

А.А. КАРПОВ

АССИСТИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ НА ОСНОВЕ АУДИОВИЗУАЛЬНЫХ РЕЧЕВЫХ ИНТЕРФЕЙСОВ

Карпов А.А. Ассистивные информационные технологии на основе аудиовизуальных речевых интерфейсов.

Аннотация. В статье обсуждаются ассистивные информационные технологии, которые представляют собой специальное программное и/или аппаратное обеспечение, предназначенное для повышения доступности информации и средств коммуникации для людей с ограниченными возможностями здоровья. Предложен комплекс ассистивных информационных технологий на базе аудиовизуальных речевых интерфейсов пользователя.

Ключевые слова: ассистивные технологии, речевые интерфейсы, многомодальные пользовательские интерфейсы, обработка аудиовизуальной речи.

Карпов А.А. Assistive Information Technologies based on Audio-Visual Speech Interfaces.

Abstract. In this paper, assistive information technologies are discussed, which are especial soft/hardware aimed for improvement of information accessibility and communication means for disabled people. A complex of assistive information technologies based on audio-visual speech interfaces is described.

Keywords: assistive technology, speech interfaces, multimodal user interfaces, audio-visual speech processing.

1. Введение. Очень многие люди, как в России, так и во всем мире ограничены в своих возможностях в связи с нарушениями слуха, зрения, речеобразования, опорно-двигательного аппарата, когнитивными нарушениями, также и многие пожилые люди испытывают подобные проблемы. Для помощи, а также социальной и профессиональной реабилитации таких людей в разных странах существуют специальные государственные программы, наиболее развитые из которых действуют в Японии, США, Израиле, Великобритании, Германии, скандинавских странах и ряде других (например, e-Inclusion и e-Accessibility). В последние годы высшее руководство России также обращает внимание на проблемы жизни инвалидов. Так, в мае 2012 года Президентом РФ была ратифицирована «Конвенция о правах инвалидов» [1], принятая несколько лет назад Генеральной Ассамблеей ООН, ключевой пункт которой состоит в том, чтобы создать условия для полноценной интеграции инвалидов в жизнь общества на всех уровнях. Недавно в России был дан старт государственной программе «Доступная среда» [2], рассчитанной пока до 2015 года, которой определяется настройка под нужды инвалидов правил работы социальных, информационных и прочих государственных служб, а также обеспечение информационной и компьютерной доступности для людей с ограниченными возможно-

стями, внедрение новых способов взаимодействия и продвижение новых товаров и услуг, использующих специальные органы и средства управления, доступные конкретным группам инвалидов. По данным Министерства здравоохранения РФ, в России насчитывается более 14 миллионов инвалидов (а это 10% населения страны), из которых около 700 тысяч детей-инвалидов, и каждый год до 1 миллиона человек впервые признаются инвалидами, что обусловлено целым комплексом причин (экологическими, техногенными, медицинскими, психологическими и др.).

Актуальная проблема общедоступности информационно-коммуникационных технологий остается первоочередной при обсуждении прав людей с инвалидностью и нетрудоспособных пожилых людей. Современные устройства коммуникации и обработки информации предполагают, что пользователь обладает способностью видеть, слышать, говорить и осязать. В результате люди с физическими, сенсорными и когнитивными недостатками не могут получить доступ к компьютерам и использовать их в силу ограниченности человеко-машинных интерфейсов, не допускающих изменений. Поэтому основное внимание в мировой практике при развитии технологий для людей с инвалидностью сфокусировано на тех аспектах, которые обеспечивали бы конструктивные решения, позволяющие осуществить альтернативное представление данных и возможность работать с устройствами ввода и вывода информации.

Для решения данной проблемы развиваются, так называемые, «ассистивные технологии» (англ. «assistive technology», от англ. «assist» — помогать, содействовать, ассистировать) или помогающие/вспомогательные технологии, обеспечивающие адаптацию управления компьютерным оборудованием, вводом данных и представлением мультимедийных информационных потоков, для людей с инвалидностью, учитывая их индивидуальные требования. Термин «ассистивные технологии» используется в «Конвенции ООН о правах инвалидов» [1], «Европейской социальной хартии», документах российской государственной программы «Доступная среда» [2], а также иных государственных административных документах и регламентах [3]. Считается, что за рубежом (в США) этот термин в его современном понимании впервые употреблялся в 1988 году в государственном документе «Technology-Related Assistance for Individuals with Disabilities Act of 1988 (The Tech Act)», с тех пор он активно используется в зарубежных странах, однако в России он стал применяться совсем недавно.

Согласно определению ЮНЕСКО [4], ассистивные/вспомогательные технологии — это устройства, продукты, оборудование, программное обеспечение или услуги, направленные на усиление, поддержку или улучшение функциональных возможностей людей с ограниченными возможностями здоровья.

Ассистивные средства и технологии могут быть самого разного характера (программные, электронные, механические, оптические и т.д.) и предназначения; это и инвалидные кресла-коляски, протезы, слуховые аппараты, оптические очки, телевизионные субтитры, роботы-помощники и роботы телеприсутствия, лифты-подъемники для колясок, звуковые сигналы светофоров, собаки-поводыри с соответствующим снаряжением, а также пандусы и направляющие на дорогах, и многое другое. Автором предложен дополняющий термин «ассистивные информационные технологии» — специальное программное и/или аппаратное обеспечения, которое повышает доступность информации и средств коммуникации для людей с ограниченными возможностями здоровья.

2. Систематизация ассистивных информационных технологий. Ассистивные информационно-коммуникационные технологии могут быть классифицированы по функциональному назначению в зависимости от категории нарушений у потенциальных пользователей:

1) технологии для людей с сенсорными нарушениями, включая:

а) ассистивные средства для лиц с нарушениями слуха (сурдоинформационные средства) [5];

б) ассистивные средства для лиц с нарушениями зрения (тифлоинформационные средства) [6];

в) ассистивные средства для лиц с нарушениями речи (голосообразующие средства);

2) технологии для людей с физическими нарушениями в работе опорно-двигательного аппарата (моторными нарушениями) [7];

3) технологии для людей с когнитивными нарушениями (умственными, психическими, нарушениями развития) [8].

4) технологии для людей с ограничениями по общемедицинским показателям (например, для пожилых или людей с серьезными заболеваниями).

На рисунке 1 показана предложенная систематизация видов ассистивных информационных технологий в зависимости от типа физических нарушений у человека.

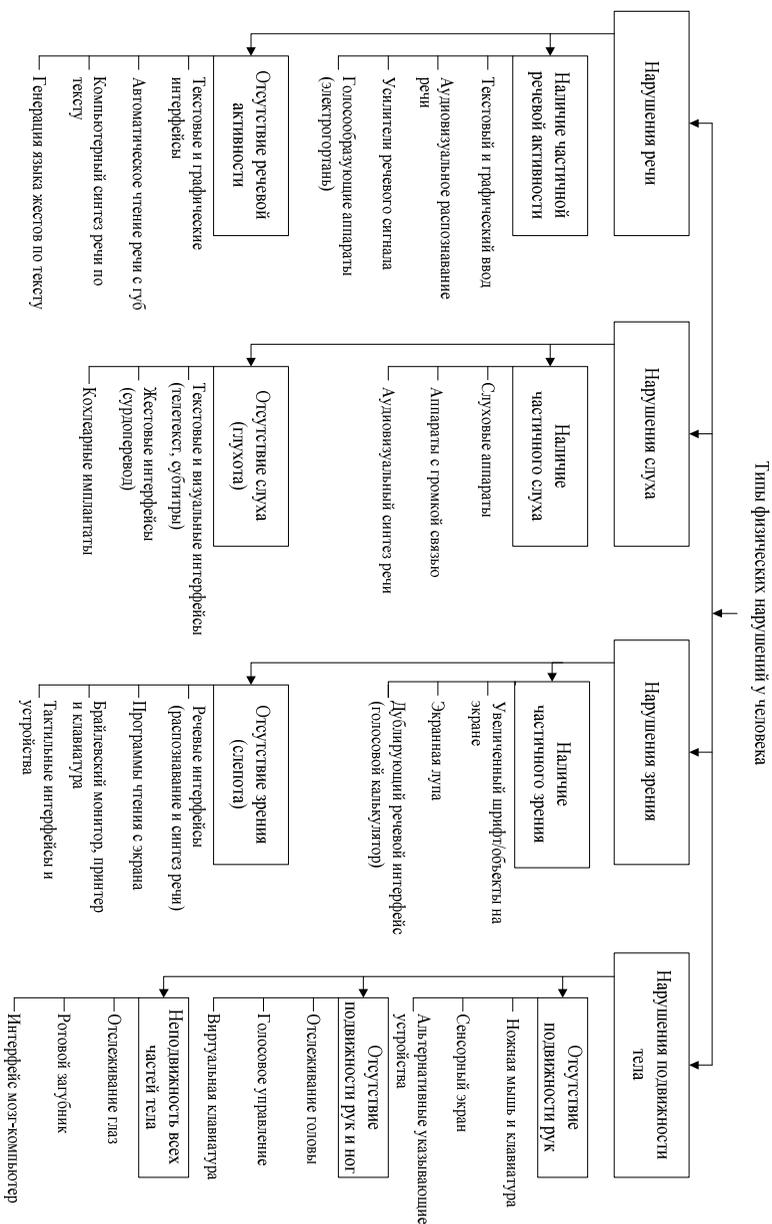


Рис. 1. Систематизация видов ассистивных информационных технологий.

В «Конвенции о правах инвалидов» определены два основных принципа создания и применения ассистивных технологий и устройств[1]:

1) Универсальный дизайн — устройство/структура предметов, обстановок, программ и услуг, призванный сделать их в максимально возможной степени пригодными к пользованию для всех людей без необходимости адаптации или специального дизайна, при этом универсальный дизайн также подразумевает и создание специализированных ассистивных устройств для конкретных групп инвалидов.

2) Разумное приспособление — внесение в случае необходимости модификаций и коррективов в существующие объекты и услуги, не становящихся несоразмерным и неоправданным бременем для общества, в целях обеспечения реализации инвалидами всех прав человека и основных свобод наравне с другими людьми. Аналогом это принципа также является адаптивная среда, применительно к которой речь идет о разумном приспособлении окружения под нужды инвалида и его потребностей, но в рамках доступных организационных, технических и финансовых возможностей их удовлетворения [4].

Современные научно-технические работы, посвященные исследованию человеко-машинного взаимодействия, направлены, в основном, на создание вычислительных машин, оборудованных большим количеством различных датчиков и сенсоров, а также сложных комплексных средств управления ими. Разрабатываемые сейчас пользовательские интерфейсы ориентированы исключительно на опытных пользователей, и в исследованиях почти не затрагиваются вопросы человеко-машинной коммуникации для лиц с ограниченными возможностями. Для организации интуитивно-понятного, универсального и естественного человеко-машинного взаимодействия наиболее перспективными являются многомодальные (особенно аудиовизуальные) интерфейсы и системы, использующие одновременно несколько способов (модальностей) обмена информации таких как: разговорная речь, мимика лица и артикуляция губ, жесты рук, головы и тела, направление взгляда пользователя и т.д. Многомодальные интерфейсы создают возможность полноценной работы и общения с компьютером не только обычным пользователям, но и тем, кому трудно работать с традиционными пользовательскими интерфейсами: детям, еще не умеющим читать и писать; непрофессиональным пользователям-новичкам (пожилым людям), а также инвалидам, имеющим нарушения двигательных функций, зрения, слуха, речеобразования и т.д.

В Таблице приведены примеры существующих в мире ассистивных информационно-коммуникационных технологий (программно-аппаратных продуктов) с разделением по типам физических/сенсорных ограничений пользователя.

Примеры ассистивных информационно-коммуникационных технологий по типам физических ограничений человека

Тип физического нарушения				Вид ассистивной технологии	Примеры программно-аппаратных средств
Зрение	Слух	Речь	Подвижность		
+				Сканирование текста с речевым выводом	ScannaR Scan N Talk Ultra
+				Экранные лупы (увеличители)	ZoomText Desktop SenseView
+				Программы чтения информации с экрана	JAWS (Job access with speech) Window-Eyes
+				Голосовые калькуляторы	Sci-Plus 300 Large Display Talking Calculator
	+	+		ПО синтеза жестового языка по тексту	iCommunicator
+		+		Синтезатор речи по тексту	Festival VitalVoice RealSpeak
			+	Отслеживание головы (трекеры)	SmartNav InterTrax
			+	Отслеживание направления взгляда/глаз	SMI BeGaze Eyegaze System iScan
			+	Ножные манипуляторы-мыши	Footime™ Foot Mouse

	+			Слуховые аппараты	PHONAK Motiva Personal FM System
		+		Голосообразующие аппараты для гортани	Servox Digital АГМ-95
	+			Кохлеарные импланты	Pulsar Sonata
	+	+		Телетайпные устройства	Compact/C FSTTY
			+	Устройства перелистывания книг	Page-Turner
+				Устройства цифрового увеличения	SmartView Graduate kReader Mobile
+				Брайлевские дисплеи	Focus Pronto SuperVario
+				Брайлевские принтеры	Basic-S Everest 4X4 PRO Tiger ViewPlus
+				Тифлокомпьютеры для незрячих	Pac Mate Omni
			+	Виртуальные клавиатуры	ScreenDoors SofType
+				Говорящие аудиокниги	DAISY (Digital Accessibility Information System)
+				Тактильные устройства	PHANTOM Desktop/Omni haptic device 3D
+			+	Распознавание/ диктовка речи	Google Voice Search Dragon Naturally Speaking SIRIUS

Следует также отметить, что в конце 2012 года компания IBM выпустила ежегодный прогноз о пяти будущих инновациях, которые изменят жизнь человека в течение ближайших 5 лет (The IBM 5 in 5) [9]. Специалисты из IBM прогнозируют новую эпоху в создании вычислительных устройств и человеко-машинных интерфейсов, связанных с автоматическим анализом компьютером всех 5 чувств, присутствующих человеку: зрение (глаза), слух (уши), вкус (язык), запах (нос), осязание/прикосновение (кожа).

3. Комплекс ассистивных информационных технологий с аудиовизуальным речевым интерфейсом. В лаборатории речевых и многомодальных интерфейсов СПИИРАН создан программно-аппаратный комплекс ассистивных информационных технологий с многомодальными (аудиовизуальными) человеко-машинными интерфейсами. Он комплексировал созданное математическое и программное обеспечение, которое позволяет производить автоматический анализ и синтез многоканальных аудио- и видеосигналов с целью организации человеко-машинных интерфейсов для ввода и вывода информации.

Общая архитектура комплекса ассистивных информационных технологий с многомодальным интерфейсом показана на рисунке 2. Программно-аппаратный комплекс состоит из набора многомодальных ассистивных систем и технологий, в том числе: 1) автоматической системы аудиовизуального распознавания русской речи; 2) компьютерной системы аудиовизуального синтеза русской речи («говорящая голова»); 3) универсальной системы синтеза русского жестового языка и аудиовизуальной речи по тексту («жестовый аватар»); 4) многомодальной системы бесконтактного взаимодействия с компьютером.

Данные системы интегрируют созданные методы, интерфейсы и программные средства (подсистемы) для автоматической обработки (как анализа, так и синтеза) аудио- и видеосигналов: автоматического распознавания звучащей речи, автоматического чтения речи по губам говорящего, аудиосинтеза речи по тексту, видеосинтеза мимики виртуальной модели головы человека, видеосинтеза жестов рук виртуальной модели человека, отслеживания движений головы пользователя.

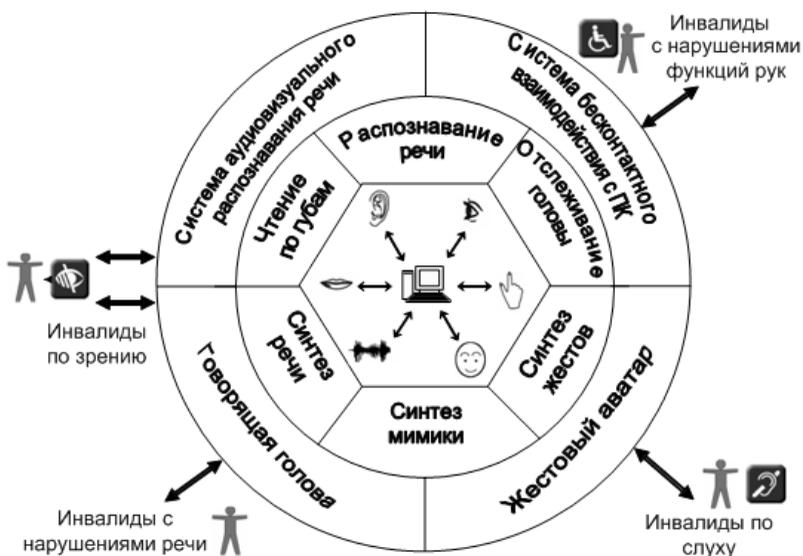


Рис. 2. Архитектура комплекса ассистивных информационных технологий.

Автоматическая система аудиовизуального распознавания русской речи [10, 11] объединяет в себе математические модели, методы и программные средства распознавания звучащей речи и автоматического чтения речи по губам говорящего. Система аудиовизуального распознавания отличается от аналогов одновременной обработкой звучащей и видимой русской речи и применением оригинального метода объединения информации на базе асинхронных моделей и весами информативности составляющих моделей. Система бимодального распознавания аудиовизуальной русской речи реализована с применением математического аппарата двоянных скрытых марковских моделей первого порядка, и позволяет производить объединение потоков векторов признаков от звуковой и визуальной модальностей речи на уровне состояний объединенной вероятностной модели. Система позволяет учитывать временное рассогласование (асинхронность) потоков соответствующих элементов звучащей речи (фоном) и визуальной речи (визем), характерное для разговорной речи, и объединять информацию о двух речевых модальностях с учетом весовых коэффициентов их информативности в зависимости от акустических условий. Автоматическая система аудиовизуального распознавания предназначена для

повышения точности и робастности распознавания слитной русской речи в акустически-зашумленных условиях эксплуатации, а также в речевых интерфейсах для организации человеко-машинного взаимодействия с людьми с ограниченными возможностями, в том числе имеющими нарушения зрения (незрячие и слабовидящие люди) и нарушения речеобразования (например, при невозможности вокализации в результате хирургических операций на гортани).

Компьютерная система аудиовизуального синтеза русской речи [12, 13] объединяет в себе виртуальные модели, методы и программные средства аудиосинтеза звучащей речи по входному тексту и видеосинтеза мимики виртуальной трехмерной модели головы человека. Система аудиовизуального синтеза (так называемая «говорящая голова») отличается от аналогов обработкой подаваемых на вход русскоязычных текстов с компьютерной генерацией слитной русской речи и применением оригинального метода синхронизации и объединения аудио- и видеомодальностей синтезируемой русской речи. Созданы два метода видеосинтеза движений виртуального лица/головы: имитационный, при котором создаётся трехмерная модель лица и настраиваются управляющие параметры для передачи мимики, выражений лица и движений губ при говорении, и компиляционный, при котором «говорящая голова» формируется путём выбора соответствующих видеофрагментов или изображений из визуальной базы данных конкретного диктора, создавая персонифицированную модель говорящей головы. Видеосинтез мимики и движений видимых органов артикуляции модели головы совмещается с синтезированным аудиосигналом, при этом предложенный метод синхронизации позволяет учесть асинхронность речевых модальностей естественную для речеобразования, а именно некоторые временные расхождения между звучащей речью и соответствующими движениями губ, что обусловлено эффектами коартикуляции и инерционностью внешних и внутренних органов речеобразования. Использование данного метода синхронизации модальностей для «говорящей головы» позволяет улучшить разборчивость синтезированной бимодальной речи и естественность модели. Система «говорящая голова» предназначена для улучшения восприятия человеком синтезируемой русской речи в акустически-зашумленных условиях эксплуатации, а также создания «очеловеченных» коммуникативных агентов и аватаров как для обычных пользователей, так и для людей с ограниченными возможностями (люди, испытывающие затруднения в речи могут использовать «говорящую голову» для замены собственной речи, а слепые люди могут ориентироваться на звуковую

модальность синтезируемой речи при получении информации от компьютера).

Универсальная система синтеза элементов русского жестового языка и аудиовизуальной русской речи по тексту [14, 15] характеризуется синхронизацией и объединением моделей, методов и программных средств аудиосинтеза звучащей русской речи и генерации визуальной артикуляции губ (составляющие системы «говорящая голова»), а также видеосинтеза элементов русского жестового языка и речи глухих людей. Основными компонентами данной системы синтеза (так называемый «жестовый аватар») являются: подсистема анализа входного произвольного русскоязычного текста для последующего аудиосинтеза звучащей речи (по словам) и видеосинтеза дактильной жестовой речи (по буквам); имитационная модель головы человека; подсистема аудиосинтеза разговорной русской речи, осуществляющая преобразование текст-речь по входному русскоязычному тексту; «говорящая голова» на основе виртуальной объемной модели головы человека и компьютерного синтеза речи; компьютерная модель верхней части тела и рук человека, в которой настраиваются параметры движений рук для синтеза элементов русского жестового языка на основе управляющих символов нотации жестов HamNoSys; многомодальный пользовательский интерфейс, интегрирующий компоненты генерации звучащей, визуальной и жестовой речи по входному русскоязычному тексту. Система «жестовый аватар» предназначена для организации универсальных человеко-машинных интерфейсов с целью коммуникации с людьми, имеющими тяжелые нарушения слуха и полностью глухими, посредством элементов русского языка жестов (калькирующей жестовой и дактильной речи, воспроизводимых жестами рук) и визуальной речи (артикуляции губ, обязательно сопутствующей жестовому языку и речи), а также вербальной коммуникации со слепыми людьми и аудиовизуальной коммуникации со зрячими пользователями.

Многомодальная система бесконтактного взаимодействия с компьютером [16, 17] отличается интеграцией методов, алгоритмов, программных средств и подсистем автоматического распознавания русскоязычных голосовых команд/речи и отслеживания движений головы пользователя для управления пользователем графическим интерфейсом компьютера без использования рук. Вместо традиционных контактных устройств (таких как клавиатура, мышь, сенсорный экран, джойстик и т.д.) для ввода информации в созданной многомодальной системе одновременно анализируются речевые команды и движения головы пользователя (жесты головой). Данная система (названная

«ICANDO: Intellectual Computer AssistaNt for Disabled Operators») предназначена для организации многомодального пользовательского интерфейса с целью бесконтактного (без использования рук) взаимодействия с компьютерными системами как обычных пользователей (например, в развивающих и игровых приложениях или же когда заняты руки в случае управления подвижным объектом), так и лиц с серьезными ограничениями в подвижности рук (например, в случае парализации верхних конечностей или нарушениях сенсомоторной системы) или же не имеющих рук/кистей вовсе.

4. Заключение. Представленный комплекс ассистивных информационных систем и технологий предназначен для помощи, реабилитации и поддержки людей с ограниченными возможностями, их интеграции в динамично развивающееся информационное общество, улучшая качество их жизни, а также делает таких людей более независимыми от помощи со стороны других лиц. Объединение современных средств вычислительной техники с речевыми и многомодальными пользовательскими интерфейсами, предназначенными для интуитивно-понятного человеку ввода/вывода информации, позволяет создавать универсальные информационно-коммуникационные технологии, в которых при взаимодействии человек-машина на первое место ставится человек с максимальным использованием его возможностей.

Литература

1. Конвенция ООН о правах инвалидов. [Электронный ресурс]. — Доступ: www.un.org/ru/documents/decl_conv/conventions/disability.shtml
2. Паспорт государственной программы РФ «Доступная среда» на 2011–2015 годы, утвержден постановлением Правительства РФ от 17.03.2011 г. N 175. [Электронный ресурс]. — Доступ: www.rosmintrud.ru/ministry/programms/3/0
3. Методика паспортизации и классификации объектов и услуг с целью их объективной оценки для разработки мер, обеспечивающих их доступность. Методическое пособие. М., 2012, 144 с.
4. Институт ЮНЕСКО по информационным технологиям в образовании. ИКТ для инклюзивного образования. [Электронный ресурс]. — Доступ: <http://iite.unesco.org/pics/publications/ru/files/3214675.pdf>
5. *Набокова Л.А.* Зарубежные "ассистивные технологии", облегчающие социальную адаптацию лиц с нарушениями развития // Дефектология. 2009. N 2. С. 84–92.
6. *Соколов В.В.* Краткий обзор современных компьютерных тифлоинформационных средств, которые могут быть использованы в процессе обучения детей с глубоким нарушением зрения // Вестник тифлологии. 2010. № 2. С. 84–87.
7. *Набокова Л.А.* Современные ассистивные устройства для лиц с нарушениями двигательного аппарата // Дефектология. 2009. N 4. С. 73–80.
8. *Набокова Л.А.* Современные ассистивные устройства для лиц с когнитивными нарушениями // Дефектология. 2009. N 3. С. 84–91.
9. The IBM 5 in 5 (2012). [Электронный ресурс]. — Доступ: www.ibm.com/smarterplanet/us/en/ibm_predictions_for_future/ideas/

10. *Карпов А.А.* Автоматическое распознавание аудиовизуальной русской речи с применением асинхронной модели // Информационно-измерительные и управляющие системы. 2010. Т. 8, N 7. С. 91–96.
11. *Карпов А.А., Ронжин А.Л., Лобанов Б.М., Цирульник Л.И., Железны М.* Разработка бимодальной системы аудиовизуального распознавания русской речи // Информационно-измерительные и управляющие системы. 2008. Т. 6, N 10. С. 58–62.
12. *Карпов А.А., Цирульник Л.И., Железны М.* Разработка компьютерной системы “говорящая голова” для аудиовизуального синтеза русской речи по тексту // Информационные технологии. 2010. Т. 9, N 8. С. 13–18.
13. *Карпов А.А.* Аудиовизуальный речевой интерфейс для систем управления и оповещения // Известия ЮФУ. Технические науки. 2010. N 3. С. 218–222.
14. *Карпов А.А.* Машинный синтез русской дактильной речи по тексту // Научно-техническая информация. Серия 2: Информационные процессы и системы. 2013. N 1. С. 20–26.
15. *Карпов А.А.* Компьютерный анализ и синтез русского жестового языка // Вопросы языкознания. 2011. N 6. С. 41–53.
16. *Карпов А.А.* Когнитивные исследования ассистивного мультимодального интерфейса для бесконтактного человеко-машинного взаимодействия // Информатика и ее применения. 2012. Т. 6, N 2. С. 77–86.
17. *Карпов А.А.* iCanDo: Интеллектуальный помощник для пользователей с ограниченными физическими возможностями // Вестник компьютерных и информационных технологий. 2007. N 7. С. 32–41.

Карпов Алексей Анатольевич — канд. техн. наук, доцент, старший научный сотрудник лаборатории речевых и мультимодальных интерфейсов СПИИРАН. Область научных интересов: автоматическое распознавание речи, мультимодальные интерфейсы, аудиовизуальное распознавание и синтез речи. Число научных публикаций — 170. karpov@iias.spb.su; СПИИРАН, 14-я линия В.О., д. 39, г. Санкт-Петербург, 199178, РФ; р.т. +7(812)328-7081, факс +7(812)328-7081.

Karpov Alexey Anatolyevich — Ph.D., assoc. professor, senior researcher, Speech and Multimodal Interfaces Laboratory, SPIIRAS. Research interests: automatic speech recognition, multimodal interfaces, audio-visual speech recognition. The number of publications — 170. karpov@iias.spb.su; SPIIRAS, 39, 14-th Line V.O., St. Petersburg, 199178, Russia; office phone +7(812)328-7081, fax +7(812)328-7081.

Поддержка исследований. Данное исследование поддержано Минобрнауки РФ (ФЦП «Исследования и разработки», госконтракт № 11.519.11.4025), Советом по грантам Президента РФ (проект № МК-1880.2012.8), а также КНВШ Правительства Санкт-Петербурга.

Рекомендовано лабораторией речевых и мультимодальных интерфейсов, заведующий лабораторией Ронжин Ан.Л., д-р техн. наук, доцент.
Статья поступила в редакцию 13.03.2013.

РЕФЕРАТ

Карнов А.А. Ассистивные информационные технологии на основе аудиовизуальных речевых интерфейсов.

В статье предложен аналитический обзор сферы ассистивных/вспомогательных технологий и технических средств реабилитации инвалидов, ориентированных на помощь людям с ограниченными возможностями здоровья, который показал, что в мире (в частности, Японии, США, Израиле, странах Европы, особенно в скандинавских странах) этой области науки и техники уделяется очень большое значение, но в рамках российской науки до настоящего времени эта область, за единичными исключениями, практически не развивалась.

Предложена систематизация ассистивных информационных технологий по их функциональному назначению в зависимости от категории нарушений у потенциальных пользователей, включающая: а) технологии для людей с сенсорными нарушениями (нарушениями слуха — сурдоинформационные средства; нарушениями зрения — тифлоинформационные средства; средства для лиц с нарушениями речи — голосообразующие средства); б) технологии для людей с физическими нарушениями опорно-двигательного аппарата (моторными нарушениями); в) технологии для людей с когнитивными нарушениями (умственными, психическими, задержкой развития); г) технологии для людей с ограничениями по общемедицинским показаниям (например, пожилых людей). Приведены примеры существующих ассистивных информационных технологий для компенсации различных типов физических и сенсорных нарушений.

Описана архитектура комплекса ассистивных информационных технологий с многомодальным пользовательским интерфейсом, интегрирующим в своем составе: а) автоматическую систему аудиовизуального распознавания русской речи; б) компьютерную систему аудиовизуального синтеза русской речи («говорящая голова»); в) универсальную систему синтеза русского жестового языка и аудиовизуальной речи по тексту («жестовый аватар»); г) многомодальную систему бесконтактного взаимодействия с компьютером. Данные системы комплексировуют разработанные методы, интерфейсы и программные средства (подсистемы) для автоматической обработки (как анализа, так и синтеза) аудио- и видеосигналов: автоматического распознавания звучащей речи, автоматического чтения речи по губам говорящего, аудиосинтеза речи по тексту, видеосинтеза мимики виртуальной 3D модели головы человека, видеосинтеза жестов рук виртуальной модели человека, отслеживания движений головы пользователя посредством компьютерного зрения.

SUMMARY

Karpov A.A. Assistive Information Technologies based on Audio-Visual Speech Interfaces.

In the given paper, an analytical survey of assistive technologies and technical means of rehabilitation for people with disabilities is presented. Such technologies are aimed to help disabled people in daily life. This area is well developed in many innovative countries all over the world (e.g., Japan, USA, Israel, European countries, especially Scandinavian ones); however, in the framework of Russian science and technology, this domain has not elaborated yet.

This paper proposes a systematization of assistive information technologies based on functional destination depending on disability types of potential users, including: assistive technologies for people with sensor disabilities (hearing impaired, vision impaired, impaired in generation of speech); assistive technologies for people with physical disabilities (motor-disabled people); technologies for people with cognitive impairments; technologies for people with medical restrictions (for instance, elderly people). Some examples of existing assistive information technologies aimed for compensation of various types of physical and sensor disabilities are given.

Architecture of a complex of assistive information technologies/systems based on audio-visual speech interfaces is described. It integrates some systems including: an automatic audio-visual Russian speech recognition system; a computer system for audio-visual Russian speech synthesis (talking head); a universal system for synthesis of Russian sign language and audio-visual speech by input text (signing avatar); a multimodal system for hands-free interaction with a computer. These systems integrate developed methods, user interfaces, and software components for automatic processing (both analysis and synthesis) audio and video signals: automatic speech recognition, automatic lip-reading, text-to-speech (audio) synthesis, video synthesis of mimics of a virtual 3D human head, visual synthesis of hand gestures (signs) of a virtual 3D human body, computer vision based head tracking.