

Акустико-семантический контент речевой коммуникации в робототехнике

Р. К. Потапова^{1,*}, В. В. Потапов^{2,†}

¹Московский государственный лингвистический университет»,
Институт прикладной и математической лингвистики. Россия, 119034, Москва, ул. Остоженка, д. 38
²Московский государственный университет имени М.В. Ломоносова, филологический факультет,
Учебно-научный компьютерный центр. Россия, 119991, Москва, Ленинские горы, д. 1, стр. 51
(Статья поступила 23.10.2019; Подписана в печать 05.11.2019)

В статье представлен оригинальный подход к исследованию акустико-семантического контента, обеспечивающего эффект речевой коммуникации реципрокного типа, позволяющий реализовать двунаправленный диалог «человек-машина» и «машина-человек», что на данной ступени развития робототехники является одной из приоритетных задач реализации обратной связи в быстродействующем экономном режиме «стимул-реакция» и «реакция-стимул» применительно к различным видам коммуникации в робототехнике, что нашло отражение в подходах, объединяемых в настоящее время в научно-исследовательскую область «Interactive Collaborative Robotics» (например, см. R. Potapova «Speech Dialogue as a Part of Human-Machine Interactive Systems» in: «Interactive Collaborative Robotics. 1st International Conference — ICR 2016», LNAI 9812, Springer International Publishing, pp. 208–218). На данном этапе активно разрабатываются различные модели человеко-машинных интерфейсов, входящих в состав интерактивных систем робототехники. Наряду с методами распознавания зрительных образов (мимики, жестов и т.д.) основополагающими остаются направления исследований в области акустико-семантического контента, включающего профильные базы данных, базы знаний, алгоритмы управления коммуникативным актом, а также типом и степенью интерактивности.

PACS: 43.60.-c

УДК: 81'33.

Ключевые слова: робототехника, интерактивные системы, акустико-семантический контент, распознавание образов, человеко-машинная коммуникация.

ВВЕДЕНИЕ

Современное технократическое состояние общества непосредственно связано с цифровизацией процессов внутренней и внешней коммуникации, что привело к созданию новой области знаний, получившей название Digital Humanities и соотносящейся с объектами и методами в сфере гуманитарных наук. В этой связи процесс цифровизации приобретает всё большее значение для развития антропоморфных систем, функционирование которых непосредственно связано с такими гуманитарными направлениями, как моделирование интеллектуально-когнитивных функций, функций речепорождения и речевосприятия, смысловой речеперепетации и т.д. Развитие Digital Humanities позволяет объединить цифровые алгоритмы с объектами и процессами гуманитарного назначения, что ведёт к созданию таких направлений, как формирование баз данных различного гуманитарного профиля, цифровое моделирование объектов и процессов гуманитарного предназначения, создание цифровых онтологий, библиотек, музеев и т.д. Существенное место в этом цифровом разнообразии занимает дальнейшая разработка проблем искусственного интеллекта, непосредственно связанных с созданием антропоморфных программ — моделей, имитирующих поведение чело-

века. Естественно, что не последнее место в решении данной задачи занимает звукоречевая коммуникация «человек-машина» [1, 2]. Следует подчеркнуть, что степень «антропоморфизации» тем выше, чем совершеннее решена задача порождения речи и интерпретации (понимания) речевых высказываний, а также предметной обратной связи с опорой на устную речь. Особое значение вышеуказанное направление цифровых разработок приобретает при решении задач, связанных с антропоморфизацией роботов, создаваемых для различных сервисных целей, что ведёт к созданию эффекта «комфортной коммуникации» «человек-машина». Применительно к решению данной задачи основой является совершенствование существующих программ синтеза устной речи и разработка новых программ с учётом требований не только разборчивости, но и натуральности, естественности звучания, речевых реакций антропоморфных сервисных систем [2, 3].

ТЕХНОЛОГИЧЕСКИЕ ОСНОВЫ АНАЛИЗА И СИНТЕЗА РЕЧЕВОГО КОНТЕНТА ПРИМЕНИТЕЛЬНО К РОБОТОТЕХНИКЕ

Как показал опыт, для решения вышеуказанных задач особую роль приобретает использование просодической информации, включающей огибающую F_0 (частота основного тона), распределение значений I_n (уровня интенсивности сигнала), t_n (временное распределение фонетических сегментов в речевой цепи (ритм, ударение, темп, перерывы в реализации речево-

*E-mail: r.potapova@linguanet.ru

†E-mail: volikpotapov@gmail.com

го потока (паузация, логико-смысловая сегментация), супraseгментный тембр (E_n), включающий изменения значений спектра гласных, зависящие от эмоционально окрашенной модальности речевого высказывания (например, готовности помочь, вежливо предложить что-либо, выразить «симпатию», вызвать улыбку пользователя, улучшив тем самым его настроение, и т.д.), что для сервисной коммуникации приобретает особое значение.

Понимание смысла воспринимаемого речевого высказывания и ответная реакция антропоморфного устройства, а также *vice-versa*, немислимы без ввода в программу определённых вероятностных значений, соответствующих синтактико-семантическим единствам и их просодическому оформлению. Данный подход включает анализ-синтез-анализ синтактико-семантических составляющих функционирующих в коммуникации «человек-машина» высказываний и их просодических коррелятов.

Цифровая модификация синтеза внедрена в робототехнические устройства, например, фирмы IBM с учётом информации о функционировании всех лингвистических уровней: акустико-фонетического, морфемно-лексического, синтактико-семантического. Для реализации синтеза и последующего использования этой информации в целях ответной вербальной реакции антропоморфного робота сервисного типа применяется метод построения мультипараметрической просодической огибающей речевого высказывания, реализуемого согласно правилам эмоционально-модальных коннотативных включений.

Используемые в программах синтеза речевых высказываний правила применительно к антропоморфному типу сервисных робототехнических устройств основаны на алгоритмах порождения фразовой интонации с включением следующих характеристик: модификаций основного тона, временного варьирования лексических составляющих, мощностно-звукового варьирования слоговых единиц, спектрально-временного варьирования вокализованных сегментов звуковой цепи. Особое значение приобретает смысловая сегментация синтезированной ответной реакции робототехнического устройства, связанной со смысловым фокусом ответной реакции. Естественно, что речь идёт не о запрограммированном диалоге «человек-машина» узкого прагматического типа, а об относительно свободной коммуникации «человек-машина» более продвинутого типа сервисного назначения.

Как показали наши исследования [1-4], особую роль играет «наложение» спектральных значений выделенных в речевом потоке гласных на просодическую огибающую речевого высказывания. Установлено также, что скорость формантных изменений непостоянна: максимальные величины в начале фразы и стремящиеся к нулю к концу артикуляции изменения реализуются по экспоненте:

$$F_i(t) = F_{a_i} + (F_{b_i} - F_{a_i}) \left(1 - e^{-\frac{t-t_0}{r}}\right) \quad (1)$$

где F_{a_i} – начальное значение F_i на участке формантного локуса; F_{b_i} – конечное значение F_i ; t_0 – время начальной реализации; r – временная константа; t – функция временной сегментации.

Установлено, что при разработке эмоциомитирующего антропоморфного робота особое значение для получения «естественной» речи имеют следующие характеристики:

- ритмическое оформление речевого высказывания с учётом языка и его ритмообразующих особенностей (например, отнесение языка к числу акцентосчитающих или слогосчитающих языков) [4];
- смысловая сегментация речевого высказывания (например, синтагматическая);
- тембральная окраска вокализма речевого высказывания;
- темповые нюансы (например, замедление/ускорение темпа);
- F0-диапазон в рамках речевого высказывания (узкий/широкий);
- голосовой регистр (верхний/средний/низкий);
- эмоционально-модальная паузация;
- слоговая/пословная длина речевого высказывания;
- просодическая выделенность «нового» в речевом высказывании, т.е. ввод в программу правил актуального членения;
- семантически оправданная паузация.

ЗАКЛЮЧЕНИЕ

Естественность звучания речи антропоморфных сервисных устройств зависит от особенностей построения модели речевой коммуникации, лежащей в основе разрабатываемой программы, с одной стороны, и особенностей конкретного языка, используемого в программе вопросно-ответного взаимодействия «человек-машина». При этом на следующих этапах разработки подобных систем прогнозируется появление более сложной задачи: условно самостоятельная и далее самостоятельная вербальная реакция антропоморфного устройства на стимулы внешней среды, что, естественно, имеет прямое отношение к дальнейшим разработкам когнитивно-речегенерирующего характера.

-
- [1] *Потапова Р.К.* Речевое управление роботом: лингвистика и современные автоматизированные системы. Изд. 2-е, перер., доп. М. 1987.
- [2] *Потапова Р.К.* // Конверсия в машиностроении. 2004. № 3. С. 92.
- [3] *Потапова Р.К.* // Конверсия в машиностроении. 2004. № 4. С. 97.
- [4] *Потапова Р.К.* // Ежегодник: акустика речи и прикладная лингвистика. Сб. трудов РАО и МГЛУ. 2002. С. 6.
- [5] *Potapova R.* Lecture Notes in Artificial Intelligence, 9812. Springer International Publishing, Switzerland. 2016. P. 208.
-

Acoustic semantic content of speech communication in robotics

R. K. Potapova^{1,a}, V. V. Potapov^{2,b}

¹*Federal state budgetary educational institution of higher education Moscow state linguistic university
Moscow 119034, Russia*

²*Scientific and Teaching Computer Centre, Faculty of Philology, Lomonosov Moscow State University
Moscow 119991, Russia*

E-mail: ^ar.potapova@linguanet.ru, ^bvolikpotapov@gmail.com

The paper presents an original approach to the investigation of acoustic semantic content that ensures the effect of reciprocal human-machine and machine-human communication, which is one of the top priority tasks for fast stimulus-reaction - reaction-stimulus model based interaction engineering in the modern research field that is now termed Interactive Communicative Robotics (see also: R. Potapova «Speech Dialogue as a Part of Human-Machine Interactive Systems» in: «Interactive Collaborative Robotics. 1st International Conference — ICR 2016», LNAI 9812, Springer International Publishing, pp. 208–218). At the present stage various human-machine interface models making part of interactive robotic systems are being developed. Alongside visual image recognition (mimics, gestures etc), acoustic semantic content investigations, including the development of problem-oriented databases, knowledge bases, algorithms for controlling the communicative act, type and degree of interactivity, is one of the essential areas forming the basis of the mentioned research field.

PACS: 43.60.-c

Keywords: robotics, interactive systems, acoustic semantic content, image recognition, human-machine communication.

Received 23 October 2019.

Сведения об авторах

1. Потапова Родмонга Кондратьевна — доктор филол. наук, профессор, директор Института прикладной и математической лингвистики; тел.: (495) 637-56-97, e-mail: r.potapova@linguanet.ru, RKPotapova@yandex.ru.
 2. Потапов Всеволод Викторович — доктор филол. наук, ст. науч. сотрудник, тел.: (917) 533-97-34, e-mail: volikpotapov@gmail.com.
-