

# Проблеми створення дружнього інтерфейсу образного комп'ютера на основі жестової, візуальної та голосової інформації

Ю.В.Крак, О.В.Бармак, А.С.Тернов, Б.А.Троценко, М.М.Шатковський

Інститут кібернетики ім. В.М. Глушкова НАН України,  
03680, МСП, Київ, проспект Академіка Глушкова, 40  
krak@unicyb.kiev.ua

## Abstract

A new concept of a user-friendly interface for a pattern computer is investigated. The concept is based on a voice, visual and human gestures information.

## 1. Вступ і постановка задачі

У доповіді „П'ята комп'ютерна революція – базовий сценарій” на відкритті Всесвітньої виставки комп'ютерних технологій СеВІТ 2008, директор фірми Microsoft Стів Балмер [1] акцентував увагу, що в наступні роки розробка природного інтерфейсу користувача буде відігравати ключову роль в прогресі комп'ютерної техніки і інформаційних технологій. Зокрема наголошувалось на важливості розробки засобів розпізнавання голосу, рукописних символів, сприйняття жестової і тактильної інформації. Відзначимо, що ці та подібні проблеми були закладені в 2000 році для вирішення в Державну науково-технічну програму „Образний комп'ютер” де під

образним комп'ютером розумілась мультимодальна система «...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.»[2]. Таким чином, українські вчені-кібернетики які працювали над виконанням завдань проекту «Образний комп'ютер» займалися тими проблемами, що на даний час визначають рівень і стан науково-технічного прогресу в області комп'ютерної техніки і технологій.

В даній доповіді досліджуються проблеми розробки природного інтерфейсу людини з комп'ютерним середовищем на основі жестової, візуальної та голосової інформації.

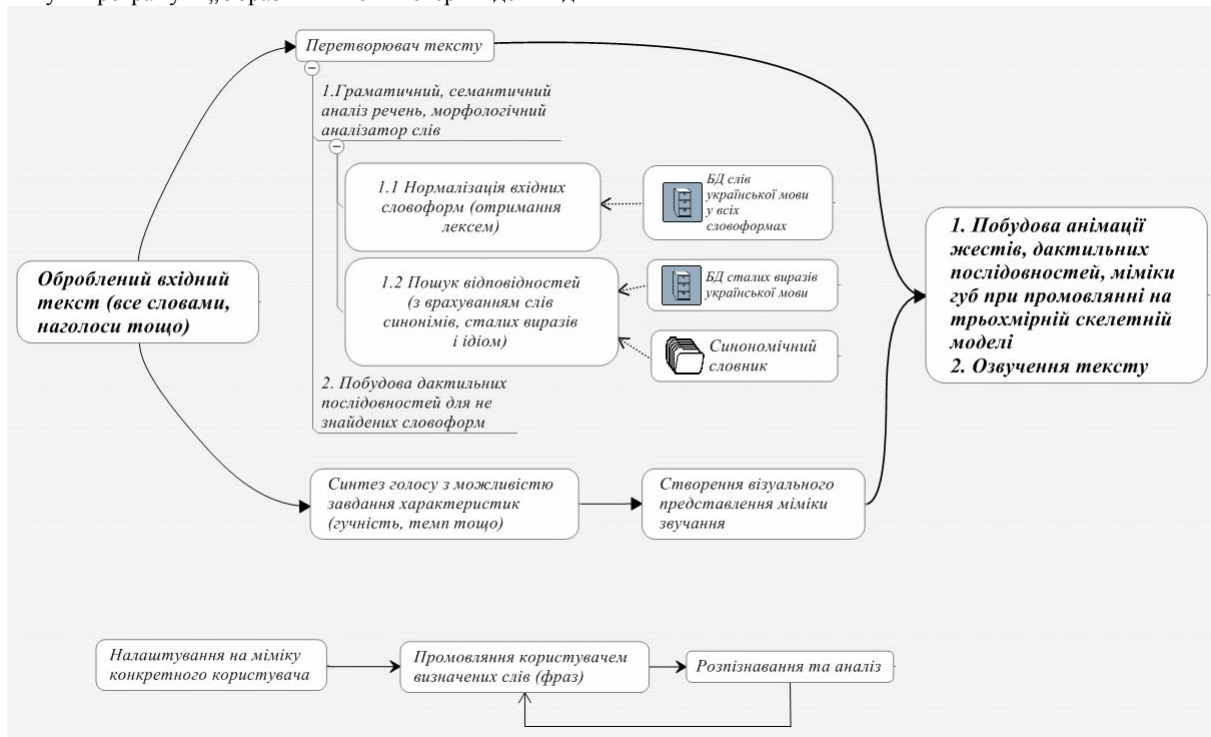


Рис. 1. Схема концепції комп'ютерної інформаційної технології природного інтерфейсу людини з комп'ютерним за допомогою мови жестів

Основна увага приділяється розробці засобів аналізу й синтезу української мови жестів, міміки і артикуляції, української розмовної мови. Такі розробки, крім комп'ютерного застосування, мають велике соціальне значення, зокрема дають можливість українцям з вадами слуху і зору активно включатись в комунікаційний процес як між собою, так із людьми без цих вад. З іншої сторони розробка і комп'ютерна реалізація української мови жестів дозволить широкому загалу людей вільно спілкуватися з нечуючими людьми [3].

Виходячи з цього сформульована наступна постановка задачі. Необхідно розробити концепцію комп'ютерної інформаційної технології природного інтерфейсу людини з комп'ютерним середовищем з можливістю реалізації у її рамках наступної функціональності:

- відображення на обличчі тривимірної комп'ютерної моделі людини мовленнєвого процесу з урахуванням емоційних складових;
- синтез рухів мови жестів глухих та дактильної абетки тривимірної моделі людини;
- синтез тексту українською мовою на природну голосову мову;
- аналіз речень української мови для синтезу їх аналогів на мові жестів;
- аналіз міміки губ мовленнєвого процесу.

## 2. Інформаційна технологія природного інтерфейсу людини з комп'ютерним за допомогою мови жестів

Після проведеного аналізу сучасного стану проблеми спілкування людей з вадами слуху з чуючими людьми була розроблена концепція комплексної інформаційної технології природного інтерфейсу людини з комп'ютером за допомогою мови жестів.

Комплексна інформаційна технологія включатиме реалізацію наступних можливостей:

- модуль перекладу звичайного тексту на мову жестів глухих, який міститиме: анімацію процесу промовляння розмовної і калькулюючої мови жестів з використанням віртуальних моделей людей;
- анімацію міміки обличчя (з урахуванням емоційних складових) при промовлянні;
- озвучення (синтез) звичайного тексту у його звуковий аналог (з використанням різних голосів);
- модуль розпізнавання по зміні міміки губ тексту, який промовляється.

Концепція експериментальної технології зображена на схемі (рис. 1).

У перетворювачі тексту основне навантаження несе граматичний, семантичний аналіз речення та морфологічний аналіз слова. При перекладі речення мови жестів у речення українською мовою, мають місце наступні проблеми при нормалізації вхідної словоформи (морфологічний аналізатор):

- вхідна словоформа має декілька варіантів слів, що належать до різних частин мови – для вирішення таких колізій необхідно попередньо провести граматичний аналіз речення, що суттєво зменшить такі колізії;
- одне і теж слово української мови має багато відтінків та переходів в значеннях і вживанні – тобто для точного визначення жесту для конкретного слова потрібно визначити семантичну структуру вхідного речення.

Схема алгоритму такої функції перетворювача зображено на рис.2.

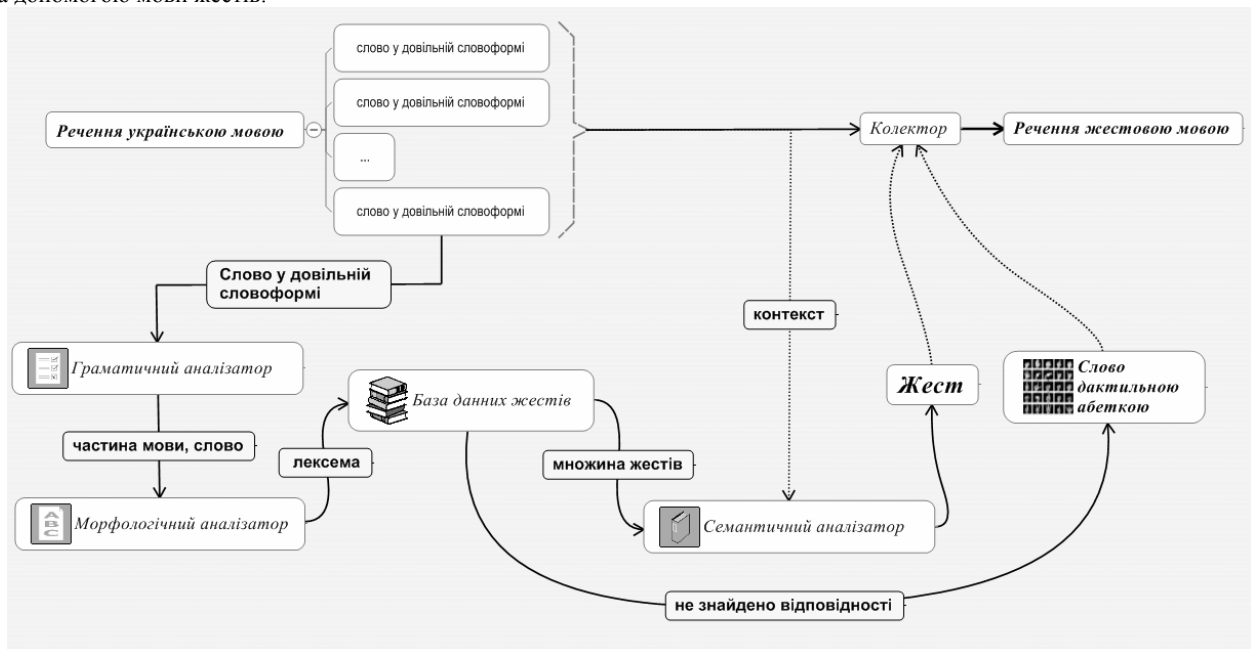


Рис. 2. Перетворювач тексту

Граматичний аналізатор визначає для кожного слова до якої частини речення воно відноситься і також частину мови до якої дане слово належить.

Морфологічний аналізатор нормалізує вхідні словоформи (отримання лексеми).

Семантичний аналізатор, враховуючи значення слова та його вживання у даному реченні, визначає жест.

У концепції експериментальної технології присутня частина, яка пов'язана з дослідженням задач розпізнавання міміки губ при промовлянні українською мовою. Для цих задач запропонована схема алгоритму подана на рис 3.

Блок 1 відповідає за попередню обробку вхідної візуальної інформації та перетворення її у простір

характеристичних ознак – координат контрольних точок апроксимуючих верхню та нижню губу NURBS кривих.

Блок 2 містить у собі алгоритми побудови базису простору характеристичних ознак та оцінки його якості. На виході будується базисна матриця векторного простору характеристичних ознак.

У третьому блоці відбувається розклад вектора характеристичних ознак, побудованого для вхідного зображення яке розпізнається, за отриманим базисом.

Результатом роботи алгоритму є вектор розкладу, на основі якого приймається рішення про відповідність вхідного вектора конкретній базовій міміці при промовлянні українською мовою.

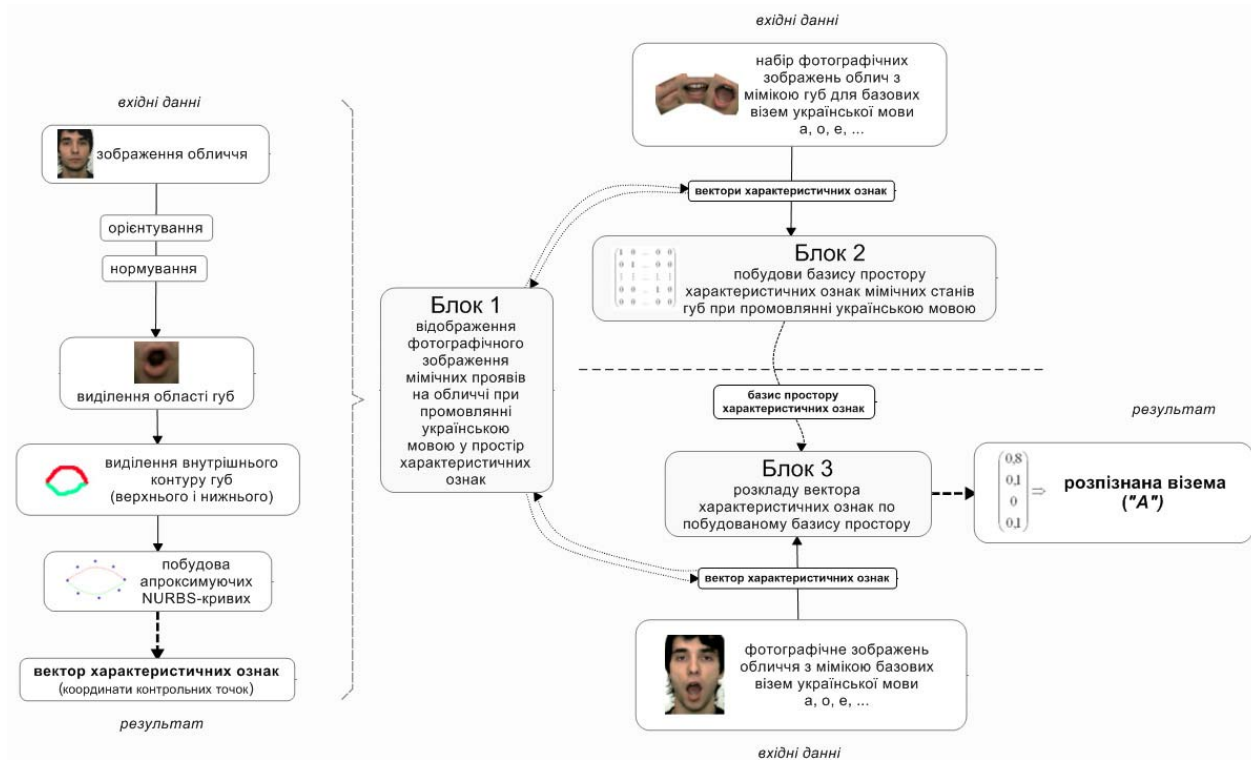


Рис 3 Схема алгоритму розпізнавання міміки при промовлянні українською мовою

### 3. Реалізація концепції

В рамках реалізації запропонованої концепції комп'ютерної інформаційної технології природного інтерфейсу людини з комп'ютером за допомогою мови жестів проведені дослідження та розроблене тестове програмне забезпечення.

Для синтезу тривимірної анімації мови жестів побудовані геометричні класи векторів-образів жестів. Побудова цих класів базувалась на використанні технології motion capture [7]. Motion capture – це технологія подання рухів, яка дозволяє перейти від характеристик руху в реальному світі, до фіксації змін параметрів для математичної моделі. Для автоматичного отримання необхідних ключових координат жесту використовувалась технологія трекінгу [8].

Для представлення жесту використовувався формат BVH, з наступним експортом отриманих даних на скелетну трьохвимірну модель (наприклад, у модулі Character Studio для 3D studio MAX або в Poser).

Запропонована реалізація технології motion capture для фіксації рухів жестової мови містить наступні етапи (рис.4):

- 1) отримання відео-потoku жестикуляції у двох ракурсах: фронт, профіль (усі необхідні параметри зйомки відомі);
- 2) обробка відео-потoku з виділенням слідів і положень (координат) рук жестикуляції в просторі;
- 3) на основі отриманих координат жестикуляції - формування BVH-файлу для синтезу трьохвимірної анімації;
- 4) застосування BVH-файлу для створення анімації (в Character Studio для 3D studio MAX або в Poser).

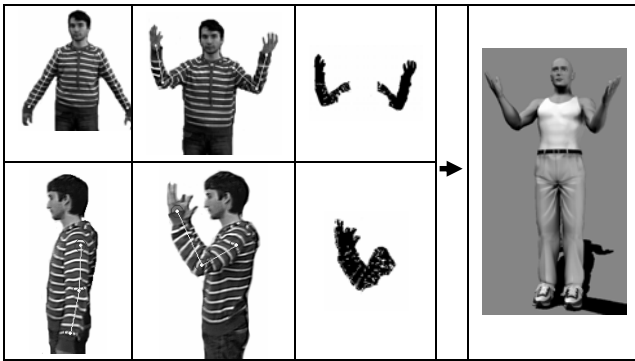


Рис.4. Побудова анімації жесту

Для обробки вхідного тексту (розміщення наголосів, виділення інфінітивів, пошуку синонімів і типових фраз мови) розроблена інформаційна модель української мови. Модель представлена у вигляді таблиць реляційної бази даних з набором збережених процедур, які реалізують необхідну, для розглянутої технології, функціональність.

Інформаційна модель української мови містить більше двох з половиною мільйонів слів (всі можливі словоформи), наголоси, синоніми, ідіоми, фрагменти слів для синтезу різними голосами, вектори-образи жестів.

Для реалізації функції візуалізації озвучення тексту розроблено синтезатор української мови. Синтезатор дозволяє створювати голосовий аналог довільного тексту різними голосами з можливістю керування характеристиками голосу (гучність, далі/ближче). Синтезатор також дає можливість побудувати візуальне представлення процесу промовляння, як за допомогою двомірних відеом, так і на трьохвимірній моделі (рис.5).

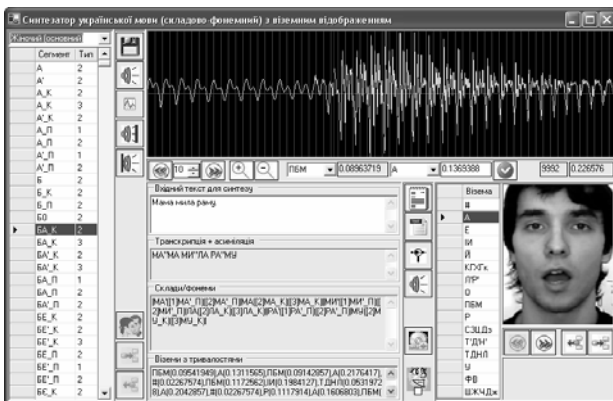


Рис. 5 Синтезатор української мови

Для комплексного тестування запропонованої технології створене відповідне програмне забезпечення для перекладу довільного тексту на калькулятивну мову жестів. Програмне забезпечення реалізує наступний алгоритм синтезу послідовності жестів:

- 1) вхідний текст синтезується у його голосовий аналог;
- 2) вхідний текст розбивається на слова;
- 3) для кожного слова у базі даних слів шукається його нормальна форма (інфінітив)
- 4) для кожної нормальної форми слова у базі даних жестів шукається відповідна послідовність рухів;

- 5) якщо для якоїсь словоформи послідовність рухів не знайдена, то слово зображається дактильною азбукою
- 6) трьохвимірною моделлю відтворює генеровані послідовності рухів та синтезований голосовий аналог.

#### 4. Висновки

В роботі запропонована концепція комплексної інформаційної технології природного інтерфейсу людини з комп'ютером за допомогою мови жестів. Отримані результати в рамках цієї концепції показали її життєздатність та виявили деякі проблеми, які потребують подальших досліджень.

Подальші дослідження направлені на реалізацію всієї множини рухів української мови жестів. Окремого розгляду потребують питання, пов'язані з граматичним та семантичним аналізом речень.

#### 5. Посилання

- [1] The Fifth Revolution - Keynote Script. Opening Speech Steve Ballmer. CeBIT 4 to 9 March 2008// [www.cebit.de/7514?pm=cb08-168-e](http://www.cebit.de/7514?pm=cb08-168-e).
- [2] Т.Вінцюк. Генеративна модель образного комп'ютера// Праці Шостої всеукраїнської міжнародної конференції «УкрОБРАЗ'2002» - К.: 2002. – С.7-14.
- [3] Ю.Крак, А.Бармак, А.Ганджа, А.Тернов, Н.Шатковский. Компьютерная система виртуального общения людей с проблемами слуха // International Journal "Information Technologies and Knowledge", Vol. 2, - 2008. - P.161-165.
- [4] Neidle C., S. Sclaroff, V. Athitsos SignStream™: A Tool for Linguistic and Computer Vision Research on Visual-Gestural Language Data, Boston University, Boston, Massachusetts, 2001, 33:3,pp. 311-320.
- [5] IBM Research Demonstrates Innovative 'Speech to Sign Language' Translation System, Extreme Blue programme HURSLEY, UK - 12 Sep 2007, IBM Media Relations <http://www.03.ibm.com/press/us/en/pressrelease/22316.wss>
- [6] Воскресенский А.Л. Хахалин Г.К. Мультимедийный толковый словарь русского жестового языка. Компьютерная лингвистика и интеллектуальные технологии: // Труды международной конференции «Диалог 2007» (Беласово, 30 мая - 3 июня 2007 г.) / Под ред. Л.Л. Иомдина, Н.И. Лауфер, А.С. Нариньяни, В.П. Селегея. - М.: Изд-во РГГУ, 2007. 658 с.
- [7] Menache A., Understanding Motion Capture for Computer Animation and Video Games, Morgan Kaufmann, 2000. 238 pp.
- [8] S. Avidan, "Support vector tracking," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, Hawaii, volume I, 2001, pp. 184-191.
- [9] Villani N., Doublestein J., Martin, Z. The MathSigner: an interactive learning tool for American sign language Adamo- Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on Volume, Issue, 14-16 July 2004 Page(s): 713-716 Digital Object Identifier, 10.1109/IV.2004.