Раздел VI. Информационно-алгоритмическая поддержка систем управления и обработки информации

УДК 004.934

А.А. Карпов

АУДИОВИЗУАЛЬНЫЙ РЕЧЕВОЙ ИНТЕРФЕЙС ДЛЯ СИСТЕМ УПРАВЛЕНИЯ И ОПОВЕЩЕНИЯ

В докладе предложена модель автоматического распознавания речи на основе анализа звуковой и визуальной информации о речи, а также модель компьютерного синтеза аудиовизуальной русской речи по произвольному тексту. Также представлен краткий обзор речевых технологий, применяемых в современной авионике.

Речевой интерфейс; аудиовизуальная речь; распознавание; синтез.

A.A. Karpov

AUDIOVISUAL SPEECH INTERFACE FOR CONTROL AND ALARM SYSTEMS

The paper proposes a model for automatic speech recognition based on an analysis of audio and visual information on speech, as well as a model for Russian text-to-audiovisual-speech synthesis. A brief survey of speech technologies applied in state-of-the-art avionics is presented.

Speech interface; audio-visual speech; recognition; synthesis.

Введение. С развитием речевых технологий (в первую очередь, систем автоматического распознавания и компьютерного синтеза речи) связывают будущее человеко-машинных интерфейсов для интеллектуальных систем управления различными техническими системами, роботами, подвижными объектами. За рубежом речевые командные системы двойного назначения уже внедряются в информационные системы, ассистирующие водителю при управлении транспортным средством (например, проект Verbmobil компании DaimlerChrysler), а также в бортовые информационные системы летательных аппаратов [1], например: Advanced Fighter Technology Integration по оснащению американских истребителей F-16A и F-35 речевой командной системой (РКС) Voice-Controlled Interactive Device; испытания РКС Avionique Speech Recognition на французских самолетах Mirage и Rafale; установка РКС Sextant ***Avionique*** TopVoice на многоцелевом вертолете Gazelle. В последних модификациях истребителя Eurofighter Typhoon также широко используется речевой интерфейс Direct Voice Input [2] для управления вспомогательным бортовым оборудованием (индикацией на приборной панели и графических экранах, выбором режимов работы информационных бортовых систем, режимами работы радара, заданием частот настройки радиоаппаратуры, выбором навигационных средств и т.д., всего 26 различных устройств), кроме управления критичными функциями и вооружением, также пилот может в ходе речевого диалога получить информацию о количестве топлива или о состоянии вооружения. Речевое управление позволяет повысить эффективность взаимодей-

ствия с бортовым оборудованием за счет бесконтактного управления информационными системами вербальным способом, что не отвлекает его от решения основных задач полета.

Конечно же, для того чтобы заменить часть ручных функций управления, голосовые команды должны обеспечивать высокую надежность и робастность распознавания при влиянии различных помех: внешних акустических шумов, неречевых звуков пилота (шум дыхания, кашель, скрежет зубов), помехи в канале связи, отражением и реверберация звука в условиях ограниченного пространства кабины. Для улучшения характеристик работы РКС в реальных условиях функционирования применяют дикторозависимое распознавание речи с настройкой на голос конкретного человека или группы дикторов и с малым словарем распознавания отдельных команд или коротких фраз (до нескольких десятков слов), используя различные методы шумоподавления и фильтрации звуковых сигналов за счет использования направленных микрофонов, многоканальной обработки сигналов и дополнительной незвуковой информации о речи, расширяя голосовые интерфейсы до многомодальных, анализирующих и распознающих одновременно несколько видов сигналов.

Автоматическое аудиовизуальное распознавание речи для ввода информации. В сложных акустических условиях функционирования типовые системы автоматического распознавания речи не способны обеспечить приемлемое качество работы даже при применении различных методов фильтрации и шумоподавления. Для того чтобы повысить робастность работы автоматических систем, в последние годы в дополнение к анализу звуковой информации начали использовать методы распознавания визуальной информации о речи на базе технологий машинного зрения (т.н. “чтение по губам”). Речь – это результат взаимосвязанной работы артикуляторных органов человека: голосовых связок, гортани, легких, губ и языка; поэтому речь от человека поступает одновременно по нескольким каналам (модальностям), в том числе по звуковому и визуальному. Сигналы от визуальных и слуховых каналов дублируют и дополняют друг друга, что помогает правильно понимать речь во многих сложных ситуациях.

В СПИИРАН разрабатывается модель бимодального распознавания русской аудиовизуальной речи, использующая передовые технологии распознавания звучащей русской речи и автоматического анализа визуальной речи (т.н. «чтение по губам») и производящая распознавание слитно произносимых фраз русской речи для малого словаря прикладной задачи со скоростью обработки, приближающейся к реальному масштабу времени. Были реализованы и апробированы две модели аудиовизуального распознавания речи: 1) синхронная модель распознавания речи, основанная на математическом аппарате многопоточных скрытых марковских моделей (МПСММ) [3]; 2) асинхронная модель бимодального распознавания речи, основанная на математическом аппарате сдвоенных скрытых марковских моделей (ССММ) [4]. В данных моделях звуковые и визуальные речевые признаки разносятся по двум разным потокам, но объединение происходит разными способами на уровне состояний соответствующих скрытых марковских моделей. Последняя модель позволяет учитывать естественные для речи временные расхождения соответствующих акустических и визуальных признаков речи, возникающие из-за определенной инертности органов речеобразования и эффектов коартикуляции. Текущее состояние двухпоточной ССММ определяется одновременно состояниями акустической и визуальной компонент модели (при этом допускается асинхронность). Акустические признаки речи основаны на спектральной обработке аудиосигнала (частота дискретизации 16 кГц, моно) с вычислением кепстральных коэффициентов и их производных [5]; для вычисления визуальных признаков речи на

каждом кадре видеосигнала (720x576 пикселей, 25 кадров в секунду) происходит детектирование лица человека каскадным методом Хаара, затем поиск прямоугольной области рта в нижней части лица (если оно найдено на предыдущем шаге), и после нормализации изображения обнаруженной области рта происходит анализ главных компонент (PCA) визуального объекта [4].

Проведено тестирование бимодальных моделей распознавания с применением предварительно подготовленного корпуса аудиовизуальной русской слитной речи (слитное произнесение последовательностей цифр с длиной фраз от 3 до 6 слов) для 6 дикторов-носителей русского языка при варьировании отношения сигнал/шум (SNR), в незашумленный аудиосигнал добавлялся белый шум или “шум толпы” различной интенсивности. Результаты экспериментов с четырьмя моделями распознавания (две одномодальные модели и две бимодальные модели распознавания) представлены на рис. 1. Бимодальное распознавание речи превосходит по точности распознавания слов одномодальное аудиораспознавание, что особенно очевидно для низких значений SNR < 15 дБ. При этом асинхронная CCMM немного опережает по точности распознавания синхронную модель на базе МПСММ. При очень низком значении SNR < 5 дБ акустическая информация становится малоинформативной и наилучшие результаты показывает одномодальная модель распознавания только по визуальным признакам речи. Таким образом, в окружающей обстановке с низким отношением акустический сигнал/шум (ниже 10 дБ) обработка визуальной речи позволяет сохранить приемлемую точность распознавания слов и фраз русской речи.

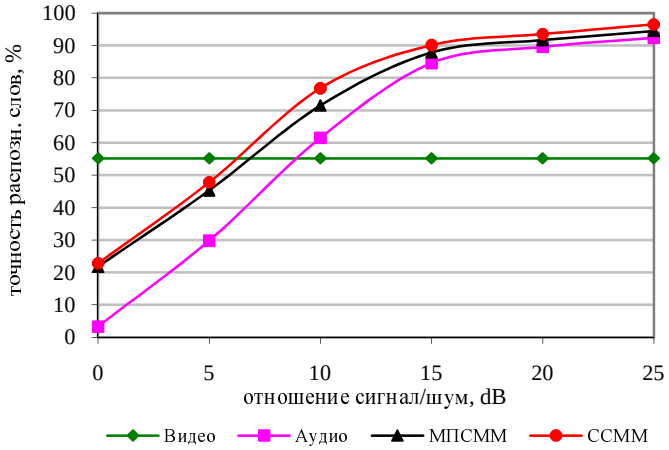

Рис. 1. Точность распознавания речи для 4-х моделей распознавания в зависимости от отношения сигнал/шум

Как известно, при плохом качестве звуковой обратной связи человек себя не слышит и стремится говорить громче, вызывая кричащую манеру произведения речи (эффект Ломбарда [6]), или же придвинуть микрофон ближе, что влечет появление отсечек звукового сигнала. Таким образом, необходимо обеспечивать для человека высококачественную обратную связь с эффективной передачей информации от машины к человеку.

Синтез аудиовизуальной речи для вывода информации. В качестве способа оповещения речевые и звуковые интерфейсы эффективнее тактильных и даже в ряде случаев эффективнее визуальных интерфейсов. Речевая модальность для информирования предпочтительнее визуальной в тех случаях, когда: сообщение однократное и не понадобится снова; касается протекания событий во времени

(например, информирование о наборе скорости); требует незамедлительной реакции на сообщение; зрительный канал восприятия человека перегружен; кроме того, в отличие от графических интерфейсов, речевые интерфейсы всенаправленные и позволяют освободить внимание пилота для выполнения основных задач полета. Объединение же визуального интерфейса с речевым в одной системе позволяет создавать многомодальные интерфейсы и добиваться наибольшей эффективности и надежности вывода информации. Визуальные сигналы очень важны для лучшего понимания произносимой речи; например, глядя в лицо собеседнику, нам легче понимать его речь.

В ходе исследований в СПИИРАН совместно с Объединенным Институтом Проблем Информатики НАН Беларуси, г. Минск и Западно-Чешским Университетом, г. Плзень, была разработана модель синтеза аудиовизуальной русской речи (т.н. «говорящая голова» [7]) по произвольному тексту. Модель синтеза представляет собой управляемую трехмерную модель лица человека, двигающую губами синхронно с синтезом соответствующей звучащей русской речи. Качество моделей синтеза речи принято оценивать по нескольким критериям: разборчивость речи (вычисляется как отношение правильно распознанных слов к общему количеству слов в высказывании), естественность и узнаваемость речи (субъективное сравнение с записями реальных дикторов). Результаты экспериментов показывают, что визуальная речевая модальность действительно помогает понимать речь лучше, особенно в зашумленных условиях, система бимодального синтеза превзошла одномодальную систему синтеза в среднем на 6% по показателю разборчивости слов в зашумленной речи, причем при анализе результатов тестирования разборчивости речи неоднократно наблюдался так называемый эффект МакГурка, когда правильное распознавание звука в слове возникает лишь при объединении акустических и визуальных сигналов, например для таких слов, как “сотка” и “сопка”; “ода” и “оба”. Конечно же, разборчивость реального голоса остается несколько выше, чем синтезированного, однако преимуществом системы синтеза речи является возможность озвучивания любого произвольного входного текста (например, названий топографических объектов, фамилий и т.д.) без необходимости предварительной записи. Кроме того, по когнитивным исследованиям известно [1], что чем более жестким и механическим звучит голос, тем скорее люди выполняют его указания, однако слишком роботизированные и неестественные голоса напрягают людей, и они их игнорируют.

Таким образом, внедрение речевого интерфейса, использующего одновременно как звуковую, так и визуальную модальности речи, для ввода и вывода информации позволяет даже в условиях зашумленной окружающей обстановки осуществлять эффективный вербальный человеко-машинный диалог и бесконтактное управление различными техническими системами как гражданского, так и военного назначения.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Кучерявый А.А.* Бортовые информационные системы / Под ред. В.А. Мишина и Г.И. Клюева. – 2-е изд., перераб. и доп. – Ульяновск: УлГТУ, 2004. – 504 с.
2. Eurofighter Typhoon, www.eurofighter.com/et_as_vt_dv.asp
3. *Карпов А., Ронжин А., Лобанов Б., Цирульник Л., Железны М.* Разработка бимодальной системы аудиовизуального распознавания русской речи // Информационно-измерительные и управляющие системы. – 2008. – № 10 (6). – С. 58-62.
4. *Nefian A., Liang L., Pi X., Xiaoxiang X., Mao C., Murphy K.* A coupled hmm for audio-visual speech recognition. Труды Международной конференции ICASSP-2002, Орландо, США, 2002.
5. *Ronzhin A., Karpov A.* Russian Voice Interface // Pattern Recognition and Image Analysis, МАИК Наука/Interperiodica. – 2007. – Т. 17. – № 2. – С. 321-336.

6. *Бондарос Ю., Колоколов А., Костюк А.* Исследование речевых сигналов в условиях кабины летательного аппарата // Вестник компьютерных и информационных технологий. – 2008. – № 4. – С. 2-10.
7. *Лобанов Б., Цирульник Л., Железны М., Крноул З., Ронжин А., Карпов А.* Система аудиовизуального синтеза русской речи // Информатика. – Минск: ОИПИ Беларуси. – 2008. – № 4 (20). – С. 67-78.

Карпов Алексей Анатольевич
Учреждение Российской академии наук Санкт-Петербургский институт информатики и автоматизации РАН.
Email: karpov@iias.spb.su.
199178, г. Санкт-Петербург, 14-я линия, д. 39.
Тел.: 88123287081.

Karpov Alexey Anatolievich
Institution of the Russian Academy of Sciences St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences.
Email: karpov@iias.spb.su
14-th Line, 39, Saint-Petersburg, 199178, Russia.
Phone: 88123287081.

УДК 004.415.53

С.В. Бирюков

РАЗРАБОТКА МЕТОДА АВТОМАТИЗАЦИИ ТЕСТИРОВАНИЯ СИСТЕМ С ИНТЕРФЕЙСОМ ПРОГРАММИРОВАНИЯ

В данной работе рассматриваются вопросы автоматизации тестирования на основе модели систем с интерфейсом программирования. Особое внимание уделяется вопросам автоматической генерации тестовых сценариев и построения тестовых оракулов. Приводится структура данных для хранения и обработки модели интерфейса с расширением для функциональных требований. Предложенный подход реализован в рамках программной среды генерации тестовых сценариев и оракулов APITest.

Интерфейс программирования; автоматизация тестирования; спецификация интерфейса; унифицированная модель.

S.V. Biryukov

THE DEVELOPMENT OF AUTOMATED TESTING METHOD FOR SYSTEMS WITH PROGRAMMING INTERFACE

The issues of model-based automated testing for systems with programming interface are considered in this paper. Particular attention is paid to the automatic generation of test scenarios and building test oracles. The data structure for storage and processing of interface model with the expansion of functional requirements is offer. The proposed approach is implemented within the software environment for generating test scripts and oracles APITest.

Programming interface; automated testing; interface specification; unified model.

При разработке и сопровождении программных систем (ПС) значительная часть усилий тратится на поиск и устранение ошибок. Самым распространённым методом поиска ошибок является тестирование, т.е. процесс выполнения программ с целью обнаружения ошибок [1]. В настоящее время необходимость систематизированного тестирования в промышленной разработке ПС общепризнанна и неоспорима. Однако в большинстве случаев тестируемые ПС связывают с наличием в нем графического интерфейса пользователя, которое служит медиа-