

УДК 14.01.85

ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ В ЭЛЕКТРОННОМ ОБУЧЕНИИ ИНОСТРАННЫМ ЯЗЫКАМ

И.Б. Тампель, Е.В. Краснова, Е.А. Панова, К.Е. Левин, О.С. Петрова

Рассматриваются способы применения в системе электронного обучения иностранным языкам технологий автоматического распознавания и синтеза речи для задач отработки произношения на сегментном и супraseгментном уровнях, тренировки коммуникативных навыков, проверки словарного запаса обучаемого, тренировки навыка понимания на слух, а также для создания системы голосовой навигации. Несмотря на некоторые ограничения, такое применение данных технологий эффективно как для упрощения задач реализации процесса обучения, так и для повышения удобства использования системы.

Ключевые слова: электронное обучение, иностранные языки, ИКТ, речевые технологии, распознавание речи, синтез речи.

Введение

Спрос на изучение иностранных языков в настоящее время, несомненно, велик. Однако из-за стремительно ускоряющегося ритма жизни человеку все меньше времени удастся уделять обучению в целом и изучению иностранных языков, в частности. В связи с этим в последнее время резко возросла необходимость в модернизации процесса обучения, которое должно быть доступным, мобильным и удобным. В последние годы Министерство образования и науки Российской Федерации всерьез озабочилось вопросом повышения уровня доступности образования и с этой целью приступило к выполнению мероприятий по созданию электронных образовательных ресурсов [1].

Электронные образовательные ресурсы создаются для широкого круга пользователей. Возможность дистанционного использования электронных ресурсов делает обучение мобильным: с помощью

персонального компьютера, смартфона или планшетного компьютера с установленным устройством доступа к вычислительной сети пользователь может самостоятельно обучаться в интерактивном непрерывном режиме. Также имеется возможность учесть предпочтения пользователя и его индивидуальные успехи и, таким образом, автоматически адаптировать курс обучения к конкретному пользователю.

Внедрение информационно-коммуникационных технологий (ИКТ) в электронные образовательные ресурсы, в частности, технологий автоматического синтеза (TTS – Text-to-Speech) и распознавания (ASR – Automatic speech recognition) речи, которым посвящена данная работа, позволяет сделать системы электронного обучения максимально удобными в использовании (например, благодаря голосовой навигации внутри курса). Что касается систем электронного обучения иностранным языкам, внедрение инновационных речевых технологий открывает также методические возможности. Очень важно отметить и социальную направленность подобных разработок: например, внедрение технологий TTS и ASR в системы электронного обучения делает эти системы доступными для людей со слабым зрением или с ограниченной возможностью движения (нарушением опорно-двигательного аппарата), т.е. обеспечивает людям с ограниченными возможностями взаимодействие с компьютером.

Цель настоящей работы – описать, каким образом с помощью технологий автоматического распознавания и синтеза речи решаются некоторые задачи системы электронного обучения иностранным языкам.

Создание программного комплекса «Лингвистический тренажер»

В настоящее время осуществляется внедрение ИКТ в систему электронного обучения иностранным языкам «Лингвистический тренажер», которая создается ООО «ЦРТ» (Центр речевых технологий). Лингвистический тренажер (ЛТ) представляет собой программный комплекс с элементами искусственного интеллекта, предназначенный для освоения навыков понимания на слух и беглого говорения на иностранных языках (предусмотрено обучение девяти языкам) с помощью компьютера, в том числе различных мобильных устройств, на основе вопросно-ответных коммуникаций.

Программная архитектура ЛТ предусматривает следующие компоненты:

- система обучения, в которой содержатся как учебные материалы (теоретический материал и упражнения), так и информация об обучаемом (о результатах обучения, об истории взаимодействия пользователя с системой и т.д.) и модуль расчетов; на основе этих данных создается экспертная система, которая позволяет модифицировать учебный курс в зависимости от результатов обучения пользователя, т.е. адаптировать курс к конкретному пользователю;
- служебные функции, обеспечивающие надежное функционирование системы в целом;
- пользовательский интерфейс, которому при разработке ЛТ уделяется особое внимание: необходимо создать максимально удобный интерфейс, так как программа создается для широкого круга пользователей [2].

Применение ИКТ в «Лингвистическом тренажере»

Использование речевых технологий в системах электронного обучения иностранным языкам (обучение произношению, чтению вслух, освоение некоторых коммуникативных навыков) началось относительно недавно [3], поэтому, а также ввиду ряда технических ограничений [4], в существующих сервисах полноценного их применения практически не наблюдается.

В ЛТ предусмотрено освоение навыков понимания на слух и говорения на девяти языках. В разработанной технологии электронного обучения иностранным языкам предполагается использование систем ASR и TTS для русской речи, таким образом, они применимы в той части ЛТ, в которой происходит обучение русскому языку, а также в интерфейсе для русскоговорящих пользователей.

Использование технологий ASR и TTS позволяет решить задачи голосовой навигации по сервису и некоторые методические задачи, а именно, отработка произношения на сегментном и супraseгментном уровнях, тренировка коммуникативных навыков, проверка словарного запаса обучаемого, тренировка навыка понимания на слух.

Использование технологии ASR

Существуют разные типы автоматического распознавания речи, все они применимы для решения разных функциональных и методических задач ЛТ.

Для голосовой навигации по сервису ЛТ, которая необходима как для мобильного использования сервиса, так и для использования его людьми с ограниченными физическими возможностями, применяется технология распознавания команд. Распознавание команд – это разновидность распознавания дискретной речи, т.е. изолированно произнесенных слов. Суть данной технологии состоит в том, что система реагирует на изолированно произнесенные слова или словосочетания из небольшого заранее заданного словаря. Применимость распознавания речи данного типа ограничивается заданным словарем, в то же время за счет этого ограничения точность распознавания достаточно высока [5].

В методических целях технология распознавания команд применима, например, для устной тренировки и проверки словарного запаса обучаемого: для этого также задается определенный словарь, состоящий из слов, которые пользователю необходимо выучить, чтобы успешно пройти урок.

Основная же область применения ASR в целях обучения иностранному языку – тренировка произношения (CAPT – Computer Assisted Pronunciation Training), на ней следует остановиться подробнее. Тренировка фонетики иностранного языка подобным способом очень востребована и удобна, поскольку пользователь может сам определять время и количество тренировок, отслеживать свои результаты, повторять пройденный материал столько раз, сколько ему необходимо, что невозможно реализовать при аудиторном обучении.

Главное отличие систем ASR, используемых для обучения иностранному языку, от стандартных систем – это распознавание акцентной речи. Системы, предназначенные для распознавания речи носителей языка, могут быть лишь частично применимы в целях обучения иностранному языку, и то лишь на более высоком уровне обучения, тогда как тренировка фонетики, в основном, требуется на начальном этапе. Более того, такие системы не будут делать различий между похожими звуками изучаемого и родного языка (например, между английским придыхательным апикально-альвеолярным /t/ и русским зубным дорсальным /т/), что не позволит разработать эффективную обратную связь для коррекции произношения. При обучении произношению на иностранном языке могут возникать и другие трудности: наличие в изучаемом языке звуков, которые отсутствуют в родном языке (например, межзубные английские /ð/ и /θ/, отсутствующие в русском языке), а также различное контекстное употребление похожих звуков (например, испанские звонкие смычные /b/, /d/, /g/ не употребляются в интервокальной позиции, а русские /б/, /д/, /г/ – в абсолютном конце (в позиции перед паузой)).

Таким образом, одним из наиболее перспективных направлений разработки систем распознавания акцентной речи является добавление информации о родном языке пользователя, что даст представление об ошибках, которые он, скорее всего, сделает. На ошибки конкретных пользователей могут быть натренированы как акустические, так и языковые модели распознавания речи (подробнее об акустических и языковых моделях см. далее) [6]. Акустические модели осуществляют распознавание ошибок на фонетическом уровне (см. приведенный пример о разнице между русским /т/ и английским /t/), языковые модели осуществляют распознавание ошибок на фонематическом уровне, например, замена носителем русского языка английского /θ/ на /t/ в слове «three» приведет к смысловой ошибке. Поскольку фонематические ошибки являются более грубыми, то при разработке системы ошибок носителей определенного языка, им должно быть уделено больше всего внимания. Наиболее эффективным материалом для отработки подобных ошибок являются упражнения на минимальные пары, подобранные специально с учетом трудностей произношения носителей определенного языка (например, при изучении английского языка трудности произношения для русскоязычных и испаноязычных пользователей будут отличаться). Как показали исследования [7], можно ранжировать носителей разных языков по количеству ошибок, которые они могут сделать в конкретном изучаемом языке, а также выявлять универсальные ошибки и ошибки, характерные для определенных языков и для определенных языковых групп. Например, при изучении произношения голландского языка меньше всего трудностей наблюдалось у англичан и носителей скандинавских языков (носителей типологически близких языков), чем у носителей славянских, романских и восточных языков [7].

Тренировка произношения на супraseгментном уровне включает в себя отработку как словесного ударения (уровень слова), так и интонации (уровень фразы). Для распознавания интонации иностранного языка требуется выделение параметров, существенно отличающихся от сегментного уровня: частоты основного тона (ЧОТ) – основного параметра в большинстве языков, а также интенсивности и длительности супraseгментных единиц. Если на сегментном уровне и уровне словесного ударения произношение можно охарактеризовать как «правильное» или «неправильное», то на интонационном уровне высказывание лучше характеризовать как «корректное» и «некорректное» [8], поскольку в данном случае нет четкого определения параметра «правильности» произношения. Исключением в данном случае являются тональные языки, в которых тон неразрывно связан со смыслом (например, китайский язык). В естественном общении критерием «корректности» является достижение коммуникативной цели, в системе распознавания подобным критерием являются значения диапазонов ЧОТ, интенсивности и длительности супraseгментных единиц. В плане описания интонационной системы наиболее удобным представляется подход С. Оде [9] при описании интонации английского, голландского и русского языков. Данный подход был разработан в автосегментной фонологии и является наглядным и удобным, особенно с точки зрения сопоставления интонационных систем разных языков.

Для отработки обучаемым коммуникативных навыков технология ASR также применима: она может использоваться при создании упражнений, представляющих собой вопросно-ответные коммуникации. Например, упражнение представляет собой диалог, где пользователю необходимо устно давать ответы на определенные вопросы, при этом в зависимости от того, что ответит обучаемый, следующий вопрос пользователю автоматически выбирается из нескольких вариантов (вопросно-ответные коммуника-

ции, представляющие собой граф). Для такого типа упражнений применима технология распознавания ключевых слов: исходя из распознанных ключевых слов, ЛТ выбирает следующий вопрос. Данная технология заключается в распознавании отдельных участков слитной речи, при этом системе распознавания не требуется полностью преобразовывать в текст произнесенную речь, достаточно распознать лишь определенные слова, на основе которых в ЛТ может строиться дальнейшая траектория взаимодействия с обучаемым [5, 10].

Нагляднее всего можно показать такой принцип работы системы на примере отработки альтернативных вопросов, ответы на которые являются полными (рисунок).



Рисунок. Пример использования технологии распознавания ключевых слов в учебном упражнении

Однако для того, чтобы в подобных и других упражнениях обеспечить полноценный автоматический анализ произнесенного пользователем текста (в том числе с точки зрения наличия/отсутствия грамматических ошибок), технологии распознавания ключевых слов недостаточно, необходимо использовать технологию распознавания слитной речи. Распознанный таким образом текст далее анализируется парсерами, т.е. в этом случае возможна отработка обучаемым правил морфологии и синтаксиса изучаемого языка [11].

Распознавание слитной речи осуществляется на большом словаре (LVCSR – Large vocabulary continuous speech recognition) и основано на акустических и языковых моделях. Акустические модели – это шаблоны звуков речи, с которыми сопоставляется входной речевой сигнал. Наибольшее распространение в системах распознавания с большим словарем получило представление акустических моделей, соответствующих акустическим событиям, например, аллофонам (монофонам, дифонам или трифонам), с помощью моделей скрытого марковского процесса [12–14]. Языковая модель – обязательная часть систем распознавания слитной речи. Не любая последовательность слов является предложением, между словами есть грамматические и семантические связи. Языковая модель позволяет узнать, какие последовательности слов в языке более вероятны, а какие – менее. Информация от языковой модели либо учитывается одновременно с информацией от акустической модели (каждая со своим весом), либо включается на втором этапе. Использование языковой модели помогает сократить пространство поиска и снять неоднозначность при выборе из нескольких близких по стоимости акустических гипотез (для русского языка, например, помогает правильно распознать слово в нужном падеже). Процесс подготовки лингвистических моделей, как правило, автоматизирован и основан на статистической обработке большой выборки текстового материала.

Качество распознавания слитной речи обеспечивается модулем оценки качества сигнала, а также модулем определения речевой активности, который позволяет выделить в звуковом потоке неречевые участки и участки, содержащие речь. Из анализа исключаются участки пауз, что позволяет сократить время обработки данных, повысить достоверность распознавания, а также реализовать адаптивные алгоритмы распознавания, настраиваемые на определенный уровень помех и искажений [15, 16]. Таким образом, система распознавания слитной речи работает достаточно эффективно и вполне применима для методических задач ЛТ.

Однако для полноценного и эффективного использования распознавания слитной речи для отработки коммуникативных навыков обучаемого также необходима настройка на акцентную речь [17]. Это условие относится к системам ASR не только для русской речи.

Эффективность работы ASR в применении к обучению иностранному языку заключается не только в высокой надежности распознавания, но и в наглядной интерпретации результата распознавания для пользователя. В данном случае необходимо учитывать, для какого круга пользователя разрабатывается система обучения языку. Для широкой аудитории обратная связь должна быть представлена в максимально простой и наглядной форме, результат распознавания может выдаваться как «правильный»/«неправильный», либо «корректный»/«некорректный» (использование системы баллов и системы «правильности» в процентах представляется неэффективным), например, могут даваться указания по основным трудностям, характерным для пользователей данного языка. Если система обучения языку разрабатывается для студентов языковых специальностей, в качестве обратной связи могут предоставляться графики спектра мощности, ЧОТ и т.д. Не только преподаватели, но и сами студенты могут сравнивать свои результаты с эталонными произнесениями носителей изучаемого языка. Исследования показывают [1, 4, 6], что подобная обратная связь эффективна в обучении.

Отдельной задачей является адаптация системы ASR в целях обучения иностранному языку для людей с речевыми нарушениями [18]. На основе опыта ООО «ЦРТ» по исследованию акустических свойств речевых нарушений в дальнейшем предполагается разработка системы распознавания дефектной речи.

Использование технологии TTS

Для голосовой навигации по ЛТ применима также технология синтеза речи. Озвучка навигационных сообщений синтезом речи не только обеспечивает удобство в использовании сервиса слабовидящими, но и исключает необходимость трудоемкой записи дикторской речи. Технология синтеза русской речи основана на методе Unit Selection (разновидность компилятивного синтеза), в котором из большой речевой базы могут выбираться элементы различной величины (аллофоны, последовательности аллофонов, предложения) [19–21]. В настоящее время эта технология позволяет достичь максимальной естественности синтезированной речи. Как пишет в своей книге «Text-to-Speech Synthesis» один из основоположников технологии Unit Selection Пол Тэйлор [19], в то время как по разборчивости синтезируемой речи эта технология сравнима с другими (также существует формантный, артикуляторный синтез, синтез на основании статистических моделей), по естественности речи она значительно их превосходит.

Принцип технологии TTS основан на том, что из большой речевой базы по полученной на ранних этапах работы синтезатора фонетико-просодической последовательности выбираются наиболее подходящие элементы. Качество синтеза напрямую зависит от размера и качества используемой речевой базы (необходимо использовать полную, сбалансированную и корректно размеченную базу).

Технология TTS также может быть использована для методических целей, например, для упражнений, представляющих собой диктовку на иностранном языке, которые необходимы для освоения навыков понимания на слух иностранной речи. В целом эффективность диктовки с помощью TTS можно оспорить, так как автоматически синтезированная русская речь на данный момент не обладает полностью естественным звучанием и отличима от реальной речи. С этой точки зрения для подобных целей было бы правильным использовать запись речи диктора, чтобы исключить возможные искажения, которые могут негативно повлиять на корректное восприятие иностранной речи обучаемым. Однако, как показали исследования, диктант посредством синтезированной речи является более эффективным на первых порах обучения, поскольку синтезированная речь имеет большую разборчивость, чем речь человека [22].

Для целей отработки произношения, например, упражнений на имитацию звуков, отдельных слов и фраз, может быть использована технология макрокомпилятивного синтеза, который обеспечивает еще более высокую естественность звучания и не допускает малейших искажений звукового сигнала. Макрокомпилятивный синтез гарантирует стыковку фрагментов непрерывной речи не на уровне аллофонов, а на уровне отдельных слов или даже фраз. Однако данный тип синтеза удобен для обеспечения идеально произнесения только отдельных, часто повторяемых фрагментов речи.

Применение TTS в электронной системе обучения иностранным языкам позволяет добавлять в ресурс новые учебные материалы, которые необходимо озвучить, и при этом снимает задачу записи речи диктора.

Использование TTS в ЛТ в методических целях планируется в перспективе.

Заключение

Информационно-коммуникационные технологии в настоящее время достигли очень высокого уровня, что позволяет использовать их в самых разнообразных сферах. Эффективным является применение информационно-коммуникационных технологий в образовании, в первую очередь – в электронных образовательных ресурсах.

Разработки в области речевых технологий можно и необходимо внедрять в системы электронного обучения, в частности – в системы обучения иностранным языкам. ООО «ЦРТ» реализует эту идею в разрабатываемом при финансовой поддержке Министерства образования и науки Российской Федерации программном комплексе «Лингвистический тренажер». Так как сервис предназначен для освоения уст-

ных коммуникативных навыков, использование в нем новейших речевых технологий логично. В работе были описаны способы решения некоторых задач лингвистического тренажера с помощью технологий автоматического распознавания и синтеза речи. Все технологии, о которых шла речь, разработаны ООО «ЦРТ».

Эффективность применения речевых технологий в лингвистическом тренажере заключается не только в упрощении задач реализации процесса обучения (применение технологий непосредственно к учебному материалу), но и в повышении удобства использования лингвистического тренажера как для широкого круга пользователей, так и для людей с ограниченными физическими возможностями (использование технологий для навигации по сервису).

В настоящее время имеются некоторые ограничения, которые не позволяют использовать системы ASR и TTS в полной мере, однако стремительное развитие информационно-коммуникационных технологий позволяет прогнозировать скорое преодоление этих ограничений. В целом же применение речевых технологий в обучении значительно повышает доступность образования, на что в данный момент нацелены мероприятия, реализуемые государством.

Литература

1. Материалы к выступлениям Министра образования и науки РФ [Электронный ресурс]. – Режим доступа: http://минобрнауки.рф/%D0%BF%D1%80%D0%B5%D1%81%D1%81-%D1%86%D0%B5%D0%BD%D1%82%D1%80?events_sections=5, свободный. Яз. рус. (дата обращения 06.12.2012).
2. Киселев В.В., Елисева О.Е., Ковалёнок Ю.И., Хитрова Ю.Н. Архитектура лингвистического тренажера для экспресс-освоения навыков общения на иностранном языке // Научно-технический вестник информационных технологий, механики и оптики. – 2013. – № 1 (83). – С. 166–170.
3. Egan K.B. Speaking: A critical skill and challenge // CALICO Journal. – 1996. – V. 16. – № 3. – P. 277–293.
4. Dodigovic M. Artificial Intelligence in Second Language Learning: Raising Error Awareness. – Toronto, 2005. – 304 p.
5. Официальный сайт ООО «ЦРТ»: Распознавание речи [Электронный ресурс]. – Режим доступа: <http://www.speechpro.ru/technologies/recognition>, свободный. Яз. рус. (дата обращения 06.12.2012).
6. Stanley Th., Nacioglu K. Improving L1-Specific Phonological Error Diagnosis in Computer Assisted Pronunciation Training. Proceeding of Interspeech-2012 [Электронный ресурс]. – Режим доступа: <http://interspeech2012.org/accepted-abstract.html?id=555>, свободный. Яз. англ. (дата обращения 12.02.2013).
7. Neri A., Cucchiarini C., Strik H. Segmental errors in Dutch as a second language: How to establish priorities for CAPT // Proceeding of the InSTIL/ICALL Symposium. – 2004. – P. 13–16.
8. Arias J.P., Yoma N.B., Vivanco H. Automatic intonation assessment for computer aided language learning // Speech Communication 523. – 2010. – P. 254–267.
9. Оде С. Заметки о понятии тональный акцент на примере русского языка // Проблемы фонетики V. – 2007. – С. 237–249.
10. Хитров М.В. Распознавание речи в системах контроля персонала // Директор по безопасности. – 2012. – № 7 – С. 38–41.
11. Cucchiarini C., Neri A., Strik H. Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback // Speech Communication. – 2009. – V. 51. – № 10. – P. 853–863.
12. Young S.J. A review of large-vocabulary continuous speech recognition // IEEE Signal Processing Mag. – 1996. – V. 13. – № 5. – P. 45–57.
13. Jelinek F. Statistical Methods for Speech Recognition. – MA.: Cambridge, 1998. – 188 p.
14. Рабинер Л.Р. Скрытые марковские модели и их применение в избранных приложениях при распознавании // ТИИЭР. – 1989. – Т. 77. – № 2. – С. 86–120.
15. T. Gandhi, V.K. Antiwal, A.K. Jain Evaluation of Smoothed Language Models. – 2011. – V. 2. – P. 487–491.
16. Wilson I. Using Praat and Moodle for Teaching Segmental and Suprasegmental Pronunciation // Proc. of the 3rd International WorldCALL Conference: Using Technologies for Language Learning. – 2008. – P. 112–115.
17. Levow G., Olsen M.B. Modeling the language assessment process and result: Proposed architecture for an automatic oral proficiency assessment. Computer mediated language assessment and evaluation in natural language processing // Proceedings of a symposium sponsored by the Association for Computational Linguistics and International Association of Language Learning Technology. – 1999. – P. 24–31.
18. Ruiters M.B., Beijer L., Cucchiarini C., Kraemer E., Rietveld T.C.M., Strik H., Van Hamme H. Human language technology and communicative disabilities: requirements and possibilities for the future // Language Resources and Evaluation. – 2012. – V. 46. – № 1. – P. 143–151.
19. Taylor P. Text-to-Speech Synthesis. – MA.: Cambridge, 2009. – 626 p.

20. Black A.W., Hunt A.J. Unit Selection in a Concatenative Speech Synthesis Using a Large Speech Database // Proc. of ICASSP 96. – 1996. – V. 1. – P. 373–376.
21. Vepa J. Join Cost for Unit Selection Speech Synthesis. Phd thesis. – University of Edinburgh, 2004. – 241 p.
22. Pellegrini T., Costa A., Trancoso I. Less errors with TTS? A dictation experiment with foreign language learners. Proc. of Interspeech–2012 [Электронный ресурс]. – Режим доступа: <http://interspeech2012.org/accepted-abstract.html?id=1105>, свободный. Яз. англ. (дата обращения 12.02.2013).

Тампель Иван Борисович

- Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кандидат технических наук, тьютор, tampel@speechpro.com

Краснова Екатерина Викторовна

- ООО «ЦРТ», младший научный сотрудник, krasnova@speechpro.com

Панова Екатерина Александровна

- ООО «ЦРТ», научный сотрудник, panova@speechpro.com

Левин Кирилл Евгеньевич

- ООО «ЦРТ», кандидат технических наук, руководитель отдела, levin@speechpro.com

Петрова Ольга Сергеевна

- ООО «ЦРТ», менеджер проектов, Petrova-o@speechpro.com