О распознавании жестов языка глухих[1]

About recognition of sign language gestures

Воскресенский А. Л. (avosj@yandex.ru), независимый исследователь,
Ильин С. Н. (mail@mocaprus.ru), «Академия фантазий», Москва
Milos Zelezny (zelezny@kky.zcu.cz), University of West Bohemia, Plzeň, Czech Republic

Обсуждаются задачи восприятия и распознавания жестов русского языка глухих в системе автоматизированного сурдоперевода. Предлагаются новый подход к морфологии жестов, метод выделения отдельных жестов жестового высказывания. Предлагается рабочее определение понятия «понимание текста».

Введение

Создаваемая система автоматизированного сурдоперевода [1] должна быть способна не только переводить воспринятый текст (и речевое сообщение) в жестовые высказывания, но и преобразовывать жестовые высказывания в слова и фразы. С последней задачей, как показывает практика, во многих случаях не справляются люди-сурдопереводчики [2]. Одной из причин этого может быть ошибочное разделение жестового выражения на составляющие жесты.

Турецкие участники проекта «Информационный киоск» [3], пытаются найти границы жестов, сопоставляя их с параллельно произносимыми диктором-сурдопереводчиком словами. Но этот подход ограниченно применим лишь при обработке «кальки» [4], он бесполезен при восприятии жестовой речи, в которой поток жестов не совпадает с потоком слов.

При создании систем машинного сурдоперевода возникает дополнительная проблема восприятия жеста системой. Например, в работах французских исследователей [5] возникали проблемы распознавания жеста при съемке одной камерой.

В данной работе описываются подходы к распознаванию жестов, лежащие в основе создания составной части системы сурдоперевода: подсистемы перевода «жесты — текст». В разделе 1 дается краткое описание используемого способа съемки жестов и преобразования их в 3D-анимацию, предлагаемый метод распознавания жестов; в разделе 2 приводятся соображения о морфологии жеста, основанные на исследовании жестов, представленных в словаре RuSLED [6], и методе разделения жестового высказывания на отдельные жесты; раздел 3 относится к пониманию словесных и жестовых высказываний.

1. Средства отображения, восприятия и распознавания жестов

Жест языка глухих представляет собой комбинацию конфигураций пальцев рук (одной или двух), положений рук относительно тела говорящего (с учетом направления движения рук) и сопутствующей мимики, передающей эмоциональную составляющую. При общении на «кальке» мимика включает в себя и артикуляцию, соответствующую произнесению (зачастую беззвучно) соответствующего слова. В «истинной» жестовой речи артикуляция используется для указания значения омонимичного жеста.

Передача столь сложной комбинации, осуществляемой в пространстве и времени, весьма затруднительна для записи. Для записи жестов используются различные варианты нотации, например Гамбургская система нотации (HamNoSys, http://www.sign-lang.uni-hamburg.de/projects/hamnosys.html). В России используется нотация, предложенная Л. С. Димскис [7].

Нотация HamNoSys использовалась в европейском проекте eSign (http://www.sign-lang.uni-hamburg.de/esign/) для управления движениями

[1] Данное исследование проводится при поддержке Министерства Образования, Молодёжи и Спорта Чешской Республики в рамках совместного проекта DIMAS-CZ, No. ME08106.

аватара Guido, демонстрирующем жесты. Мимику этот аватар не отображал.

В настоящее время эта нотация используется в проекте «Информационный киоск», в котором от России участвует СПИИРАН. Аватар, демонстрирующий жесты, разработан в университете Западной Богемии (Чешская республика). Несмотря на то, что дополненная версия HamNoSys имеет знаки для отображения некоторых элементов мимики, при управлении аватаром они не используются. Версия аватара, модифицированная СПИИРАН, артикулирует русские слова.

Словарь русского жестового языка RuSLED дополнен функцией поиска жеста по его описанию. Задача состоит в том, что нужно найти жест, который человек видел, но не знает его значение. Для этого используется упрощенная нотация, зашитая внутри словаря, для пользователя доступны списки возможных значений — текстовые для описания места исполнения жеста, текст с рисунком — для конфигураций пальцев. На основе выбранных пользователем значений формируется поисковый запрос и выдается набор жестов, отвечающих этому запросу, из которого пользователь выбирает интересующий его жест.

Демонстрация жестов в новой версии словаря осуществляется анимированным персонажем — аватаром, для записи жестов используется методика «захвата движений» (motion capture). Запись жестов проводится студией «Академия фантазий» (www.mocaprus.ru). Движения демонстратора, фиксируемые с помощью 12 камер и отражателей на костюме (рис. 1), преобразуются в 3D-модель (рис. 2), используемую для формирования облика аватара, который может быть помещен в любую сцену (рис. 3).

Движения пальцев демонстратора фиксируются с помощью специальных перчаток. Для снятия мимики и артикуляции используются фиксируемые на лице отражатели (рис. 4). Их сигналы преобразуются в трехмерную модель мимики лица (рис. 5).

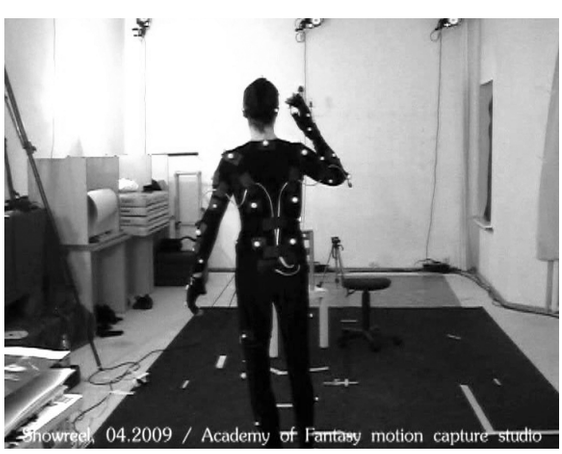

Рис. 1. Демонстратор в костюме с отражателями

Рис. 2. 3D-модель

Рис. 3. Аватар, построенный на основе 3D-модели и помещенный в виртуальную сцену

Реализация подобного преобразования позволит существенно ускорить наполнение словаря за счет использования нескольких сурдопереводчиков для демонстрации жестов, сохраняя при этом единство действия, выраженное единым обликом виртуального демонстратора жестов. Сформированный таким образом словарь позволит компоновать жестовые высказывания из хранящейся в словаре коллекции жестов, сохраняя, как указывалось выше, единство действия, что важно для восприятия жестовых высказываний человеком.

Студийная запись жестов, позволяя формировать исходный словарь, очевидно, не может быть средством коммуникации с глухими людьми.

Для распознавания жестов, например, при преобразовании жестовых высказываний в текст, предполагается разработка средств преобразования снятых видеокамерой растровых изображений сурдопереводчика в векторное изображение. Это преобразование включает в себя распознавание существенных для данной задачи деталей изображения: голова, кисти рук (и положение каждого пальца),

туловище. Указанные детали изображения преобразуются в эллипсы и прямоугольники, координаты которых сопоставляются с параметрами скелета виртуального демонстратора (аватара).

Рис. 4. Демонстратор с наклеенными на лицо отражателями

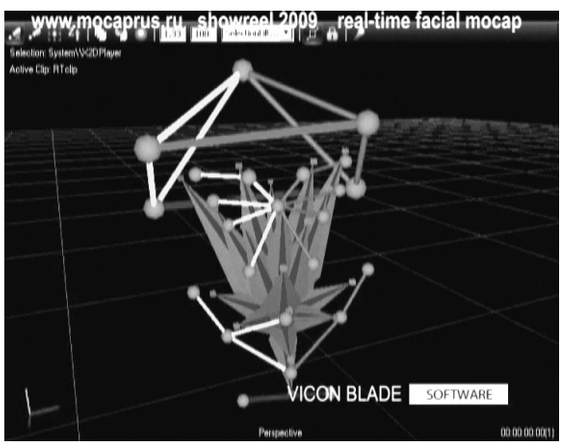

Рис. 5. Трехмерная модель лицевой мимики

Преобразование воспринятых изображений в векторную форму позволяет существенно сократить требования к объему памяти интеллектуальной системы и ускорить процедуры сравнения с эталонами. Представляется, что подобная методика может быть использована не только для распознавания жестов, но и для более широкого круга задач.

Методы преобразования изображения, которые предполагается использовать, близки к тем, которые используются при предварительной обработке изображений в системах распознавания символов (см., например, [8]).

Наиболее близким аналогом являются работы отдела Artemis Национального института телекоммуникаций Франции (http://www-artemis.it-sudparis.eu/web2.0/index.php?pid=1).

Отличием является то, что в разработках отдела Artemis ставится задача наиболее близкого отображения облика сканируемых объектов, тогда как в данном проекте внешний вид сканируемого объекта заменяется заранее заданным обликом аватара, но при этом ставится задача наиболее точного отображения положений пальцев, рук, туловища в целом. Информацию о точном положении в пространстве, например, рук, отсутствующую в двумерном растровом изображении, полученном от одной видеокамеры, планируется извлекать из знаний о возможных и допустимых взаимных положений различных частей тела человека. Для точного определения позы аватара будут применяться соответствующие геометрические построения, обеспечивающие наиболее близкое совпадение проекций аватара на плоскость отображения с исходным растровым изображением.

В случае достаточно надежного распознавания жестов с помощью одной видеокамеры (желательно добиться качественного распознавания с помощью типовых веб-камер) и создания системы сурдоперевода возможно будет обеспечить оперативную коммуникацию глухих с представителями администрации и общественности, что является одной из функций «электронного правительства».

2. О морфологии жестов

Считается, что впервые морфологию жестового языка описал W. Stockoe [9]. Соответственно, всякий жест этого языка (функционально близкий морфеме) складывается из хирем (от греч. χείρ — рука), делящихся на три класса — табы указывают на место исполнения жеста, дезы — на конфигурации руки, а сиги — на характер движения. Хиремы функционально эквивалентны фонемам, но в отличие от фонем, выстраивающихся в морфеме в линейную последовательность, в жесте-морфеме одновременно присутствует хирема каждого из трех классов. Общее количество хирем сопоставимо с числом фонем в звуковых языках — в ASL (американском жестовом языке) имеется 12 табов, 19 дезов и 24 сига, в шведском жестовом языке, соответственно, 18, 22 и 24, в языке глухих южной Франции — 16, 17 и 20 и т. д. W. Stockoe разработал для ASL систему записи жестов как последовательности таба, деза и сига. Эти положения лежат в основе разработки систем нотационной записи жестов.

Однако, как значения слов не всегда определяются составляющими морфами, завися от контекста, так и при анализе жестовых высказываний нужно учитывать, что многие жесты являются составными, содержат в себе комбинацию нескольких жестов и предшествующих дактильных знаков, модифицирующих значение данного жеста. Несколько примеров приведено в табл. 1.

Таблица 1. Примеры составных жестов

Наименование жеста	Предшествующие дактилемы	Составляющие жесты
абитуриент	—	учиться, войти
абрикос	а	Имитация разламывания плода
автомат (робот)	—	механизм, задание
азбука	а, б, в	—
акварель	а	краска
алый	—	красный, яркий
алюминий	а	металл
античность	а	старый, было
аттестат	а	печать, выдать
аудитория (помещение)	—	учиться, комната
аудитория (слушатели)	—	слушать, все
бронза	б	металл
девочка	—	женщина, маленький
душа	д	я
ненаглядный	—	смотреть, любить
озеро	—	вода, площадь (место)
океан	о	море
она (об отсутствующей при разговоре)	—	женщина, вы
пословица	—	говорить, предложение
раскаиваться	—	ошибка, ум
снег	—	белый, падать

Когда нужно передать, например, падежные окончания, вслед за жестом показываются соответствующие дактилемы.

Число составных жестов велико, но статистику приводить в настоящее время преждевременно. Это связано со сравнительно малым объемом имеющихся словарей жестов, что может привести к существенному смещению статистических оценок.

Приведенный весьма ограниченный (и, возможно, не самый показательный) набор примеров показывает, что жестовый язык изобилует образными и идиоматическими выражениями, значение которых не всегда совпадает с суммой значений составных элементов. Возможно, именно это приводит к ошибкам сурдопереводчиков при переводе жестов в текст и речь.

Жестовая речь не содержит пауз между отдельными жестами. Паузами разделяются лишь фразы. Это привносит дополнительные сложности при осуществлении автоматизированного сурдоперевода, напоминающие те, которые встречаются при разработке систем распознавания слитной речи.

Учитывая составной характер жестов, разделение жестовой фразы на отдельные жесты следует вести, выбирая из словаря соответствующие жесты, имеющие наибольшую длину, и анализируя семантику получаемого высказывания. В случае, если его значение не соответствует дискурсу, можно приступить к поочередному расщеплению «длинных» жестов на составные элементы, пытаясь получить высказывание, содержание которого соответствует дискурсу.

Это лишь предварительные предположения, которые подлежат экспериментальной проверке.

3. Сопоставление значений словесных и жестовых высказываний

При сурдопереводе, как и в других случаях перевода с одного языка на другой, основной задачей является правильная передача значения переводимого сообщения. Для этого необходимо понимать исходное сообщение, что является сложной психологической задачей [10].

В данной работе принято следующее рабочее определение понимания текста:

Результатом понимания текста должно быть опознание объектов, описанных в тексте, их пространственного положения, а также изменений их характеристик, действий и положения в соответствии с изменением времени текста.

Это определение было выработано для задачи перевода текста в жесты [6]. Однако оно действительно и для задачи перевода жестовых высказываний в текст. Даже еще в большей степени, учитывая необходимость определения объектов, описываемых одинаковыми жестами. Так, например, местоимения «*он*», «*она*», «*оно*» передаются одним и тем же жестом (в случае присутствия в месте разговора). Для нахождения правильного значения в этом случае необходим анализ всего дискурса [11].

Опознание объектов включает в себя не только выделение именных групп, описывающих тот или иной объект, но и распознавание того, встречался ли ранее в тексте данный объект, совпадают ли объекты, имеющие одинаковые имена [12].

В [13] указывается, что классическая линейная схема процесса обработки речи (рис. 6) с 60-х годов двадцатого столетия признается психолингвистами нереалистичной. Её предлагается заменить схемой, учитывающей взаимодействие различных этапов обработки и обратные связи между ними (рис. 7).

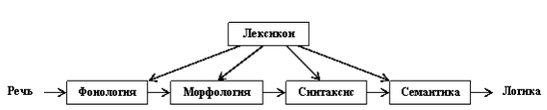

Рис. 6. Классическая схема обработки речи

При переводе с жестового языка из полученного в результате распознавания набора наименований жестов необходимо генерировать грамматически правильные фразы, отражающие содержание жестового высказывания. Это многоступенчатый процесс анализа получаемых вариантов, в ходе которого возможны возвращения на предшествующие этапы обработки, вплоть до выбора иного разделения жестовой фразы на составляющие фразы. Схема подобного процесса может быть близка к схеме, изображенной на рис. 7.

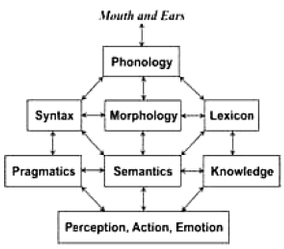

Рис. 7. Когнитивистский подход к пониманию речи

Как показано в [14], разрешение дейктической референции при обработке жестовых высказываний требует знания о пространственном положении объекта. Анализ пространственного положения объектов, описываемых в тексте, подразумевает способность интеллектуальной системы представлять визуальные образы, сопоставлять их друг с другом, расставлять на пространственной шкале «ближе» — «дальше». В [15] описан подход к созданию прототипа интеллектуальной системы, способной распознавать и сопоставлять визуальные и текстовые объекты. При этом каждая подсистема (в том числе концептуальной графики, обработки и синтеза текста) участвует в работе совместно с другими подсистемами, представляя общий результат на едином концептуальном языке прикладной онтологии, являющейся общим описанием окружающего мира для всех подсистем.

Заключение

Описываемая работа далека от завершения, однако и на данном этапе ожидаются полезные результаты. Использование методов motion capture позволяет быстро формировать словарь жестового языка. Так, например, во время одной из пробных съемок за час работы были сняты три серии по 30 отдельных жестов. Это позволяет надеяться, что пополнение словаря жестов может вестись со скоростью до 50 жестов в день. В качестве демонстраторов жестов используются глухие носители жестового языка, что позволяет снять замечания по содержанию словаря RuSLED, вызванные тем, что включенные в него жесты демонстрировались слышащими людьми, не вполне правильно изображавшими некоторые жесты.

Формирование объемного словаря жестового языка позволит уточнить сведения по морфологии жестов, а также послужит основой формирования библиотеки шаблонов для планируемой работы по распознаванию жестов.

Благодарность

Авторы выражает искреннюю благодарность Н. А. Чаушьян, чьи замечания позволили обратить внимание на содержащиеся в словаре RuSLED ошибки и принять меры к их исправлению. Её весьма полезные замечания по жестовому языку позволили обратить внимание на особенности, не всегда заметные исследователю, не являющемуся носителем языка.

Литература

1. *Voskressenski A.* Signs and speech: two forms of human communication // Proceedings of the Ninth International Conference «Speech and Computer» SPECOM'2004. Saint-Petersburg, Russia, 2004, P. 666–669.
2. *Овсянникова Л. А.* Проблемы жестового перевода на телевидении // Русский жестовый язык и проблемы перевода: Материалы конференции. — М., 2001.
3. *Hruz M., Campr P., Karpov A., Santemiz P., Aran O. and Zelezny M.* Input and Output Modalities Used in a Sign-Language-Enabled Information Kiosk // Proceedings of the 13-th International Conference "Speech and Computer" SPECOM'2009. — St. Petersburg: SUAI, 2009. — P. 113 — 116.
4. *Зайцева Г. Л.* Жестовая речь. Дактилология: Учебное пособие для ВУЗов. — М.: ВЛАДОС, 2000.
5. *T. Zaharia, F. Prêteux.* Video archiving and sign language indexation within the AMIS platform // Proceedings IASTED Conference on Signal Processing, Pattern Recognition and Applications (SPPRA'02), Crete, Greece, 2002, p. 396–401.
6. *Воскресенский А. Л., Гуленко И. Е., Хахалин Г. К.* Словарь RuSLED как инструмент семантических исследований. // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27–31 мая 2009 г.). Вып. 8 (15).– М.: РГГУ, 2009. — С. 64–68.
7. *Димскис Л. С.* Изучаем жестовый язык: Учеб. пособие для студ. дефектол. фак. высш. пед. учеб. заведений. — М.: Издательский центр «Академия», 2002. — 128 с.
8. *Масалович А. А.* Кластеризация изображений графем на основе непрерывного гранично-скелетного представления. // Математические методы распознавания образов ММРО-12: Доклады 12-й Всероссийской конференции. — М., 2005. — С. 374–378.
9. *Stockoe W. C.* Sign language sructure: An outline of the usual communication system of the American Deaf / Buffalo 14, New York University of Buffalo, 1960. — 79 p.
10. *Лурия А. Р.* Понимание компонентов речевого высказывания // Языки сознание / Ред. Е. Д. Хомской. — М.: Изд. МГУ, 1979. — С. 217–226.
11. *Ван Дейк Т. А., Кинч В.* Стратегии понимания связного текста. // Новое в зарубежной лингвистике. — Вып. 23. Когнитивные аспекты языка. — М., 1988.
12. *Kazi Z., and Ravin. Y.* Who's Who? Identifying Concepts and Entities across Multiple Documents.// Proceedings of the 33rd Hawaii International Conference on System Sciences — 2000. (0-7695-0493-0/00).
13. *Majumdar A., Sowa J., and Stewart. J.* Pursuing the Goal of Language Understanding // Proceedings of the 16th ICCS /P. Eklund and O. Hammerlé, eds. — LNAI 5113, Springer, Berlin, 2008, pp. 21–42.
14. *Кибрик А. А., Прозорова Е. В.* Референциальный выбор в русском жестовом языке. // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2007» (Бекасово, 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. — С. 220–230.
15. *Хахалин Г. К., Воскресенский А. Л.* Мультизадачное использование прикладной онтологии. // Одиннадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2008 (28 сентября — 3 октября 2008 г., г. Дубна, Россия): Труды конференции. В 3 т. — М.: ЛЕНАНД, 2008. — Т. 1. С. 112–123.