

**УПРАВЛЕНИЕ ДИНАМИКОЙ РЕЧЕВОГО ПОВЕДЕНИЯ  
ВИРТУАЛЬНЫХ КОМПЬЮТЕРНЫХ АГЕНТОВ**  
**CONTROLLING DYNAMIC SPEECH BEHAVIOUR OF VIRTUAL  
COMPUTER AGENTS**

*Котов А.А. (kotov@harpia.ru)*

*Российский государственный гуманитарный университет*

В докладе представлена модель для управления речевым поведением компьютерного агента (героя компьютерной игры, компонента интерфейса или, в перспективе, подвижного робота). Модель позволяет имитировать «динамику настроения» агента, что определяет поведение агента в коммуникации. В частности, модель строит из набора шаблонов высказываний короткие монологи, отражающие «переживания» агента, и обеспечивает переключения в диалоге между несколькими собеседниками.

Виртуальные компьютерные агенты могут представлять из себя простые программные модули, которые имитируют поведение отдельного организма в большом сообществе – и создаются в исследовательских целях [Slovan, 2001]. Такие программные модули могут также обладать интерфейсом и использоваться для создания правдоподобных персонажей в компьютерных играх или даже в детских матерчатых игрушках [Paiva, Chaves et al., 2003]. Простые технологии имитации эмоционального поведения и эмоциональной речи используются для разработки интерфейсов компьютерных программ – компьютерных персонажей, которые развлекают пользователя своим поведением, являются учителями иностранного языка, наставниками по фитнесу или даже помогают социализации замкнутых школьников и детей иммигрантов [Hall, Woods et al., 2005]. Сложное речевое поведение и «адекватные» эмоциональные реакции агента на собственные неудачи или неудачи пользователя несомненно повысят интерес человека к таким агентам и позволят создать новые типы интерфейсов, общение с которыми подобно общению с живым человеком – такие интерфейсы снизят нагрузку на пользователя и избавят его от предварительного обучения работе с интерфейсом. Подобные исследования в этих областях представлены в работах и сборниках [Picard, 2000; Cassel, Sullivan и др., 2000; Andre, Dybkjær и др., 2004; Tao, Tan и др., 2005; Paiva, Prada и др., 2007]. Есть надежда, что в перспективе подобные речевые технологии будут использоваться в бытовых роботах, взаимодействующих с пользователем в реальном мире, распознающих эмоции пользователя и отвечающих эмоциональными высказываниями в неловких ситуациях.

Ранее мы рассматривали модель компьютерного агента, который, попадая в эмоциональную ситуацию, реагирует на неё одним высказыванием [Котов, 2006] – (см. Рис. 1). Это высказывание может выбираться в зависимости от имитируемого темперамента или настроения агента<sup>1</sup> – однако по одному высказыванию пользователь не всегда сможет определить состояние агента, и не во всех случаях такой короткий ответ будет достаточен для коммуникации. Для более адекватного взаимодействия агент должен уметь строить небольшой монолог, соответствующий своему эмоциональному настроению: предлагать подряд несколько высказываний, пусть даже не объединённых сюжетными отношениями. Ещё более интересная задача состоит в том, чтобы динамически менять эмоциональное состояние агента в монологе (см. Рис. 2). Например, после негативного эмоционального события агент может расстраиваться, ругать себя, потом «нападать» на адресата, успокаиваться и, наконец, предлагать обсудить решение возникших трудностей. Если такая замена состояний происходит быстро, она производит комический эффект, что часто используется в мультипликации для создания привлекательных персонажей. Наша задача несколько сходна с задачей мультипликаторов, с той разницей, что смена состояний агента, последовательность и структура его высказываний, а также итоговый сценарий – формируются алгоритмически и могут различаться в зависимости от параметров внешнего воздействия и самого агента. Перспективная задача состоит в том, чтобы создать агента, который сможет правдоподобно себя вести в потенциально бесконечном количестве ситуаций (пусть не слишком интеллектуально, но «понятно» с эмоциональной точки зрения), строя при этом потенциально бесконечное число высказываний.

<sup>1</sup> Далее во всех случаях, когда упоминается эмоциональное состояние агента – речь будет идти о таком эмоциональном состоянии, которые мы пытаемся имитировать с помощью речевого поведения агента и которое, в случае успеха, будет приписано агенту наблюдателем.

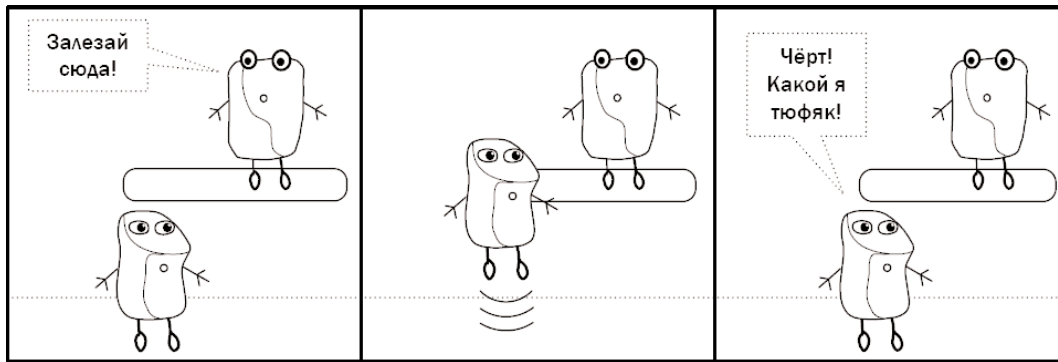


Рис. 1. Базовая ситуация, в которой виртуальный агент переживает неудачу (испытывает фрустрацию)

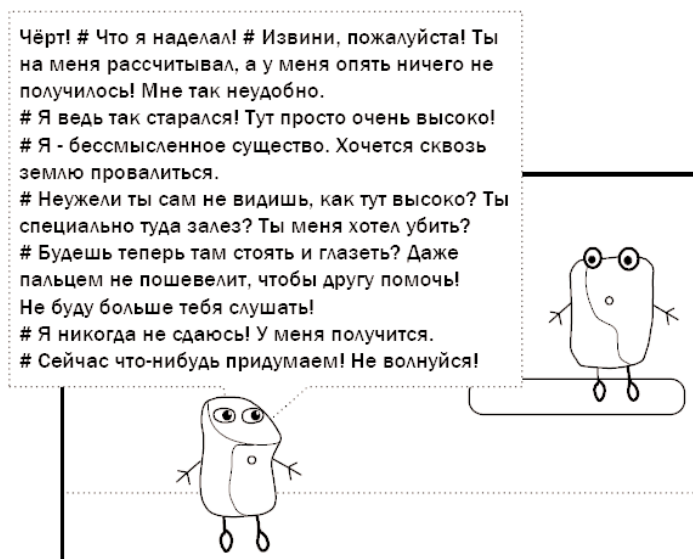


Рис. 2. Смена микросостояний позволяет строить монологи, в которых заметна динамика внутренних состояний агента

В настоящее время предлагаемая нами теоретическая модель реализована в виде достаточно простого компьютерного агента, напоминающего героя мультфильмов или компьютерных игр (интерфейс сделан на технологии Flash). Лингвистическая и поведенческая части агента пока достаточно «негибки»: агент оперирует набором фиксированных фраз (шаблонов) в наборе фиксированных тестовых ситуаций. Вместе с тем, набор используемых фраз достаточно широк и собран на большом экспериментальном материале (около 500 испытуемых разных профессиональных и возрастных групп [Kotov, 2005]), для некоторых фраз предусмотрены альтернативы в выборе лексики для некоторых синтаксических позиций (например, замена местоимений при обращениях к разным участникам ситуации и выбор нейтральной или инвективной лексики – в зависимости от степени «филтрации»).

Хотя такая модель достаточно ограничена, она, тем не менее, хорошо применима для оценки адекватности различных поведенческих стратегий: используемый инвентарь фраз достаточен, чтобы «разыгрывать» разнообразные коммуникативные стратегии и проверять их адекватность с помощью экспертов или в экспериментальном исследовании.

В имеющейся базе размечены соответствия между высказываниями и теоретическими компонентами модели, что открывает возможности для более сложного варьирования высказываний, а также для построения новых высказываний в соответствии с теоретическими правилами.

#### Архитектура агента

Попытка описать динамику эмоционального поведения агента предпринималась в нескольких научных проектах. В работе [Allen, 2001], посвящённой программной реализации эмоциональных агентов, для этой цели предлагалось использовать набор так называемых «контрольных состояний»: это *планы, намерения, желания, эмоции* (краткосрочные контрольные состояния), а также *предпочтения, навыки, правила и личностные характеристики* (долгосрочные контрольные состояния). Предполагается, что конкретное действие агента определяется конкуренцией в иерархии контрольных состояний, каждое из которых обладает разной степенью активизации. В указанной работе эта теоретическая архитектура частично реализована программными средствами для имитации поведения агента (задача имитации речевого поведения при этом не ставится). В нашей работе мы

## Управление динамикой речевого поведения виртуальных компьютерных агентов

также следуем представлению о том, что выбор высказывания определяются рядом конкурирующих контрольных состояний, например, личностными характеристиками или краткосрочными эмоциями. В частности, мы используем наблюдения над личностными характеристиками разных групп реальных испытуемых, чтобы сформировать и уточнить набор контрольных состояний более низкого уровня – «эмоций» или «настроений».

В другом научном проекте [Besker, Корр и др., 2004] сделана попытка имитации речевого поведения трёхмерного компьютерного агента, причём агент может менять свои эмоции в ходе диалога с пользователем. Речевое поведение агента контролируется тремя общими переменными: *pleasure* (степень удовольствия или неудовольствия), *arousal* (степень возбуждения) и *dominance* (субъективное представление о контроле над ситуацией). Значения этих переменных задают трёхмерное пространство, в котором различные регионы (сферы) отвечают за конкретные эмоции и связанные с ними формы поведения. Например, *страх* определяется как неприятная эмоция, связанная с высоким возбуждением и ощущением потери контроля над ситуацией: [P = -80, A = 80, D = -100]. Входящие стимулы могут менять значения этих переменных (например, оскорбления «расстраивают» агента), что вызывает формы поведения, характерные для нового состояния (агент может выражать недовольство с помощью мимики или даже уходить).

Используемый нами агент обладает двухуровневой структурой. Первой уровень состоит из набора *сценариев* и отвечает за отдельные речевые реакции. Второй уровень состоит из набора *микросостояний* и позволяет агенту отвечать в ситуации сразу множеством высказываний (монологом, не содержащим сюжетных структур) и распределять высказывания во времени, имитируя смену кратких эмоциональных состояний. Представим эти два уровня несколько подробнее.

Простые речевые реакции агента моделируются с помощью набора *сценариев* – правил типа «если – то». Множество сценариев подразделяется на д-сценарии (доминантные сценарии), описывающие речевые или поведенческие эмоциональные реакции, и р-сценарии (рациональные сценарии), описывающие рациональное поведение в эмоциональной ситуации или «рациональные» речевые стратегии, например, ситуация может рассматриваться агентом как задача, требующая решения, или агент может стремиться выработать новое правило для поведения в аналогичных ситуациях. Д-сценарии представлены фиксированным набором – 35 единиц<sup>2</sup>, в тестовых протоколах используется около 10. Р-сценарии отражают «рациональные» знания агента о ситуациях реального мира или стратегиях этикета; таким образом, для любой ситуации число потенциально возможных р-сценариев бесконечно, но представленные в протоколах ответы можно объединить в группы – всего представлено около 20 классов рациональных ответов. Остальные д-/р-сценарии могут использоваться для синтеза оригинальных или ироничных ответов, которые теоретически возможны, но не представлены в нашей базе (если мы разбили чужую вазу, мы можем сказать: *Как эта ваза красиво шлёпнулась и разлетелась на части!* – однако такой ответ или его варианты не засвидетельствованы в протоколах).

Начальные условия д-сценариев содержат набор стандартных валентностей. Так, негативные д-сценарии включают валентности AGGR (каузатор негативной ситуации) и VICT (жертва негативной ситуации). Если агент произносит фразу *Ты меня чуть не убил*, то такая фраза является следствием активизации д-сценария ОПАСН, в котором агент (говорящий) относит себя к валентности VICT, а адресата – к валентности AGGR. Таким образом, различное распределение валентностей между участниками коммуникации определяет несколько коммуникативных схем. Для негативных д-сценариев это, например, «жалоба», «речевое воздействие», «конфликт», «речевая агрессия», а для позитивных д-сценариев, например, «комплимент», «реклама» и «хвастовство».

Итоговое высказывание, полученное после выбора д-сценария и коммуникативной схемы, подвергается фильтрации. В нашей модели фильтрации подвергаются (а) все высказывания из отдельного класса (например, все фразы, связанные с негативными д-сценариями, могут быть запрещены в этикетном диалоге) и (б) отдельные лексические единицы в составе высказываний – ср. *Куда ты побежал?* (допустимо) *Куда ты понёлся?* (стилистически маркировано и часто некорректно). Фильтрация может не только подавлять некорректные высказывания, но и повышать предпочтение этикетных ответов, именно такая стратегия ответа засвидетельствована в тестовых протоколах абитуриентов (с сравнении с группой «взрослых»).

Степень фильтрации может быть глобальной переменной (агент будет одинаково фильтровать свои высказывания в одной ситуации при обращении к разным собеседникам), или дифференцированной, когда агент будет «официально» говорить с одним присутствующим, и открыто – с другим. В этих случаях используется различный уровень *локальной фильтрации* по отношению к разным участникам ситуации.

Поведение агента во времени и выбор коммуникативных стратегий контролируется несколькими переменными. Среди этих переменных: (а) общее напряжение, сила возбуждения при эмоциональном событии и скорость угасания возбуждения, (б) предпочтение позитивных или негативных реакций (д-сценариев) – агент может быть «оптимистом» или «пессимистом», предпочтение «рационального» обсуждения ситуации (р-сценариев),

<sup>2</sup> Список д-сценариев представлен на странице: <http://www.harpia.ru/d-scripts.html>

(в) предпочтительная коммуникативная схема – агент может любить жаловаться или ругаться, (г) степень фильтрации исходящих высказываний – агент может быть «открытым» или «зажатым», (д) ориентация на себя или на адресата – агент может преимущественно говорить о себе или об адресате (см. Схему 1).

### Переключение в коммуникации



Рис. 3. Фильтрация подавляет некорректные высказывания, и они отображаются в виде «мыслей»

Благодаря фильтрации определённых классов высказываний в игровом интерфейсе можно разделить «слова» агента и его «мысли» (высказывания, появляющиеся в «облачке») – это позволяет, к примеру, создать «лицемерного» агента, который на словах ободряет адресата, но сам при этом расстраивается, или наоборот – успокаивает адресата, но сам при этом – злится на него (см. Рис. 3).

Различная степень фильтрации высказываний по отношению к разным участникам ситуации позволяет имитировать переключения в коммуникации, когда агент обращается то к одному, то к другому оппоненту. Если агент испытывает общее напряжение, но при этом из-за требований этикета (высокой локальной фильтрации) не может открыто выразить своё состояние адресату, он может выполнить переключение в коммуникации и обратиться с высказыванием к другому участнику ситуации, для которого уровень локальной фильтрации ниже (см. Рис. 4).

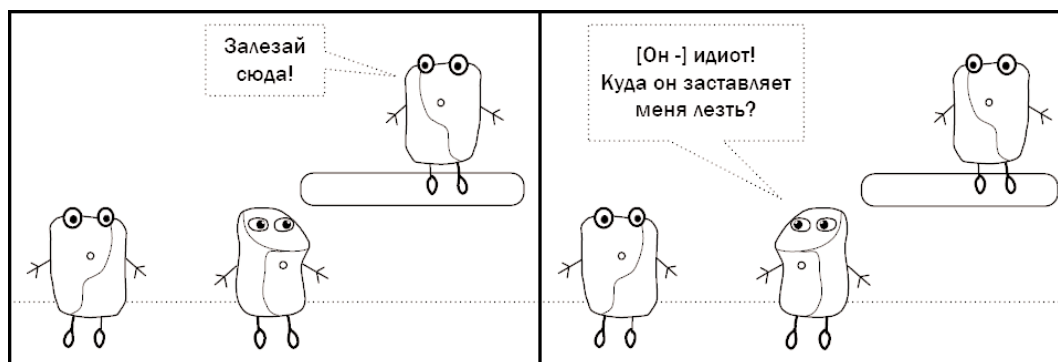


Рис. 4. Переключение в коммуникации: сильный уровень локальной фильтрации заставляет обращаться к другому агенту

### Микросостояния

Набор значений общих управляющих переменных, приведённых выше, может быть связан с некоторым коротким состоянием, например, агент может быть сильно возбуждён и на всех ругаться – это состояние будет характеризоваться (а) высоким возбуждением, (б) предпочтением негативных д-сценариев, (в) предпочтением коммуникативной схемы «конфликт» и (г) низкой фильтрацией. Такие краткосрочные состояния, выражающиеся в коммуникации, мы обозначаем термином *микросостояния*. В структуре модели микросостояние определяется как вектор значений управляющих переменных. Каждое микросостояние обладает некоторым уровнем активности – самое активное микросостояние контролирует текущий речевой выход агента. Если микросостояние выражается в речи, то его активизация последовательно снижается, и оно может уступить место другим микросостояниям. Конкуренция двух микросостояний (последовательное проявление то одного, то другого



## Управление динамикой речевого поведения виртуальных компьютерных агентов



Схема 1. Управление внутренними компонентами модели для имитации динамического речевого поведения

Обозначения:

- Управление
- ▷ Подразделение
- ➔ Передача данных

(1) Сегменты ситуации перебираются в определённой последовательности, что создаёт иллюзию «переключения внимания» агента – агент сначала говорит о частных («что со мной случилось?»), а затем – о более общих фрагментах ситуации («я доставил тебе неприятности»).

Микросостояния определяются как вектор переменных, контролирующих разные компоненты модели: (2) выбор эмоционального или рационального способа реакции, выбор оптимистичной или пессимистичной эмоциональной реакции, (3) распределение участников ситуации в коммуникативной схеме (выбор в ситуации «главного вредителя» или «объекта восхищения») и предпочтение определённой коммуникативной схемы (агент может быть склонен ругаться или жаловаться), и (4) степень глобальной фильтрации высказываний, а также степень локальной фильтрации по отношению к каждому из контрагентов. (5) Предпочтение определённого способа реагирования может влиять на выбор аспекта ситуации (сегмента) – агент может «обращать внимание» на свои действия или на действия других – например, агент может быть «злым», и во всех ситуациях искать свидетельство заговора.

микросостояния в речи) создаёт эффект эмоционального колебания, например, когда агент то хочет *всё бросить и уйти*, то хочет *проучить адресата*. При подобном колебании снижающаяся активизация микросостояний приводит к выбору всё более «нейтральной» лексики в монологе.

В обозначение микросостояния включаются основные управляющие переменные: указание на сегмент ситуации<sup>3</sup>, например, **s5-7** (сегменты 5 и/или 7), **dneg/dpos** — предпочтение негативных или позитивных д-сценариев, **al/ah** — низкое или высокое напряжение (активизация), **fl/fh** — низкая или высокая фильтрация выхода. Нерелевантные для каждого конкретного микросостояния параметры не указываются. Например:

**m\_dneg** — предпочтение негативных д-сценариев: агент эмоционален – ругается или печалится; степень фильтрации или сегменты ситуации не определены;

**m\_s4-5-7\_dneg** — агент расстроен/агрессивен и винит адресата (в сегментах 4, 5 и 7 адресат занимает валентность AGGR);

**m\_s4-5-7\_dneg\_al\_fl** — в эмоциональной ситуации агент может флегматично заметить адресату: *Послать бы тебя куда подальше!* В этом микросостоянии агент предпочитает негативные д-сценарии, в которых адресат

<sup>3</sup> Сегменты указывают на структуру предиката, который выделяются в текущей ситуации, например, сегмент 1: «я сделал P<sub>1</sub>», сегмент 3: «я сделал P<sub>1</sub>, что затронуло тебя», сегмент 7: «ты сделал P<sub>2</sub>, что затронуло меня». Почти в любой ситуации можно выделить любой сегмент, таким образом, предпочтение сегментов будет в значительной степени зависеть от внутреннего состояния агента, а не от структуры наблюдаемой ситуации.

занимает валентность AGGR, при этом у этих сценариев низкая активизация и низкая фильтрация — мы не сильно переживаем и открыто выражаем наши не слишком сильные эмоции в речи. Данное микросостояние было выделено на основе наблюдений ответов группы практикующих врачей: в эмоциональной ситуации (Вы разби́ли чужую вазу, женщина-адресат Вас в этом укоряет) более 27% ответов на вопрос *Что бы вы сделали, если бы Вам не нужно было себя сдерживать?* в этой экспериментальной группе имели вид *Послал бы её куда подальше!* (или аналогичные формулировки, относящиеся к тому же классу). В других группах доля ответов этого типа составляла от 2,75% до 6,6%. Из других данных протоколов также следовало, что практикующие врачи демонстрируют достаточно низкий уровень эмоциональной активизации и фильтрации исходящих ответов. Таким образом, эмоциональные высказывания в протоколах врачей вызваны не сильными переживаниями, а низким уровнем фильтрации.

В целом, наблюдение за статистическими различиями тестовых групп позволяет выделить некоторые особенности речевого реагирования. Эти особенности являются неизменными характеристиками речевого реагирования отдельной группы испытуемых, но в нашем случае эта же характеристика может быть преобразована в отдельное микросостояние, которое будет определять специфику поведения агента в течение достаточно короткого времени (2-10 фраз).

Микросостояния могут сменять друг друга, формируя стандартные последовательности. Например, агент может расстроиться и ругаться, но потом — успокоиться и сказать *Ничего страшного* или постыдиться за своё поведение — *Я вёл себя ужасно! Я доставил вам неприятности*. Смена микросостояний происходит за счёт применения ряда правил, микросостояния могут: (а) возбуждать друг друга, (б) подавлять друг друга, (в) одно микросостояние может истощаться и давать место другому микросостоянию.

### *Экспериментальные исследования*

Исследование статистических различий между разными группами испытуемых составляет один из способов выделения и оценки микросостояний. Другой метод — это экспериментальные исследования на группировку эмоциональных высказываний или создание «сценария» из заданного набора эмоциональных фраз.

Мы провели два эксперимента, для которых выбрали по 2-3 характерных примера для каждого из выделенных нами классов высказываний — всего 115 высказываний. Участники каждого эксперимента получали эти высказывания в виде набора карточек.

В первом эксперименте участники должны были разделить высказывания на группы (эксперимент проведен А. Майгуровой под руководством Ю.Е. Кравченко). Для наших целей важны такие протоколы испытуемых, где группировка высказываний осуществляется не по предметному сходству, а по подобию эмоциональных переживаний, связанных с конкретными высказываниями. В этих случаях сходство между протоколами позволяет выделить «центральные» или «ядерные» высказывания для каждой группы, а различия в протоколах позволяют определить «периферийные» или «многозначные» высказывания, которые жёстко не связываются с одним микросостоянием, или которые могут выражать различные микросостояния (например, могут употребляться серьёзно или иронично).

Задача испытуемых во втором эксперименте состояла в том, чтобы сложить из карточек монолог героя мультфильма, который пытается запрыгнуть на верхнюю платформу (Рис.1). Для каждого высказывания испытуемые могли использовать ремарки «подумал» или «сказал», кроме того, испытуемые могли вставлять в свой протокол события «прыгает» (герой делает ещё одну попытку запрыгнуть наверх) или «пауза».

Испытуемые могли написать на протоколе свой адрес электронной почты или оставить протокол анонимным — специальная инструкция разъясняла, что исследование личностных характеристик испытуемых не входит в задачи исследования. В эксперименте приняли участие 30 студентов, 3 протокола были анонимными, на основании остальных протоколов были сделаны мультфильмы, в которых герои разыгрывали представленную испытуемым последовательность фраз и действий. Каждый мультфильм был отослан автору соответствующего протокола для проверки — при этом мы хотели убедиться, что авторы считают свой сценарий адекватным не только на бумаге, но и в форме мультфильма, то есть в ситуации, более приближенной к реальному поведению.

Мы исходили из того, что стандартные группировки и повторяющиеся последовательности фраз в протоколах будут указывать на микросостояния (несколько высказываний, часто упоминаемых вместе, соответствуют одному «внутреннему состоянию») или на регулярные смены микросостояний (близость в тексте фраз, указывающих на «грусть» и «воодушевление» может указывать на то, что «воодушевление» стандартно сменяет выражение в тексте негативных переживаний). Анализ результатов с помощью кластерного анализа позволил в целом подтвердить ранее выделенные микросостояния, а также уточнить степень членства конкретного высказывания в определённом микросостоянии.

В целом, проективные и творческие экспериментальные исследования (классификации высказываний или

---

*Управление динамикой речевого поведения виртуальных компьютерных агентов*

---

создание сценариев), могут служить дополнительным инструментом для выделения новых микросостояний или их стандартных последовательностей. Кроме этого, как можно предположить, не менее эффективными инструментами могут стать наблюдение за эмоциональной речью и стратегиями поведения в реальных ситуациях, а также анализ стандартных выразительных средств для передачи эмоционального состояния героя в мультипликации и художественной литературе. Эти методы могут позволить создать агентов, демонстрирующих гибкое и правдоподобное эмоциональное речевое поведение при взаимодействии с человеком.

**Список литературы**

1. Allen S. R. Concern Processing in Autonomous Agents, Ph.D thesis.- University of Birmingham, 2001.
2. Andre E., Dybkjær L., Minker W., и др. (Eds.). Affective Dialogue Systems.- Berlin: Springer-Verlag.-2004.
3. Becker C., Kopp S., Wachsmuth I. Simulating the Emotion Dynamics of a Multimodal Conversational Agent // Affective Dialogue Systems / Ред. E. Andre, L. Dybkjær, W. Minker, P. Heisterkamp.- Berlin, Heidelberg, New York: Springer, 2004.- С. 154-165.
4. Cassel J., Sullivan J., Prevost S., и др. (Eds.). Embodied Conversational Agents.- Cambridge, London: MIT Press.- 2000.
5. Hall L., Woods S., Aylett R., Newall L., Paiva A. Achieving empathic engagement through affective interaction with synthetic characters // J. Tao, T. Tan, and R.W. Picard (Eds.) ACII 2005, LNCS 3784.- Berlin, Heidelberg: Springer-Verlag, 2005.- 731-738.
6. Kotov A. A. Application of Psychological Characteristics to D-Script Model for Emotional Speech Processing // J. Tao, T. Tan, and R.W. Picard (Eds.) ACII 2005, LNCS 3784.- Berlin, Heidelberg: Springer-Verlag, 2005.- С. 294-302.
7. Paiva A., Chaves R., Piedade M., и др. SenToy: a tangible interface to control the emotions of a synthetic character // Proceedings of the second international joint conference on Autonomous agents and multiagent systems.- NY: ACM Press, 2003.- С. 1088 - 1089.
8. Paiva A., Prada R., Picard R. W. (Eds.). Affective Computing and Intelligent Interaction. Lecture Notes in Computer Science.- Berlin, Heidelberg: Springer.-2007.
9. Picard R. Affective Computing.- Cambridge, London: MIT Press, 2000.
10. Sloman A. Beyond Shallow Models of Emotion // Cognitive Processing.- 2001.- 2, № 1.- С. 177-198.
11. Tao J., Tan T., Picard R. (Eds.). Affective Computing and Intelligent Interaction.- Berlin, Heidelberg, New York: Springer.-2005.
12. Котов А. А. Модель эмоционального речевого поведения для виртуального агента ролевой компьютерной игры // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2006».- М: РГГУ, 2006.- С. 285-289.