

УПРАВЛЕНИЕ РОБОТОТЕХНИЧЕСКИМ СБОРОЧНЫМ КОМПЛЕКСОМ С ИСПОЛЬЗОВАНИЕМ РЕЧЕВОГО ИНТЕРФЕЙСА

И.Ф. Ревонченкова (ИПУ им В.А. Трапезникова РАН)

Показано, что использование речевого интерфейса при управлении робототехническим комплексом сборочной линии является актуальной задачей. Приведен обзор алгоритмов распознавания речи, применяемых в робототехнических комплексах. Предложен подход к разработке речевого интерфейса на русском языке.

Ключевые слова: речевой интерфейс, робототехнический комплекс, сборочная линия, словари.

Введение

В промышленности конечная цель — это получение дополнительной прибыли за счет увеличения выпуска продукции, снижение затрат на единицу готового изделия, повышение качества и снижение брака в готовом продукте. Реальные затраты на автоматизацию производственного процесса должны приводить к окупаемости приобретенного оборудования. При этом цена этого оборудования может быть достаточно высокой, но если в конечном счете она быстро окупается, то этот фактор определяет выбор того или иного оборудования в пользу решения «дороже но качественней». И чем быстрее происходит процесс внедрения автоматизированной робототехнической системы, оснащенной качественным программным обеспечением, тем выше будет рентабельность производства.

Сборочный процесс является наиболее сложным и ответственным этапом при автоматизации производственного процесса. Внедрение робототехнических комплексов в сборочный процесс уменьшает общую энергоемкость производства и увеличивает производительность труда в несколько десятков раз. Особенно это касается мелко- и среднесерийного производства, которое характеризуется высокой себестоимостью изделий, большими затратами на проектирование системы управления производственным процессом, разработку ПО и сам сборочный процесс.

Управление всем сборочным процессом теоретически можно разбить на следующие подзадачи:

- управление технологическими режимами выполнения отдельных операций в каждой технологической цепочке в отдельности, входящей в общую структуру технологического процесса (ТП);
- управление отдельными технологическими цепочками, входящими в структуру ТП;
- управление процессом программирования и перепрограммирования оборудования;
- реализация всех допустимых операций ТП в целом;
- речевое управление ТП оператора слежения для контроля производимых операций и задания определенных условий управляющей системе в процессе производства в зависимости от текущей ситуации и изменения в режиме on-line;
- система «искусственного» интеллекта для управления системой подсказок, экспертной оценки и анализа ситуации, управления системой выводов и пр.

Алгоритмы распознавания речи для робототехнической системы

Разработка пользовательских интерфейсов человеко-машинного взаимодействия является актуальным

направлением исследований в современном мире. Наиболее перспективным считается создание речевых интерфейсов как наиболее естественных для человека [1, 2]. В то же время разработка подобных интерфейсов связана с большим числом проблем, одной из которых является выбор между использованием записанного живого голоса и синтезированного для воспроизведения информации при взаимодействии с пользователем. Исследованиями подтверждено, что многие люди предпочитают слышать живые голоса, а не синтезированные. Поэтому рекомендуется использовать именно их, а при необходимости — переключаться на синтезированную речь, заранее предупреждая об этом пользователя [3].

Некоторые АСУ включают не только систему распознавания речи, но и систему распознавания образов, формируют 3D-модели готовых и сопрягаемых деталей.

При проектировании управляющей системы робота можно ограничиться системой распознавания речи с минимальным набором команд, необходимых для управления, созданием ограниченного словаря, и использовать специальные термины, воспринимая отдельные слова как минимально значимые единицы. Предварительно составляется база используемых слов с возможностью в дальнейшем ее дополнять новым словам. Практически все используемые в настоящее время способы распознавания речи основаны на сравнении произносимых слов с эталоном.

В настоящее время чаще всего применяются два способа распознавания речи.

1. Алгоритм динамического искажения времени (Dynamic Time Warping, DTW), позволяющий вычислить степень схожести речевого фрагмента с существующим в системе эталоном. Подразумевается, что одна и та же фраза, произнесенная одним и тем же диктором, может различаться по длительности произношения, кроме того, различную длительность будут иметь и составляющие фразу (слово) звуки. С учетом данного факта метод подразумевает выравнивание двух речевых сигналов по временной шкале при максимальном совмещении между собой сегментов с одинаковыми звуками. После этого происходит вычисление оценки различия сегментов двух данных произнесений. Суть метода заключается в поиске такой оптимальной последовательности пар сегментов, которой бы соответствовала минимальная суммарная оценка различия. Визуально работу данного подхода можно представить с помощью матрицы, по горизонтали и вертикали которой отложено время, по оси ординат откладывается фраза-эталон, а по оси абсцисс — фраза, произнесенная для распознавания [4].

Преимущество данного подхода заключается в простоте установления временного соответствия между проверяемым и эталонным речевым фрагментом, что делает возможным нахождение меры различия между ними. Основной недостаток данного подхода следует из особенностей его работы, а именно, требований по наличию эталонов для всех распознаваемых единиц (для всех дикторов). Данный факт ограничивает область использования метода только случаями распознавания небольшого словаря команд и требует при этом предварительного обучения системы [4].

2. Марковские модели (Hidden Markov Models, НММ) — статистическая модель, имитирующая работу процесса, похожего на Марковский процесс, с неизвестными параметрами. Метод пользуется большой популярностью благодаря простоте и удобству его применения на практике. Одним из главных недостатков данного подхода является наличие условия о независимости последовательности векторов признаков, вычисленных для речевого фрагмента, что делает невозможным отслеживание закономерностей и взаимных влияний в сигнале на достаточно больших промежутках времени. Однако для компенсации влияния данного фактора были разработаны методы вычисления векторов признаков речевого сигнала, несущих в себе, помимо абсолютных значений характеристик, и значения того, как быстро они изменяются во времени относительно предыдущих фрагментов [4].

Увеличить число операторов, команды которых будет распознавать система, можно с помощью так называемого метода многих эталонов. Представляет интерес использование этого метода, совместно с методами динамического искажения времени, основываясь на использовании сложных решающих правил и композиций алгоритмов распознавания [5].

Для сегментации речевого потока применяются также искусственные нейронные сети (Artificial Neural Networks, ANN) — алгоритмы на основе нейросетей отличаются хорошей устойчивостью к наличию посторонних шумов, низкому качеству записи, делают возможным моделирование длительных по времени взаимосвязей в речи. Это позволяет широко использовать при вычислениях параллельное программирование. Среди недостатков нейросетевого подхода выделяют сложность его адаптации для применения в условиях, когда анализируемый сигнал может иметь различную длительность, проблемы выбора начальной структуры и характеристик сети, а также практическую невозможность извлечения данных о выявленных нейросетью закономерностях [4].

Альтернативным направлением разработки методов распознавания речи может быть построение n-граммных моделей различных типов [5]. Это новый подход к построению компьютерной системы передачи и обработки текстовой информации на основе n-граммной языковой модели. N-грамма — последовательность из n элементов. В области обработки естественного языка N-граммы используются в основном

для предугадывания на основе вероятностных моделей. N-граммная модель рассчитывает вероятность последнего слова N-граммы, если известны все предыдущие. При использовании этого подхода для моделирования языка предполагается, что появление каждого слова зависит только от предыдущих слов [6].

Разработка речевого интерфейса на русском языке

Большинство присутствующих на отечественном рынке робототехнических систем разработано в зарубежных странах. Их речевой интерфейс использует узкоспециализированный язык профессионалов на основе формализованного языка страны производителя. Для работы с этим оборудованием требуется проведение обучения операторов, что влечет за собой дополнительные материальных и временных затрат со стороны промышленных предприятий.

Поэтому необходимо разрабатывать отечественные робототехнические системы, применяя системы речевого управления на русском языке. Это производственная необходимость и этого требует современный уровень производства. Для решения проблемы предлагается осуществлять перевод инструкций, заложенных в программный продукт фирмы-изготовителя на русский язык. Наиболее логично использование программы-переводчика и формализатора, преобразующего запрос, сформулированный на естественном языке оператора, в формальный запрос конкретной БД. При этом используется система подбора формальных слов из словаря переводчика (STYLUS, SOKRAT и пр.). Один тип перевода основывается на отдельном использовании программы-переводчика, а второй — на включении программы-переводчика в БД.

На сервере и в отдельных ПК на станциях сборки, объединенных локальной сетью, должны быть заложены словари-формализаторы. В них содержится определенный набор ключевых слов, по которым формируется исходящее или входящее сообщение, на основе которого блоки информации из базы знаний позволяют расширять сообщение, связанное с формальным сообщением двусторонней связью. Далее составляется системное задание на обработку детали или формируется команда на операцию устранения сбоя или неполадки. На выходе при подтверждении оператором правильности выбора команды формируется управляющий сигнал на контроллер с помощью блока модуляции, встроенного в ПК. Потом в обратной последовательности из сообщения контроллера о произведенных тех или иных действиях системы.

Единственная, оставшаяся окончательно нерешенной проблема для речевого управления процессом производства, — наличие различных механических шумов, наложенных на речевое сообщение. Шумы могут искажать речевой сигнал до неузнаваемости, процесс переключения затягивается и производство останавливается на неопределенное время. Существует несколько методов, позволяющих отфильтровать голосовой сигнал от шумовых наводок и посторонних сигналов. Все они

применяются на практике, но не позволяют полностью очистить речевой сигнал от шума без потери его информационной составляющей.

Эта проблема в определенной мере может быть решена с помощью достаточной звукоизоляции оперативного помещения и специальным оборудованием для шумоподавления. Использование вычислительных математических методов спектрального анализа и фильтрации речевого сигнала от посторонних шумов различной этиологии закрепит полученный результат.

Обработка звукового сигнала осуществляется на основе математического моделирования технологического процесса и алгоритмов компьютерной обработки сигнала в пределах приближения к физической реальности производственного цикла сборки. Критерии точности воспроизведения речевого сигнала оцениваются с помощью методов математической статистики и теории вероятности: вычисление коэффициента разборчивости сигнала, эквивалентное отношение мощности сигнала к мощности стационарного шума, среднее значение мощности сигнала в пределах конечного интервала, а также среднее значение по множеству случайных переменных.

В процессе распознавания слов и предложений задействуются различные словари, в том числе спектральный и орфографический словари и словарь синтаксических правил, накопитель орфографии слов, база слов и словосочетаний, словарь лексических правил. Кроме того, проводится статистический анализ и формируется счетная матрица слов и словосочетаний. Для этого используется динамическое программирование и задействуется блок предсказаний.

Выделим следующие подзадачи управления в общей схеме работы автоматизированной системы управления роботом с помощью речевого интерфейса:

- составление грамотного, четкого приказа — требования;
- отсечение шумов и наводок из акустического сигнала;
- составление иерархического списка команд в БД;
- формирование фонетического и акустического словарей;
- составление предложения из слов;
- синтез запроса системы;
- составление и накопление знаний в БЗ всех возможных вариантов запросов;
- управление многоярусной системой диалога;
- определение начала и конца диалога;
- размыкание последовательности диалога в сложной ситуации;
- фиксация и контроль рабочей ситуации;
- возможные варианты отката в режиме работы робота;

- составление и синтез ответа системы при выполнении той или иной операции;
- формирование комментариев к выполненной команде;
- составление формы отказа от выполнения операции и комментарии к нему;
- формирование ответа выполненных операций.

Заключение

Эффективность всего производства заключается в эффективности использования оборудования и средств управления отдельными технологическими цепочками, в том числе таким трудоемким, как сборочный процесс. За счет максимальной автоматизации ручного труда роботизированные системы управления производственным процессом с речевым интерфейсом позволяют уменьшить численность рабочих в сборочном процессе, повысить коэффициент использования оборудования, сократить время перенастройки, переналадки и простоя оборудования, ускорить процесс сборки в десятки раз. При этом даже очень трудоемкие операции становятся безопасными за счет высокой степени автоматизации. Создание роботизированных комплексов, управляемых речевым интерфейсом на русском языке позволяет улучшить процесс взаимодействия человека и робота, помочь оператору в освоении современной техники.

Таким образом, совмещение проектирования сложных операций сборки деталей и стыковки различных узлов с использованием речевого интерфейса для управления этим процессом в режиме on-line — это перспективное направление применения системы распознавания речи и САПР.

Список литературы

1. *Oviatt S.L.* Ten myths of multimodal interaction // Communications of the ACM. 1999. Vol. 42 (11), P. 74-81.
2. *Карпов А.А., Ронжин А.Л.* Многомодальные интерфейсы в автоматизированных системах управления // Изв. вузов. Приборостроение. 2005. Т. 48. № 7. С. 9-14.
3. *Прищеп М. В., Баранов К.Ю.* Особенности разработки пользовательского интерфейса мобильного информационного робота // Изв. вузов. Приборостроение. 2012. Т. 55. №11. № 7. С. 9-14.
4. *Желтов П.В., Желтов В.П., Семенов В.И.* Применение вейвлет-преобразования при сегментации речи // Современные проблемы науки и образования. 2015. № 1 (Ч. 1).
5. *Ющенко А.С.* Диалоговое управление роботами с использованием нечетких моделей // Тр. V междунар. научно-практич. конф. «Интегрированные модели и мягкие вычисления в искусственном интеллекте». г. Коломна; МГТУ им. Н.Э. Баумана; УлГТУ. 2009.
6. *Jurafsky, D. and Martin, J.H.* Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Pearson Prentice Hall. 2009. 988 p.

*Ревонченкова Инна Федоровна — научный сотрудник ИПУ им Трапезникова РАН.
Контактный телефон (495) 334- 90-60.*