

© 2011 г. А.А. КАРПОВ

КОМПЬЮТЕРНЫЙ АНАЛИЗ И СИНТЕЗ РУССКОГО ЖЕСТОВОГО ЯЗЫКА

В статье описываются характеристики жестового языка, используемого в коммуникации с неслышащими людьми, даны его подтипы и функции, приведен анализ особенностей русского жестового языка и его отличий от иных жестовых языков мира, кратко описаны свойства его лексических элементов. Также приводится аналитический обзор компьютерных средств обработки жестового языка, в том числе систем жестовой нотации и существующих в мире систем жестового синтеза и преобразования «текст-в-жесты». Предложена идея создания универсальной многомодальной системы аудиовизуального синтеза звучащей и жестовой речи по тексту на русском языке.

ВВЕДЕНИЕ

Жестовый язык (ЖЯ) является одним из основных способов межчеловеческой коммуникации, в котором для передачи информации используются различные кинетические средства: жесты рук, мимика, артикуляция губ. Согласно общепринятому определению, «жест» (от лат. *gestus* 'движение тела') – это некоторое действие или движение человеческого тела или его части (например, рук или головы), имеющее определенное значение в конкретной ситуации. Например, жестом головой является утвердительный кивок, отрицательное покачивание, наклон головы при приветствии и т. д., причем у разных народов такие распространенные жесты могут иметь различные толкования, иногда даже полностью противоположные. Многие жесты выполняют функцию целого высказывания: указание пальцем или рукой на некоторый объект, улыбка, рукопожатие, воздушный поцелуй. Однако совокупность всех этих жестов не составляет самостоятельную коммуникативную систему, оставаясь вспомогательным компонентом вербального языка. О языке жестов можно говорить только в том случае, когда жесты являются основным способом коммуникации.

В настоящее время жестовые языки известны как средство коммуникации людей с нарушениями речи и слуха. ЖЯ является вторым государственным языком межчеловеческого общения в США, Финляндии, Испании и др., что закреплено конституцией этих стран. В России ЖЯ до сих пор официально не признан национальным языком общения, хотя дискуссии об этом идут в связи с присоединением России в сентябре 2008 г. к Конвенции ООН о правах инвалидов, в которой ЖЯ отводится важная роль по обеспечению доступности информации и равных возможностей для людей с нарушениями слуха.

* Данное исследование проводится при поддержке Совета по грантам Президента РФ (проект МК64898.2010.8), Министерства образования и науки РФ в рамках ФЦП «Кадры» (госконтракт 14.740.11.0357), а также Комитета по науке и высшей школе Правительства Санкт-Петербурга.

1. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ РУССКОГО ЖЕСТОВОГО ЯЗЫКА

Основным способом межчеловеческой коммуникации в среде глухих людей является национальный жестовый язык, в котором каждому понятию (или группе синонимичных понятий) соответствует определенный уникальный жестовый эквивалент. Не существует одного универсального ЖЯ, они различны в разных странах, так как возникают и развиваются естественным путем в разных локальных сообществах и изменяются со временем в связи с появлением новой лексики. ЖЯ состоит из разных знаковых систем, совмещая в себе язык неслышащих людей и жесты, используемые в обществе в целом. На рисунке 1 приводятся примеры жестов из разных ЖЯ, обозначающие одно и то же понятие – «СТОП» (остановка, завершение). Здесь показаны финальные статические стадии выполнения данных жестов, так как сам жест выполняется в динамике. Это понятие в РЖЯ (русском жестовом языке) изображается как остановка некоторого механизма поворотом умозрительного рычага в направлении от диктора. В других ЖЯ это понятие может, например, показываться как установка определенного ограждения (стены или шлагбаума) одной или двумя руками.

Очевидно, что различия в ЖЯ вызывают трудности как при экспертном переводе с языка на язык, так и при разработке автоматических компьютерных систем обработки и синтеза ЖЯ. Анализ показывает, что лишь менее 30 % наиболее распространенных жестов являются одинаковыми для разных ЖЯ, это в основном простые жесты, указывающие на объекты в поле зрения человека («Я», «ТЫ», «ГОЛОВА», «НОС» и т. д.), тогда как жесты для абстрактных понятий зависят от национальных и культурных особенностей. Производились попытки создать универсальный ЖЯ, например, «Жестуно» – это ЖЯ, сконструированный специально для общения неслышащих из разных стран мира [Гейльман 1980] и являющийся фактически аналогом эсперанто. Этот язык обладает своей собственной грамматикой и словарем, однако широкого распространения за границами России он не получил.

Помимо ЖЯ глухие люди в своем общении используют две иные коммуникационные системы, которые сами по себе полноценными и самостоятельными языками не являются: дактильную азбуку и калькирующую жестовую речь. Дактильная азбука служит для жестовой передачи букв алфавита посредством пальцев рук. В силу специфики передаваемого материала, она используется не как основное средство коммуникации неслышащих людей, а как важный вспомогательный инструмент. В том случае, если говорящий не знает жеста для какого-либо понятия или ему необходимо передать точное звучание некоторого слова, он прибегает к дактильной азбуке, которая используется также для передачи различных имен собственных, аббревиатур, редко встречающихся слов, слов иностранного происхождения. Можно выделить характерные отличия русской дактильной азбуки от, например, британской: 1) все 33 буквы русского алфавита показываются одной рукой, а буквы британского – двумя руками; 2) ряд букв русской азбуки отображается посредством динамических жестов, в то время как жесты британ-



Рис. 1. Жесты, означающие понятие «СТОП» в различных ЖЯ мира (слева направо: русский, британский, французский, испанский и чешский языки) [ЭР1]

ской азбуки статические; 3) конфигурации кисти и пальцев рук в русском дактильном алфавите более сложны. Большинство стран имеет свою оригинальную дактильную азбуку, зависящую от социокультурных особенностей народов, причем одноручные азбуки (в Германии, Франции, США, Китае, Индии и др.) преобладают в мире над двухручными (в Великобритании, Турции, Чехии, Новой Зеландии и др.) [ЭР2].

Калькирующая жестовая речь (КЖР) является иным способом передачи письменного или устного языка в среде глухих. «Калькирующая жестовая речь» (данный термин предложен Г.Л. Зайцевой) представляет собой систему жестов, предназначенную для визуального воспроизведения (калькирования) устной речи. При этом жесты выступают как эквиваленты слов, а порядок их следования аналогичен расположению в предложении, жестам в КЖЯ соответствуют слова в начальной форме. При использовании КЖР дополняется жестами из дактильного алфавита, применяемыми для показа неизвестных человеку жестов-слов или для слов, пока не имеющих аналогов в жестовом словаре, различных аббревиатур и др. КЖЯ не имеет собственной грамматики, а копирует структуру общепринятого письменного и устного языка, поэтому является вторичной знаковой системой [Зайцева 2000]. В лексике КЖР выделяются два основных класса жестов: 1) жесты, заимствованные из разговорного РЖЯ (например, «ЖЕНЩИНА», «ШКОЛА»), которые составляют подавляющее большинство в КЖР; 2) жесты, существующие исключительно в КЖР, которые в свою очередь подразделяются на три подкласса: а) собственно слова-жесты («ДИСКРИМИНАЦИЯ»); б) слова русского языка, воспроизводимые с помощью дактилологии («К-И-Б-Е-Р-Н-Е-Т-И-К-А»); в) слова, включающие жест РЖЯ и несколько букв из дактильного алфавита («ЧТО-Б-Ы»). Основное функциональное назначение КЖР состоит в использовании в ситуациях «официального» общения, тогда как РЖЯ является разговорным языком бытовой коммуникации в среде неслышащих людей [Пенин 2005]. Также КЖР весьма удобна для сопровождения фильмов и телепередач; так, в некоторых странах, например в Турции, теленовости дублируются сурдопереводчиками исключительно с помощью КЖР с некоторыми упрощениями [Aran et al., 2008].

2. ИНФОРМАЦИОННЫЕ ЛЕКСИЧЕСКИЕ РЕСУРСЫ РЖЯ

Исторически РЖЯ происходит от французского ЖЯ и относится к французской семье ЖЯ, куда входят также американский (так называемый Амслен ASL), ирландский, нидерландский, испанский, мексиканский, бразильский и некоторые другие национальные ЖЯ. РЖЯ, помимо французского наследия, также имеет определенный пласт лексики из немецкого и австрийского ЖЯ, вследствие чего лексика РЖЯ близка многим жестовым языкам мира. Однако как и другие ЖЯ, РЖЯ имеет много слов, заимствованных из национального звучащего языка или трансформировавшихся под его влиянием. Жесты для некоторых понятий основаны на дактилировании букв, с которых начинаются эти слова в русском языке, и, следовательно, эти жесты уникальны для РЖЯ. В республиках бывшего СССР РЖЯ распространялся через создание школ для глухих. В результате сформировалось несколько схожих друг с другом «диалектов» РЖЯ¹. В последние годы наиболее активные научные исследования диалектов русского жестового языка проводились в МГПУ под руководством А.А. Комаровой.

Существование различных диалектов РЖЯ, различающихся в основном лексикой, является определенной трудностью для неслышащих людей и сурдопереводчиков, возникают проблемы при переводе и понимании жестовой речи дикторов из других регионов и стран СНГ. Значительные проблемы встают и перед разработчиками компьютерных систем и мультимедийных обучающих словарей ЖЯ. На рисунке 2 представлено разнообразие жестов, означающих понятие «ОДИН» (цифра), взятых из различных словарей РЖЯ (пяти мультимедийных словарей и печатного). Можно отметить три основных и отличных друг от друга варианта жестикуляции для данной цифры, принятые

¹ Википедия <http://ru.wikipedia.org>, статья «Русский жестовый язык».

в разных локальных обществах глухих в России. Зачастую члены одного общества не принимают жесты, принятые в другом сообществе, хотя многие из них интуитивно понятны. Сейчас в России доступно несколько печатных книг и электронных ресурсов и обучающих систем, содержащих словари РЖЯ. Среди существующих мультимедийных компьютерных систем можно отметить следующие:

- Видеословарь «Тематический словарь русского жестового языка» создан Московской городской организацией ВОГ при поддержке Департамента социальной защиты населения Москвы по правительственной целевой программе «Социальная интеграция инвалидов Москвы на 2004–2006 годы» [ЭР3]. Содержит несколько тысяч наиболее употребительных жестов, показанных московскими носителями языка, состоит из четырех DVD дисков и имеет высокое качество видеоматериалов (рисунок 2а).

- Мультимедийный видеословарь «Толковый словарь русского жестового языка – RuSLED» [Воскресенский и др. 2009], руководитель проекта А.Л. Воскресенский, содержит более 2500 жестов. Достоинством данного ресурса можно назвать наличие дополнительных текстовых разъяснений, относящихся к манере исполнения жеста и описывающих смысловую интерпретацию жестов. Программа является интерактивной оболочкой к видеословарю РЖЯ (носителей ЖЯ Санкт-Петербурга и Ленинградской области), записанному в 2002 г. в Межрегиональном центре реабилитации лиц с проблемами слуха г. Павловска [Специфические средства... 2002], (рисунок 2б).

- «Русский жестовый язык. Базовый курс» – мультимедийная электронная обучающая система (ЭОС) 2001 г. создана Центром «Истина» при поддержке Центрального управления ВОГ, руководитель проекта А.И. Ракитов. Основывается на словаре нормированной жестовой речи и содержит несколько тысяч жестов [ЭР4] (рисунок 2в).

- Интерактивный онлайн видеословарь «Русский жестовый язык – DigitGestus» [ЭР5] содержит несколько сотен жестов-слов, создан в Новосибирске в 1996–97 гг. (рисунок 2г).

- Интерактивный онлайн словарь жестового языка европейского проекта Spreadthesign («Распространим жест») [ЭР1], проект поддерживается Евросоюзом, координатор от России – А.А. Комарова. Содержит несколько тысяч жестов-слов на РЖЯ, а также нескольких других европейских ЖЯ (рисунок 2е).

- Интерактивный онлайн словарь РЖЯ [ЭР6] собран носителями языка, работающими в Стэнфордском университете США, содержит несколько сотен жестов.

- Интерактивный онлайн словарь проекта «Сурдосервер» [ЭР7], разработан в ИПУ РАН им. В.А. Трапезникова, содержит те же видеоматериалы, что и ЭОС «Русский жестовый язык. Базовый курс» 2001 г., а также оригинальный жестовый видеословарь по направлению «Информационные технологии», состоящий из нескольких десятков жестов [Мясоедова и др. 2010].

Среди печатных словарей РЖЯ, содержащих фотографии и описания жестов, необходимо прежде всего отметить работы И.Ф. Гейльмана (Санкт-Петербург) [Гейльман 1979] (рисунок 2д), Г.Л. Зайцевой (Москва) [Зайцева 2000], Р.Н. Фрадкиной (Москва) [Фрадкина 2001], Л.С. Димскис (Минск) [Димскис 2002] и некоторых других исследователей РЖЯ. Все эти мультимедийные и печатные словари РЖЯ достаточно полно отражают его лексику.

Существующие отличия в диалектах РЖЯ можно связать и с историей возникновения первых школ сурдопедагогов. Первая сурдопедагогическая школа была открыта в начале XIX века в Павловске около Санкт-Петербурга, работала она по французской методике. В результате РЖЯ оказался в родстве с ЖЯ Франции и Америки. В Москве же первая сурдопедагогическая школа открылась только во второй половине XIX века, и она работала уже по иной – немецкой – методике. Последствия конкуренции этих двух методик отражаются в диалектах РЖЯ до сих пор. По результатам проведенного анализа мультимедийных жестовых словарей можно сказать, что крупнейшие диалекты РЖЯ – петербургский и московский – различаются для 30–40 % жестов-слов, хотя иногда расхождения незначительны и значение жеста понятно. Это хорошо видно на рисунке 2, где жесты из словарей петербургского диалекта представлены в сред-

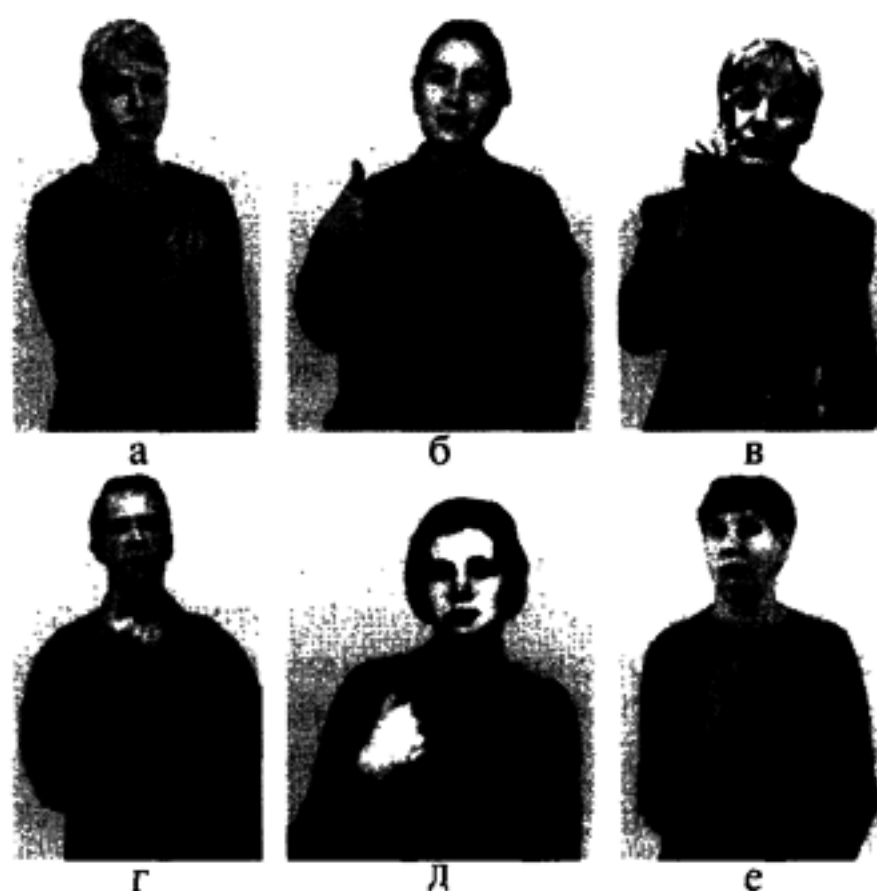


Рис. 2. Разнообразие жестов, означающих понятие «ОДИН», в словарях РЖЯ

ней колонке, данный жест визуально практически аналогичен жесту для иного понятия («ХОРОШИЙ») московского варианта РЖЯ. И даже в московском регионе разные словари РЖЯ не всегда унифицированы и нормализованы, т. е. существуют поддиалекты в отдельных общинах, например жест на рисунке 2в из словаря ЭОС «Русский жестовый язык», который разработан Центральным правлением ВОГ, отличается от более распространенного московского варианта; подобная ситуация наблюдается и в других регионах. Таким образом, разработка компьютерных средств анализа и синтеза русской жестовой речи не может быть универсальна, но можно говорить об адаптации к региональным особенностям произношения и регионально ориентированной компьютерной системе.

В отличие от КЖР РЖЯ обладает собственной грамматикой и выразительными средствами, причем по своей структуре грамматика ЖЯ значительно отличается от грамматики литературного или разговорного языка. Подобно тому как ЖЯ существуют и развиваются отдельно от вербальных, так и их грамматики слабо связаны друг с другом. В настоящее время грамматика и синтаксис РЖЯ еще недостаточно изучены и формализованы, чтобы вести разговоры о полностью автоматическом переводе текста на ЖЯ. В последние годы изучением структуры и грамматики РЖЯ наиболее активно занимается коллектив под руководством А.А. Кибрика на филологическом факультете МГУ [Кибрик, Прозорова 2007; Прозорова 2007]. Некоторые исследования по автоматизированному сурдопереводу ведутся в МГТУ им. Н.Э. Баумана совместно с ИПУ РАН [Мясоедова, Филиппович 2010], в том числе в рамках развития интернет-портала «Сурдосервер». Ряд исследователей [Воскресенский 2009] отмечает, что серьезные различия в семантико-синтаксической структуре письменного и жестового языков не позволяют выполнять однозначный машинный перевод русскоязычных текстов на РЖЯ, поэтому действующих автоматических систем сурдоперевода на данный момент не существует. Для создания такой полноценной модели необходимо производить глубокий семантический анализ письменных фраз, а это пока возможно лишь на поверхностном уровне из-за несовершенства алгоритмов и баз знаний. Однако КЖР отражает грамматику и лексику разговорной речи, поэтому машинный перевод здесь возможен.

3. СИСТЕМЫ ФОРМАЛИЗАЦИИ И НОТАЦИИ ЖЕСТОВОГО ЯЗЫКА

Лексические элементы ЖЯ — жесты — во всех типах жестовых систем формируются приблизительно одинаково и могут быть описаны схожим образом: по конфигурации (форма руки или рук), по месту исполнения (где находятся руки, выполняющие жест) и по характеру движения (как именно движутся руки; что они делают). Такое описание было впервые предложено американским исследователем жестовой речи У. Стоуки [Stokoe 1960]. В своем описании американского ЖЯ (Амслен) он применял «фонологический» принцип для транскрибирования жестов. Под фонологическим принципом понимается разложение любого жеста на минимальные регулярно повторяющиеся жестовые примитивы (хиремы). Иными словами, хиремы — это минимальные единицы, из которых строится жест. Однако специфика структуры жеста заключается в своеобразии связей между его компонентами: все они воплощаются в жесте одновременно. Этим жест принципиально отличается от слова, в котором звуковые элементы речи (фонемы) реализуются последовательно во времени. Хиремы, по концепции У. Стоуки, определяются по тому, где они жестикулируются (например, у лба, у щеки, на уровне груди и т. п.) и как именно они жестикулируются (например, рука движется вперед, совершает волнообразные или зигзагообразные движения и т. п.). Кроме того, для хирем необходим третий признак — форма руки/рук.

Специфика организации жестового общения человека с ЭВМ состоит в том, что ЖЯ и жестовый словарь должны быть определенным образом формализованы, чтобы компьютер мог обрабатывать и синтезировать жесты. Для описания жеста по его визуальным признакам существует несколько различных систем нотации (жестовой транскрипции), позволяющих зафиксировать представление жеста. Одна из самых известных систем нотации, широко распространенная в Америке (в частности, для Амслена), называется Sign Writing [ЭР8]. Знаки в SignWriting обозначают движения рук, тела человека и пишутся сверху вниз. Эта система отличается хорошей проработанностью и однозначностью записи, однако обладает чрезвычайно большим инвентарем знаков, что делает ее малоприменимой для использования в обучении и компьютерных технологиях. Намного практичнее сократить количество частных случаев конфигурации руки, описывая их как переменные (например, количество показываемых пальцев) при одной константе (форме ладони). Такому требованию, например, удовлетворяет система нотации жестов, предложенная Л.С. Димским [Димский 2002]. В ее нотации выделяется более 30 конфигураций рук, около 50 характеристик места исполнения жеста и более 70 типов движений рук.

Наиболее широкое распространение в мире в последнее время получила «Гамбургская система нотации» (основатель и основной разработчик Т. Ханке), более известная под сокращенным названием HamNoSys (англ. «Hamburg Notation System»). В настоящее время актуальна HamNoSys версии 4.0 [ЭР9]. Эта система отличается наибольшей проработанностью и пригодна для использования в компьютерных приложениях за счет того, что ее знаки переведены в компьютерную систему кодировки Юникод с соответствующими компьютерными шрифтами. Инвентарь HamNoSys позволяет представить практически любой жест, что делает эту систему универсальной и пригодной для любого ЖЯ мира. Фактически по своим функциям HamNoSys является аналогом международного фонетического алфавита МФА (IPA) и по умолчанию используется лингвистами, занимающимися исследованием ЖЯ.

В системе HamNoSys при записи любого жеста сначала задается форма кисти активной руки (или рук), затем ее ориентация по двум параметрам (направление пальцев и разворот кисти), место, где располагается рука во время жеста, и характер самого движения. Жесты могут быть одnorучными и двуручными, поэтому во втором случае в запись добавляются операторы для двуручных жестов, позволяющие совместить действия обеих рук в рамках одного выражения. Форма кисти, т.е. конфигурация пальцев и ладони, задается несколькими простыми знаками, символически изображающими ладонь и пальцы, примеры представлены на рисунке 3.

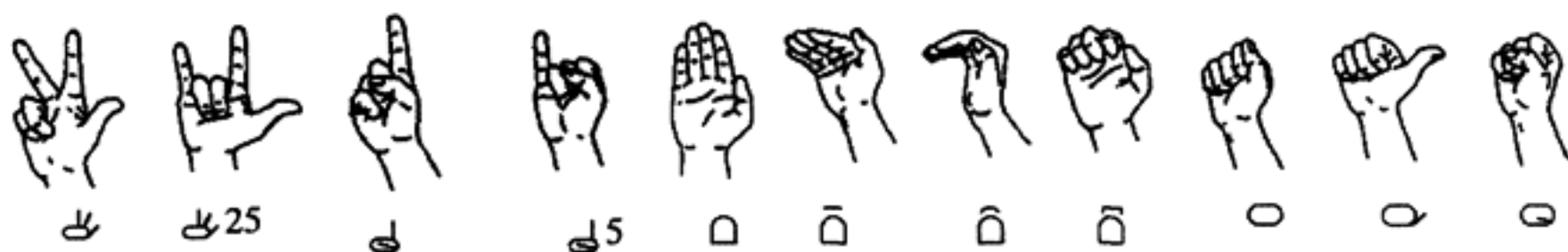


Рис. 3. Примеры кодирования конфигурации пальцев и кистей рук в HamNoSys

Стоит особо отметить способы указания отдельных пальцев на руке. В базовых символах нет элементов, например, для отдельно отставленного безымянного или указательного пальцев. В этом случае используется числовое обозначение пальцев рук: 1 – большой палец; 2 – указательный; 3 – средний; 4 – безымянный; 5 – мизинец; а также их комбинации. Соответственно для точной кодировки пальцев следует всего лишь задать их количество и точное значение цифрами (рисунок 3). Такой подход позволяет сократить инвентарь знаков, сделать его менее громоздким, удобным в обучении и применении по сравнению с системой Sign Writing. Кроме того, задается ориентация кисти руки в пространстве по кончикам пальцев; она определяется в трех плоскостях, в зависимости от специфики жеста (рисунок 4).

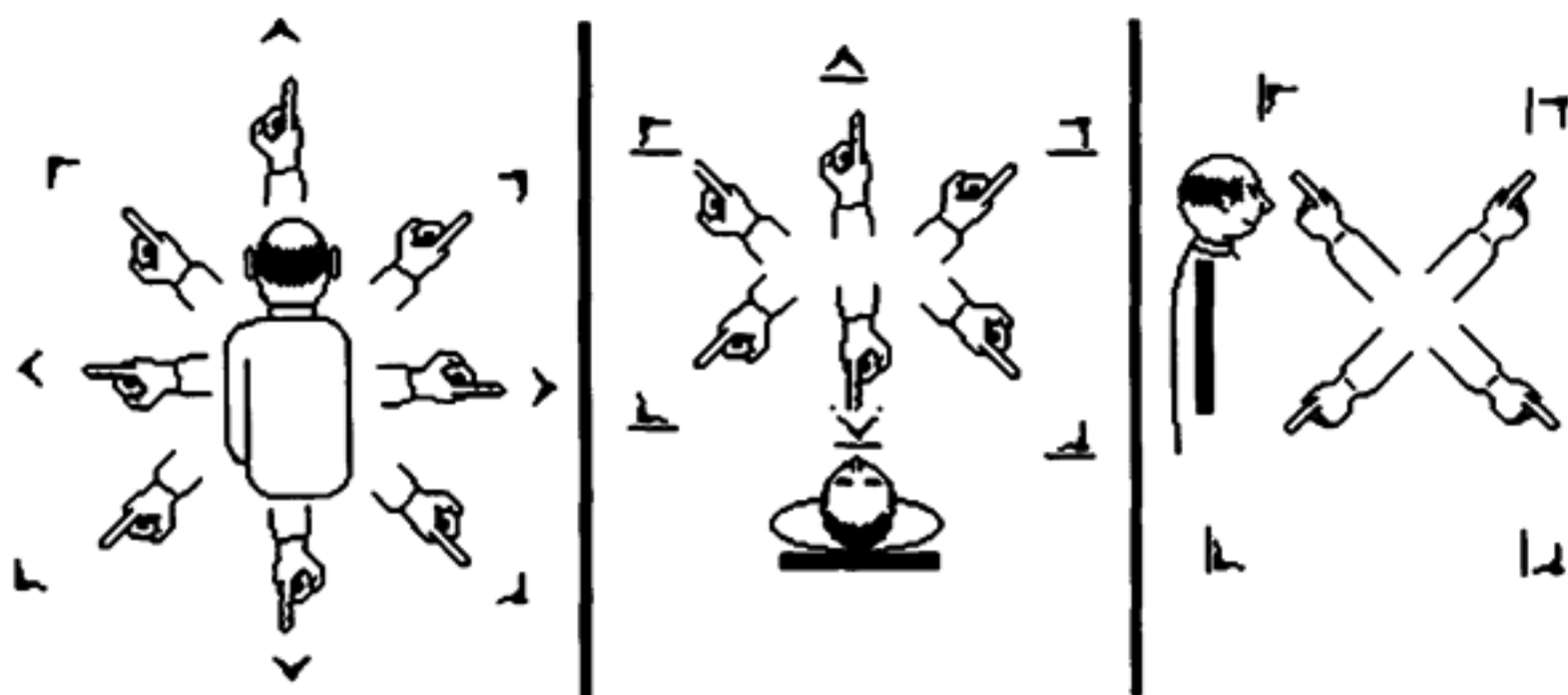


Рис. 4. Описание ориентации руки в нотации HamNoSys

Наконец, в HamNoSys существуют разнообразные знаки практически для всех возможных движений, которые могут осуществлять руки, а также жестовых локаций. Их довольно много (около сотни в версии 4.0 [ЭР9]), и этот инвентарь постоянно расширяется, чтобы соответствовать новой лексике, появляющейся в ЖЯ, который является открытой коммуникативной системой. В качестве примера использования данной нотации возьмем жест для числительного «ОДИН» (рисунок 2а), каждая характеристика которого записана при помощи знаков HamNoSys. Данный статический жест записывается следующей цепочкой знаков: $\cup \wedge \circ \square \square \uparrow$. Подобным образом, используя систему жестовой нотации, можно графически формализовать весь словарь жестов автоматической системы.

4. КОМПЬЮТЕРНЫЙ ВИДЕОСИНТЕЗ ЖЕСТОВОЙ РЕЧИ

Глобальной целью компьютерных систем, основанных на жестовом языке и речи, является обеспечение равноправной коммуникации слышащих и глухих (не немых, а разговаривающих на ЖЯ) людей с тяжелыми нарушениями слуха. Одним из самых эффективных средств обучения являются мультимедийные компьютерные системы, по-

этому создание информационных приложений, способных взаимодействовать с ЖЯ, – одна из приоритетных задач при работе с глухими и слабослышащими людьми. Особый интерес представляют системы компьютерного синтеза русского жестового языка и речи по тексту.

Наиболее простая система синтеза ЖЯ может быть создана на основе компьютерного видеословаря жестов, целиком показанного одним человеком-демонстратором. В такой системе видеосфрагменты нужных слов из словаря последовательно компилируются в составные фразы, в качестве примеров можно привести системы компьютерного синтеза РЖЯ, разработанные в последние годы в Новосибирске [ЭР10], а также в Минске [ЭР11]. Серьезным недостатком любых подобных систем можно назвать то, что такой синтез ЖЯ не является слитным (непрерывным), где жесты естественным образом перетекают из одного в другой, а лишь воспроизводит последовательность изолированных друг от друга жестов. Дело в том, что в словаре все жесты-слова показываются целиком от начала до конца и независимо друг от друга, начинаясь всегда с определенного нейтрального положения рук и, как правило, им же и заканчиваясь. В случае слитного синтеза ЖЯ последовательные жесты должны не проходить через это нейтральное положение, а плавно перетекать из одного жеста в другой без остановки, как это делается человеком. Иными словами, студийная видеозапись жестов, позволяя сформировать исходный словарь, не может быть средством коммуникации с глухими людьми, поэтому для компьютерного синтеза слитной жестовой речи предпочтительнее использование анимированных персонажей или виртуальных помощников-людей (аватаров). Словарь анимированных трехмерных жестов позволяет компоновать высказывания из хранящейся в словаре коллекции жестов, сохраняя единство действия, что важно для восприятия жестовых высказываний человеком. Системы синтеза ЖЯ, создаваемые с использованием трехмерных аватаров, обладают рядом достоинств, в частности:

- Позволяют просматривать видеосинтез жестовой речи с разных сторон и углов обзора, что помогает лучше воспринимать пространственную информацию, например, степень удаленности рук от туловища и друг относительно друга (в отличие от двухмерных моделей).

- Дают возможность относительно легко пополнять и корректировать словарь жестов, так как вместо видеозаписей реального человека в словаре присутствуют компьютерные анимированные аватары, соответственно для расширения словаря не обязательно записывать того же самого человека-демонстратора жестов в той же самой одежде и с той же прической, а также уровнем освещенности.

- Позволяют выполнять слитный синтез ЖЯ, в котором отдельные слова во фразах стыкуются бесшовно, т. е. не видны явные границы между соседними словами.

- Дают возможность заменять виртуальный персонаж, используя новые высокореалистичные модели людей (мужчин или женщин, а также любых вымышленных героев).

- Позволяют воспроизводить жестовую фразу на экране с любой необходимой скоростью, как замедляя, так и ускоряя видеоряд.

Одним из возможных способов создания высокореалистичной модели компьютерного синтеза жестовой речи является технология цифрового «захвата движений» (от англ. «motion capture»), использующая систему из нескольких видеокамер или радиопередатчиков. По такому пути под руководством А.Л. Воскресенского в московской анимационной студии создается видеословарь жестов московского варианта РЖЯ и трехмерный анимированный персонаж для системы автоматического синтеза слитной жестовой речи [Воскресенский и др. 2010]. Движения демонстратора – носителя ЖЯ, фиксируемые с помощью двенадцати синхронизированных цифровых видеокамер комплекса Vicon MX и десятков цветных меток на специальном костюме, преобразуются в трехмерную модель человека с использованием новейших возможностей компьютерной графики. Далее эта модель обрабатывается для формирования визуального облика аватара, который затем может быть помещен в любую виртуальную сцену, управляясь игровым движком. Движения пальцев демонстратора фиксируются с помощью специальных перчаток с метками, а для снятия мимики лица и артикуляции используются

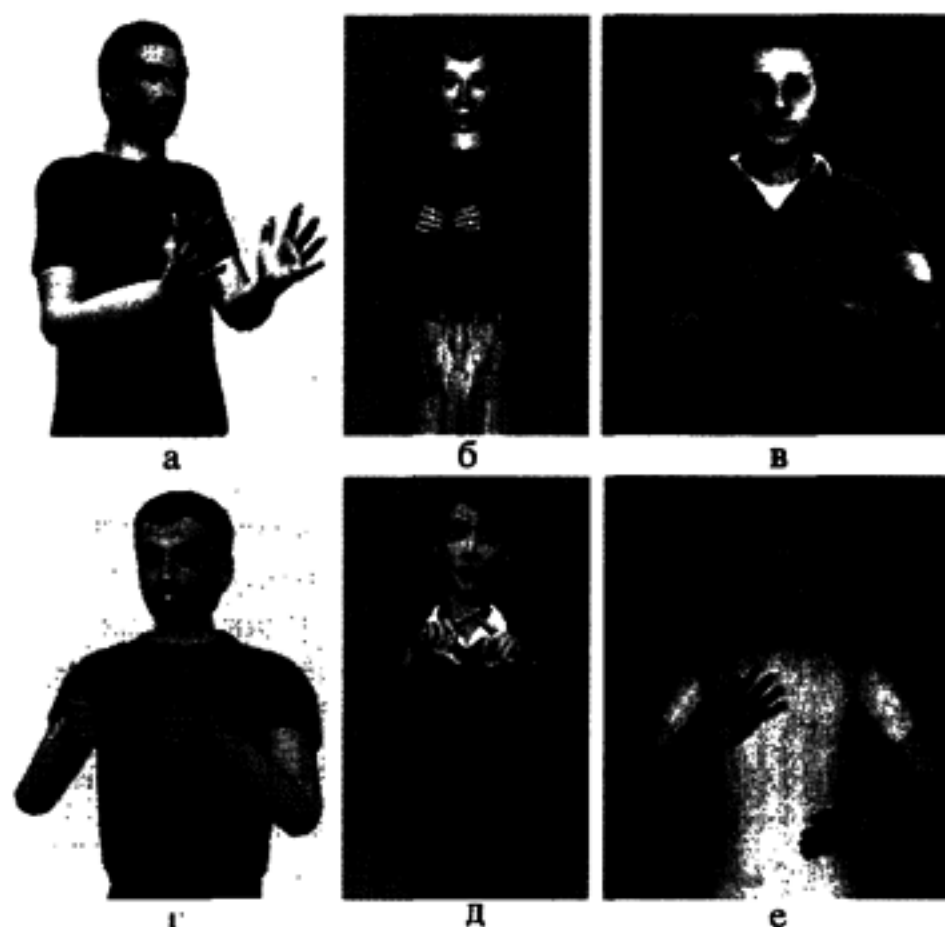


Рис. 5. Примеры трехмерных компьютерных аватаров, предназначенных для синтеза жестовой речи

фиксируемые на лице отражатели, сигналы от них преобразуются в трехмерную динамическую модель лица. Подобным образом разрабатывают свою модель синтеза украинского диалекта РЖЯ и киевские коллеги [Крак и др. 2008]. Отличия в создании этой модели состоят в том, что используется система всего из трех цифровых видеокамер, расположенных как фронтально, так и справа и слева от демонстратора жестов.

Достоинством таких разработок является реалистичность виртуальных персонажей, а также естественность и плавность их движений, которой добиваются, используя системы из нескольких видеокамер с большой частотой кадров. Недостаток же подобных разработок – дороговизна оборудования видеостудии и анимации трехмерного аватара при помощи игрового движка (программного обеспечения), а также сложность пополнения словаря, так как для этого необходимо производить повторные видеозаписи демонстратора в студии и затем обрабатывать их.

Другим возможным вариантом использования трехмерных виртуальных персонажей для синтеза ЖЯ можно назвать их управление посредством символов нотации жестов, которые описывают требуемые конфигурации рук и различные типы движений. Словарь жестов в такой системе представляет собой цепочку символов в выбранной нотации, поэтому может легко модифицироваться и расширяться без использования специального оборудования. Применяя такой подход, в начале XXI века на волне создания ассистивных (предназначенных для помощи людям) компьютерных технологий был разработан ряд моделей компьютерного синтеза жестовой речи для нескольких ЖЯ, включая американский, британский, французский, чешский и др. Среди известных зарубежных компьютерных систем синтеза жестовой речи, использующих различные анимированные аватары, следуют отметить системы, разрабатываемые в рамках текущих интеграционных проектов 7-й рамочной программы Евросоюза Dicta-Sign [ЭР12], SIGNSPeAK [ЭР13]; завершенные европейские проекты ViSiCAST с аватаром Visia [ЭР14] (рисунок 5д) и проект eSign с аватаром vGuido [ЭР15] (рисунок 5а), исследовательский проект Musslap [ЭР16] (рисунок 5е) Западночешского университета, проект коммерческой американской компании Vcom3D с аватарами Sign Smith [ЭР17] (рисунок 5б) и Sign4Me (рисунок 5в), а также британский проект SiSi (Say It Sign It) корпорации IBM [ЭР18] (рисунок 5г) в рамках более крупного проекта по машинному переводу английской речи в Амслен.

5. РАЗРАБОТКА СИСТЕМЫ СИНТЕЗА РУССКОЙ ЖЕСТОВОЙ РЕЧИ

С 2009 года СПИИ РАН совместно с Западночешским университетом г. Пльзень в рамках совместного проекта (по программе двусторонней поддержки науки РАН и Министерства образования, науки и спорта Чехии) ведет разработку и исследование системы синтеза РЖЯ и КЖР. За основу взята модель синтеза чешского ЖЯ [Zelezny et al. 2006] (рисунок 5е), использующая международную систему жестовой транскрипции HamNoSys. Была создана не просто система видеосинтеза жестов рук персонажа, а универсальная многомодальная система и интерфейс человеко-машинного взаимодействия [Ронжин, Карпов 2010], где кроме этой обработки производится также аудиовизуальный синтез звучащей русской речи. Общая архитектура компьютерного синтезатора представлена на рисунке 6. Система аудиовизуального синтеза русской речи (так называемая «говорящая голова») разработана ранее в сотрудничестве с коллегами из ОИПИ НАН Белоруссии [Лобанов и др. 2008] и представляет собой компиляционный синтез звучащей разговорной речи по тексту, совмещенный с движениями губ и лицевых органов трехмерной виртуальной головы человека. Для синтеза ЖЯ необходимо объединить визуальную модель головы человека с моделью туловища и рук (включая пальцы). Одной из основных задач исследования является создание и экспертное наполнение словаря РЖЯ с использованием специального редактора, анализирующего на входе символы HamNoSys и трансформирующего их в соответствующие движения аватара.

Предлагаемая система нацелена не только на пользователей компьютерных систем с ослабленным слухом, но также и на обычных пользователей, являясь универсальной многомодальной системой синтеза. Русская жестовая речь складывается из комбинации динамических жестов, воспроизводимых обеими руками (либо одной рукой), и артикуляции губ, проговаривающих показываемое слово. Многие глухие люди, которые ранее обладали слухом, способны «читать речь по губам» собеседника даже без использования жестов, поэтому такая речевая модальность должна быть неотъемлемым элементом компьютерной системы синтеза жестовой речи. В предлагаемой системе артикуляция губ, которые являются частью органов речеобразования, сопровождается также звучащим речевым сигналом, который может и не быть доступен полностью глухим людям, однако для слышащих людей аудиовизуальная система синтеза речи доступна и даже необходима для повышения разборчивости и естественности синтезируемой компьютером речи. Таким образом, основными компонентами универсальной многомодальной системы синтеза аудиовизуальной звучащей русской речи и калькирующей жестовой речи являются:

– имитационная модель головы человека, в которой настраиваются управляющие параметры для передачи мимики, выражения лица и движений губ при говорении, синхронизируемая с системой синтеза русской речи [Zelezny et al. 2006];



Рис. 6. Архитектура универсальной многомодальной системы синтеза аудиовизуальной разговорной речи и калькирующей жестовой речи

– компьютерная система аудиосинтеза разговорной русской речи, осуществляющая преобразование текста в речь по произвольному входному тексту [Лобанов и др. 2008];

– русскоязычная бимодальная система «говорящая голова» на основе виртуальной объемной модели головы человека и компьютерного синтеза речи по произвольному тексту, моделирующая естественную асинхронность звуковой и визуальной модальностей речи при синтезе [Карпов и др. 2010];

– компьютерная модель верхней части тела и рук человека, в которой настраиваются параметры движения рук для синтеза элементов РЖЯ на основе управляющих символов международной нотации HamNoSys;

– многомодальная система синтеза, синхронизирующая и объединяющая компоненты генерации визуальной речи, звучащей речи и калькирующей жестовой речи для аудиовизуального синтеза.

Синхронизация звучащей речи и жестов КЖР в системе осуществляется на основе временных меток начала и конца слов звучащей речи, синтезированных системой по тексту. Так как звучащая речь имеет более высокий темп воспроизведения, чем жестовая речь, то виртуальный аватар последовательно проговаривает и артикулирует с невысокой скоростью изолированные слова звучащей речи, дожидаясь окончания жестиксации слова (может включать в себя несколько элементов дактильной азбуки), плавно переходя к следующему жесту слитной КЖР.

Применение такой полноценной компьютерной системы синтеза актуально во многих сферах, где КЖР используется для официального общения и коммуникации, в частности, на собраниях, конференциях, лекциях, а также при автоматическом сурдопереводе телевизионных передач.

ЗАКЛЮЧЕНИЕ

В статье приведен анализ основных характеристик РЖЯ, который в значительной степени отличается от ЖЯ, применяемых в других странах, а, кроме того, ввиду территориальной обширности России в нашей стране распространены несколько различных диалектов РЖЯ. По этой причине компьютерные средства обучения, анализа и синтеза РЖЯ должны быть адаптированы к региональным особенностям языка.

Компьютерная обработка русской жестовой речи является новым междисциплинарным научным направлением, развитие которого требует как глубокого изучения принципов функционирования и развития РЖЯ и КЖР, так и разработки эффективных технологий, связанных с анализом и синтезом аудио- и видеоинформации. На данной стадии исследования проанализирована специфика РЖЯ, а также подготовлена демонстрационная версия универсального многомодального синтезатора аудиовизуальной звучащей русской речи и жестов калькирующей жестовой речи (см. демонстрацию показа времени на сайте лаборатории речевых и многомодальных интерфейсов СПИИ-РАН [ЭР19]).

СПИСОК ЛИТЕРАТУРЫ

- Воскресенский и др. 2009 – А.Л. Воскресенский, И.Е. Гуленко, Г.К. Хахалин. Словарь RuSLED как инструмент семантических исследований // Компьютерная лингвистика и интеллектуальные технологии. Труды междунар. конф. «Диалог'2009». Бекасово, 27–31 мая 2009 г. М., 2009.
- Воскресенский и др. 2010 – А.Л. Воскресенский, С.Н. Ильин, М. Железны. О распознавании жестов языка глухих // Компьютерная лингвистика и интеллектуальные технологии. Труды междунар. конф. «Диалог'2010». Бекасово, 26–30 мая 2010 г. М., 2010.
- Гейльман 1979 – И.Ф. Гейльман. Специфические средства общения глухих. Дактилология и мимика. Т. 1–4. Л., 1979.

- Гейльман 1980 – *И.Ф. Гейльман*. Изучаем жестуно. Л., 1980.
- Димскис 2002 – *Л.С. Димскис*. Изучаем жестовый язык. М., 2002.
- Зайцева 2000 – *Г.Л. Зайцева*. Жестовая речь. Дактилология: Учебник для студентов вузов. М., 2000.
- Карпов и др. 2010 – *А.А. Карпов, Л.И. Цирульник, М. Железны*. Разработка компьютерной системы «говорящая голова» для аудиовизуального синтеза русской речи по тексту // Информационные технологии. 2010. № 8.
- Кибрик, Прозорова 2007 – *А.А. Кибрик, Е.В. Прозорова*. Референциальный выбор в русском жестовом языке // Компьютерная лингвистика и интеллектуальные технологии. Труды междунар. конф. «Диалог'2007». Бскасово, 30 мая - 3 июня 2007 г. М., 2007.
- Крак и др. 2008 – *Ю. Крак, А. Бармак, А. Ганджа, А. Тернов, Н. Шатковский*. Компьютерная система виртуального общения людей с нарушениями слуха // Advanced studies in software and knowledge engineering. 2008. № 4.
- Лобанов и др. 2008 – *Б.М. Лобанов, Л.И. Цирульник, М. Железны, З. Крноул, А. Ронжин, А. Карпов*. Система аудиовизуального синтеза русской речи // Информатика. 2008. № 4.
- Мясоедова, Филиппович 2010 – *М.А. Мясоедова, Ю.Н. Филиппович*. Система автоматизированного перевода жестовой речи // Материалы научно-методического семинара научно-образовательного кластера CLAIM. М., 2010.
- Мясоедова и др. 2010 – *М.А. Мясоедова, Д.Н. Душкин, З.П. Мясоедова, Н.В. Петухова, М.П. Фархадов*. Разработка интернет-портала «Сурдосервер» с ресурсами русского жестового языка // Труды IV междисциплинарного семинара «Анализ разговорной русской речи АР-3». СПб., 2010.
- Пенин 2005 – *Г.Н. Пенин*. Калькирующая жестовая речь: особенности, сфера употребления // Сб. докладов конф. «Покровские педагогические чтения 2004–2005 гг.». СПб., 2005.
- Прозорова 2007 – *Е.В. Прозорова*. Российский жестовый язык как предмет лингвистического исследования // ВЯ. 2007. № 1.
- Ронжин, Карпов 2010 – *А.Л. Ронжин, А.А. Карпов*. Проектирование интерактивных приложений с многомодальным интерфейсом // Доклады ТУСУР. 2010. № 1.
- Специфические средства... 2002 – Специфические средства общения глухих России: Видеокурс в 3-х ч. Павловск, 2002.
- Фрадкина 2001 – *Р.Н. Фрадкина*. Говорящие руки: Тематический словарь жестового языка глухих России. М., 2001.
- ЭР1 – Европейский проект Spreadthesign. www.spreadthesign.com/ru (электронный ресурс)
- ЭР2 – Дактильные алфавиты мира. www.deafnet.ru/dn/abc.phtml
- ЭР3 – Московская городская организация Всероссийского общества глухих. www.deafmos.ru/info.phtml?c=24&id=105
- ЭР4 – Электронная обучающая система «Русский жестовый язык. Базовый курс». <http://istina.inion.ru/NIOT/rgy.htm>
- ЭР5 – Интерактивный онлайн видеословарь «Русский жестовый язык DigitGestus». www.digitgestus.com
- ЭР6 – Интерактивный онлайн словарь РЖЯ Стэнфордского университета. www.stanford.edu/dept/lc/rsi/
- ЭР7 – Проект «Сурдосервер». <http://surdoserver.ru>
- ЭР8 – Система жестовой нотации «Sign Writing». www.signwriting.org
- ЭР9 – Система жестовой нотации «Hamburg Notation System». www.sign-lang.uni-hamburg.de/projects/hamnosys.html
- ЭР10 – Система компьютерного синтеза жестового языка, разработанная в Новосибирске. www.vesti.ru/doc.html?id=358385&p=35&cid=1
- ЭР11 – Система компьютерного синтеза жестового языка, разработанная в Минске. http://ont.by/news/our_news/0062482
- ЭР12 – Европейский FP7 проект Dicta-Sign. www.dictasign.eu
- ЭР13 – Европейский FP7 проект SIGNSPEAK. www.signspeak.eu/en
- ЭР14 – Европейский проект ViSiCAST. www.visicast.co.uk
- ЭР15 – Европейский проект eSign. www.sign-lang.uni-hamburg.de/esign
- ЭР16 – Чешский проект Musslap. <http://musslap.zcu.cz>
- ЭР17 – Проект Sign Smith компании Vcom3D. www.vcom3d.com

ЭР18 – Американский проект SiSi. www-03.ibm.com/press/us/en/pressrelease/22316.wss

ЭР19 – Мультимедийная демонстрация синтезатора РЖЯ СПИИРАН.

www.spiiras.nw.ru/speech/demo/signlang.avi

Aran et al. 2008 – *O. Aran, I. Ari, L. Akarun, E. Dikici, S. Parlak, M. Saraclar, P. Campr, M. Hruz.* Speech and sliding text aided sign retrieval from hearing impaired sign news videos // *Journal on multimodal user interfaces*. 2008. № 2.

Stokoe 1960 – *W.C. Stokoe.* Sign language structure: An outline of the visual communication systems of the American deaf // *Studies in linguistics: Occasional papers*. 1960. № 8.

Zelczyn et al. 2006 – *M. Zelezny, Z. Krnoul, P. Cisar, J. Matousek.* Design, implementation and evaluation of the Czech realistic audio-visual speech synthesis // *Signal processing*. 2006. № 12.

Сведения об авторе:

Алексей Анатольевич Карпов

Санкт-Петербургский институт информатики и автоматизации РАН

karpov@iias.spb.su