**Разговор с машиной: мифы и реалии речевого управления**

Почуев Сергей Иванович, д.т.н., профессор.

На научно-популярном уровне изложены некоторые новые технико-эргономические психологические принципы и подходы к созданию и использованию естественно языкового интерфейса “человек-машина” для управления сложными техническими системами.

**Введение**

С момента появления первых “интеллектуальных” машин (под “интеллектуальной” машиной далее понимается объект искусственного происхождения, способный воспринимать и рационально использовать информацию, передаваемую ей в любой форме человеком, для выполнения определенных целевых действий) люди стремятся к максимальному удобству управления ими. Это стремление находит свое выражение как в конкретных технических разработках, так и в неосознанных и осознанных мечтах людей о думающих, понимающих человека и говорящих с ним помощниках, имеющих искусственную природу. За примерами далеко ходить не надо. Достаточно вспомнить Али-Бабу, дистанционно управлявшего дверями пещеры с сокровищами с помощью сакраментальной речевой команды “Сезам откройся!”, героев многочисленных научно-фантастических романов или даже пушкинскую царицу, периодически консультирующуюся с чудо зеркальцем по поводу собственной внешности. В последнем случае мы, по-видимому, имеем дело с гениально предугаданным прообразом системы глобального мониторинга женщин, сопряженной с экспертной системой определения уровня их красоты и оснащенной естественно языковым интерфейсом….

В настоящее время попытки создания подобного интеллектуального речевого интерфейса “человек-машина” все более перемещаются из мира сказок в реальную жизнь, а соответствующие системы стремятся занять свое место под солнцем в кабинах самолетов, рубках кораблей, трубках мобильных телефонов, “мозгах” роботов и т.п.

Вместе с тем объем использования естественно-речевых технологий в практике управления сложными техническими системами до настоящего времени весьма невелик, несмотря на очевидную привлекательность такого способа общения с “интеллектуальной” машиной.

Более того, все мы являемся свидетелями того, что разнообразные, постоянно совершенствуемые электронные приборы как бытовые, так и используемые в различной профессиональной деятельности людей, продолжают “обрастать” все новыми многофункциональными кнопками, значения которых не в состоянии быстро запомнить и эффективно использовать, особенно в условиях возможного дефицита времени, ни один нормальный человек. В результате, возрастающие функциональные возможности техники часто оказываются нереализованными, а освоение новых электронных устройств человеком-оператором превращается в мучительный процесс запоминания многотомных инструкций и наставлений!

Что надо сделать для того, чтобы люди могли и хотели разговаривать с машинами, а машины отвечали им взаимностью?

Отвечая на этот непростой вопрос, современные ученые часто уповают на трудно преодолимые технические сложности надежного распознавания естественной слитной человеческой речи и ее смысловой интерпретации машиной. С данными утверждениями во многом следует согласиться. Это, прежде всего, касается существенных ограничений созданных к настоящему времени алгоритмов искусственного интеллекта, программ распознавания слитной речи и, в ряде случаев, недостаточных реальных возможностей аппаратной части современных компьютеров.

Однако, когда мы имеем дело с управлением сложной технической системой и не требуем от нее порождения новых знаний или оперативного освоения (понимания) новых команд, ситуация упрощается. В рассматриваемом случае нам нужен не мудрый думающий собеседник, а толковый и дружелюбный исполнитель нашей воли, который при решении той или иной функциональной задачи без ущерба для конечного успеха может быть ограничен в своих словах и действиях.

Как здесь не вспомнить наших четвероногих любимцев, которых мы привыкли считать умными, если они знают десяток команд и попусту не лают ... Автор далек от мысли проводить буквальное сравнение, чтобы не обидеть ни собак, ни машины. Каждые по-своему хороши! Вместе с тем, абсолютное большинство современных технических систем строиться именно так, чтобы на бесконечное число ситуаций внешнего мира реагировать большим, но конечным количеством действий. Это, в первую очередь, касается т.н. разовых команд, определяющих режимы работы того или иного технического устройства. Вспомните рычаг указателя поворота на Вашем автомобиле, кнопку “START” на видеокамере и т.п. Находясь в любой точке земного шара и, будучи дисциплинированным водителем, Вы будите действовать одинаково, указывая поворот. Ваши стандартные манипуляции с видеокамерой для начала съемки также никак не зависят от выбранного сюжета, а определяются ее конструкцией.

Таким образом, даже работая с фиксированным словарем речевых команд, можно добиваться достаточно впечатляющих результатов с точки зрения повышения комфортности и оперативности управления различными техническими устройствами и системами, в полной мере наслаждаясь иллюзией их интеллектуальности.

Данный вывод не претендует на новизну, т.к. еще лет 30 назад, а может быть и более, специалисты предрекали блестящее будущее подобным системам. Но, повторюсь, предсказания не оправдались, несмотря на значительные успехи в области повышения надежности распознавания речи, в особенности команд из ограниченного набора.

Мифы о речевом управлении так и не стали широкой повседневной практикой! Почему?

В качестве главных причин подобного явления обычно называют:

-существенную зависимость надежности распознавания от индивидуальных особенностей диктора (например, он шепеляв и совсем не тот, кто ранее обучал машину);

-подчиненность результатов распознавания текущему психофизиологическому состоянию человека (например, оператор “с большого бодуна”, простужен или сильно напуган приближением к Земле очередного астероида);

-сильное влияние акустических помех, присутствующих в ситуациях реальной жизни (например, я говорю, а рядом стонет мотор старых “Жигулей”).

-наличие пользовательского стереотипа управления (например, аргументы оператора типа: “меня так учили, мне так удобнее, я двадцать лет успешно нажимаю на кнопки и не собираюсь произносить Ваши дурацкие команды!”).

Рискну предположить, что дело здесь не только, а может быть и не столько в этих причинах (последняя причина, кстати сказать, скорее не причина, а следствие низкого уровня эргономичности немногочисленных реально существующих систем речевого управления).

Организация речевого диалога с машиной это не только борьба за высокую надежность распознавания человеческой речи в разнообразных условиях, но и придание самому процессу общения некой осмысленности, удобства, взаимности, доверия и своего рода психологической совместимости со стороны человека!

Но этой стороне проблемы речевого управления, как мне кажется, в настоящее время учеными и конструкторами не уделяется должного внимания. Поэтому, абстрагируясь в дальнейшем от упомянутых ранее “чисто распознавательных” аспектов речевого управления, сосредоточимся на некоторых новых технико-эргономических и психологических принципах и подходах, использование которых может позволить ускорить процесс внедрения “речевых” технологий в техническую практику.

Дабы излишне не “грузить” читателя, также оставим в стороне и важную смежную проблему синтеза и использования речевых информационных сообщений, генерируемых машиной, независимо от речевых команд человека. Благо, данная проблема более изучена и приятные женские голоса, сообщающие человеку-оператору плохие новости, уже достаточно давно звучат в наушниках пилотов самолетов и салонах “крутых” автомобилей.

Итак, будем полагать, что мы имеем дело с некоторой сложной технической системой, управление которой осуществляется в реальном масштабе времени с использованием фиксированного набора разовых (дискретных) речевых команд. Далее речь пойдет именно о таких командах, т.к. человеку осуществлять непрерывное управление технической системой, как правило, удобнее с помощью рук, ног и органов управления, на которых находятся эти и, как знать, возможно, и другие части тела.

По мнению автора, при построении эффективной и дружелюбной системы речевого общения необходимо помнить и правильно использовать следующее.

Речевой канал не должен быть единственно возможным для решения любой из функциональных задач, возлагаемых на управляемую техническую систему и должен использоваться только тогда, когда это действительно удобно человеку, например при дефиците времени или большом количестве рутинных тактильных (ручных) операций.

Словарь машины должен создаваться и модифицироваться на стадиях проектирования и модернизации технической системы, но не в процессе управления ею, исходя из конструктивно заложенной в системе логики работы и реально имеющихся органов управления и визуальной индикации.

Машина должна быть толерантна к условиям распознавания, таким как: диктор, его психофизиологическое состояние, внешние шумы.

В общем случае не все априорно известные (содержащиеся в словаре) машине команды будут доступны для выполнения в произвольно выбранный момент процесса управления. Машина в любой момент процесса управления должна быть способна самостоятельно без участия человека идентифицировать подмножество доступных команд на множестве априорно известных, которым ее заранее обучили.

Машина должна “интеллектуально” или, по крайней мере, “человекоподобно”, реагировать на результаты распознавания команд своими ответными словами и действиями. Для этого она должна иметь возможность самостоятельно контролировать надежность распознавания каждой речевой команды и рационально использовать эту информацию, например, для того, чтобы переспросить человека в сомнительных случаях или предупредить его о возможной ошибке.

Необходимо обеспечить возможность априорной и оперативной (как минимум) или адаптивной (как максимум) настройки речевого интерфейса под конкретного пользователя.

В первом случае до начала или в процессе управления человек сам указывает машине то, как она должна вести себя в процессе общения с ним, например, быть молчаливым исполнителем команд, активным полемистом или бюрократом, который в ответ на каждую команду запрашивает подтверждение на ее выполнение.

Во втором случае машина сама должна подстраиваться под человека, анализируя его психофизиологическое состояние и уровень профессиональной подготовки.

Человек легко прощает себе собственные ошибки управления, но не склонен прощать аналогичные ошибки машине, как впрочем и другим людям… Потеря доверия человека к речевому каналу управления вследствие частых ошибок машины или ее в принципе правильного, но неудобного для человека поведения, на практике приводит к отказу от использования данного канала. Поэтому необходима высоко надежная и простая процедура сертификации человеком реализуемых с помощью речевого канала управленческих решений.

Человек может забыть известную и произнести неизвестную (не содержащуюся в словаре) команду;

Человек может забыть логику работы управляемой им технической системы и произнести недоступную для выполнения в данный момент времени, но априорно известную для машины команду;

Человек не всегда рационален и, тем более, не всегда оптимален в своих решениях (произносимых речевых командах). В необходимых случаях желательна возможность автоматического формирования машиной подсказки для человека, о том какие команды в текущий момент времени являются доступными, а какие из доступных команд рациональными для сложившейся ситуации управления сложной технической системой. В то же время, машина не должна иметь возможность блокировать волеизъявление человека, выраженное посредством доступной, но не рациональной по ее мнению, речевой команды.

Машина может принимать различные правильные решения по распознаванию. Число возможных вариантов при ранее принятом допущении о фиксированном наборе речевых команд конечно:

“человеком произнесена неизвестная команда --> неизвестная команда правильно распознана машиной”;

“ человеком произнесена известная доступная команда --> известная доступная команда правильно распознана машиной ”;

“ человеком произнесена неизвестная недоступная команда --> неизвестная недоступная команда правильно распознана машиной ”.

Машина по своей вине или по вине оператора может допускать ошибки распознавания 6-ти типов. Число вариантов также конечно.

“Неизвестная команда --> известная недоступная команда” - (оператор ошибся, машина не смогла распознать ошибку);

“Неизвестная команда --> известная доступная команда ” - (оператор ошибся, машина не смогла распознать ошибку);

“Известная доступная команда ---> неизвестная команда ” (ошиблась машина);

“Известная недоступная команда --> неизвестная команда ” - (оператор ошибся, машина не смогла распознать ошибку);

“Известная доступная команда --> известная недоступная команда ” (ошиблась машина);

“Известная доступная команда --> известная доступная команда распознана неправильно ” (ошиблась машина).

Поясним изложенные положения и определим возможные направления их практического учета и применения.

Деление известных команд на доступные и недоступные должно осуществляться, исходя из располагаемого конечного множества режимов работы управляемой технической системы, которые, как правило, закладываются уже на этапе ее создания. (Речь идет об использовании ранее упомянутого принципа конечного числа реакций машины на бесконечное множество ситуаций внешнего мира.)

Для простоты восприятия читателем, режим в данном случае, если угодно, - это совокупность нажатых кнопок управления, а доступная известная команда это очередная кнопка (одна из нескольких возможных), которая может быть “нажата голосом” оператора.

Например, на многих моделях автомобилей нельзя включить “мигалку” указателя поворота при вынутом ключе зажигания, но в той же ситуации можно подать звуковой сигнал. В данном иллюстративном примере “вынутый ключ зажигания” - это режим, “мигалка” - это недоступная известная команда, “гудок” - это доступная известная команда для данного режима.

В наиболее общем случае режим – это не только совокупность нажатых кнопок управления, но и совокупность значимых для процесса управления событий, наступление которых машина способна самостоятельно идентифицировать без участия человека.

Например, пусть обзорная система ледокола “Титаник-II” способна самостоятельно фиксировать факты обнаружения появляющихся в зоне обзора айсбергов, нумеровать их и переходить к непрерывному сопровождению наиболее опасных при получении команды от капитана судна. В данном примере наступление значимого события - “обнаружение айсберга” без непосредственного участия человека инициирует возможность использования нового режима – “сопровождение айсберга”. Для данного режима характерна известная машине, но ранее недоступная для выполнения (айсбергов раньше просто не было видно!) речевая команда капитана – “следи за наиболее опасным айсбергом №3”.

Необходимо отметить, что априорное деление известных речевых команд на доступные и недоступные с использованием содержания понятия “режим” позволяет сократить множество альтернатив и тем самым повысить надежность распознавания. Альтернативы (доступные команды) сравниваются в рамках режимов, в которых они выполнимы. Затем на “межрежимном” уровне сравниваются “альтернативы – победительницы”. Результат распознавания – альтернатива победительница для межрежимного уровня.

В условиях стрессовой нагрузки или вследствие слабой подготовки (и такое, оказывается, случается, как пишут в газетах!) оператор может произнести неизвестную для системы команду, например, выругаться в сердцах “по матушке”. В данном случае машина должна самостоятельно, в соответствии с используемыми критериями (например, пороговыми или иными, в зависимости от используемого алгоритма распознавания) идентифицировать команду как неизвестную. Чтобы поведение машины (произносимые ей ответные слова и выполняемые действия) выглядело дружелюбным и “интеллектуальным” возможны следующие основные (перечень возможных реакций машины здесь и далее ограничивается лишь уровнем естественного интеллекта разработчика “софта”, располагаемыми вычислительными ресурсами и прагматической целесообразностью) варианты ее реакции:

промолчать и ничего не делать;

сообщить человеку о возможной ошибке и попросить повторить команду;

предложить помощь, произнося доступные в данном режиме команды.

Вариант реакции можно, например, задать заранее при проектировании системы речевого управления в зависимости от априорной важности того или иного режима в процессе управления “интеллектуальной” машиной. Реализация такого подхода, естественно требует анализа профессиональной деятельности оператора, учета циклограммы его работы в процессе управления технической системой и детального знания режимов ее работы.

Более сложный, но и более “интеллектуальный” вариант - оперативная оценка психофизиологического состояния оператора и использование специализированной экспертной системы - помощника, формирующей апостериорную (de facto) подсказку в реальном масштабе времени с учетом “человеческого фактора” и результатов наблюдений с помощью датчиков информации за изменяющимися условиями внешнего мира. Проектируя подобную экспертную систему, необходимо помнить, что далеко не все доступные речевые команды оптимальны или, по крайней мере, рациональны для конкретной ситуации управления. Более того, выполнение некоторых из них при негативном развитии событий, в принципе, может привести к печальным или трагическим (вспомните Чернобыль!) последствиям.

Здесь, однако, уместно замечание по поводу невозможности блокирования выполнения машиной доступных речевых команд. Современный уровень развития систем искусственного интеллекта и, в частности, экспертных систем в настоящее время таков, что пока не позволяет эффективно решать гуманитарные проблемы и проблемы связанные с порождением новых знаний непосредственно в процессе управления сложной технической системы. Например, сегодня машину сложно научить находить разницу между потерей инстинкта самосохранения и осознанным самопожертвованием в действиях пилота боевого самолета, или отличие непрофессионализма от нетривиального (не содержащегося в инструкции для данной ситуации), но спасительного для сотен людей управленческого решения оператора крупного железнодорожного узла.

Поэтому представляется, что система речевого управления правильно и надежно интерпретирующая волю человека посредством выполнения доступных речевых команд и в необходимых случаях формирующая для него “толковые” подсказки может считаться вполне “интеллектуальной”. При этом ответственность за последствия нерациональной деятельности дилетанта за рычагами управления или перед “ушами” системы речевого управления должен нести он сам или люди, которые его туда допустили, а никак не конструктор “интеллектуальной машины”…

В случае, когда распознанная команда идентифицируется как известная доступная, возможны следующие основные варианты реакции машины:

выполнить команду;

выполнить команду и информировать оператора голосом, или как-то иначе, о ее выполнении;

запросить подтверждение команды у оператора и не выполнять команду до получения подтверждения;

проинформировать оператора о нерациональности его действий, возможных негативных их последствиях и оптимальной с точки зрения машины команде.

В этом месте необходимо сделать небольшое, но важное, “лирическое отступление”, касающееся облика интерфейса экспертных систем, сопрягаемых с командно-речевыми системами. Нужно помнить, что человек “устроен” так, что выступать со своими альтернативными советами и соображениями машина должна крайне осторожно. Работать с “машиной–занудой”, особенно когда у тебя плохое настроение, “запахло жареным” и нет времени на раздумье, или когда имеется собственное и, конечно же, единственно правильное мнение не склонен ни один хомо сапиенс. Вспомните, как поступала пушкинская царица со своим чудо зеркальцем в проблемных случаях. Вспомнили? Правильно! В первый сунула под лавку, а во второй – разбила об пол…

Отдельного рассмотрения заслуживает упомянутый вариант диалогового подтверждения команды. Машина и человек, зная о том, что они обоюдно на 100 % не застрахованы от ошибок (см. приведенные выше возможные типы ошибок), должны иметь “супернадежную” совместную процедуру сертификации принятия решения на выполнение речевой команды.

В основу такой процедуры может быть положен принцип разделения команд по длительности. Например, человек в ответ на распознанную и оглашенную машиной команду может ответить короткое “Да” или произнести длинное “Отставить” (или, возможно, что-то другое длинное и грубое) в случае своего несогласия. Подобный предельно простой принцип распознавания с использованием только 2-х возможных хорошо различимых команд представляется наиболее надежным в отличие от использования любых других известных принципов, предполагающих применение “навороченных” математических методов типа формантного анализа или нейрокомпьютерных технологий.

После получения подтверждения или отказа от выполнения команды возможны следующие основные варианты реакции машины:

На команду человека - “Да”:

выполнить распознанную и “озвученную” команду;

выполнить распознанную и “озвученную” команду и информировать оператора голосом о ее выполнении.

На команду человека - “Отставить”:

не выполнять распознанную и “озвученную” команду;

не выполнять и “озвученную” команду и сообщить об этом оператору.

Кстати, реализуя данную процедуру, совсем не лишним будет заранее “договориться” с машиной о том, как она будет интерпретировать молчание хозяина:

молчит - выполнить команду;

молчит - не выполнять команду;

молчит - у хозяина отказали мозги или пропал голос, запрошу еще раз, а там видно будет…

Особый интерес представляет случай, когда распознанная команда идентифицируется как известная недоступная. В данном случае наступает “пикантная” ситуация, связанная с ошибкой оператора, по каким-то причинам забывшего режимы и логику работы управляемой им технической системы.

Возможны следующие основные варианты реакции машины:

промолчать и ничего не делать;

сообщить оператору о невозможности выполнения команды;

сообщить оператору о невозможности выполнения команды и проинформировать его о причинах этого;

сообщить оператору о невозможности выполнения команды, причинах этого и предложить список доступных команд или наиболее рациональную, по мнению машины, команду.

Рассмотрим еще один безусловно важный с эргономической и психологической точек зрения вопрос, связанный с процессом инициации работы системы речевого управления. Должна ли она работать постоянно, пока осуществляется управление технической системой или включаться вручную на время ввода команды?

На первый взгляд более удобен первый вариант, не требующий никаких дополнительных тактильных действий со стороны оператора, т.к. ничего не нужно нажимать или переключать. Однако не все так просто и для человека и для машины.

Люди в процессе управления имеют обыкновение разговаривать не только с машиной, но и со своими коллегами. Некоторые вообще разговаривают сами с собой (есть и такие!) или поют песню “Вечерний звон” как в фильме “Небесный тихоход”… Постоянное осознание того, что в это время машина как преданный пес пытается уловить известную ей команду и может неверно интерпретировать не обращенные к ней речи хозяина, нервирует человека-оператора и мешает ему нормально работать. Не легче и машине, которая захлебывается в потоке незнакомых слов и звуков. Как уже отмечалось ранее, сегодня существуют серьезные трудности с обеспечением приемлемой вероятности распознавания слитной речи. Поэтому, по мнению автора настоящей статьи, более рациональным является второй вариант – использование единой кнопки ввода, которую, например, следует нажимать перед началом ввода любой речевой команды и отпускать после ее окончания. Это, кстати, дополнительно дисциплинирует оператора, заставляя предварительно подумать над тем, что он на самом деле хочет сказать машине.

Ежели Вы все же склонны остановиться на первом варианте как более “продвинутом”, осмелюсь рекомендовать использование некоторой предварительной уникальной хорошо распознаваемой в различных условиях команды – обращения к машине. И после того как Ваш верный железный Мухтар “пролает” в ответ на обращение к нему о том, что он Вас услышал и ждет команду, можете смело произносить исполнительную команду: “Фас!”. Последующее выделение исполнительной команды из слитной речи может быть эффективно осуществлено по временному признаку: сделайте небольшую паузу до и после исполнительной команды (Твикс при этом можно не кушать!).

Как показывает анализ известных автору коммерческих программ распознавания речи, первый вариант оказывается неплохим для студийных условий с использованием предварительных команд конкретного диктора компьютеру типа: “Wake up!” и “Go to sleep!”, но в реальной жизни не работоспособен.

**Заключение**

Предложенный в статье подход к созданию систем речевого управления в части психолого-эргономических аспектов проектирования интерфейса “человек – машина” для работы на ограниченном наборе речевых команд был достаточно успешно апробирован автором при создании компьютерных прототипов систем речевого управления общего назначения с использованием собственного оригинального программного обеспечения процедуры распознавания, которая в данной статье не рассматривалась.

Статья является оригинальной разработкой: при ее написании не использовались материалы никаких отечественных и зарубежных организаций или частных лиц.

Изложение материала сознательно облечено в научно-популярную форму, чтобы не перегружать статью “заумными” формулами и ссылками на “приличествующие” случаю обзорные монографии, не “выплеснуть ребенка”, а также не детализировать до уровня возможности практического использования имеющиеся авторские know-how. Вместе с тем, как мне кажется, объем приведенных в статье соображений вполне достаточен для того, чтобы составить общее представление и впечатление об актуальности, новизне и практической значимости предлагаемых решений, равно как и о заблуждениях и ошибках автора.

Буду весьма признателен, получив Ваши отзывы, критические замечания, советы, пожелания и, возможно, предложения о дальнейшей совместной работе. Пожалуйста, если можно, критикуйте и советуйте не по форме изложения материала, а по существу обсуждаемой проблемы!

Весьма интересны для автора мнения психологов, специалистов в области эргономики и конечных пользователей, на практике управляющих сложными техническими системами (моряков, летчиков, операторов систем УВД и АЭС, “телекоммуникационьщиков”), а также специалистов в области компьютерного распознавания речи, робототехники и всех остальных кто “в теме”:

Нужны ли подобные системы для автомобилей, кораблей, самолетов, в быту и т.п.? Как Вы представляете их использование?

Приходилось ли Вам сталкиваться с элементами изложенного подхода, и как Вы его оцениваете (достоинства, слабые места)?

Целесообразен ли форум или телеконференция по этой проблематике?

Какие вопросы в рамках рассматриваемой проблематики Вас волнуют?

Что следует почитать автору для ликвидации текущей безграмотности?