

Н.Б. Васильева, О.Н. Сухоручкина, В.В. Яценко

Особенности построения модели речевого общения пользователя с многофункциональным сервисным мобильным роботом

Описано формирование базы данных и знаний для генеративной модели интерпретации речевого сигнала в рамках предметных областей. Предложен подход для упрощения пошагового диалога человека и многофункционального мобильного робота. Рассмотрена структура, позволяющая оперировать предметными областями, языками, типами смыслов и предложений в результате генерирующие соответствующие естественно языковые тексты.

Ключевые слова: мобильный робот, человеко-машинный интерфейс, предметная область, типы смыслов, типы предложений, акустическая модель, фонемная ошибка.

Описано формування бази даних та знань для генеративної моделі інтерпретації усномовного сигналу в межах предметних областей. Запропоновано підхід до спрощення покрокового діалогу між людиною та багатофункціональним мобільним роботом. Розглянуто структуру, яка дозволяє оперувати предметними областями, мовами, типами сенсів та речень, які в результаті генерують відповідні природномовні тексти.

Ключові слова: мобільний робот, людинно-машинний інтерфейс, предметна область, типи сенсів, типи речень, акустична модель, фонемна помилка.

Введение. Речевое общение человека и технической системы постепенно входит в нашу повседневную жизнь. Например, подсистема *Siri* в устройствах на базе *iOS* и подсистема *Cortana* в *MS Windows* и *Android* помогают находить всевозможную информацию как в Интернете, так и на самом устройстве. В основе моделирования такого общения лежат технологии анализа, распознавания и интерпретации речевого сигнала в пределах предметных областей и устного диалога [1, 2]. Развиваются сервисные службы с использованием диалоговых систем, позволяющие получать информацию о расписании движения транспорта, текущем состоянии банковского счета, бронировании билетов и т.д. [3].

Особый интерес разработчиков и потребительского рынка направлен на сервисные мобильные роботы (МР) для непрофессионального использования. Такие МР, являясь сложными техническими системами, требуют наличия развитого человеко-машинного интерфейса (ЧМИ), обеспечивающего пользователям без специальной подготовки простые и комфортные условия эксплуатации устройства.

В продаже уже имеются роботы-пылесосы, мойщики полов и окон, газонокосилки, роботы

для очистки бассейнов [4]. Однако каждая из названных и им подобных систем имеет строго определенное функциональное назначение, которое для системы управления (СУ) сводится, как правило, к одной целевой задаче. Пользователю доступен выбор вариантов и некоторых параметров ее исполнения. ЧМИ в данном случае ограничен узкой предметной областью и, в случае использования речевого интерфейса, словарь может содержать лишь несколько имен команд, связанных с возможными действиями робота, и, возможно, несколько имен параметров для уточнения условий выполнения задачи. Практическая реализация речевого интерфейса при этом не содержит лингвистической сложности, связанной с анализом смыслового наполнения и связанности слов речевых высказываний пользователя, поскольку использование речи в ЧМИ большинства технических устройств остается на так называемом сигнальном уровне, когда речь используется лишь в роли простого условного звукового возбудителя для СУ [5–6]. Технической системе достаточно воспринять звуковой сигнал, и в этом сигнале идентифицировать конкретную команду из известных ей заранее обусловленных команд языка интерфейса, заданных в виде аку-

стических моделей. После чего СУ активизирует исполнение соответствующей реакции системы в виде ее конкретного действия. Задача модуля речевого интерфейса в данном случае – идентификация известных системе звуковых сигналов.

Особый интерес представляют роботы, способные автономно выполнять широкий репертуар *сформулированных в общем виде* заданий в не полностью детерминированном динамически изменяющемся окружающем пространстве, что часто порождает непредсказуемую вариабельность ситуаций, в которые попадает техническая система и ее пользователь. Многофункциональные роботы в виде прототипов и экспериментальных образцов сегодня демонстрируются самыми разнообразными коллективами разработчиков на всевозможных выставках и через средства массовой информации [7–9]. Роботы–помощники для людей пожилого возраста, инвалидов, офисные помощники, роботы–официанты и музейные гиды, роботы для гостиничного обслуживания, мобильные роботы как системы дистанционного присутствия, используемые в том числе для дистанционного обучения либо консультирования, домашние роботы для досуга – таков спектр деятельности, который непрофессиональные пользователи ожидают от робототехнических систем сервисного типа.

Отметим, что сервисные роботы бытового назначения в дальнейшем будут обладать все большим уровнем автономности как способностью технической системы выполнять без вмешательства человека сложные задания в различных внешних условиях. Предметная область и словарь возможного человеко-машинного общения при этом существенно расширяются, смысловые связи и конструкции допустимых речевых высказываний усложняются. Весьма затруднительно предусмотреть, создать и хранить в памяти системы акустические модели (АМ) всех возможных речевых высказываний пользователя, на которые СУ робота должна адекватно реагировать. Кроме того, в реальных условиях пользователь на самом деле не должен быть ограничен исключительно словарем предметной обла-

сти. Поэтому предлагается использовать фоновое распознавание слитной речи как на словесном, так и на морфемном уровнях.

Базовая постановка задачи

Комфортность человеко-машинного общения с сервисными МР для непрофессионального использования может быть повышена благодаря организации в ЧМИ автоматической системы ведения контекстно-зависимого диалога со смысловой интерпретацией речи пользователя при формулировании заданий технической системе [1, 10]. Помимо технологии ведения речевого диалога, ЧМИ дополняется технологией синтеза речевых сообщений – автоматически генерируемых роботом докладов о своем состоянии, месте нахождения, окружающей обстановке, стадиях выполнения заданий и возможных обращениях к пользователю, например, при возникновении нештатных ситуаций. Такие интерфейсные решения обеспечивают поддержку двусторонней передачи сведений между пользователем и технической системой в вербальном звуковом виде. Для успешного человеко-машинного общения важна понятность этих сведений обеим сторонам [11]. В математической лингвистике и в искусственном интеллекте как научных дисциплинах задачам автоматической смысловой интерпретации речи уделяется огромное внимание. Некоторые проблемы построения семантической модели речи и интеграции ее в ЧМИ рассмотрены в [12].

Конечной целью создания интеллектуальной системы устного диалога, реализованной на основе задач распознавания и смысловой интерпретации слитой речи, используемой в едином взаимосвязанном процессе *речь–речь*, есть определение смысла сообщения, его интерпретация и соответствующая реакция с учетом состояния МР.

Человек формулирует задание или ставит вопросы МР, который должен понимать эти задания, задавать при необходимости уточняющие вопросы, докладывать о ходе выполнения задания и извещать о завершении, реагировать и отвечать на запросы пользователя. Таким образом, для обеспечения речевого общения человека и робота задачи распознавания

и понимания речевого сигнала должны исполняться в одном взаимосвязанном процессе.

Понятийное множество взаимодействия – база семантических составляющих элементов словаря человеко-машинного интерфейса робота

Условно схему информационного взаимодействия пользователя с роботом как объектом управления (ОУ) можно представить рис. 1.

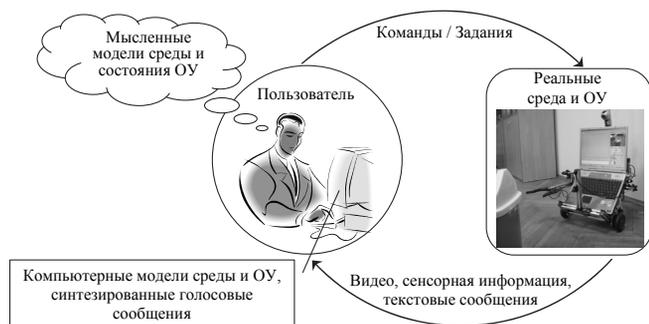


Рис. 1. Схема человеко-машинного общения

Семантическая модель мира с соответствующей моделью языка ЧМИ для одного типа технических систем часто не применима для других систем. Для каждого автономного робота существует не реальный внешний мир, а его отражение в виде модели мира, действительной только для данного типа систем [13]. Роботы, оснащенные различными аппаратно-программными средствами восприятия внешнего мира и интерпретации получаемой о мире информации, воспринимают мир по-разному. Понятийное множество конкретного автономного МР с соответствующими семантическими составляющими элементами словаря ЧМИ опирается на совокупность программных структур верхнего уровня управляющей системы робота, описывающих пространственную модель воспринимаемого внешнего мира, модель собственного устройства и текущего состояния робота, соответствующий словарь имен объектов и их характеристик, воспринимаемых бортовой сенсорной подсистемой, библиотеку формализованных правил выполнения определенных целенаправленных действий робота с соответствующими именами возможных действий МР и их параметров [14]. Т.е. субъективность мировосприятия МР связана с конкретным набором

органов восприятия и их чувствительностью, используемыми в управляющей системе методами анализа воспринятой информации и построения причинно-следственных связей, способами организации памяти и извлечения из нее необходимых данных, а также с репертуаром доступных для исполнения действий.

МР с интеллектуальным управлением способны автономно выполнять задания пользователя, сформулированные в виде высказываний с довольно общим смыслом. В подсистеме ЧМИ, отвечающей за ведение диалога, существенное значение имеет проверка соответствия содержащегося в высказывании пользователя желаемого целевого состояния робота возможностям его управляющей системы по переводу робота из текущего состояния в задаваемое. Иными словами, модуль распознавания и смысловой интерпретации речи пользователя должен постоянно взаимодействовать с понятийным множеством управляющей системы в виде семантической модели мира робота. Последовательность уточняющих вопросов, генерируемых подсистемой ведения диалога для минимизации расхождений в понимании пользователем и управляющей системой робота формулируемого задания, должна приводить к полной определенности всех параметров, однозначно описывающих целевое состояние робота (рис. 2).

Моделирование смысловой интерпретации речевых сообщений при взаимодействии человека и мобильного робота

Рассмотрим подход к распознаванию и смысловой интерпретации речи при взаимодействии человека и робота в рамках генеративной модели исходя из иерархии речевых образов [1].

Предметная область (ПО) состоит из конечного множества типов смыслов (ТС). В каждый ТС входит множество эквивалентно содержательных типов предложений (ТП), которые описываются *LISP*-структурами [1, 3]. Тип предложения – это конструкция, экономно задающая множество предложений, полученных из *LISP*-структуры независимыми допустимыми заменами, перестановками и опусканием подсловарей. Причем, в роли подсловарей могут служить параметры (метаслова) [2].

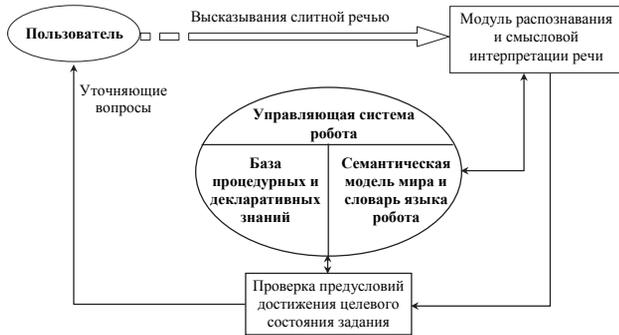


Рис. 2. Структура подсистемы ведения диалога

Рассмотрим пример ТС, который задает императив перемещения к некоторому объекту. Этот ТС предлагается выразить двумя ТП:

$$\left(\left(\begin{array}{c} \text{подойди} \\ \text{переместись} \end{array} \right) \left(\begin{array}{c} \text{пожалуйста} \\ * \end{array} \right) \right) \times \\ \times \left(\left[\begin{array}{c} \text{объекту} \\ \text{предмету} \end{array} \right] (obj_n) \right), \\ \left(\left(\begin{array}{c} \text{подойди} \\ \text{переместись} \end{array} \right) \left(\begin{array}{c} \text{пожалуйста} \\ * \end{array} \right) \right) (к\ obj_d),$$

где obj_n и obj_d – параметры (метаслова), задающие имя объекта–цели в именительном и дательном падежах соответственно, примерами которых в именительном падеже являются: *стол, карта Киева, розы*. Им соответствуют имена объектов–цели в дательном падеже: *столу, карте Киева, розам*.

В круглых скобках () указаны подсловари, которые можно переставлять местами, а в квадратных [] – нельзя переставлять. Переставлять подсловари можно только внутри старших скобок. Символ * означает пустое слово.

Нетрудно убедиться, что приведенный тип предложения, заданный первой структурой, задает $3! \cdot 2 \cdot 2 \cdot 2 \cdot 1 = 48$ различных предложений, допустимых в языке диалога и таких, которые выражают один и тот же ТП относительно ТС императива перемещения к объекту. Среди этих предложений будут, например, и такие:

Подойди, пожалуйста, к объекту карта Киева. Пожалуйста, к объекту карта Киева подойди. Подойди к предмету карта Киева. К карте Киева, пожалуйста, переместись.

Очевидно, что приведенные ТП дают возможность генерировать многие синтаксически

допустимые предложения разговорной речи. Отметим, что среди этих предложений могут быть образованы предложения, например второе, которые не есть типичными для разговорной речи, но, тем не менее, допустимыми.

В процессе интерпретации речевого сигнала (многозначный) результат преобразования текста в речь сравнивается со сгенерированными предложениями исходя из всех ТП для выбранного ПО и по результатам сравнения формируется ответ интерпретации в виде номера типа смысла и значений его параметров (метаслов) [1]. Описание ПО входит в базу данных и знаний в общем случае системы устного диалога или же модуля распознавания и интерпретации речи для речевого взаимодействия с МР.

Структура базы данных и знаний

В системе устного диалога типа *человек–МР–человек* взаимодействие сторон происходит наиболее естественным путем – голосом. Устный диалог между человеком и МР рассматривается в рамках определенной ПО. Представленная информация должна быть связной и структурированной. Именно такая информация формирует базу данных и знаний, с которой может работать эксперт–лингвист и которая есть информационной моделью ПО [15].

Рассмотрим общую структуру системы устного диалога человека и МР (рис. 3). Мы предлагаем рассматривать диалог как заполнение один за другим ТП или постепенно, путем озвучивания уточнений одной стороны ко второй.

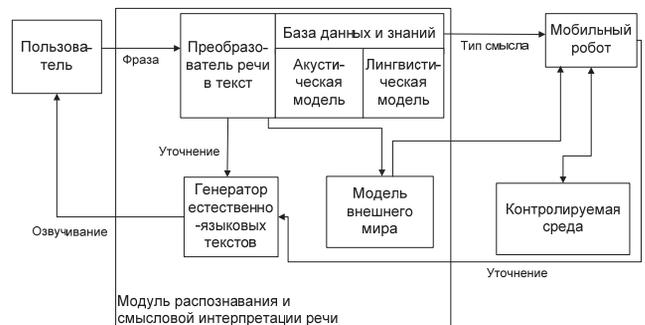


Рис. 3. Общая структура модуля распознавания и смысловой интерпретации речи для взаимодействия пользователя с МР

Уточнения в виде канонических форм передаются на генератор естественно-языковых текстов, озвучиваемых далее. Для генерирования фраз и

уточнений как человека, так и робота предлагается использовать единую базу данных и знаний. Характерное отличие устного диалога от голосового управления устройством – это то, что МР сам формирует уточнения с учетом изменений контролируемой им среды, например, трехмерной сцены.

МР при выполнении задания задает уточняющие вопросы, докладывает о состоянии его выполнения и о завершении работы. Генерирование соответствующих текстов происходит в соответствии со спецификацией ТС, являющейся общей с системой интерпретации речевого сигнала.

Пример многоязычной спецификации ТС приведен в табл. 1. Ограничения на следование подсловарей моделируется лингвистической моделью, поэтому скобки для спецификации допустимых перестановок подсловарей игнорируются, что облегчает работу эксперта при создании базы данных и знаний.

Приведем пример, объясняющий работу предложенной структуры. Пусть человек говорит в микрофон такую фразу в рамках заданной ПО, как «Перейди вперед». Речевой сигнал поступает в модуль распознавания и интерпретации речи, в результате чего получаем ответ распознавания и интерпретации в виде имени ТС *MID* (*Meaning Type Identification*) и распознанного текста. В нашем случае ТС принимает значение «navigate» (*MID* = «navigate»).

Таблица 1. Пример спецификации типов смысла

<i>MID</i>	MT <i>Ru</i>	MT <i>Ua</i>	MT <i>Eng</i>
<i>take_obj</i>	возьми/брать [объект] <i>\$obj</i> [<i>\$obj</i>] [<i>\$direction</i>]
<i>come2obj</i>	подойди к [объекту] <i>\$obj</i> [<i>\$obj</i>] [<i>\$direction</i>]		
...			
<i>navigate</i>	перейди/идти <i>\$where_to</i> [<i>\$valid</i>]		
<i>put_obj</i>	открой схват		
...			

Система обращается к базе данных и знаний, где в таблице ТС находит соответствующее представление ТП: «Перейди *\$where_to* [*\$valid*]» (табл. 1). Это выражение содержит метаслова, т.е. параметры команды технической системы, представленные в виде переменных *\$where_to* и

\$valid. Наличие метаслов в ТП требует уточнения в том случае, если фраза не была произнесена или распознана в полной мере. Отметим, что метаслова, поданные в квадратных скобках, не обязательны при формировании ТП, поэтому уточнения не требуют.

Кроме того, система обращается с вопросом к метасловам и, при отсутствии необходимых метаслов, задает уточняющий вопрос согласно табл. 2. Допустим, что пользователь не указал направление движения. Тогда, согласно таблице, система устного диалога должна задать вопрос *куда?* или *в каком направлении?*. Пользователь должен дать корректный ответ на этот вопрос, например, сказать: *вперед* или *к столу*. Тогда переменная примет значение *\$where_to = вперед*. Вопрос к необязательному неинициализированному параметру *\$valid* не задается. В зависимости от состояния контролируемой среды и технической системы, она может инициировать *необязательные* уточнения, например, *к какому столу?* В процессе выполнения задания МР с определенной долей детализации комментирует выполняемые действия.

Таблица 2. Пример спецификации вопроса к метасловам

<i>ID</i>	<i>Q: ru</i>
<i>\$where_to</i>	Куда?
	в каком направлении?
...	
<i>\$obj</i>	Какой объект?
	назовите объект
	что за объект?
	где находится объект?
<i>\$obj_n</i>	Какой объект?
...	

Приведем пример обратной связи с МР:

Человек:	Перейди к столу		
		Определяю маршрут движения	:MP
		Начинаю движение	
		Подхожу к столу	
		Задание выполнено	

Из примера следует, что ТП доклада о работе МР совпадают с императивами с точностью до формы глагола. Поэтому в базу данных и знаний предлагается включить словарь словоформ для подбора необходимой формы глагола.

Пример спецификации имен и значений переменных приводится в табл. 3. Переменные,

которые передаются диалоговой системе, могут рекурсивно определяться через другие переменные. В этом случае переменная $\$where_to$ определяется через переменные $\$position$, $\$direction$ и $\$range$.

Таблица 3. Пример рекурсивного определения переменных

ID	\$val	Ru
$\$where_to$		$\$position \$direction \$range$
$\$position$	$pos01$	в заданное положение
	$pos02$	в начальное положение
	$pos03$	в стартовое положение
	...	
$\$direction$	$dir01$	вперед
	$dir03$	к объекту
	$dir04$	назад
	...	
$\$range$		$\$linear \$angular$
	...	
$\$wobj_n$	$wobj_n1$	первый обнаруженный
	$wobj_n2$	ближайший
		все возможные

Система устного диалога в каждый момент получения i -го ответа распознавания и интерпретации сохраняет предысторию:

$$History\ cash_j = \{ (MID_j, params_j) \mid 0 < j < i \},$$

где MID_j и $params_j$ соответственно имя ТС и имя параметров, полученных вследствие j -го ответа распознавания и интерпретации. В этой памяти сохраняется информация, гипотетически позволяющая уменьшить количество уточнений.

В момент, когда техническая система в полной мере определяет ТС, происходит реакция системы в виде речевого сообщения и/или действия на объекты оперативной среды.

Выбор акустической модели распознавания речи

При преобразовании *Речь-в-Текст* особо важную роль играет входящая в ее базу данных и знаний АМ, поскольку от этого зависит надежность как распознавания, так и смысловой интерпретации. Поэтому, следует рассмотреть вопрос, какую АМ распознавания речи выбрать для эффективной работы технической системы и каким образом ее формировать [16].

Параметры АМ оцениваются на основании речевого корпуса, состоящего из структурированного множества речевых фрагментов, текстового описания этих фрагментов, а также

инструментария для оперирования всем множеством данных корпуса.

Одним из способов формирования речевого корпуса – это запись диктора, зачитывающего определенный текст, в котором представлено все фонетическое разнообразие украинской речи, описанный в [11]. Это позволяет избежать этапа ручного транскрибирования и сегментирования, а также одновременно создать текстовый корпус в соответствующей предметной области. В его основе – электронные тексты, которые находятся в свободном доступе в Интернете. Другой способ – записать эфирную речь и транскрибировать его. Таким образом создан многодикторный корпус речи, описанный в [17]. Оба из предложенных корпусов имеют свои преимущества и недостатки.

Для исследований, описанных в [18], элементарной единицей, использованной при построении предложенных АМ, была фонема. Проанализированы такие варианты АМ распознавания: модель, построенная только на слитой речи; модель, объединяющая слитую речь и изолированные слова; модель, которая не учитывает или учитывает частичную ударность гласных.

В экспериментальных исследованиях оценивались показатели фонемной ошибки (PER – *Phoneme Error Rate*), которые отображают отношение между разницей правильно распознанных фонем и ошибочных вставок к общему числу фонем.

При распознавании использована грамматика свободного порядка следования субсловных (морфемных) элементов: фонем, слогов и т.д. Ряд экспериментов проведен наложением ограничений на последовательности элементов использованием биграммной лингвистической модели (ЛМ) на фонемно-морфемном уровне.

Проведены предварительные эксперименты на многодикторной АМ, элементарной единицей которого была фонема-трифон. Использованная в исследованиях ЛМ построена на текстовом корпусе, с которого выбирались тексты для однокдикторной речевой базы и тексты контрольных выборок (КВ). В табл. 4 приведены результаты распознавания АМ, построенной на одно- и многодикторных речевых кор-

пусах для КВ слитой речи на разных речевых образах, используя биграммные ЛМ.

Сравнивая результаты распознавания АМ, построенных на одно- и многодикторных речевых корпусах, можем сделать выводы, что эффективность распознавания также зависит от использованной ЛМ. В рамках одной предметной области результаты распознавания, даже для биграммной ЛМ, обнадеживающие. Таким образом, можно предположить, что созданные ЛМ способом, описанным в [18], откроют перспективы использования акустической модели многодикторного корпуса слитой речи, элементарной частью для обучения будет не только фонема, но и контекстно-зависимая фонема-трифон.

Таблица 4. Показатели фонемной ошибки PER (%) распознавания акустической модели, построенной на одно- и многодикторном речевом корпусе для КВ слитой речи на основании разных речевых образов

Название КВ	Фоне-ма	Откры-тый слог	Слог по прави-лам сло-годе-ления	Фонема		
				Откры-тый слог	Слог по прави-лам сло-годе-ления	Слог по прави-лам сло-годе-ления
	Однодикторный корпус			Многодикторный корпус		
Случайная	24,8	27,5	24,8	47,3	45,9	37,5
Частотная	27,7	24,2	22,0	50,9	38,5	45,2
Википедия	28,2	31,8	28,2	49,5	40,9	40,2

Заключение. Показано, что подход к смысловой интерпретации речи в рамках генеративной модели [1] может служить основой для создания технологий и систем устного диалога на естественном языке между человеком и технической системой, в частности с МР.

Предложенный способ формирования базы данных и знаний систем устного диалога позволяет эффективно создавать такие системы для разных предметных областей, в которых работает МР, с привлечением экспертов-лингвистов.

Надежность распознавания речи, а соответственно и корректность интерпретации зависят от выбранного речевого образа (фонемы, слога, морфемы). Улучшение результатов распознавания происходит как путем усложнения лингвистической части, так и подбором наиболее соответствующих АМ.

Планируется исследовать влияние лингвистической и акустической составляющих технологии распознавания речи на улучшение результатов смысловой интерпретации высказываний в системе устного диалога.

1. Винцюк Т.К. Анализ, распознавание и смысловая интерпретация речевых сигналов. – Киев. Наук. думка, 1987. – 262 с.
2. Яценко В.В. Параметризація типів речень предметної області для системи усного фразника-перекладача // Штучний інтелект. – 2011. – № 4. – С. 134–142.
3. Сажок М., Яценко В. Система усного перекладу на основі інтерпретації мовленнєвого сигналу в межах предметних областей // Праці Міжнар. конф. «Укр-Образ'2010». – Київ. – 2010. – С. 103–106.
4. Бенецкий С. 9 домашних роботов, которые можно купить уже сегодня. – <http://www.forbes.ru/stil-zhizni/pokupki/46726-10-domashnih-robotov-kotoryh-mozhno-kupit-uzhe-segodnya/slide/1>
5. Бернштейн Н.А. Физиология движений и активность. – М.: Наука, 1990. – 496 с.
6. Ушакова Т.Н. Функциональные структуры второй сигнальной системы. Психофизиологические механизмы речи. – М.: Наука, 1979. – 248 с.
7. Care-O-bot. – <http://www.care-o-bot.de/en/care-o-bot-4.html>
8. Evolution of ASIMO. – <http://world.honda.com/ASIMO/technology/index.html>
9. Synergy SWAN. – http://rbot.com/?page_id=19&lang=ru
10. Понов Э.В. Общение с ЭВМ на естественном языке. – М.: Наука, 1982. – 360 с.
11. Васильева Н.Б. Дослідження невідповідності шкали акустичної та лінгвістичної моделей розпізнавання злитого українського мовлення // Штучний інтелект. – 2012. – № 3. – С. 118–125.
12. Шамис А.Л. Поведение, восприятие, мышление: проблемы создания искусственного интеллекта. – М.: Едиториал УРСС, 2005. – 224 с.
13. Рапопорт Г.Н., Герц А.Г. Искусственный и биологический интеллект. Общность структуры, эволюция и процессы познания. – М.: КомКнига, 2005. – 312 с.
14. Сухоручкина О.Н. Структуры функциональной организации интеллектуализированного управления мобильной системой // УСиМ. – 2007. – № 3. – С. 26–33, 63.
15. Яценко В. Автоматизовані засоби формування лінгвістичної бази даних і знань для системи усного перекладу // Штучний інтелект. – 2012. – № 4. – С. 211–220.
16. Васильева Н.Б. Використання граматики вільного порядку слідування фонем і складів для пофонемного розпізнавання злитого мовлення // Там же. – 2011. – № 4. – С. 80–86.

Окончание на стр. 28

17. *Ukrainian Broadcast Speech Corpus Development* / V. Pylypenko, V. Robeiko M. Sazhok et al. // Міжнар. конф. «Speech and Computer – SpeCom», Kazan, RF. – 2011.
18. *Васильєва Н.Б., Федорин Д.Я.* Проблеми створення систем розпізнавання мовлення для різних комп'ю-

терних платформ // Штучний інтелект. – 2013. – № 4. – С. 158–167.

E-mail: n.vassilleva@gmail.com, sazhok@gmail.com, sukhoru@gmail.com, yatsenko.valya@gmail.com
© Н.Б. Васильєва, О.Н. Сухоручкіна, В.В. Яценко, 2015

UDC 004.934

N.B. Vasylieva, O.N. Sukhoruchkina, V.V. Yatsenko

Features of Constructing a Spoken Communication Model for a Multi-Functional Service Mobile Robot

Keywords: mobile robot, HMI, subject area, meanings types, sentences types, acoustic model, Phoneme Error Rate.

Introduction. Service mobile robots for everyday operation are in great interest of developers and market, which brings interaction between user and a machine to the foreground. The most natural way of human communication is a spoken dialogue. Meanwhile, current robotic systems may provide only formalized command spoken elements. Therefore, we propose a way to introduce natural speech man–machine interaction based on the experience of a speech technology development.

The purpose of research is to provide a spoken communication between a person and a mobile robot by means of generative model for the speech signal understanding within subject areas.

Results. We consider exchange of the utterances as a step-by-step sentence/meaning type initialization. These constructions are specified by an expert, who forms the data and knowledge base for a subject area. We introduced a way to specify a common structure imperatives, questions and state reports. The result of a speech understanding is delivered to a robot, and the robot, in turn, passes back to the system a request for further task detalization or reports on its state. Acoustic model efficiency is crucial for both speech-to-text conversion and speech understanding even more. Therefore, we explore the influence of the several speech patterns on recognition accuracy.

Conclusions. It is shown that the generative model for speech understanding can be applied to human-machine natural communication system development. Speech recognition/understanding accuracy may be improved on account of the speech pattern base model selection.

