

О ПОСТРОЕНИИ ИНТЕРФЕЙСОВ ЧЕЛОВЕКО-МАШИННОГО ВЗАИМОДЕЙСТВИЯ© 2019 *А. П. Преображенский, Ю. П. Преображенский, О. Н. Чопоров**Воронежский институт высоких технологий (Voronezh, Russia)
Воронежский государственный технический университет (Voronezh, Russia)*

В работе обсуждаются проблемы построения интерфейсов человеко-машинного взаимодействия. Обсуждаются возможности распознавания речевых сигналов.

Ключевые слова: человеко-машинное взаимодействие, распознавание речи, управление системами.

В настоящее время можно наблюдать активное развитие интерфейсов, базирующихся на человеко-машинном взаимодействии [1, 2], в том числе – на распознавании речи.

Если рассматривать задачи, в которых распознаются отдельные слова, то они гораздо проще по сравнению с теми, в которых автоматическим образом распознаётся разговорная речь. Это связано с несколькими причинами. Прежде всего, наблюдается определенное варьирование при произношении слов. При этом транскрипции слов будут представляться фонетическим образом. Произносимые слова в большинстве случаев не будут совпадать.

Допустимые фразовые конструкции, будут описываться на базе соответствующей языковой модели. По мере того, как развиваются технологии, связанные с речевой обработкой, а также синтезом речевых сигналов, происходит создание различных человеко-машинных интерфейсов [3, 4]. Уже сейчас можно наблюдать, как подобные интерфейсы могут быть внедрены в состав интерактивных диалоговых и запросно-ответных систем. В качестве примера можно привести работу автоматических диалоговых телефонных контакт-центров. Вследствие бесконтактного управления есть возможности при речевом управлении повышать эффективность управления в информационных системах. Системы с автоматическим распознаванием речи применяются при

бесконтактном вводе голосовых команд, когда реализуются процессы управления.

Речевые сигналы, которые идут от диктора, говорящего в микрофон, система должна автоматическим образом распознать. Подобная задача относится к задачам распознавания образов. Для таких проблем необходимо, чтобы входной образ был соотнесен с отдельными классами. Возможность задания классов осуществляется в зависимости от того, какой тип анализируемых данных. При распознавании и необходимо получать оценку по выходным классам, выходным гипотезам, относительно которых данные будут принадлежать к классам в рамках наибольшей вероятности. Последовательность слов должна быть правильным образом идентифицирована в системе распознавания речи. При этом можно опираться на некоторый оптимальный критерий, который связан с формулой Байеса [5, 6].

Прежде всего, осуществляются процессы обучения с использованием моделей, описывающих акустические сигналы. В таких моделях анализ осуществляется на основе построения соответствующих векторов признаков. При этом их вычисление осуществляется на базе речевых данных, а также тех транскрипций, которые являются известными.

Построение языковой модели осуществляется за счёт того, что происходит оценка по частотам появления комбинаций слов по текстам соответствующих предметных областей. Входные сигналы будут преобразованы в последовательности векторов признаков. Далее, на базе того, что используются заранее обученные звуковые и речевые модели или грамматики, осуществляются процессы поиска по наиболее вероятным гипотезам. Также поиск идет по подмножествам гипотез, являющихся лучшими. Проведем анализ по

Преображенский Андрей Петрович – Воронежский институт высоких технологий, д. т. н., профессор, app@vvt.ru.

Преображенский Юрий Петрович – Воронежский институт высоких технологий, к. т. н., retrovich@vvt.ru.

Чопоров Олег Николаевич – Воронежский государственный технический университет, д. т. н., профессор, choporov_oleg@vvt.ru.

разным возможностям классификации в системах, связанных с речевыми сигналами. Размеры словарей могут меняться от малого – тогда число слов не превышает 100, до неограниченного, когда число слов стремится к бесконечности. То есть, могут быть использованы самые разные слова, которые встречаются в языке. Классификация осуществляется по размеру словарей. Данная классификация необходима, для того, чтобы по голосу пользователи была осуществлена предварительная настройка.

Тогда рассматривают дикторонезависимые системы. С точки зрения практики, такие системы представляют интерес, когда происходят процессы сравнения ранее зафиксированных эталонных сигналов, созданных из слов, а также входных сигналов, которые есть в системах, связанных с несанкционированным доступом. По заданному словарю наборы слов, произнесенные пользователями, вносятся в память вычислительных устройств [7]. Можно наблюдать линейный рост времени обучения, по мере того, как происходит увеличение объемов словарей. Нет необходимости в том, чтобы вести настройки по голосам пользователей для дикторонезависимых систем.

Тогда нет необходимости в проведении обучающих процессов в системе. Не нужно осуществлять и частичную подстройку, в которой происходит произношение не всего словаря диктором, а определенного адаптационного текста. В системе ведется процесс настроек [8, 9], требуемых акустических данных. Это особенно важно, если объемы словарей являются большими. В подобных системах, за счёт того, что осуществляется статистическое моделирование, по речевым и звуковым процессорам обеспечивается поддержка дикторонезависимости. Требуется при этом привлечение весьма больших объемов акустических данных, на базе которых ведется создание стохастических моделей. После проведения обучения систем, основанных на речевых блоках, можно говорить о том, что в них будет достаточно точно сформирован речевой сигнал, относящийся к среднестатистическому диктору.

Существует зависимость в моделях распознавания речевых сигналов от того, какой тип речи будет применяться. Речь может быть разговорной, слитной, изолированной и др. Бывают варианты, когда встречается спонтанная речь. С другой стороны, может быть и необходимость в распознавании изолированной речи. Задачи распознавания то-

гда могут быть решены, когда словарь имеет малый объём. Модели слов или фразы рассматриваются в виде словарных элементов. Также в словарь необходимо поместить и модель паузы. Она требуется для того, чтобы были обозначены речевые границы. Модуль распознавания будет проводить моделирование соответствующих классов для возможных слов. Происходит вычисление оценок распознавания по каждому из таких классов.

Размер словарей оказывает существенное влияние на то, какое будет качество распознавания. Неточности говорящих людей могут быть скомпенсированы тем, что существуют различные варианты произношений. Но, при этом точность распознавания будет снижена вследствие близких по произношению слов при росте размерности словарей. В таких случаях информационная система сможет распознать жестким образом задаваемые фразовые конструкции, которые будут известны для систем. Допустимые последовательности слов будут описаны внутри каждого класса. Проблемы могут наблюдаться, если будет пропуск или добавление некоторых слов. Задачи также усложняются, когда требуется распознать слитную речь. Необходимо осуществлять распознавание по отдельным фразам, которые формулируются из ключевых слов, существующих в словарях. Связанные друг с другом и осмысленные слова образуют последовательности. Именно таким образом и рассматриваются речевые сигналы. Внутри класса происходит моделирование каждого возможных слов. Произвести запись и обработку спонтанной речи гораздо сложнее, чем прочитанной речи. В последнем случае можно говорить о постоянном темпе и артикуляции. Не наблюдаются запинки, а также исправления. Если они даже возникнут, то будут удалены из речевых блоков на этапах обработки. Большей частью исследователи ориентируются на то, чтобы были использованы простые и короткие высказывания.

Для прочитанных речевых фрагментов характерными будут четкость произношения фонем, монотонность. Кроме того, отсутствуют интонации.

Для других типов речи, помимо разговорной, можно указать множество особенностей. Они обуславливают ухудшение качества распознавания между сессиями записей. В том числе, это характерно для разных дикторов. Можно наблюдать изменения в темпе речи, в артикуляции, в длинах предложений.

Произношение слов заметным образом будет варьироваться, когда наблюдается спонтанная речь. Можно говорить о появлении явлений редукции, ассимиляции звуков. При этом требуется, чтобы были сформированы альтернативные словарные транскрипции. В результате расширяется словарь произношений.

Вывод. В работе дан анализ характеристик систем, базирующихся на человеко-машинном взаимодействии. Показаны особенности распознавания речевых сигналов.

ЛИТЕРАТУРА

1. Кострова, В. Н. Особенности процессов цифровизации в современном обществе / В. Н. Кострова, Т. А. Цепковская // Инновационные доминанты социально-трудовой сферы: экономика и управление. Материалы ежегодной международной научно-практической конференции по проблемам социально-трудовых отношений. Редакционная коллегия: А. А. Федченко, О. А. Колесникова. – 2019. – С. 143-145.
2. Львович, И. Я. Интеллектуальные системы в промышленном производстве / И. Я. Львович // Прогрессивные технологии и процессы. Сборник научных статей 6-й Всероссийской научно-технической конференции с международным участием. Курск. – 2019. – С. 172-174.
3. Преображенский, Ю. П. Построение информационной интеллектуальной системы / Ю. П. Преображенский // Прогрессивные технологии и процессы. Сборник научных статей 6-й Всероссийской научно-технической конференции с международным участием. – Курск. – 2019. – С. 222-224.
4. Преображенский, Ю. П. Некоторые проблемы автоматизации процессов / Ю. П. Преображенский // Техника и технологии: пути инновационного развития. Сборник научных трудов 8-й Международной научно-практической конференции. Юго-Западный государственный университет. – Курск. – 2019. – С. 62-64.
5. Преображенский, Ю. П. Возможности построения компьютерных моделей физических процессов / Ю. П. Преображенский // Современные инновации в науке и технике сборник научных трудов 9-й Всероссийской научно-технической конференции с международным участием. Курск. – 2019. – С. 279-282.
6. Львович, И. Я. О проектировании информационно-измерительных систем / И. Я. Львович // Молодежь и системная модернизация страны. сборник научных статей 4-й Международной научной конференции студентов и молодых ученых. – 2019. – С. 110-112.
7. Завьялов, Д. В. О применении информационных технологий. / Д. В. Завьялов // Современные наукоемкие технологии. – 2013. – № 8-1. – С. 71-72.
8. Львович, И. Я. Основы информатики / И. Я. Львович, Ю. П. Преображенский, В. В. Ермолова. – Воронеж. – 2014. – 339 с.
9. Lvovich, I. Management of the internet of things system based on decision-making and optimization approaches / I. Lvovich, Ya. Lvovich, A. Preobrazhenskiy, O. Choporov // Conference of Open Innovations Association, FRUCT. – 2019. – № 24. – С. 688-696.

ABOUT CONSTRUCTION OF INTERFACES OF HUMAN-MACHINE INTERACTION

© 2019 A. P. Preobrazhenskiy, Yu. P. Preobrazhenskiy, O. N. Choporov

*Voronezh Institute of High Technologies (Voronezh, Russia)
Voronezh State Technical University (Voronezh, Russia)*

The paper discusses the problems of constructing interfaces of human-machine interaction. Speech recognition capabilities are discussed.

Key words: human-machine interaction, speech recognition, systems management.