

А.А. КАРПОВ, И.А. КАГИРОВ
**ФОРМАЛИЗАЦИЯ ЛЕКСИКОНА
СИСТЕМЫ КОМПЬЮТЕРНОГО СИНТЕЗА
ЯЗЫКА ЖЕСТОВ**

Карпов А.А., Кагиров И.А. **Формализация лексикона системы компьютерного синтеза языка жестов.**

Аннотация. В статье дан аналитический обзор особенностей русского жестового языка и калькирующей жестовой речи, в том числе жестовых лексиконов и грамматических конструкций данного языка, а также возможных методов формализованного представления элементов словаря жестов. В ходе междисциплинарных исследований была адаптирована для задачи синтеза русского жестового языка виртуальная трехмерная модель аватара человека и предложена модель универсального многомодального аудиовизуального синтезатора русской звучащей речи и калькирующей жестовой речи по тексту.

Ключевые слова: жестовый язык глухих, аудиовизуальный синтез, ассистивные информационные технологии.

Karpov A.A., Kagirov I.A. **Lexicon Formalization for a Computer System of Sign Language Synthesis.**

Abstract. In this paper, we give an analytical survey of peculiarities of the Russian sign language and calc signed (gesture) speech, including state-of-the-art of sign lexicons and grammatical constructions, as well as methods for formalization of sign lexicon items. In the course of the multidisciplinary research, a virtual 3D model of human being's avatar has been adopted for the Russian sign language synthesis, and a multipurpose model of multimodal audio-visual synthesizer aimed for text to Russian auditory speech and calc signed speech transformation.

Keywords: sign language of deaf, audio-visual synthesis, assistive information technologies.

1. Введение. Жестовый язык (ЖЯ) является способом межчеловеческой коммуникации, в котором для передачи информации используются только визуально-кинетические средства (жесты рук, артикуляция губ, мимика и эмоции, проявляющиеся на лице). В настоящее время жестовые языки используются в основном как специфическое средство коммуникации людей с нарушениями речи и слуха. ЖЯ не является универсальным во всех странах мира, так как он возникает и развивается естественным путем в разных локальных сообществах и изменяется со временем с появлением новой лексики. ЖЯ состоит из разных знаковых систем, совмеща в себе язык глухих людей и жесты, используемые в обществе в целом. Русский жестовый язык (РЖЯ) объединяет в себе несколько разновидностей жестовой коммуникации. Выделяют так называемую «ручную азбуку» (дактильную азбуку), использующуюся для жестовой передачи букв русского алфавита, и собственно жестовую речь, ориентированную на передачу не графем

(букв), а слов, языковых конструкций и семантических концептов. Дактильная азбука (воспроизводящая буквы посредством пальцев) используется, как важный вспомогательный инструмент. В том случае, если говорящий не знает знака для какого-либо понятия или ему необходимо передать точное звучание некоторого слова, он прибегает к дактильной азбуке, она используется также для передачи различных имен собственных, аббревиатур и редко встречающихся слов, слов иностранного происхождения. На рис. 1 слева показана дактильная азбука РЖЯ, а справа — дактильная азбука британского варианта ЖЯ. Американский вариант ЖЯ отличается от британского полностью, в отличие от английской разговорной и письменной речи; такая же ситуация и в испаноязычных латиноамериканских странах, каждая из которых имеет свой ЖЯ.

Отметим характерные различия данных алфавитов:

- все буквы русской азбуки показываются одной рукой, а британской — двумя руками;
- ряд букв русской азбуки отображается посредством динамических жестов, в то время как жесты британской азбуки — статические;
- конфигурации кисти и пальцев рук в русском дактиле более сложны.

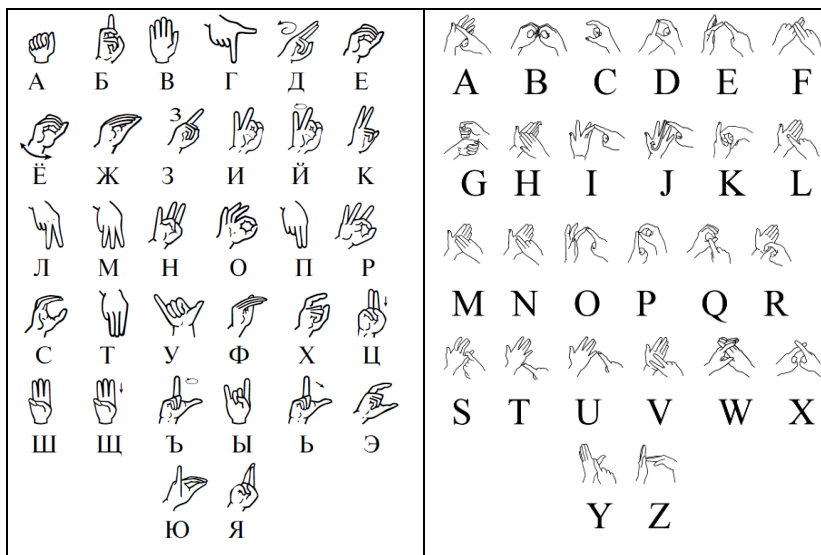


Рис. 1. Русский дактильный алфавит (слева) и британский алфавит (справа) [2].

Большинство стран имеет свою оригинальную дактильную азбуку, зависящую от социокультурных особенностей их народов, причем по численности одноручные азбуки (в Германии, Франции, Америке, Китае, Индии и т. д.) преобладают над двуручными (в Великобритании, Турции, Чехии, Новой Зеландии и т. д.) [7].

Основным же способом межчеловеческой коммуникации в среде глухих является непосредственно ЖЯ, в котором каждому смысловому понятию (или группе синонимичных понятий) соответствует определенный уникальный жестовый эквивалент. Причем система ЖЯ распадается на две подсистемы жестовой речи: 1) национальную и 2) калькирующую [3]. Калькирующая жестовая речь полностью копирует речь звучащую, в которой один-в-один воспроизводятся слова разговорной речи и их порядок в предложениях.

Различия в ЖЯ вызывают неоднозначности при экспертном переводе с языка на язык, а также при разработке автоматических компьютерных систем синтеза и сурдоперевода. Анализ показывает, что менее 30 % наиболее распространенных жестов одинаковы для разных языков, это большинство простых жестов, показывающих объекты в поле зрения («Ты», «Голова», «Нос»), абстрактные же понятия зависят от культур и обычаев обществ. Вышесказанное о функциях ЖЯ можно подытожить табл. 1.

Таблица 1. Жестовые системы, их функции и основные характеристики

Жестовая система	Сфера применения	Характеристика
Дактильная азбука	Применяется во всех жестовых системах как средство передачи редких и незнакомых слов; тесно связана с вербальным языком и письменностью	Аналог национального письменного алфавита
Калькирующая жестовая речь	Точная передача естественного вербального языка и речи при помощи жестов	Вторичная по отношению к вербальному языку система
ЖЯ	Основное средство общения между глухонемыми	Отдельный язык со своей грамматикой и лексикой

Основное, что объединяет все жестовые системы, — это способ передачи и восприятия информации. В качестве базовой смысловой единицы здесь служит жест (визуально-кинетический акт), в котором участвуют в первую очередь руки, а также часто мимика лица и арти-

куляция губ. Кинетическая природа жеста и его визуальное восприятие обуславливают особенности ЖЯ: возможность определенным образом расположить жест в пространстве (ближе к телу, дальше, правее, левее и т. д.), исполнить одновременно два жеста двумя руками и т. д.

2. Грамматические особенности РЖЯ. Калькирующая жестовая речь не имеет собственной грамматики, она копирует структуру вербального разговорного языка, поэтому является вторичной знаковой системой. В отличие от нее, 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 развиваются и существуют отдельно от вербальных, так и их грамматики слабо связаны друг с другом. Например, в [3] приведен пример фразы: «В левом углу находятся стул и торшер, стул находится под торшером» на РЖЯ, где понятия «Торшер» и «Стул» демонстрируются одновременно во времени и пространственно одно под другим на разных уровнях с левой стороны от диктора.

РЖЯ обладает также средствами для передачи разнообразных морфологических значений и четко оформленным синтаксисом, отличающимся от синтаксиса вербального русского языка. В качестве примера морфологии ЖЯ можно привести глагольные показатели. Глагол в ЖЯ, так же как и в вербальном языке, обладает грамматическими категориями времени, аспекта и модальности [12]. Для выражения временных значений используются вспомогательные слова «Было», «Будет», т. е., к примеру «Читать было» — читал, «Читать» — читает, «Читать будет» — прочитаю. Также довольно широко используются такие темпоральные модификаторы глаголов, как «Сегодня», «Вчера», «Скоро», «Недавно», «Давно». В целом темпоральную систему РЖЯ можно сравнить с так называемыми предметно-ориентированными языками [12], в которых временные категории передаются при помощи лексических средств, т. е. они не являются частью грамматики языка, а относятся только к словарю.

Аспектуальные категории распадаются на два семантических класса (аспекта): 1) количественный и 2) фазовый. Фазовые значения передают различные стадии действия (завершенность, начало и т. п.), а количественные — повторяемость, характер длительности действия. Для фазовых значений используются слова-модификаторы «Готово», «Закончено», «Еще не» (например, «Писать готово» — написал, «Писать еще не» — еще не написал и т. п.), а количественные значения выражаются простым повторением жеста нужное число раз. Модаль-

ные значения также передаются лексически, при помощи таких вспомогательных слов, как «Бы», «Если» и т. п.

Синтаксис является наиболее специфической чертой РЖЯ. Упрощенно можно считать, что фразы РЖЯ коротки и просты по структуре, а синтаксис описывается всего тремя типами структур:

- 1) SOV (субъект—объект—глагол);
- 2) SVO;
- 3) VSO.

Однако некоторые исследователи отмечают [3], что классический анализ высказываний РЖЯ в терминах субъекта, предиката, объектов, адьюнктов и т. п. невозможен в силу нелинейности синтаксиса РЖЯ. Так в высказывании «Стул находится под торшером» точно сказать, где находится субъект, а где объект, просто невозможно. В качестве альтернативы выдвигается анализ на основе функционального подхода и учета специфики субстанции жеста, равно как и естественной констатиивности ЖЯ.

На настоящее время грамматика РЖЯ еще недостаточно изучена и формализована, чтобы вести разговоры об автоматическом сурдопереводе из произвольного русскоязычного текста на ЖЯ. Ряд исследователей РЖЯ утверждают [1], что серьезные различия в семантико-синтаксической структуре письменного и жестового языков не позволяют выполнять однозначный машинный перевод русскоязычных текстов на РЖЯ, и действующих автоматических систем сурдоперевода на данный момент не существует. Для создания такой полноценной модели необходимо производить глубокий семантический анализ и разбор письменных фраз, а это пока возможно лишь на поверхностном уровне из-за несовершенства алгоритмов и баз знаний. Однако калькирующая жестовая речь непрямую отражает разговорную звучащую речь, поэтому компьютерный синтез калькирующей жестовой речи намного проще. Примером компьютерной системы распознавания перевода разговорной речи (английской) в калькирующую жестовую речь с элементами ЖЯ (Амслен) может служить американская разработка iCommunicator [8], лексикон которой состоит более чем из 9 тыс. видеофрагментов жестов.

3. Лексикон русского РЖЯ. РЖЯ распространен не только в России, но и на территории бывшего Советского Союза, в Белоруссии, Казахстане, Украине, причем в разных регионах страны существуют несколько различающихся диалектов РЖЯ, но использующих практически единую дактильную азбуку. Причем зачастую члены одного общества не приемлют жесты, принятые в другом сообществе, хотя

многие из них являются интуитивно понятными. По результатам анализа мультимедийных жестовых словарей можно сказать, что крупнейшие диалекты РЖЯ — петербургский и московский — различаются примерно в 30–40 % жестов-слов, хотя иногда расхождения незначительны.

Среди существующих мультимедийных компьютерных словарей РЖЯ отметим следующие:

1) мультимедийный видеословарь «Толковый словарь русского жестового языка — RuSLED». Программа является интерактивной оболочкой к видеословарю РЖЯ, записанному в 2002 г. в Межрегиональном центре реабилитации лиц с проблемами слуха г. Павловск;

2) видеословарь «Тематический словарь русского жестового языка», созданный Московской городской организацией Всероссийского общества глухих (ВОГ);

3) мультимедийная электронная обучающая система «Русский жестовый язык. Базовый курс», созданная в 2001 г. центром «Истина» при поддержке центрального правления ВОГ;

4) интерактивный видеословарь «DigitGestus», созданный в Новосибирске в 1996–1997 гг., содержит несколько сотен элементов;

5) интерактивный видеословарь жестового языка европейского проекта Spreadthesign («Распространим жест»), поддерживается Европейской комиссией в лице Шведского международного отделения образования и обучения;

6) интерактивный видеословарь проекта «Сурдосервер 2.0» [7], содержащий те же самые видеоматериалы, что и электронная обучающая система (ЭОС) «Русский жестовый язык. Базовый курс» 2001 г., а также оригинальный жестовый словарь по направлению «Информационные технологии».

Лексические жестовые элементы и в самом РЖЯ, и в дактильной азбуке формируются практически одинаково и могут быть формализовано описаны сходным образом:

- по конфигурации (форма руки или рук),
- по месту исполнения (где находятся руки, выполняющие жест),
- по характеру движения (как именно движутся руки; что они делают).

Такое описание впервые предложено американским исследователем жестовой речи У. Стоуки в 1960-х гг. [17]. В своем описании американского ЖЯ (Амслен) он применял фонологический принцип для транскрибирования жестов. Под фонологическим принципом понимается разложение любого жеста на минимальные регулярно повторяю-

щиеся 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 — фонемы реализуются последовательно во времени. Прямому отождествлению хиремы с фонемой мешает множество различных факторов, среди которых на первом месте стоит кардинальное различие акустического и кинетического планов человеческой коммуникации, однако и хиремы описываются по тем же принципам, что и фонемы. Так, традиционно фонемы характеризуются по месту образования и способу образования (традиция, заложенная Н.С. Трубецким [11]), т. е. по тому, в какой области речевого тракта находится язык и тому, как именно артикулируется фонема. Хиремы, по концепции 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 движения и т. п.). Кроме того, для хирем необходим третий признак — форма руки или рук. Эти три признака (место образования, способ образования и форма руки) используются во всех современных системах нотации ЖЯ, которые предназначены для формализации элементов жестового лексикона.

4. Способы формализации лексикона ЖЯ. Специфика человеко-машинного взаимодействия состоит в том, что ЖЯ и жестовый словарь должны быть определенным образом записаны, чтобы компьютер мог обрабатывать и синтезировать жесты. Для описания жеста по его признакам существуют несколько различных систем нотации (жестовой транскрипции), позволяющих зафиксировать представление жеста в записи. Первая система нотации, разработанная У. Стоуки, базировалась на латинской системе письма. Для кодирования хирем использовались буквы латинского алфавита, вспомогательными элементами были арабские цифры и различные диакритики. Однако предложенная запись плохо подходит для образовательных целей, так как совершенно не наглядна, поэтому разрабатываются другие системы нотации ЖЯ, использующие символы, более наглядно демонстрирующие положения рук и их перемещения.

Насколько известно, в России на сегодня не существует законченных систем транскрибирования и формализации РЖЯ. Многие исследователи обращаются к уже зарекомендовавшим себя универсальным зарубежным системам нотации, применяя их опыт на русском матери-

але. Среди таких систем нотации популярностью пользуется программа транскрипции ELAN [13], с помощью которой экспертами аннотируются и транскрибируются корпуса аудиовизуальных и жестовых данных, в том числе артикуляция и эмоциональные характеристики.

Для практического, в том числе компьютерного применения в мире разработаны несколько систем нотации, обладающих наглядностью и в то же время унифицированностью для быстрого ввода жестов в словарь. Одна из самых известных систем нотации, широко распространенная в Америке (в частности, для Амслена), называется SignWriting [10] и разработана В. Саттон. Знаки в SignWriting символизируют мимику, движения рук и тела ЖЯ и пишутся сверху вниз. На рис. 2 показан пример записи известного английского четверостишия Матушки Гусыни «Jack and Jill went up the hill to fetch a pail of water. Jack fell down and broke his crown and Jill came tumbling after» в данной нотации [10].

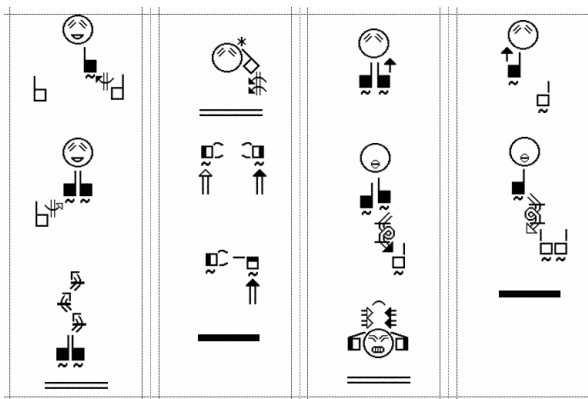


Рис. 2. Пример записи четверостишия в нотации SignWriting [10].

SignWriting базируется на принципах, предложенных У. Стоуки, однако имеет конкретное иконическое воплощение жестов при записи. Этот ЖЯ отличается большей проработанностью и однозначностью записи. Однако у него весьма велик набор знаков (свыше тысячи), что делает его малоприменимым для применения в компьютерных технологиях даже с использованием возможностей Unicode-кодировки. Например, в нем каждой специфической форме руки соответствует отдельный знак (рис. 3). Поэтому запомнить и использовать на практике их трудно.

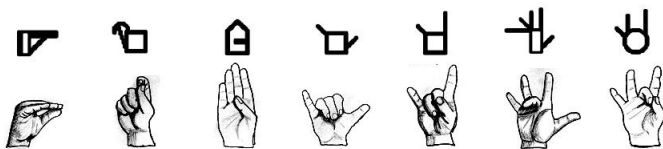


Рис. 3. Пример кодировки конфигурации руки в SignWriting [10].

Но намного практичнее сократить число частных случаев конфигурации руки, описывая их как переменные (например, число показываемых пальцев) при одной константе (форме ладони). Такому требованию, например, удовлетворяет система нотации жестов, предложенная Л. Димскис [2]. В ее нотации выделяется более 30 конфигураций рук, около 50 характеристик места исполнения жеста и более 70 характеристик локализации. В нотации Димскис структура жеста состоит из отдельных элементов и имеет следующие постоянные характеристики:

- 1) форму руки (рук);
- 2) место расположения жеста (жестовое пространство);
- 3) характер движения.

Первый элемент — форма руки предполагает строго определенное положение ладоней и направление пальцев, он состоит из трех частей — формы руки, положения ладони, направления пальцев, необходимых для правильного кодирования (записи) и декодирования (воспроизведения) отдельных хирем.

Второй элемент структуры — место расположения жеста в пространстве. Жест может исполняться над головой, у лица, на уровне плеч, шеи и т. д. Локализация каждого жеста строго постоянна и изменение жестового пространства влияет на смысловое значение жеста в РЖЯ.

По характеру исполнения жесты подразделяются на одноручные; двуручные с одинаковой формой рук; двуручные с разной формой рук. На рис. 4 приведен пример кодирования двух числительных посредством данной нотации [2]. Для числительного «Четыре» выполняется следующее:

- первый знак (буква П) указывает на то, что жест выполняется правой рукой;
- второй знак иконографически указывает форму руки — четыре оттопыренных пальца;
- третий символ схематически изображает разворот ладони — она направлена к себе, к корпусу говорящего;

— четвертый знак указывает на ориентацию ладони — пальцами вверх;

— последний символ означает, что жест выполняется на уровне груди человека.

Для числительного «Двадцать» различие состоит лишь в нескольких элементах:

- 1) неподжатыми остаются только два пальца, а не четыре;
- 2) кисть указывает налево;
- 3) в конце записи добавлен элемент, описывающий кинетику жеста: в конце исполнения жеста кисть руки опускается вниз, делая «стряхивающее» движение.



Рис. 4. Пример записи двух жестов-чисел в нотации Димкис: «Четыре» (слева) и «Двадцать» (справа) [2].

Наиболее широкое распространение в мире за последние годы получила гамбургская система нотации (основатель и основной разработчик — Т. Ханке), более известная под сокращенным названием HamNoSys (от англ. *Hamburg Notation System*) [16], разработанная в конце XX века в Германии. На начало 2011 г. актуальна система HamNoSys версии 4.0 [9]. Она отличается наибольшей проработанностью и пригодна для использования в компьютерных приложениях за счет того, что ее знаки переведены в систему кодировки Unicode с соответствующими компьютерными шрифтами. Инвентарь HamNoSys позволяет записать практически любой жест, что делает эту систему универсальной и пригодной для любых ЖЯ мира. Фактически, по своим функциям HamNoSys является аналогом международного фонетического алфавита (МФА) и используется многими лингвистами, занимающимися исследованием ЖЯ. В системе HamNoSys сначала задается форма кисти; 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 одноручными 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, представленными на рис. 5, символически изображающими ладонь и пальцы.



Рис. 5. Символы HamNoSys, предназначенные для формальной записи конфигураций пальцев и ладони руки [16].



Рис. 6. Примеры записи конфигурации кисти руки в HamNoSys [9].

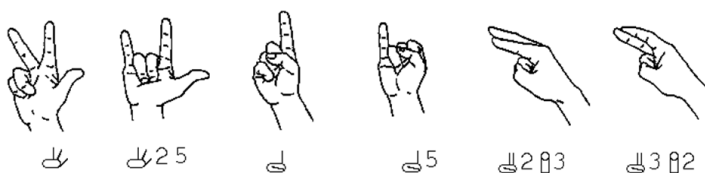


Рис. 7. Примеры записи конфигурации пальцев руки в HamNoSys [9].

На рис. 6 показаны несколько примеров кодирования конфигурации кисти и пальцев руки, в том числе большого пальца, который задается отдельно. При помощи особых диакритических символов можно задать форму ладони, характер складывания пальцев и т. п.

Отметим способы указания отдельных пальцев на руке. В базовых символах нет элементов для, например, отдельно отставленного безымянного или указательного пальцев. В этом случае используется числовое обозначение пальцев рук:

- 1 — большой палец,
- 2 — указательный палец,
- 3 — средний палец,
- 4 — безымянный палец,
- 5 — мизинец.

Соответственно, для точной кодировки пальцев следует всего лишь задать их число и точное значение цифрами (рис. 7). Такой подход позволяет сократить набор знаков, сделать его менее громоздким и удобным в использовании по сравнению с SignWriting. Ориентация кисти руки задается по кончикам пальцев; она определяется в трех плоскостях в зависимости от специфики жеста (рис. 8). Центром координат всегда является сам человек-демонстратор.

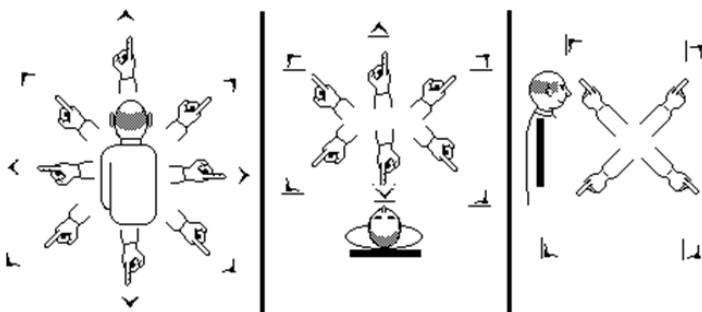


Рис. 8. Описание ориентации руки в нотации HamNoSys [9].

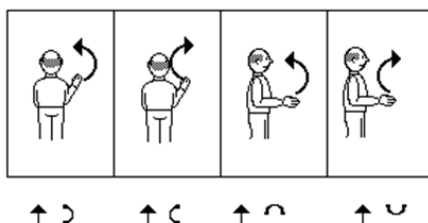



Рис. 9. Примеры записи нелинейных дугообразных движений в HamNoSys [9].

Кроме того, в HamNoSys существуют разнообразные знаки практически для всех возможных движений, которые могут осуществлять руки, для примера на рис. 9 показаны некоторые нелинейные дугообразные движения. Их более сотни в версии 4.0, и этот набор постоянно расширяется и совершенствуется разработчиками [9], чтобы соответствовать новой лексике, постоянно появляющейся в ЖЯ.

Таблица 2. Запись числительного «Один» РЖЯ в нотации HamNoSys

Изображение жеста	Нотационная запись HamNoSys	
	Форма кисти — выставлен указательный палец	—
	Ориентация кисти — пальцы вверх, кисть правой руки обращена к себе	
	Место выполнения — на уровне правого плеча	

	Тип движения руки — вверх	TR
--	---------------------------	----

В качестве примера использования данной нотации рассмотрим числительное «Один» общепринятого варианта РЖЯ, каждая характеристика которого расписана при помощи знаков HamNoSys (табл. 2). Это понятие записывается следующей цепочкой знаков:



5. Компьютерный синтез жестовой речи. Одним из наиболее эффективных средств обучения и человеко-машинного взаимодействия являются мультимедийные компьютерные программы, поэтому, создание информационных приложений, способных работать с ЖЯ (синтезировать и распознавать), является одной из приоритетных задач при работе с глухими и слабослышащими людьми. Возможным вариантом компьютерного синтеза ЖЯ является использование анимированного аватара человека, который управляется посредством символов жестовой нотации, описывая требуемые конфигурации рук и различные типы движений. Жесты из лексикона в такой системе синтеза представляют собой цепочку символов в выбранной нотации, поэтому словарь может легко модифицироваться и расширяться без использования специального оборудования.

Лаборатория речевых и многомодальных интерфейсов СПИИРАН совместно с Западно-Чешским университетом в рамках совместного проекта ведет разработку и исследование системы синтеза РЖЯ (калькирующей жестовой речи). За основу в системе взята модель синтеза чешского ЖЯ [15], использующая систему жестовой транскрипции HamNoSys. Нами разрабатывается не просто система синтеза жестов из видеофрагментов, а универсальная многомодальная система [4] для аудиовизуального синтеза русской звучащей и жестовой речи. Архитектура системы синтеза представлена на рис. 10.

Основными компонентами многомодальной системы синтеза являются:

- 1) имитационная модель головы/лица человека, в которой настраиваются управляющие параметры для передачи мимики, выражения лица и движений губ при говорении [18];
- 2) компьютерная система акустического синтеза разговорной русской речи, осуществляющая преобразование текст-речь по произвольному входному русскоязычному тексту [6];
- 3) русскоязычная бимодальная система «говорящая голова» на основе виртуальной 3D-модели головы/лица человека и компьютерного синте-

за речи по произвольному тексту, моделирующая естественную асинхронность звуковой и визуальной модальностей речи при синтезе [5];



Рис. 10. Архитектура универсальной многомодальной системы аудиовизуального синтеза русской звучащей и жестовой речи.

4) виртуальная 3D-модель верхней части туловища и рук человека, в которой настраиваются управляющие параметры движения рук для синтеза элементов РЖЯ глухих на основе управляющих символов международной нотации HamNoSys [15];

5) многомодальная система синтеза, объединяющая компоненты генерации русской визуальной речи, звучащей речи и калькирующей жестовой речи.

На текущем этапе разработки производится создание и экспертное наполнение словаря ЖЯ системы (дактильной азбуки, цифр и наиболее употребительных жестов РЖЯ) посредством специального редактора [14], анализирующего на входе символы HamNoSys и трансформирующего их в соответствующие движения органов виртуального 3D-аватара.

5. Заключение. В данной статье проанализирована специфика РЖЯ и калькирующей жестовой речи, включая словари жестов, грамматические особенности данного языка и способы формализованного представления лексикона.

В ходе исследований для РЖЯ и разговорной речи адаптирована виртуальная трехмерная модель компьютерного синтеза и подготовле-

на демонстрационная версия универсального многомодального синтезатора аудиовизуальной русской речи и РЖЯ.

Мультимедийную демонстрацию показа текущего времени в формате ХХ час(-а,-ов) YY минут(-а,-ы) на РЖЯ можно посмотреть на веб-сайте лаборатории речевых и многомодальных интерфейсов СПИИРАН: www.spiiras.nw.ru/speech/demo/signlang.avi

Литература

1. *Воскресенский А.Л., Гуленко И.Е., Хахалин Г.К.* Словарь RuSLED как инструмент семантических исследований // Материалы Междунар. конф. «Диалог-2009», 2009. С. 64–68.
2. *Димскис Л.С.* Изучаем жестовый язык. М.: Академия, 2002. 128 с.
3. *Зайцева Г.Л.* Жестовая речь. Дактилология: Учебник для студентов. М.: ВЛАДОС, 2000. 192 с.
4. *Карпов А.А., Ронжин А.Л.* Многомодальные интерфейсы в автоматизированных системах управления // Известия высших учебных заведений. Приборостроение. 2005. Т. 48, №7. С. 9–14.
5. *Карпов А.А., Цирульник Л.И., Железны М.* Разработка компьютерной системы “говорящая голова” для аудиовизуального синтеза русской речи по тексту // Информационные технологии. 2010. №8, т. 9. С. 13–18.
6. *Лобанов Б.М., Цирульник Л.И., Железны М., Крнoul З. и др.* Система аудиовизуального синтеза русской речи // Информатика. 2008. № 4. С. 67–78.
7. Проект «Сурдосервер 2.0» // Электронный ресурс [<http://surdoserver.ru>].
8. Проект iCommunicator» // Электронный ресурс [<http://www.icommunicator.com>].
9. Система жестовой нотации Hamburg Notation System // Электронный ресурс [<http://www.sign-lang.uni-hamburg.de/projects/hamnosys.html>].
10. Система жестовой нотации SignWriting // Электронный ресурс [<http://www.signwriting.org/about/>].
11. *Трубецкой Н.С.* Основы фонологии. М.: Аспект Пресс, 2000. 352 с.
12. *Bybee J.L., Perkins R., Pagliuca W.* The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World. University of Chicago Press, 1994.
13. *Hellwig B. et al.* ELAN — Linguistic Annotator (version 4.0). 2010. 236 p.
14. *Kanis J., Krñoul Z.* Interactive HamNoSys Notation Editor for Signed Speech Annotation // Proc. of the 6th Intern. Conf. on Language Resources and Evaluation LREC-2008. Paris: ELRA. 2008. P. 88–93.
15. *Krñoul Z., Kanis J., Železný M., Müller L.* Czech Text-to-Sign Speech Synthesizer // Proc. of the Intern. Conf. on Machine Learning for Multimodal Interaction MLMI-2007. LNCS 4892. 2008. P. 180–191.
16. *Prillwitz S., Hanke T., et al.* HamNoSys. Version 2.0; Hamburg Notation System for Sign Languages. An introductory guide. Hamburg: Signum, 1989. 46 p.
17. *Stokoe W.C.* Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf // Studies in Linguistics: Occasional papers. 1960. №8. 78 p.
18. *Železný M., Krñoul Z., Cisar P., Matousek J.* Design, implementation and evaluation of the Czech realistic audio-visual speech synthesis // Signal Processing. 2006. Vol. 86, N 12. P. 3657–3673.

Карпов Алексей Анатольевич — к.т.н., старший научный сотрудник лаборатории речевых и мультимодальных интерфейсов Учреждения Российской академии наук Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН). Область научных интересов: автоматическое распознавание речи, мультимодальные интерфейсы, аудиовизуальное распознавание и синтез речи. Число научных публикаций — 120. karpov@iias.spb.su; СПИИРАН, 14-я линия В.О., д. 39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-7081, факс +7(812)328-7081.

Karpov Alexey Anatolyevich — Ph.D., senior researcher, Laboratory of Speech and Multimodal Interfaces St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: automatic speech recognition, multimodal interfaces, audio-visual speech recognition. The number of publications — 120. karpov@iias.spb.su; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-7081, fax +7(812)328-7081.

Кагиров Ильдар Амирович — младший научный сотрудник лаборатории речевых и мультимодальных интерфейсов Учреждения Российской академии наук Санкт-Петербургского института информатики и автоматизации РАН (СПИИРАН). Область научных интересов: компьютерная лингвистика, язык жестов глухих. Число научных публикаций — 10. kagirov@iias.spb.su; СПИИРАН, 14-я линия В.О., д. 39, Санкт-Петербург, 199178, РФ; р.т. +7(812)328-7081, факс +7(812)328-7081.

Kagirov Ildar Amirovich — junior researcher, Laboratory of Speech and Multimodal Interfaces St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS). Research interests: computer linguistics, sign language of the deaf. The number of publications — 10. kagirov@iias.spb.su; SPIIRAS, 39, 14th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-7081, fax +7(812)328-7081.

Поддержка исследований. Данное исследование поддержано Советом по грантам Президента РФ (проект №МК-64898.2010.8); Министерством образования и науки РФ в рамках ФЦП «Научные и научно-педагогические кадры инновационной России на 2009–2013 гг.» (госконтракты №П2579 и №П2360); фондом РФФИ (проект №09-07-91220-СТ), а также Комитетом по науке и высшей школе Правительства Санкт-Петербурга.

Рекомендовано лабораторией речевых и мультимодальных интерфейсов, зав. лабораторией д.т.н., доц. А.Л. Ронжин.

Статья поступила в редакцию 11.03.2011.

РЕФЕРАТ

Карнов А.А., Кагиров И.А. Формализация лексикона системы компьютерного синтеза языка жестов.

Язык жестов является способом межчеловеческой коммуникации глухих людей, в котором для передачи информации используются визуальнокинетические средства (жесты рук, артикуляция губ, мимика и эмоции, проявляющиеся на лице). Язык жестов не является универсальным во всех странах мира, так как он возникает и развивается естественным путем в разных локальных сообществах и изменяется со временем с появлением новой лексики. Русский жестовый язык объединяет в себе несколько разновидностей жестовой коммуникации: дактильную азбуку, используемую для жестовой передачи букв русского алфавита; калькирующую жестовую речь, которая копирует речь звучащую и один-в-один воспроизводит слова разговорной речи и их порядок в предложениях; и собственно жестовый язык, в котором каждому смысловому понятию (или группе синонимичных понятий) соответствует определенный уникальный жестовый эквивалент. В отличие от калькирующей жестовой речи, русский жестовый язык обладает собственной грамматикой и выразительными средствами, причем по своей структуре грамматика жестового языка значительно отличается от грамматики литературного или разговорного языка. По результатам анализа мультимедийных жестовых словарей можно сказать, что крупнейшие диалекты русского жестового языка (петербургский и московский) различаются примерно в 30–40 % жестов-слов, хотя иногда расхождения незначительны.

Лексические элементы и в языке жестов, и в дактильной азбуке формируются практически одинаково и могут быть формализованы сходным образом: по конфигурации (форма рук), по месту исполнения (где находятся руки, выполняющие жест) и по характеру движения (как именно движутся руки; что они делают). В статье описаны и проанализированы возможные методы формального представления, записи и транскрибирования элементов жестового лексикона: международные системы нотации жестов HamNoSys, SignWriting, а также оригинальные системы Л. Димскис и У. Стоуки.

В ходе ведущихся междисциплинарных исследований нами адаптирована для задачи синтеза русского жестового языка виртуальная трехмерная модель аватара человека и предложена программная модель универсального мультимедийного аудиовизуального синтезатора русской звучащей речи и калькирующей жестовой речи по тексту. На текущем этапе разработки производится создание и экспертное наполнение словаря языка жестов системы (дактильной азбуки, цифр и наиболее употребительных жестов) посредством специального редактора, анализирующего на входе символы HamNoSys и трансформирующего их в соответствующие движения рук виртуального объемного аватара.

SUMMARY

Karpov A.A., Kagirov I.A. Lexicon Formalization for a Computer System of Sign Language Synthesis.

Sign language is a way of inter-human communication of the deaf, where only visual-kinetic means for information transmission (hand gestures, lips articulation, mimics and emotions of facial organs) are used. Sign languages is not universal in all the counties of the world, because it is appeared and developed by a natural way in local communities and updated time-to-time with new words. Russian sign language combines some types of gestural communication: fingerspelling alphabet, which is used for gestural representation of the written national alphabet; calc signed speech, which imitates natural auditory speech and translates words of conversational speech including their order in phrases one-to-one; sign language, where each language concept (or a group of concepts) has own unique gestural representation. In contrast to the calc signed speech, the Russian sign language possesses own grammar and expression means, moreover, structure of the sign language grammar essentially differs from the grammar of spoken or written Russian language. In the result of analysis of available multimedia sign lexicons, we can conclude that the main dialects of the Russian sign language (Petersburg and Moscow dialects) differ approximately in 30-40% of gestures for corresponding words, though sometimes these differences are minor.

Lexical items both in sign language and in fingerspelling alphabet are formed in the same manner and can be formalized by the following components: configuration (shape of the hands), location of execution (places of the signing hands) and movement character (direction and manner of hands motion). In the present paper, we describe and analyze some possible methods for formal representation and transcribing of elements (gestures corresponding to concepts) of sign lexicon: International notation systems HamNoSys, SignWriting as well as original systems proposed by L. Dimskis and W. Stokoe.

As the result of the ongoing multidisciplinary research, we have adopted a virtual 3D model of human being's avatar for synthesis of the elements of the Russian sign language as well as we have proposed a software model for universal multimodal synthesis of audio-visual Russian speech and calc signed speech from an input text. At the current stage of the work, we are filling the sign vocabulary with more elements (fingerspelling alphabet, digits and the mostly used gestures and words) using an especial editor, which analyses input HamNoSys symbols and translates them into corresponding movements of the hands of the virtual 3D avatar.