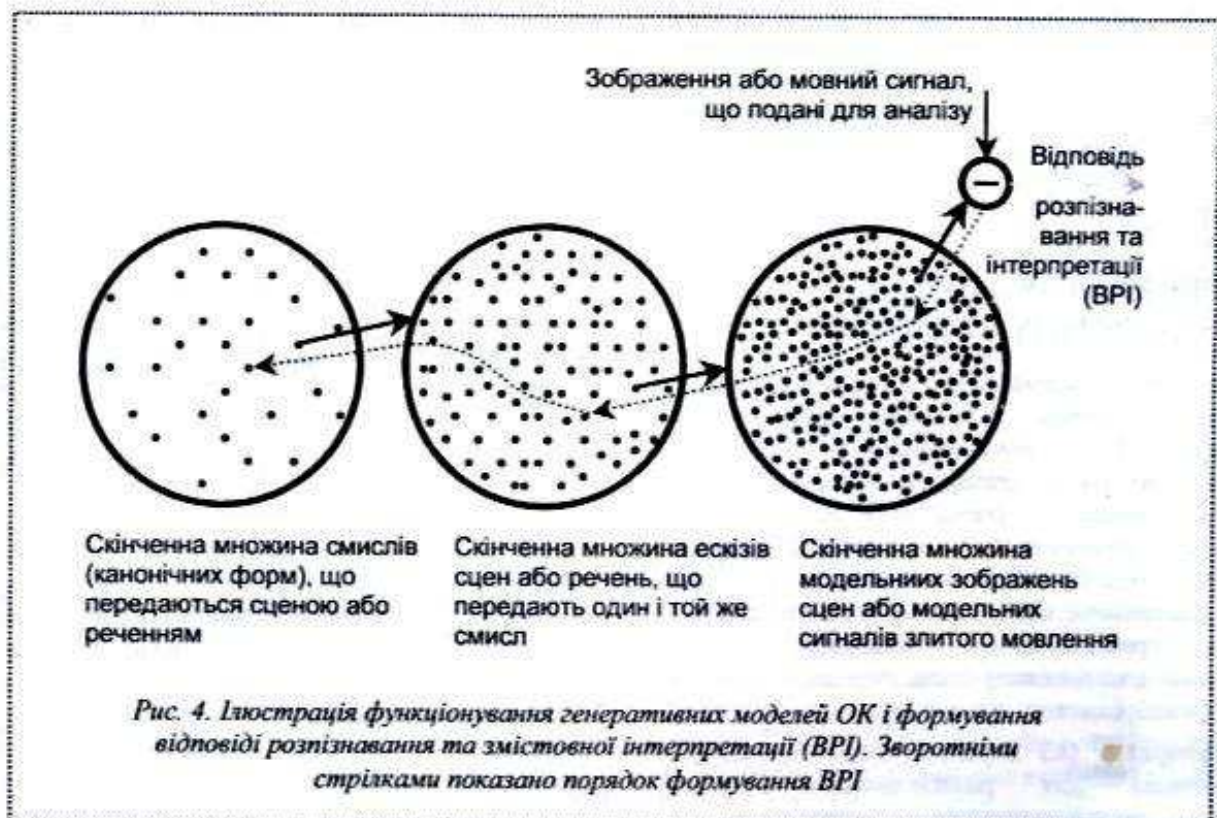
множина всіх можливих модельних сигналів або зображень. На цьому ж рис. 4 пунктирними стрілками також показано як формуються відповіді розпізнавання та інтерпретації (BPI).

ГЕНЕРАТИВНА МОДЕЛЬ ПУД-МАШИНИ

Узагальнена структура диктувальної (Д) машини та машини усного (У) перекладу (П) — ПУД-машини, що ґрунтується на генеративній моделі, подана на рис. 5. Генеративна модель ПУД-

машини утворюється об'єднанням генеративних моделей розуміння мовного сигналу для окремих мов. На рис. 5 зображені дві мови: канал природної мови 1 та канал природної мови 2.

Диктувальна машина. Проблема автоматичного редагування та друкування текстів під диктування на природній мові, наприклад 1, вирішується так. Для усномовного сигналу, що пред'явлений для оброблення, спершу знаходимо (шляхом направленої синтезу та відбору) такий модельний сигнал зв'язного мовлення, який є в певному сенсі найбільш схожим на розпізнаваний



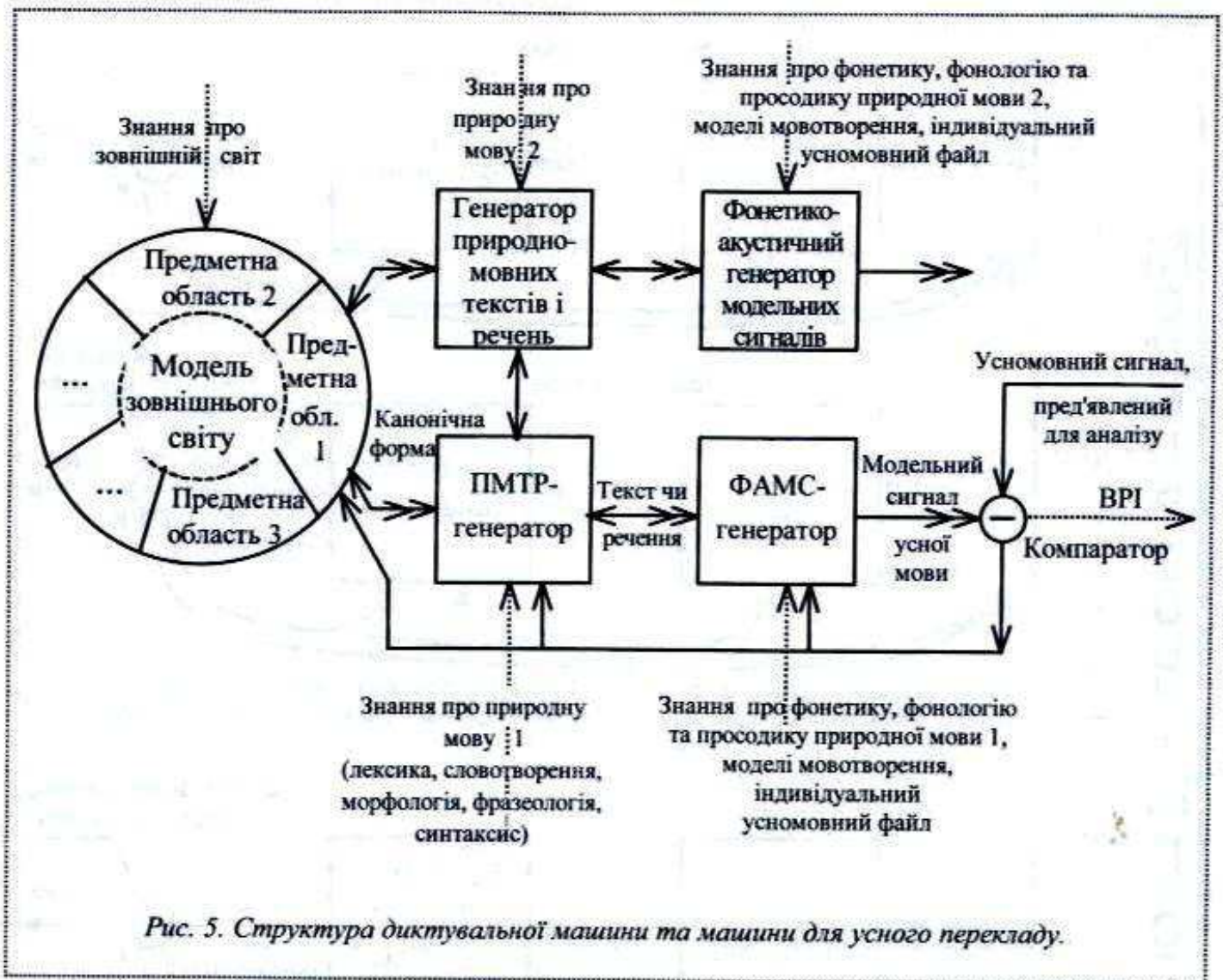


Рис. 5. Структура диктувальної машини та машини для усного перекладу.

сигнал. 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. Це і буде диктувальна машина.

Машинна усного перекладу. Якщо ж скористатись канонічною формою смислу, що був переданий мовним сигналом на природній мові 1, і звернутись до ЗС-моделі та каналу природної мови 2, то цій канонічній формі може бути поставлений у відповідність текст на природній мові 2, з тим же смислом, що і на природній мові 1 (перекладально-диктувальна машина), а сам текст на природній мові 2 може бути озвучений модельним сигналом природної мови 2 – машина усного перекладу.

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 лінгвістичні 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.

ГЕНЕРАТИВНА МОДЕЛЬ ОБРАЗНОГО КОМП'ЮТЕРА

Доповнимо генеративну модель ПУД-машини (див. рис. 5), зокрема, приєднаємо до моделі зовнішнього світу канал генерації зображень та просторових сцен, що складається з композитора зображень і сцен та генератора модельних зображень

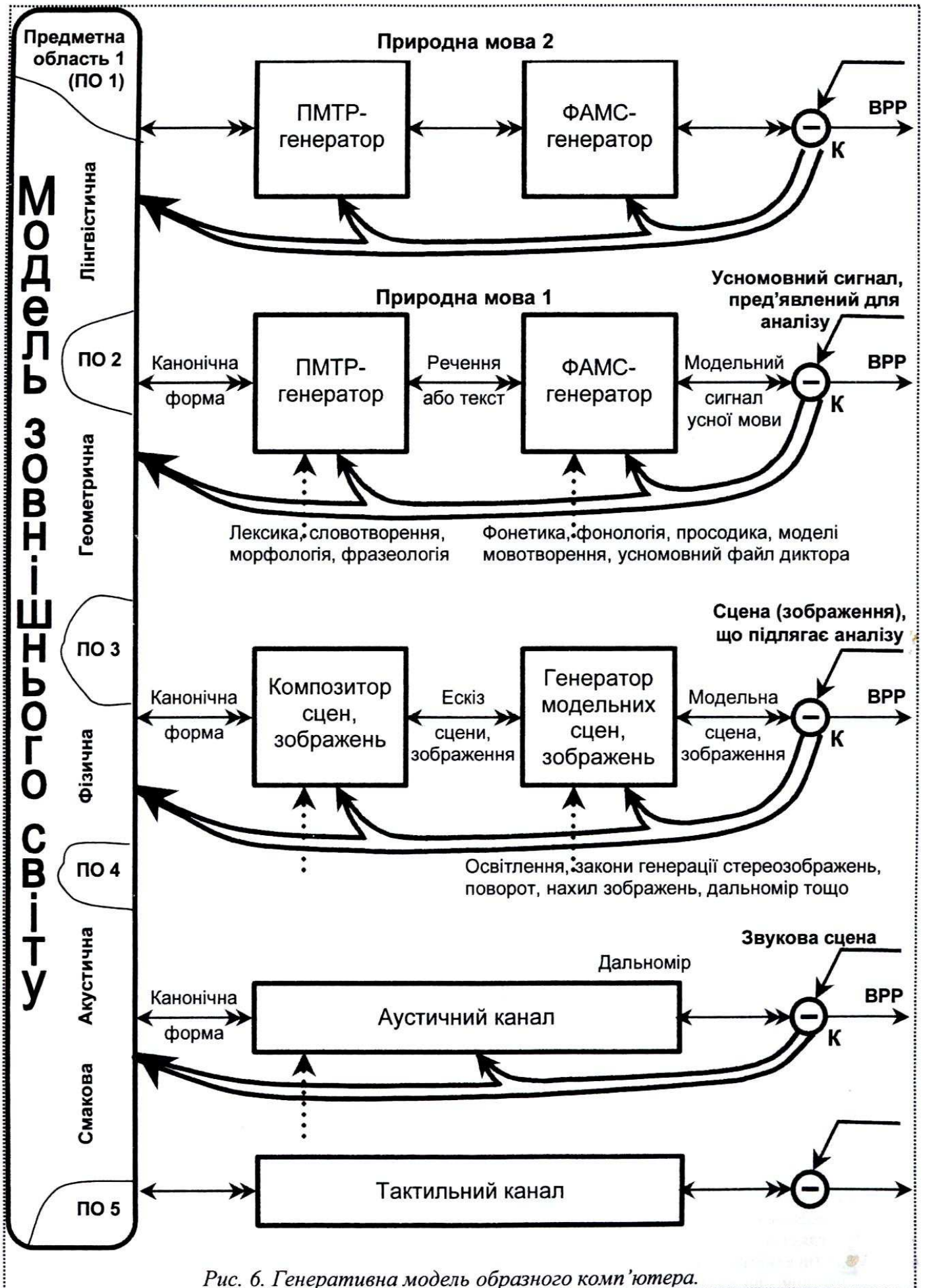


Рис. 6. Генеративна модель образного комп'ютера.

і сцен. Так само приєднаємо канал сприйняття звукових (неусномовних) сцен — акустичний канал, тактильний, нюховий та смаковий канали, які також організуємо за принципом генеративної моделі.

В результаті отримаємо генеративну модель образного комп'ютера, яка 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, озвучує тексти, сприймає акустичні образи та звукові сцени, аналізує тактильну, нюхову та смакову інформацію.

Генеративна модель ОК схематично зображена на рис. 6. Взаємодія між каналами сприйняття інформації (на рис. 6 ця взаємодія не показана) та образне прийняття рішень організується образним операційним середовищем (ООС), в яке власне і “погружається” генеративна модель ОК.

На рис. 6 образне операційне середовище зображається полем прямокутної рамки цього рисунку.

РЕАЛІЗАЦІЯ ГЕНЕРАТИВНОЇ МОДЕЛІ

Виникає ряд принципових питань щодо реалізації генеративної моделі образного комп'ютера.

Розглянемо ці питання на прикладі реалізації диктувальної машини та машини усного перекладу — ПУД-машини, що є чи не однією з найскладніших підсистем у складі ОК.

Якщо говорити про ПУД-машину як тільки автономну підсистему, що не взаємодіє з іншими каналами ОК, то досить обмежитись тільки такою моделлю зовнішнього світу або мовою канонічних форм передаваного смислу, які є зорієнтованими на лінгвістику — на передачу смислу мовами (орфографічними текстами або усномовними сигналами).

Для спрощення викладок, розіб'ємо за передаваним смислом всі можливі в мові речення на предметні області, а всі можливі речення предметної області — 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 зручно задавати списковими структурами, наприклад, *LISP*-структурами. Отже, типи смислів специфікуються списком (переліком) своїх *LISP*-структур.

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 утворюються 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.

Не менш важливе питання — 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 направлено 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 (*HCDP*)—технологією автоматичного розпізнавання, розуміння та синтезу усномовних сигналів [1–3, 5, 7–8]. Перша вимога задовольняється використанням ієрархічно ($I(H-Hierarchy)$) структурованих автоматних породжувальних графіків, які синтезують (композиція ($K(C-Composition)$)) модельні сигнали, а друга — заснована на динамічному програмуванні (ДП (*DP-Dynamic Programming*)).

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 *HCDP*- чи *HMM*[9]-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, що передаються пред'явленим усномовним сигналом [1–3, 5, 7–8].

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.

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 (ГУМФ) диктора [1–3, 5, 7–8].

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, ГУМФ котрого є в бібліотеці дикторів ПУД-машини. Ці модельні сигнали відрізняються нелінійно змінюваним темпом та інтенсивністю вимовляння, просодичними характеристиками тощо [1–3, 5, 7–8].

ПРИКІНЦЕВІ ПОЛОЖЕННЯ

Генеративна модель розпізнавання усномовних сигналів (інша назва – ІКДП-технологія) була запропонована автором у 1966 році. Перша публікація з’явилась у 1968 році [2]. Ці та подальші публікації автора були визнані піонерними в світі (див., наприклад, [9]). З 1975 року генеративна модель розпізнавання мовних сигналів в аналогічній інтерпретації від імені інших авторів поширюється також під назвою НММ-модель [10].

В кінці 60-х автор спробував узагальнити генеративну модель і на розпізнавання зображень, використовуючи прийоми двовимірних генеративних процесів та відповідного їм деякого розширеного динамічного програмування [3]. Пізніше більш глибокі теоретичні розробки стосовно двовимірних породжувальних грамастик та їх використання в

обробленні зображень були виконані М.І.Шлезінгером [4, 6].

В 1997 році з’явилась ідея об’єднання досліджень з автоматичного розпізнавання образів (сприйняття слухової, зорової, акустичної, смакової, тактильної тощо інформації). Ця ідея почала реалізовуватись у 2000 році після схвалення ДНТП “Образний комп’ютер”.

ЛІТЕРАТУРА

- [1] Тарас Вінцюк, “Образний комп’ютер”, *Зб. наук. праць “Сучасні проблеми в комп’ютерних науках”*, Вид-во Нац. ун-ту “Львівська політехніка”, Львів, 2000, с. 5-14.
- [2] Т.К. Винцюк, “Распознавание слов устной речи методами динамического программирования”, *Кибернетика*, 1968, № 1, с. 81-88.
- [3] Т.К. Винцюк, “Распознавание рукописных знаков методами динамического программирования”, *Сб. «Кибернетика и вычислительная техника», Вып. 3, «Распознавание образов»*, Киев: “Наукова думка”, 1969, с. 52-77.
- [4] М.И.Шлезингер, “Синтактический анализ двумерных зрительных сигналов в условиях помех”, *Кибернетика*, 1976, № 4, с. 113-130.
- [5] Т.К. Винцюк, “Анализ, распознавание и интерпретация речевых сигналов”, Киев: “Наукова думка”, 1987, 264 с.
- [6] М.И.Шлезингер, “Математические средства обработки изображений”, Киев: “Наукова думка”, 1989, 198 с.
- [7] 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.
- [8] Taras K.Vintsiuk, “Two Approaches to Create a Dictation/ Translation Machine”, *Proc. Second Intern. Workshop “Speech and Computer”*, Cluj-Napoca, 1997, pp 1-6.
- [9] S.E.Levinson, “Structural Methods in Automatic Speech Recognition”, *Proc. of the IEEE*, Vol. 73, No. 11, Nov. 1985, pp 1625-1650.
- [10] F.Jelinek, “Continuous Speech Recognition by Statistical Methods”, *Proc. IEEE*, Vol. 64, Apr. 1976, pp 532-556.