

СТВОРЕННЯ АУДІОВІЗУАЛЬНОЇ БАЗИ ДАНИХ НЕМАНУАЛЬНИХ СКЛАДОВИХ ЖЕСТОВОЇ МОВИ

Ю.В.Крак, А.С.Тернов, В.О. Кузнєцов, Ю.В. Барчукова

Інститут кібернетики ім.В.М. Глушкова НАН України,
03680, МСП, Київ, проспект Академіка Глушкова, 40
Yuri.krak@gmail.com, anton.ternov@gmail.com

Анотація

Discussed here is the process of audiovisual database creation for Ukrainian Sign Language, obtaining parameters, related to nonmanual components of the sign language, in purpose to find the structure and relations between manual and nonmanual components.

1. Вступ і постановка задачі

Дані дослідження присвячені проблемі створення аудіовізуальної бази даних (АВБД) компонент Українського жестової мови (УЖМ), а саме немануальних компонент. Жестова мова – складна взаємозалежна структура, що характеризується багатопоточністю (кілька каналів передачі інформації), своєю граматикою і своїми засобами передачі інформації – жестами (мануальні компоненти) і мімікою, рухами тіла, поглядом, поворотами голови та іншими елементами (немануальні компоненти) [1].

Ряд досліджень [2-4], присвячених жестовим мовам (ЖМ) показали, що кожен з немануальних компонент може передавати: експресію, логічне членування речень, ступені порівняння прикметників тощо. У той же час, через складність локалізації та маркування, знання про немануальні компоненти, що використовуються в УЖМ, особливо міміку, недостатньо вивчені і несистематизовані. Подібні дослідження проводилися для американської (ASL) [5], французької (FSL) [6], німецької (DGS) [7] та інших жестових мов.

З метою отримання даних про немануальні компоненти УЖМ було вирішено створити АВБД, яка б дозволила, на основі множини різних прикладів, визначити базис немануальних компонент, а також вивести правила їх використання на основі досить великого набору вихідних даних. У кінцевому підсумку це дозволило б використовувати отримані знання, у тому числі, і для інших досліджень. Виходячи з зазначеного, була сформульована наступна постановка задачі.

Необхідно розробити АВБД, яка виконувала такі функції:

- відображення аудіовізуальних зразків немануальних компонент;
- містила метадані для маркування немануальних компонент у використовуваних аудіовізуальних зразках;
- надавала можливість пошуку і отримання статистичних характеристик для певного набору компонент;

- дозволяла отримувати чисельні характеристики змін кожної з немануальних компонент, що містяться в аудіовізуальному зразку, для можливості подальшого використання при моделюванні, синтезі і розпізнаванні міміки (що неможливо без навчального набору даних).

2. Вихідні дані

Для отримання вихідних даних для АВБД було проведено ряд досліджень, які стосувались різних аспектів отримання параметрів немануальних компонент. В якості джерел даних використовувалась різноманітна інформація: зображення, відеозаписи на УЖМ, словники жестової мови та інше. Це дозволило: по-перше, досліджуючи словників ЖМ, побачити найбільш часто уживані немануальні компоненти, які використовуються разом з жестами, по-друге, при дослідженні відеозаписів УЖМ, оцінити ступінь використання немануальних компонент в досліджуваній жестовій мові і співставити однакові компоненти з різних жестових мовах.

3. Створення текстових описів

Для опису аудіовізуальних даних немануальних компонент на УЖМ використовувалася текстова система опису у формі лінгвістичних предикатів. Для цього обличчя і тулуб умовно розбиваються на області, в яких відбувалася зміна певних параметрів (форма, контур та інше). Описи у вигляді предиката містять у собі декілька елементів: об'єкт, дія, характер дії. Об'єкт характеризує змінну частину немануального компоненту (іменник). Дія може бути виражена як незавершеною дією (дієслово), так і завершеною (дієприслівник). Елемент "характер дії" може містити ступені вираження: для дієслова - швидкість, інтенсивність, для дієприслівника - порівняльні ступені. Предикат може містити як і кілька об'єктів дії, так і декілька дій, утворюючи складний предикат. Представлені таким чином вихідні дані дозволяють співставити різні немануальні компоненти, при порівнянні їх описів між собою.

Для опису аудіовізуальних фрагментів, що містять відео, використовувалося спеціальне програмне забезпечення для лінгвістів Elan [8], яке дало змогу отримувати описи, а також часові характеристики немануальних компонент у XML форматі (рис. 1).

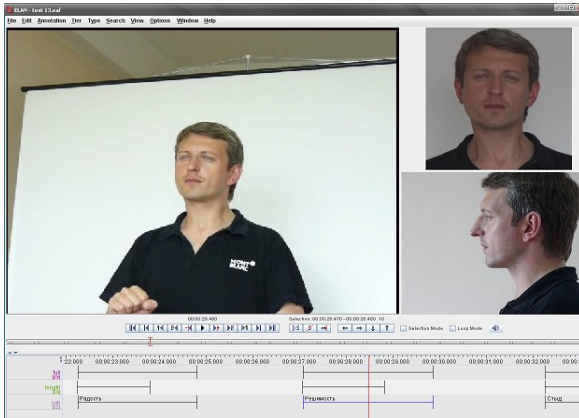


Рис 1: Маркування відеофрагментів в Elan.

Отримані текстові описи було перенесено в таблицю, яка містить такі елементи: назва жесту, опис, а також часові характеристики (тільки для відеофрагментів).

Для забезпечення подальшої роботи з описами немануальних компонент необхідно отримати відповіді на наступні питання:

- чи є немануальні компоненти цільними виразами, як, наприклад, емоційні мімічні вирази в розмовній мові чуючих людей;
- в яких жестах використовуються немануальні компоненти, а в яких жестах їх присутність необов'язкова;
- як співвідносяться між собою мімічні вирази і артикуляція фонем жестової мови.

За основу були взяті результати досліджень жестових мов LSF [9], ASL [5] і УЖМ [10], представлених у вигляді ілюстрованих словників жестових мов. Також, було проведено дослідження описів набору виразів в ASL і УЖМ з метою перевірки ступеня відповідності немануальних компонент в УЖМ і ASL жестовою мовою на основі гіпотези походження першого з другого. Було виконано порівняння на вибірці немануальних компонентів з УЖМ і ASL.

В результаті були знайдені елементи, що входять до складу кожної з цих мов, що дозволило шукати аналогічні немануальні компоненти й в Українській жестовій мові.

4. Отримання списку інтегральних (неподільних) немануальних компонент

Візуальні описи, представлені в словниках у формі зображень та описів немануальних компонент, переносяться у таблицю, яка містить наступні параметри: назва жесту, тип немануального компоненту і лінгвістичний опис зміни немануальних компонент в жесті.

Тип немануального компонента приймає значення «експресивний», «нейтральний» чи «граматичний» (до них належить група компонент, що описують час, прикметники, тему, наголос та інші елементи граматики жестової мови).

Для отримання параметра «тип немануального

компонента» для жестів, немануальні параметри яких не містили описів в словниках FSL і УЖМ, були проаналізовані дані про відомі немануальні компоненти зі словника ASL [5]. З цих даних була сформована таблиця, яка містить:

- назву жесту;
- назву немануального компонента;
- тип немануального компонента (емоційний або граматичний);
- назву групи жестів, в яких присутній даний немануальний компонент.

На підставі даної таблиці було отримано кілька груп жестів, що містять немануальні компоненти:

- жести, що містять експресію (емоційні мімічні вирази);
- жести, що вказують час і місце (містять рухи тулуба);
- жести, що містять напрям погляду;
- жести, що містять граматичні вирази (повороти голови, міміка і артикуляція): кількісні і якісні прикметники, логічні операції.

Це дозволило поширити знання про жести з ASL і на дані зі словників FSL і УЖМ. Для цих цілей аналізувалися одночасно декілька параметрів жесту:

- назва жесту, на основі якого робилося судження про наявність ознак емоційного відтінку у слова, що описує жест, або про наявність граматичного значення, при наявності жестів-синонімів, що мають немануальні компоненти;
- лінгвістичний опис, на основі якого робився висновок про можливість належності немануального компонента з вказаними параметрами до класу немануальних компонент.

Після цього було отримано набір жестів, що містять міміку, а також окрема таблиця, яка містить набір і типи немануальних компонент.

5. Отримання набору ознак, що входять у вирази

Після отримання набору інтегральних немануальних компонент виникла необхідність розбити кожен з лінгвістичних описів цих компонент на складові. В якості критерію виступали: об'єкт дії (область прояву) і дія (характер і ступінь прояву).

Кожен із структурних елементів лінгвістичних описів був закодований за допомогою систем нотації мануальних і немануальних компонент жестової мови HamNoSys [7], SignWriting [11] і FACS [12], що дозволяє представити складний лінгвістичний опис у вигляді вектора ознак.

6. Отримання кількісних значень змін інтегральних немануальних компонент

У постановці задачі дослідження був зазначений пункт «отримання чисельних значень змін немануальних компонент». Необхідність такої функції в базі даних була викликана кількома причинами:

- в АВБД присутня різномірна інформація - фото та відео з різних джерел, що не дозволяло коректно виконувати співставлення різних прикладів аудіовізуальних даних одного і того ж немануального компонента;
- числові значення дозволяють більш точно співставляти одні й ті ж немануальні компоненти, при цьому відсутній вплив геометричних спотворень, що вносяться фото і відеоапаратуру;
- необхідність уніфікації зразків немануальних компонент, що дозволило б представляти набір даних з однієї множини в розмірностях іншої (варто зазначити, що подібна задача існує і в комп'ютерному моделюванні - retargeting рухів 3D моделей);
- використання числових значень немануальних компонент дозволило б використовувати їх у подальших розробках, пов'язаних з моделюванням, синтезом і розпізнаванням міміки;
- використання числових значень спільно з 3D моделлю дозволило б істотно скоротити обсяг аудіовізуальної бази даних, оскільки відео і зображення займають більше місця в пам'яті комп'ютера, ніж 3D модель і текстура.

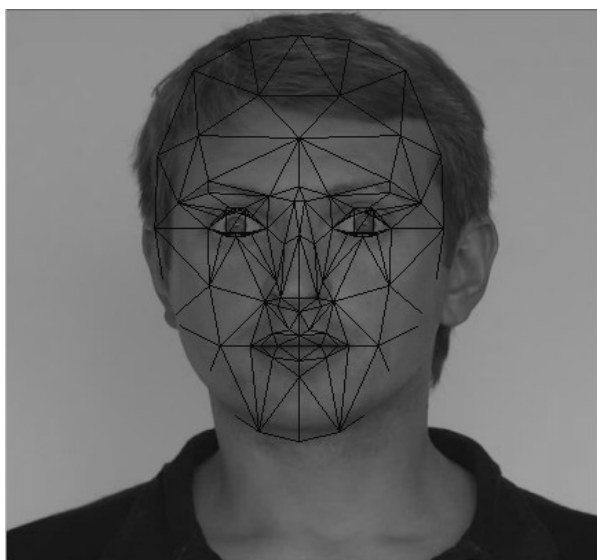


Рис. 2. Отримання числових значень параметрів моделі особи за допомогою моделі CANDIDE-3

Для цих цілей була обрана як математична модель міміки CANDIDE-3 [13], що представляє зміни точок 3D обличчя на основі гнучких шаблонів, в одиницях змін компонент по системі FACS, а також дозволяє робити ці описи на основі стандартних компонент точок, які використовуються в стандарті кодування відео MPEG-4 (див. рис.2).

Таким чином, отримані в результаті аналізу опису набори ознак, що входять в інтегральні компоненти, можна представити як у вигляді змін, що описуються кодами FACS, так і адаптувати запропоновану модель до

наявних аудіовізуальних даних.

Дана модель дозволяє змінювати геометричні параметри обличчя (15 параметрів), а також застосовувати зміни, пов'язані з мімікою (10 параметрів), що дає можливість описати форму і зміни довільного людського обличчя, наявного в отриманій аудіовізуальній базі даних, з досить високим ступенем достовірності.

7. Аналіз отриманих результатів досліджень

При аналізі словника УЖМ з'ясувалося, що сам словник, окрім мімічних немануальних компонент, містить також артикуляцію слів, які позначають жест в розмовній мові. Однак, ці дані не дозволяють зробити однозначного висновку щодо співвідношення міміки і ЖМ. Пояснення даного феномена пов'язане з тим, що артикуляція в ЖМ та розмовної мови має різні призначення: в ЖМ – для показу слів у калькуючій жестовій мові, показу незнайомого для співрозмовника жесту або для позначення одразу цілого слова. У звичайній розмовній мові артикуляція використовується для модуляції звукового потоку і (частково) для передачі експресії, що міститься в структурі слова.

Наявність в жестовій мові одночасно і фонем розмовної мови при артикуляції слів, і міміки в жестах, що містять граматичний відтінок, призводить до нерозуміння, оскільки фонеми розмовної мови мають схожість з деякими мімічними виразами. Це вимагає додаткових досліджень для встановлення причинно-наслідкових зв'язків, оскільки у досліджуваних матеріалах (зображення) відсутня динаміка.

Для того щоб перевірити достовірність описів зміни компонент було вирішено адаптувати тривимірну модель обличчя до декількох зображень з АВБД, шляхом отримання кодів змін параметрів в системі FACS і порівняння їх з кодами бази даних. Попередні дослідження на тестових прикладах дали позитивні результати.

В подальшому передбачається провести їх на всьому наявному наборі аудіовізуальних даних.

8. Висновки та подальші дослідження

У результаті досліджень була побудована база даних аудіовізуальної інформації. В якості бази даних використовувалася БД MS Access з СУБД Microsoft Access з виведенням вмісту таблиць у форми.

На рис. 3 показана архітектура таблиць, створених у базі даних.

На даний момент база даних охоплює близько тисячі жестів, що містять міміку, в тому числі: 344 жестів французької жестової мови, 526 жестів американської жестової мови і 105 жестів української жестової мови, 97 інтегральних немануальних компонентів, які включають: емоційних – 70, граматичних – 27, а також 88 елементів з яких складаються немануальні компоненти.

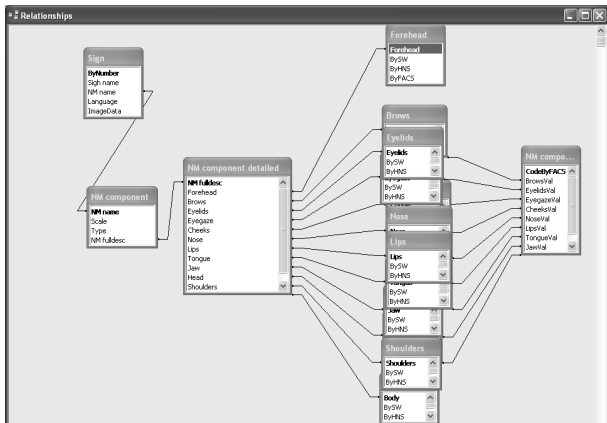


Рис. 3. Структура таблиц бази даних у вікні Relationships СУБД MS Access

У результаті проведених досліджень вдалося систематизувати наявні дані про немануальні компоненти Української жестової мови, визначити їх кількість, локалізацію та структуру, а також визначити подальші напрямки дослідження. Необхідно більш детально вивчити взаємодію міміки, жестів і артикуляції та інших немануальних компонент, з використанням дослідження динаміки процесу на основі параметричної моделі CANDIDE. Такий підхід дозволить отримати часові характеристики для кожної зі складових немануальних компонентів у процесі демонстрації жестової інформації.

Часові характеристики, а також характер зміни параметрів моделі можуть бути безпосередньо використані для моделювання, синтезу і розпізнавання інформації, що передається жестовою мовою, включаючи міміку і артикуляцію.

9. Література

- [1] Кульбіда С.В. Українська жестова мова як природна знакова система / С. В. Кульбіда // збірник наукових праць «Жестова мова й сучасність»: К. : Педагогічна думка, 2009. – С. 218-239.
- [2] Зайцева Г.Л. Жестовая речь. Дактилология: Учеб. для студ. высш. учеб. заведений / Зайцева Г.Л. – М.: Гуманит. изд. центр ВЛАДОС, 2000. – 192 с.
- [3] Киммельман В. Нерегулярные отрицания в русском жестовом языке // Материалы конференции МСКЛ-2. — М., 2007. — С. 89–103.
- [4] Ткачева Т. Жестовый язык глухих как обличья знаковая коммуникативная система (на материале русского и французского жестовых языков) // Ткачева Т., Ахмедышева Н. – Материалы Международной научно-практической конференции «Приоритеты и интересы современного общества». — Астрахань: АГУ, 2010. — С. 361–365.
- [5] Tennant R. A. The American Sign Language Handshape Dictionary // Tennant R. A., Marianne Gluszk Br. - Washington, Gallaudet University Press, 1998. - 407 p.
- [6] Chetelat-Péle E. Annotation of Non Manual Gestures: Eyebrow movement description / Chetelat-Péle E., Braffort A., Veronis J / In: Crasborn, Otto et al. (2008, eds.) 3th Workshop on Representation and Processing of

Sign Languages. Workshop proceedings. International Conference on Language Resources and Evaluation (LREC), Marrakesh, Morocco. – Paris, ELRA, 2008. – P.28-32.

- [7] Interface Definitions, ViSiCAST Deliverable D5-1./ [Hanke T., Langer G., Metzger C. et al] – Hamburg, Insitute of German Sign Language and Communication of the Deaf, 2002. – 74 p.
- [8] Електронний ресурс Max Planck Institute for Psycholinguistics.
Режим доступу: <http://www.lat-mpi.eu/tools/elan/>
- [9] Companys M. Dictionnaire 1200 signes // Companys M., Tourmez F. - Angers, Éditions Monica Companys, 2000. - 150 p.
- [10] Український жестівник для батьків // С.В. Кульбіда, І.І. Чепчина, Н.Б. Адамюк, Н.В. Іванюшева; НАПН України, Інститут спеціальної педагогіки, Лаб. жестової мови, Укр. т-во глухих. — К.: СПКТБ УТОГ, 2011. — 380 с.
- [11] Sutton V. Lessons in SignWriting / Sutton V. – La Jolla, Center for Sutton Movement Writing, 2002. – 195p.
- [12] Miller F. P. Facial Action Coding System. / Miller F. P., Vandome A. F., McBrewster J. – Beau Bassin, International Book Marketing Service Ltd, 2011. – 80 p.
- [13] Ahlberg J. CANDIDE-3 – an updated parameterized face / Ahlberg J. – Linkoping, Ed. Of Linkoping University, 2001 – 16p.