

Ю.В. Крак, А.С. Тернов, В.А. Кузнецов

Разработка аудиовизуальной базы данных немануальных компонент жестовой речи

Рассмотрен процесс создания аудиовизуальной базы данных украинского жестового языка, получения характеристик немануальных компонент жестовой речи, с целью выделения структуры и связей между мануальными и немануальными компонентами.

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 is discussed.

Розглянуто процес створення аудіовізуальної бази даних української жестової мови, отримання характеристик немануальних компонентів жестової мови, з метою виділення структури і зв'язків між мануальними та немануальними компонентами.

Введение. Предлагаемые исследования посвящены созданию аудиовизуальной базы данных (АВБД) компонент Украинского жестового языка (УЖЯ), а именно немануальных компонент. Жестовый язык (ЖЯ) – сложная структура, характерная многопоточностью (несколько каналов передачи информации), определенной грамматикой и своими средствами передачи информации – жестами (мануальные компоненты), мимикой, движениями, взглядом и другими элементами (немануальные компоненты) [1].

Ряд исследований [2–4] ЖЯ показали, что каждый из немануальных компонент может передавать экспрессию, логическое членение предложения, степени сравнения прилагательных и другое. В то же время, из-за сложности локализации и маркировки немануальных компонент, особенно мимики, недостаточно изучены и не систематизированы знания о таких компонентах, используемых в УЖЯ. Подобные исследования проводились для американского (ASL) [5], французского (FSL) [6], немецкого (DGS) [7] и других жестовых языков.

Постановка задачи

С целью получения данных о немануальных компонентах УЖЯ было решено создать АВБД, которая позволила бы на основе различных примеров определить базис немануальных компонент и вывести правила их использования на основе набора исходных данных, что позволило бы применять полученные знания и для других исследований. Исходя из сказанного, была сформулирована следующая постановка задачи.

Разработать АВБД, которая выполняла бы следующие функции:

- отображение аудиовизуальных примеров немануальных компонент;
- содержала бы метаданные для маркировки немануальных компонент в используемых аудиовизуальных примерах;
- предоставляла бы возможность поиска и получения статистических характеристик для определенного набора компонент;
- получать численные характеристики изменений каждой немануальной компоненты аудиовизуального примера для дальнейшего использования при моделировании, синтезе и распознавании мимики (что невозможно без обучающего набора данных).

Методика исследований

Для получения исходных данных для АВБД проведены исследования различных аспектов получения параметров немануальных компонент. В качестве источников данных использовалась разнородная информация: изображения, видеозаписи на УЖЯ, словари жестового языка и др. Это позволило: *во-первых*, исследуя словари ЖЯ, увидеть наиболее часто употребляемые немануальные компоненты с жестами; *во-вторых*, исследуя видеозаписи на УЖЯ, оценить степень использования немануальных компонент в исследуемом жестовом языке и сопоставить одинаковые компоненты в разных ЖЯ.

Схема этапов проведения исследований для получения исходных данных для АВБД показана на рис. 1.

Создание лингвистических описаний

Для описания аудиовизуальных данных немануальных компонент на УЖЯ использована текстовая система описания в форме лингвистических предикатов. Для этого лицо и туловище

ще условно делились на области, в которых происходило изменение каких-либо параметров (форма, контур и др.). Исходные данные, представленные в предикатах, таким образом позволяли сопоставить различные немануальные компоненты, сравнивая их описания.

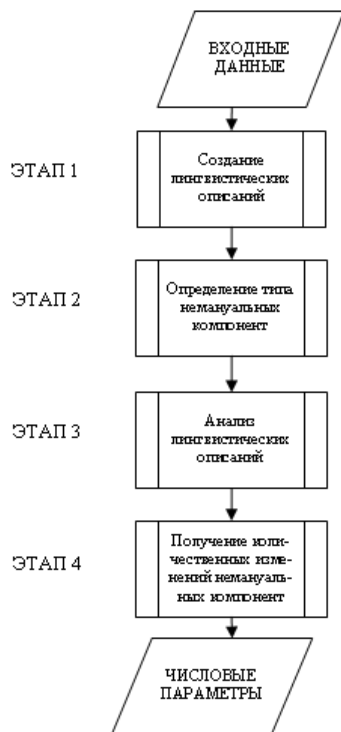


Рис. 1. Схема этапов обработки данных

Для описания аудиовизуальных фрагментов, содержащих видео, использовано специальное программное обеспечение для лингвистов *Elan* [8]. Данное программное обеспечение позволяет получать описания и временные характеристики немануальных компонент (рис. 2).

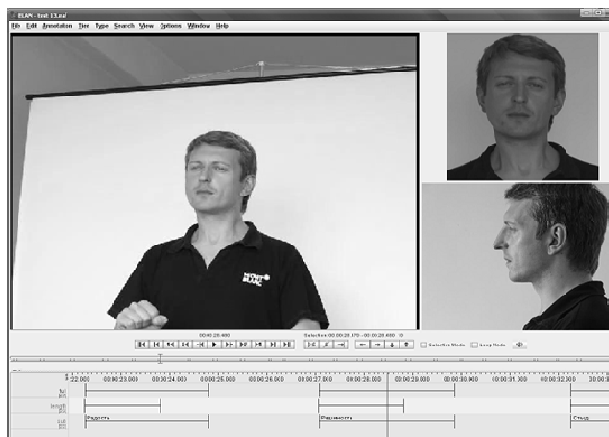


Рис. 2. Маркировка видеофрагментов в *Elan*

Для дальнейшей работы с описаниями немануальных компонент необходимо получить ответы на следующие вопросы:

- являются ли немануальные компоненты цельными выражениями, как, например, эмоциональные мимические выражения в разговорной речи слышащих людей;
- в каких жестах используются немануальные компоненты, а в каких их присутствие не обязательно;
- как соотносятся между собой мимические выражения и артикуляция фонем жестовой речи.

За основу были взяты результаты исследований жестовых языков *ASL* [5], *FSL* [9] и *УЖЯ* [10], представленных в виде иллюстрированных словарей ЖЯ. Было проведено также исследование описаний набора выражений в *ASL* и *УЖЯ* с целью проверки степени соответствия немануальных компонент в *УЖЯ* и *ASL*-жестовом языке на основе гипотезы происхождения первого из второго. Выполнено сравнение на выборке немануальных компонент из *УЖЯ* и *ASL*.

В результате найдены элементы, входящие в эти компоненты, что позволило искать аналогичные немануальные компоненты и в *УЖЯ*.

Получение списка интегральных (неделимых) немануальных компонент. Для получения типа немануального компонента для жестов, параметры которых не описаны в словарях *FSL* и *УЖЯ*, были проанализированы данные об известных немануальных компонентах из словаря *ASL* [5].

На основании этих данных получено несколько групп жестов:

- содержащих экспрессию (эмоциональные мимические выражения);
- указывающих время и место (содержат движения туловища);
- содержащих направление взгляда;
- соответствующих грамматическим выражениям (повороты головы, мимика и артикуляция): количественные и качественные прилагательные, логические операции.

Это позволило распространить знания о жестах из *ASL* и на данные из словарей *FSL* и *УЖЯ*.

Для этих целей анализировались одновременно несколько параметров жеста:

- название жеста, на основе которого делалось суждение о наличии признаков эмоционального оттенка слова, описывающего жест, либо о наличии грамматического значения при наличии жестов-синонимов, имеющих немануальные компоненты;

- лингвистическое описание, на основе которого сделан вывод о возможности принадлежности немануального компонента с указанными параметрами к классу немануальных компонент.

Затем был получен набор жестов, содержащих мимику, а также отдельная таблица, содержащая набор и типы немануальных компонент.

Получение набора признаков, входящих в выражения. После получения набора интегральных немануальных компонент возникла необходимость разбить каждое из лингвистических описаний этих компонент на составляющие. Критериями были: объект действия (область проявления) и действие (характер и степень проявления).

Каждый из структурных элементов лингвистических описаний был закодирован с помощью систем нотации мануальных и немануальных компонент ЖЯ *HamNoSys* [7], *SignWriting* [11] и *FACS* [12], что позволило представить сложное лингвистическое описание в виде вектора признаков.

Получение количественных значений изменений интегральных немануальных компонент. В постановке задачи исследования обозначен пункт «получение численных значений изменений немануальных компонент». Необходимость такой функции в базе данных вызвана несколькими причинами:

- в АБД представлена разнородная информация – фото и видео из различных источников, что не позволяет корректно выполнять сопоставление различных примеров аудиовизуальных данных одного и того же немануального компонента;

- числовые значения позволяют более точно сопоставлять одни и те же немануальные ком-

поненты без влияния геометрических искажений, вносимых фото и видеоаппаратурой;

- необходимостью унификации образцов немануальных компонент, что позволило бы представлять набор данных из одного множества в размерностях другого (подобная задача существует и в компьютерном моделировании – *retargeting* движений 3D-моделей);

- использование числовых значений немануальных компонент позволило бы использовать их в дальнейших разработках, связанных с моделированием, синтезом и распознаванием мимики;

- использование числовых значений совместно с 3D-моделью позволило бы существенно сократить объем аудиовизуальной базы данных, так как видео и изображения занимают больше места в памяти компьютера, чем 3D-модель и текстура.

Для этих целей авторами выбрана как математическая модель мимики *CANDIDE-3* [13], представляющая собой изменение точек лица 3D на основе гибких шаблонов, в единицах изменений компонент по системе *FACS*, а также позволяющая делать эти описания на основе стандартных компонент точек, используемых в стандарте кодирования видео *MPEG-4*.

Таким образом, полученные в результате анализа описания наборы признаков, входящих в интегральные компоненты, можно представить в виде изменений, описываемых кодами *FACS*, и адаптировать предложенную модель к имеющимся аудиовизуальным данным.

Данная модель позволяет изменять геометрические параметры лица (15 параметров), а также применять изменения, связанные с мимикой (10 параметров), что дает возможность описать форму и изменения любого человеческого лица в полученной аудиовизуальной базе данных с достаточно высокой степенью достоверности.

Анализ полученных результатов исследований. При анализе словаря УЖЯ, выяснилось, что собственно словарь, кроме мимических немануальных компонент, содержал также артикуляцию слов, обозначающих жест в разговорной речи. Однако эти данные не по-

зволюали сделать однозначного вывода относительно соотношения мимики и ЖЯ. Данный феномен объясняется тем, что артикуляция в ЖЯ и разговорной речи имеет различные назначения: в ЖЯ – для показывания слов в калькирующей жестовой речи, показывании незначимого для собеседника жеста или для обозначения сразу целого слова. В обычной разговорной речи артикуляция служит для модуляции звукового потока и (частично) для передачи экспрессии в произнесенных словах.

Наличие в жестовой речи одновременно и фонем разговорной речи при артикуляции слов, и мимики в жестах, содержащих грамматический оттенок, приводит к непониманию, поскольку фонемы разговорной речи имеют сходство с некоторыми мимическими выражениями. Это требует дополнительных исследований для установления причинно-следственных связей, поскольку отсутствует динамика в исследуемых материалах (изображения).

Для проверки достоверности описания изменения компонент, было решено адаптировать трехмерную модель лица к нескольким изображениям из АБД, получив коды изменений параметров в системе *FACS* и сравнив их с кодами в самой базе данных. Предварительные исследования на тестовых примерах дали положительные результаты. В дальнейшем предлагается провести их на всем имеющемся наборе аудиовидеоданных.

Модель базы данных

В качестве базы данных было предложено использовать БД *MS Access* с СУБД *Microsoft Access* с выводом содержимого таблиц в формы. На рис. 3 показана физическая структура прототипа базы данных.

Информация, полученная на этапе 1 (см. рис. 1), заносилась в таблицу *Sign*, которая содержала поля: название жеста (*Sign Name*), название немануального компонента (*NM Name*), язык (*Language*), *OLE* объект, содержащий видео или изображение (*ImageData*).

Информация, полученная на этапе 2, заносилась в таблицу *NM Component*, содержащую следующие поля: название немануального компонента (*NM Name*), тип немануального ком-

понента (*Type*), эмоциональный оттенок (*Scale*) и лингвистическое описание изменения немануальных компонент в жесте (*NM Fulldesc*).

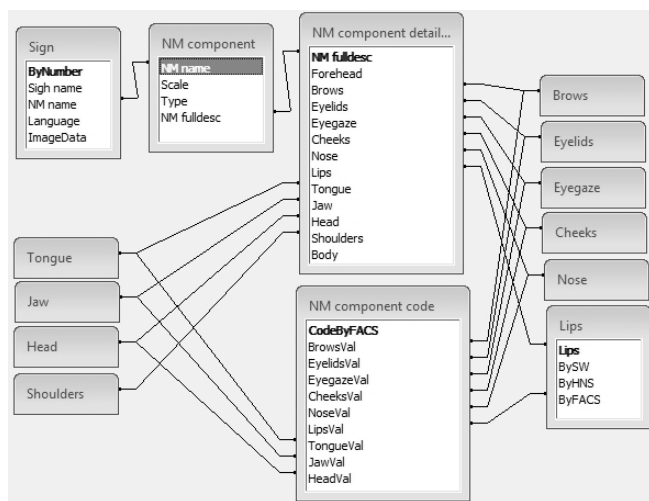


Рис. 3. Физическая модель данных в окне *Relationships* СУБД *MS Access*

Информация, полученная на этапе 3, заносилась в таблицу *NM Component Detailed*, содержащую поля: лоб (*Forehead*), брови (*Brows*), веки (*Eyelids*), направление взгляда (*Eyegaze*), щеки (*Cheeks*), нос (*Nose*), губы (*Lips*), язык (*Tongue*), челюсть (*Jaw*), голова (*Head*), плечи (*Shoulders*); а также в таблицы с именами, соответствующими именам полей таблицы *NM Component Detailed*, которые содержали списки соответствия описания каждой из компонент в системах *FACS*, *SignWriting* и *HamNoSys*.

Числовые параметры, получаемые на этапе 4 были источниками данных для таблицы *NM Component Code*, содержащей числовые параметры для соответствующих полей таблицы *NM Component Detailed*.

В настоящий момент база данных охватывает 975 жестов, содержащих мимику, в том числе: 344 (*FSL*), 526 (*ASL*), 105 (*УЖЯ*), 97 интегральных немануальных компонент, а также: эмоциональных – 70, грамматических – 27 и 88 составляющих немануальных компонент.

Отметим, что АБД все еще не совершенна, так как некоторые рутинные операции необходимо выполнять в полуавтоматическом режиме. При увеличении объема и разнообразия поступающих данных, как например при охвате

всего набора жестов в УЖЯ (несколько тысяч) [1], понадобится очень большое количество корректировок, связанных с уточнением уже существующих данных. Кроме того, необходимы средства поиска с наличием вывода в формы и элементов управления. Однако на данном этапе АВБД удовлетворяет поставленным требованиям, поскольку удалось проанализировать большой объем разнородных данных и получить на основе знаний о других языках знания, применимые к Украинскому жестовому языку и некоторым другим естественным языкам.

Заключение. В результате исследований удалось систематизировать имеющиеся данные по немануальным компонентам Украинского жестового языка, определить их количество, локализацию и структуру, а также определить дальнейшие направления исследования. В дальнейшем необходимо более подробно изучить взаимодействие мимики, жестов и артикуляции, а также других немануальных компонент, используя динамику процесса на основе параметрической модели *CANDIDE*, что позволит получить характеристики изменения во времени для каждой из составляющих немануальных компонент в процессе показывания жеста. Данные характеристики и свойства параметров модели могут быть непосредственно использованы для моделирования, синтеза и распознавания мимики.

1. Кульбіда С.В. Українська жести́ва мо́ва як приро́дна знакова система: Зб. наук. пр. «Жести́ва мо́ва й сучасність». – К. : Педагогічна думка, 2009. – С. 218–239.
2. Зайцева Г.Л. Жестовая речь. Дактилология: Учебник. – М.: ВЛАДОС, 2000. – 192 с.

3. Киммельман В. Нерегулярные отрицания в русском жестовом языке // Материалы конф. МСКЛ-2, Москва, 13 апр. 2007. – М.: Изд-во МГУ, 2007. – С. 89–103.
4. Ткачева Т., Ахмедышева Н. Жестовый язык глухих как особая знаковая коммуникативная система (на материале русского и французского жестовых языков) / Материалы Междунар. науч.-практ. конф. «Приоритеты и интересы современного общества». – Астрахань: АГУ, 2010. – С. 361–365.
5. Tennant R.A., Gluszak M. The American Sign Language Handshape Dictionary. – Br. – Washington: Gallaudet University Press, 1998. – 407 p.
6. Chetelat-Péle E., Braffort A., Veronis J. Annotation of Non Manual Gestures: Eyebrow movement description / Ed. by. O. Crasborn et al. // 3th Workshop on Representation and Processing of Sign Languages. Workshop proc. Int. Conf. on Language Resources and Evaluation (LREC), Marrakesh, Marrocco. – Paris: ELRA, 2008. – P. 28–32.
7. Interface Definitions, ViSiCAST Deliverable D5-1 / T. Hanke, G. Langer, C. Metzger et al. – Hamburg, Institute of German Sign Language and Communication of the Deaf, 2002. – 74 p.
8. <http://www.lat-mpi.eu/tools/elan/>
9. Companyys M., Tourmez F. Dictionnaire 1200 signes. – Angers: Éditions Monica Companyys, 2000. – 150 p.
10. Український жести́вник для батьків / С.В. Кульбіда, І.І. Чепчина, Н.Б. Адамюк та ін. // НАПН України, Ін-т спец. Педагогіки. Лаб. жести́вої мо́ви, Укр. т-во глухих. – К.: СПКТБ УТОГ, 2011. – 380 с.
11. Sutton V. Lessons in SignWriting. – La Jolla: Center for Sutton Movement Writing, 2002. – 195 p.
12. Miller F.P., Vandome A.F., McBrewster J. Facial Action Coding System. – Beau Bassin: Int. Book Marketing Service Ltd, 2011. – 80 p.
13. Ahlberg J. CANDIDE-3 – an updated parameterized face. – Linkoping: University, 2001 – 16 p.

Тел. для справок: +38 044 526-4118, 526-7418 (Киев)
 E-mail: krak@unicyb.kiev.ua, anton.ternov@gmail.com,
kuznetsow.wlad@gmail.com

© Ю.В. Крак, А.С. Тернов, В.А. Кузнецов, 2013