

Система речевого контроля состояния машиниста

Д. Н. Бабин, В. В. Дементенко, И. Л. Мазуренко,
В. И. Миргородский, Д. С. Михайлов, А. Б. Холоденко,
А. В. Уранцев

Предлагается система контроля состояния оператора сложной технической системы, ориентированная на анализ его речевого ответа. Для возможности работы в сильных шумах предусматриваются дополнительные каналы ввода информации, а также специальный сценарий диалога оператора с компьютером.

Ключевые слова: распознавание речи, сильный шум, самосравнение.

Создание систем автоматического контроля функционального состояния машиниста, в которых используется «общение» человека с системой, связано с рядом трудностей. Зрительный и механический каналы общения обычно заняты в управлении локомотивом. Кроме того, затруднено использование высоконадежных медицинских датчиков, таких как у летчиков или космонавтов, так как применение этих датчиков мешает машинисту работать. Мы предлагаем использовать канал речевого общения.

В ответ на определённые действия системы или в ответ на возникновение определённых ситуаций в окружающем пространстве оператор должен произнести одну из ключевых фраз из утвержденного регламента переговоров, которая будет означать, что он увидел внешнее воздействие и отреагировал на него должным образом. Для повышения вероятности получения правильного ответа при работе в неблагоприятных условиях (например, при высоком уровне акустических шумов) в разработанную систему добавлены несколько каналов обработки звуковой информации, а также каналы с неакустической информацией. В качестве таких каналов в рассматриваемой

системе могут использоваться датчик скорости воздушного потока, расположенный в непосредственной близости рта оператора, а также инфракрасный датчик, определяющий ширину открытия рта. Для повышения точности распознавания дополнительные датчики могут быть смонтированы на гарнитуре, как вариант они могут быть расположены в шлемофоне с дополнительной системой защиты слуха оператора. В результате предлагается метод надёжного определения факта речевого ответа оператора в условиях сильных акустических помех.

Использование традиционных методов автоматического ввода и анализа звуковой информации в реальных условиях транспортного средства, как показали исследования, не может гарантировать высокой надёжности работы речевой системы безопасности. Связано это с тем, что канал звукового общения чрезвычайно зашумлен, и отношение сигнал/шум в этом канале может достигать 0 дБ и даже ниже.

Акустическая шумочистка, традиционно выполняющаяся с помощью параллельного использования нескольких микрофонов для ввода речи, не может привести к полному очищению звукового канала от акустических помех. Задача усложняется еще и тем, что традиционные методы распознавания речевых команд даже в условиях отсутствия шума работают с уровнями надёжности 0.99 и ниже. Мы рассматривали три типичных случая: применение дополнительного близкорасположенного микрофона, применение дополнительного удаленного микрофона и применение массивов микрофонов.

В случае использования близкорасположенного микрофона даже для очень высокого уровня шумов речевой сигнал будет присутствовать в ближнем микрофоне. Используя разность энергий сигналов в близком и дальнем микрофонах, начало речевого ответа можно определить, так как сильные импульсные помехи от ударов колес вносят отклики с одинаковым уровнем энергии в обоих микрофонах. Внешний источник речевых помех (наличие говорящего близкорасположенного человека, включенного радиоприемника, ради и т. п.) также вносят отклики с одинаковым уровнем энергии в обоих микрофонах.

В случае использования удаленного микрофона можно успешно использовать тот факт, что в удаленном микрофоне будет отсутствовать речевой сигнал. Используя результат сравнения энергий сигналов с двух микрофонов в разных частотных полосах, можно повысить

надежность идентификации факта речевой активности. При этом дополнительный удаленный микрофон будет использоваться для оценки уровня внешних помех.

Массивы микрофонов используются для получения речевого сигнала, очищенного от помех. Используя простейший метод «суммирования с задержкой» или более сложные методы с использованием метода наименьших квадратов, удается смоделировать микрофон с переменным лучом направления, позволяющим настраиваться на источник речи, и в результате значительно повысить отношение сигнал/шум.

Составной частью системы является речевой информатор, ведущий диалог с машинистом. В определенные моменты машинист должен подтвердить речевой командой адекватное восприятие факта сообщенного речевым информатором. Например, на запрос о сигнале светофора должен быть дан ответ соответствующий реальному цвету. На основе результатов анализа ответа машиниста система принимает решение: продолжить работу в обычном режиме, выдать предупреждающий сигнал или даже остановить поезд.

Задача, решаемая в рамках заявленного проекта, принципиально отличается от традиционной задачи распознавания речи тем, что здесь требуется значительно более высокая надежность, а распознавание происходит в условиях нестационарных железнодорожных шумов и речевых помех от радиостанции и других источников речи.

В экспериментах использовалась гарнитура с микрофоном, на которой были установлены линейка из 8 светодиодов и фотодиодов для измерения коэффициента отражения от губ диктора. Для независимости работы системы от внешнего освещения светодиод питается импульсным током. Фотодатчик является важным источником информации, не зависящим от уровня внешних акустических и световых помех. Для идентификации начала и конца речевого ответа удобно использовать, определяемую по фотодатчику ширину раскрытия рта и ее производную — скорость движения губ. При этом удается отличать движения губ, характерные для речевого ответа, от других ситуаций, при которых человек открывает рот. Заметим, что открытый рот отражает свет гораздо хуже, чем закрытый, поэтому на графике (рис. 1) изображена величина, которая во время речевого ответа уменьшается. Показания фотодатчика не зависят от акусти-

ческого канала, поэтому его использование на порядок увеличивает надежность обнаружения начала речевого ответа.

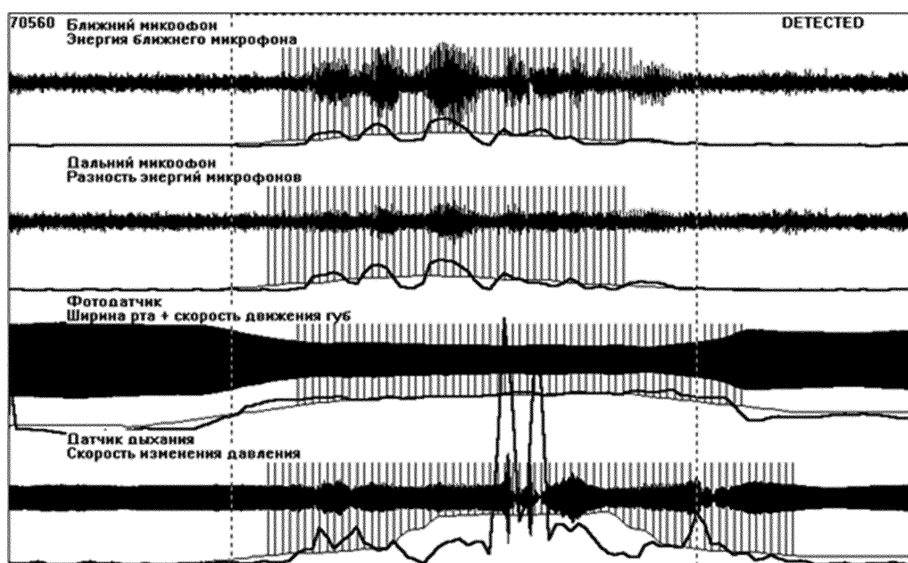


Рис. 1. Информация с различных датчиков при произнесении в шуме фразы «вижу зеленый сигнал».

В качестве датчиков уровня воздушного потока перед ртом диктора использовались низкочастотный микрофон, и температурный датчик воздушного потока. Этот тип датчиков надежно фиксирует факт вдыхания и выдыхания воздуха, причем по характеру изменения показателей датчика обычный выдох отличается от выдоха воздуха при речевом ответе. Однако датчик воздушного потока реагирует на внешние потоки воздуха (например, сквозняк из открытого окна в транспортном средстве). Датчик воздушного потока позволяет надежно выделять факт наличия взрывных согласных в речевой фразе, что может значительно повысить надежность распознавания.

В качестве ларингофона мы использовали специальный низкочастотный микрофон, закрепляемый на шее диктора в районе адамова яблока. Сигнал с этого микрофона практически не зависит от внешних акустических помех, и эквивалентен речевому сигналу, пропущенному через низкочастотный фильтр с частотой среза на уровне 1–2 кГц. В результате из речевого сигнала выделяются только глас-

ные звуки. Несмотря на это показатели этого датчика могут использоваться для высоконадежной идентификации факта речевого ответа, так как гласные звуки присутствуют в любой речевой фразе.

По описанным датчикам можно надежно определить факт речевого ответа, но наша задача в том, чтобы отличить правильный ответ от неправильного.

В самом деле, возможны разные ситуации:

- машинист выведен из строя, спит, потерял сознание, умер;
- машинист жив, но дремлет или недостаточно бодр;
- машинист достаточно бодр, но не выполняет работу по управлению поездом;
- машинист достаточно бодр, но неправильно распределяет время между разными работами по управлению поездом;
- машинист достаточно бодр и правильно выполняет работы по управлению поездом.

Имеет смысл вести диалог машиниста с бортовым компьютером, задавая машинисту вопросы, ответы на которые известны бортовому компьютеру и очевидны для правильно выполняющего свои обязанности работника. Например, какая скорость поезда, какой сигнал светофора, сколько времени. В этом случае мы находимся в ситуации определения, к какому из заранее записанных ответов настоящий ответ ближе. Речь — это не телефонный номер и не пин-код.

Теперь подробнее остановимся на механизме распознавания команд. Разные произнесения одного и того же слова могут сильно отличаться.

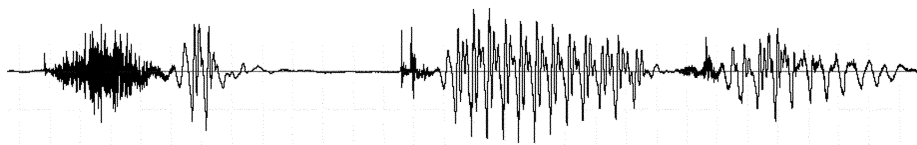


Рис. 2. График звукового давления при произнесении слова «четыре».

На рис. 2 хорошо видны куски графика, соответствующие гласным «Е» и «Ы» в слове «четыре», и черные куски графика, соответствующие шипящей «Ч», взрывной согласной «Т» и фонеме «Р». Эти куски графика без потери правильного распознавания могут быть произвольно растянуты, поэтому поэлементное сравнение звуковых

графиков (вернее их спектров Фурье) для распознавания речевых команд не используется.

Для решения этой проблемы традиционно используется процедура динамической деформации времени (метод скрытых Марковских моделей или процедура Витерби), которая, являясь разновидностью динамического программирования, состоит в интеллектуальном переборе возможных растяжений графика с целью сравнения наилучшего из растяжений с эталоном.

В нашем случае сильных нестационарных низкочастотных шумов сложность использования типового алгоритма динамической деформации времени заключается в том, что эталоны, однажды записанные в каких-то определенных шумах (или без шумов), будут использоваться для распознавания в совсем других сильных нестационарных низкочастотных шумах. Для решения задачи нами был предложен метод самосравнения звуковых записей, когда отдельные части звукового сигнала (вернее их спектры Фурье) сравниваются с другими частями того же самого звукового сигнала. Шумы в звуковом сигнале одни и те же, и не влияют на отличия частей сигнала друг от друга. Получается не так много разных матриц самосравнения с точностью до одинаковых столбцов и строк. Их можно просто накопить и идентифицировать речевой ответ по его матрице самосравнения. Пример визуализации такой матрицы приведен на рис. 3.

Как показали эксперименты, каждый из вышеперечисленных источников информации о речевом ответе независимо может использоваться для решения задачи идентификации факта речевого ответа со своей надежностью.

Однако надежность идентификации удастся повысить значительно, если использовать несколько разнородных датчиков в совокупности. Это связано с тем, что некоторые датчики могут считаться попарно независимыми; типичные помехи для датчиков разных типов имеют разную природу и вероятность их одновременного появления невелика; даже для зависимых датчиков (например, микрофон и ларингофон) уровень внешних помех может быть неодинаковым (в нашем примере высоким для микрофона и низким для ларингофона).

В экспериментах авторов погрешность распознавания речевого ответа удалось понизить в 10 раз. Датчики могут быть встроены в макет диалоговой системы речевого взаимодействия, состоящий из гарнитуры, надеваемой на голову машиниста, и оборудованной на-

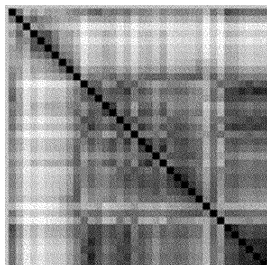


Рис. 3. Матрица самосравнения звукового сигнала. Видно, что на диагонали матрицы стоят нули.

ушниками, несколькими микрофонами, фотодатчиками, следящими за губами машиниста, датчиками дыхания и другими необходимыми приборами, что повысит контроль за состоянием машиниста, не отвлекая его от выполнения своих обязанностей.

Работа выполнена при поддержке гранта РФФИ 12-07-13136 офи_м_РЖД-2012.

Список литературы

- [1] Бабин Д. Н., Мазуренко И. Л., Уранцев А. В., Холоденко А. Б. Способ идентификации факта речевой активности оператора. Патент РФ № 2148505 от 17.02.1999.
- [2] Бабин Д. Н., Мазуренко И. Л. Распознавание изолированных слов ограниченного словаря. Свидетельство об официальной регистрации программы для ЭВМ № 940412.