

О роли аппарата эмоций как системообразующего фактора в адаптивных системах управления *

А.А. Жданов

Аннотация. Обосновывается роль аппарата эмоций организма, как основного системообразующего фактора, определяющего строение и функционирование нервной системы и ее имитационных моделей. Показывается, что такой аппарат является: источником внутренней активности, целевой функцией управления, средством качественного оценивания общего текущего состояния и отдельных элементов знаний – образов и действий, внутренними часами системы, а также каналом передачи оценок качества информации при общении. Предлагается математический формализм основных функций аппарата эмоций.

В работах П.К. Анохина по теории функциональных систем было введено понятие *системообразующего фактора (результата системы)*, под которым понимался полезный приспособительный эффект в соотношении «организм – среда», достигаемый при реализации системы [1]. На роль такого в функциональной системе предлагался «полезный приспособительный результат» - будущее, ожидаемое состояние системы.

Нам представляется, что такое понимание системообразующего фактора является лишь частичным. Действительно, остается непонятным: а) что ставит системе цели, б) какие цели могут быть поставлены перед системой, в) почему система должна прилагать усилия к достижению этих целей, г) для достижения каких именно целей приспособлена нервная система организма? Мы предлагаем расширить и конкретизировать представление о системообразующем факторе, организующем и направляющем функционирование такой системы, как организм животного, и на эту роль предлагаем его аппарат эмоций [2]. Рассмотрим основные функции и возможности аппарата эмоций.

1. Эмоции – источник внутренней активности системы

Когда создается любая машина, всегда предусматривается источник ее движущей энергии, должный приводить в движение части этой системы для достижения цели, поставленной перед нею. Выбор источника энергии для построения конкретной машины, зависит от сторонних обстоятельств.

Источник энергии должен использовать некоторый подходящий естественный процесс, протекающий в определенном направлении. В машине создается механизм преобразования энергии в требуемую от данной машины работу, а также механизм контроля и регулирования расходования энергии. Так, естественное «стремление» электронов от отрицательного к положительному полюсу источника электрического напряжения используется как движущая сила для всех электрических приборов. Следует различать цели, которые ставил перед машиной ее создатель, и «внутренние» целевые функции, непосредственно приводящие машину в действие. Так часы идут не потому, чтобы давать информацию о времени, а потому, что их шестерни подталкивает пружина. Отключим источник энергии, и функционирование любой машины останавливается, так как отключается его основная непосредственная «мотивация».

В живой организм его «создателем» также должен был быть введен некий источник «энергии» внутренней активности. Это должна быть универсальная «движущая сила», заставляющая активно действовать любой организм, независимо от его сложности, возраста и интеллекта. Здесь также следует различать две цели: цель «создателя» организма, и целевую функцию, «защитную» в конструкции организма.

Мы не знаем, кто или что в Природе создавало живые организмы, был ли это творческий целенаправленный акт, либо результат самоорганизации материи. Мы не знаем и истинных целей феномена жизни, а можем только строить гипотезы, предполагая, например, что целью организмов является сбор и накопление информации о мире, или эволюционный поиск удачных форм жизни, или некоторое преобразование мира. Однако в иерархической системе целей обязательно должны присутствовать две: а) выживание организма, как целостной системы, способной длительное время сохранять в биологическом мозге накопленные знания и способствовать их логически непрерывному развитию, и б) последовательное накопление знаний, необходимых для выживания организма. Выживанию способствует прогнозирование будущих событий, основанное на накопленных эмпирических знаниях, и совершение упреждающих действий, обеспечивающих переход в благоприятные для выживания состояния. Обе эти цели требуют активного поискового характера поведения. Мы видим в аппарате эмоций следующий механизм, задающий активное поведение.

В аппарате эмоций, прежде всего, имеет место множество объектов, раздражение которых сопровождается приятными, либо неприятными для организма ощущениями, включая самые предельные их степени (Рис. 1). Это множество ощущений и их носителей-объектов упорядочено по степени этого качества – «приятности» ощущения. Пусть в каждый момент времени выделен один и только один из этих объектов, который назовем *эмоциональной оценкой текущего состояния системы*, и обозначим как $S(t)$. Соответственно, эта оценка может принимать одно из следующего множества возможных значений

$$S = \{s_1, s_2, \dots, s_z, \dots, s_z\},$$

* Работа выполнена при поддержке гранта РФФИ № 03-01-00323

где, например, $s_1 = \text{«}5\text{»}$, $s_2 = \text{«}4\text{»}$, ..., $s_z = \text{«}5\text{»}$. Значение $S(t) = s_z \in S$ функционально зависит от совокупности эмоциональных оценок $p_{z,j}$ множества распознанных в текущий момент образов O_j . Здесь возможные значения эмоциональных оценок образов также могут быть равны $p_{1,j} = \text{«}5\text{»}$, $p_{2,j} = \text{«}4\text{»}$, ..., $p_{z,j} = \text{«}5\text{»}$. В множестве образов, которые система уже научилась распознавать, что составляет ее «Память образов», распознанные в текущий момент образы пометим символом «1», т.е. если образ O_j распознан в момент t , то будем писать $O_j(t) = 1$, а если не распознан, то будем писать $O_j(t) = 0$. Эмоциональная оценка $p_{z,j}$ образа является атрибутом образа. Пусть функция, определяющая оценку $S(t)$, есть, например, просто средняя величина от эмоциональных оценок всех распознанных в текущий момент образов, округленная до ближайшего значения из множества S . Такой аппарат можно представить в виде своего рода термометра, показывающего среднюю текущую «эмоциональную температуру тела».

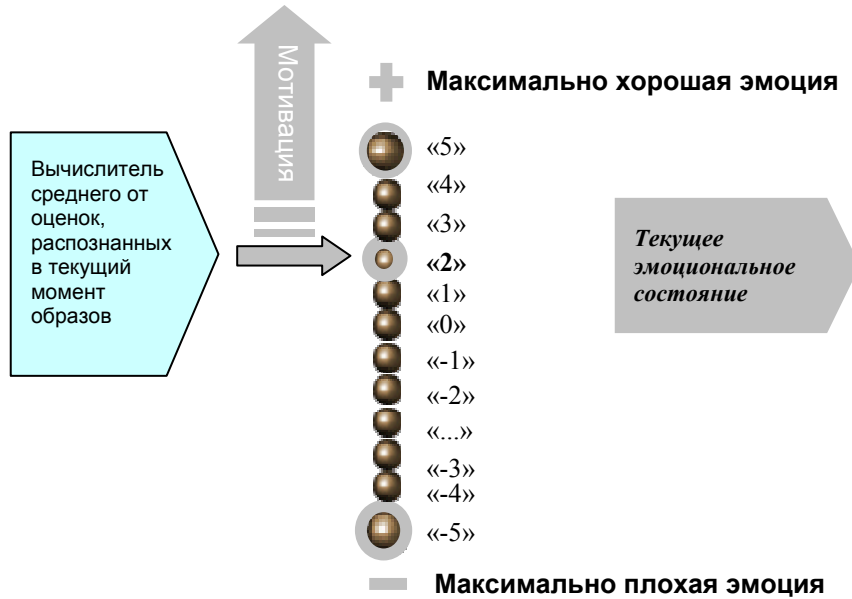


Рис. 1. Модель аппарата эмоций. Одна из его функций состоит в вычислении оценки текущего эмоционального состояния объекта – одной из множества возможных эмоциональных оценок. Другая его функция есть задание универсальной мотивации – постоянного стремления управляющей системы к повышению эмоционального состояния.

Жестко «зашьем» в управляющую систему целевую функцию максимизировать эмоциональную оценку текущего состояния системы $S(t)$. Другими словами, организм должен бояться и бежать неприятных эмоциональных оценок, и стремиться к приятным оценкам. Средством

достижения этой цели должно быть только инициирование актуаторов, которыми располагает организм. Эти актуаторы должны воздействовать не непосредственно на аппарат эмоций (в отличие от известной экспериментальной крысы, получившей возможность нажимать на педаль и возбуждать свой центр приятных ощущений через вживленный в него электрод, чем она и занималась до смерти), но на окружающую среду, обеспечивая возможные способы взаимодействия данного организма со средой. И только вызвав в среде появление таких объектов, процессов или явлений, воздействующих на сенсоры системы, которые будут распознаны этой же системой как образы, вместе с их атрибутами - определенными эмоциональными оценками, нервная система сможет тем самым повлиять на эмоциональную оценку текущего состояния. Тем самым, природа заставила организм, стремящийся лишь к получению приятных ощущений и к избежанию неприятных, взаимодействовать со средой, что приводит к накоплению новых знаний о свойствах системы «среда-организм», и повышает шансы организма на выживание. Если отключить аппарат эмоций, то организм перестанет что-либо делать, поскольку у него полностью пропадет мотивация к активности, и он погибнет в полном безразличии.

Очевидно, что природе следует позаботиться о том, чтобы нервной системе были приятны те воздействия среды, которые объективно полезны для организма, способствуют его выживанию и накоплению знаний, а вредные воздействия были бы неприятны. Эта задача решается в природе естественным отбором. Успешнее выживают организмы, у которых образы получают более адекватные оценки, а при неправильных оценках, организм стремится к вредным воздействиям и погибает.

Описанный механизм побуждения к активности через аппарат эмоций является универсальным и в целом не зависит от конкретного содержания образов. Обычно при рассмотрении функциональных систем говорят о наличии определенных «мотиваций», как о стремлении удовлетворить определенные «потребности», которые тут же перечисляются, например, потребность к сытости, к размножению и еще несколько других. Однако описанное нами стремление к получению положительных эмоциональных оценок и к избежанию отрицательных оценок является более общим. Так, неважно, какой образ вызвал резко отрицательную оценку – от срабатывания рецепторов в пустом желудке или от падения курса акций на финансовой бирже, нервная система будет стараться инициировать действия, направленные на вытеснение неприятных ей эмоциональных оценок, а единственным способом для нее является совершение воздействий на среду и такое изменение ее состояния, которое вызовет вытеснение неприятных образов.

2. Выработка эмоциональных оценок образов

Вначале следует сказать о понятии образа. Всякая нервная система получает информацию из окружающей ее среды через специализированные сенсоры – рецепторы, в виде стандартизованных бинарных сигналов, представленных

нервными импульсами (не будем здесь рассматривать восприятие информации нервной системой через другие возможные каналы – химические вещества или полевые воздействия). Практически вся поверхность тела, отделяющая его от окружающей среды, а также многие внутренние участки тела организма снабжены рецепторными клетками, суммарное число которых огромно. Нервная система не сможет справиться с таким объемом входных данных без их конвергенции, агрегирования и обобщения. Обобщение входной информации осуществляют нейроны, соединенные в конвергентные сети. При этом, сети выделяют также существенную для управления информацию, что осуществляется в процессе автоматической кластеризации сигналов (в поиске статистически зависимых, коррелирующих компонент) в закономерные паттерны – образы. Каждый нейрон, на наш взгляд, пропускает дальше по сети только ту информацию, которая для него не случайна, т.е. представляет собой паттерн – образ, который этот нейрон научился распознавать. При этом в отношении каждого отдельного образа можно говорить, что он либо еще *не сформирован*, либо уже *сформирован*, а об уже сформированном образе можно говорить, что он либо *распознан* в текущий момент, либо *не распознан*. Некоторые образы важнейших для организма данного биологического вида явлений и состояний являются сформированными изначально к моменту рождения (например, образ голода), или формируются по прецедентам (образ родителя для некоторых животных). Остальные образы формируются в процессе жизни. Итак, пусть нервная система оперирует информацией в виде образов. Конечными выходными сигналами нервной системы, как сети нейронов, являются сигналы, инициирующие актуаторы. Такие сигналы, являющиеся результатом решения, принятого нервной системой, будем называть *действиями*. Зависимость образов и действий, найденная нервной системой, есть ее *знания*, отражающие функциональные свойства системы «среда-организм». Однако эти знания нельзя использовать для управления без задания целей и критериев, т.е. без окрашивания данной информации в разные цвета качества, что и делает аппарат эмоций.

Пусть сформирован некоторый новый образ O_j . Эмоциональную оценку $p_{z,j}$ для этого образа нервная система может формировать в процессе одновременного наблюдения за эмоциональной оценкой текущего состояния системы $S(t)$ в моменты, когда данный образ $O_j(t)=1$ распознан или в некоторых окрестностях этих моментов времени. Если каждый раз при распознавании данного образа оценка $S(t)$ падает, это является основанием для приписывания этому образу низкой оценки, и, соответственно, наоборот. Конкретно, оценка $p_{z,j}$ образа O_j может быть равна среднему арифметическому от оценок $S(t)$ в моменты, когда данный образ распознан, т.е. когда $O_j(t)=1$. Практически, сразу после формирования образа его оценка нейтральна, но с накоплением статистики она может измениться.

Оценка каждого вновь сформированного образа в свою очередь начинает участвовать в выработке эмоциональной оценки текущего состояния системы $S(t)$, поэтому наше состояние начинает определяться и вкладом новых сформиро-

ванных нами образов. Этот итерационный процесс формирования оценок образов начинается с оценок образов, изначально сформированных в организме, т.е. переданных ему по наследству от предков данного вида, и приспособленных к усредненным условиям существования организмов данного вида.

Эмоциональные оценки образов, во всяком случае - ярко выраженные оценки, по-видимому, достаточно инерционны и долго хранятся в памяти нервной системы.

3. Выраженные эмоциональные оценки

Эмоциональные оценки как отдельных образов, так и общего текущего состояния носят сравнительно объективный характер и могут существенно отличаться для разных организмов даже одного вида. Такие оценки не слишком удобны для принятия решений, особенно при общении с себе подобными. Поэтому можно наблюдать существование дополнительной шкалы относительных оценок, используемых как для непосредственного управления, так и при общении между организмами. Назовем такого рода оценки *выраженными эмоциональными оценками* и будем обозначать такие оценки символом $B(t)$.

Формально упорядоченная шкала оценок $B(t)$ подобна шкале оценок $S(t)$, с той разницей, что она содержит меньше оценок, т.е. они более грубы, однако имеют (в первую очередь у людей) уже словесное выражение. Так, можно весьма правдоподобно определить эту шкалу следующим множеством оценок: $B = \{b_1, b_2, \dots, b_n, \dots, b_H\}$, где $b_1 =$ "невыносимо плохо", $b_2 =$ "очень плохо", $b_3 =$ "плохо", $b_4 =$ "так себе", $b_5 =$ "хорошо", $b_6 =$ "очень хорошо", $b_7 =$ "прекрасно", $b_8 =$ "недостижимо прекрасно". В каждый текущий момент t инициирована только одна из оценок $B(t) = b_n \in B$.

Существует некоторое пропорциональное отображение ζ из множества S в множество B , $\zeta: S \rightarrow B$, такое, что всегда текущая оценка $S(t)$ отображается в текущую оценку $B(t)$. Однако не все множество S может быть испытано нервной системой данного индивида к текущему моменту, но только оценки из диапазона $[S_{min}, S_{max}]$. Поэтому отображение ζ переводит S_{min} в b_1 , а S_{max} в b_H , что дает индивиду возможность пользоваться всем диапазоном оценок B . Такая шкала объясняет некоторые эффекты различного индивидуального восприятия явлений. Например, пусть индивид всю жизнь находился в очень благоприятных условиях и узнал только диапазон оценок $[S_{min1}, S_{max1}]$, а другой индивид жил в плохих условиях и узнал только оценки из диапазона $[S_{min2}, S_{max2}]$, пусть $S_{min2} < S_{min1}$ и $S_{max2} < S_{max1}$. Тогда один и тот же образ с оценкой $p_{z,j}$ может оцениваться первым индивидом, например, оценкой $b =$ "плохо", а вторым индивидом – оценкой "очень хорошо". Другой эффект – если у некоторого индивида вдруг упала оценка S_{min} , то благодаря отображению ζ произойдет переоценивание образов, и оценки всех образов автоматически возрастут, то, что оценивалось как плохое, станет хорошим, в том числе и оценка текущего состояния индивида.

4. Роль эмоциональных оценок в принятии управляющих решений

Алгоритм принятия решений, основанный на эмоциональных оценках, описан нами в [2-4]. Его идея состоит в следующем. В каждый текущий момент подсистема распознавания образов сообщает нервной системе, какие образы распознаны ею в текущий момент. Тем самым в базе знаний управляющей системы отделяются те ее области, которые адекватны данной ситуации, т.е. выделяется множество действий, которые вообще можно совершить в данной ситуации, согласно знаниям, накопленным нервной системой. В этих отделенных областях базы знаний нервной системой отыскиваются действия, которые обещают привести систему «среда-организм» в состояния, имеющие максимальные из возможных в данной ситуации эмоциональные оценки, либо в состояния, когда прекратят распознаваться образы с максимально плохими оценками. Найдя такое действие, нервная система дает команду исполнителям совершить его. Этот способ принятия решений одновременно активен и реактивен. Например, при отсутствии всякого внешнего стимула, животное может встать и отправиться на поиски чего-то приятного, воспоминание о чем хранится у него в памяти, а в текущей ситуации он не видит для этого препятствий. Если же на животное подействовал некий раздражающий фактор, оно начнет совершать действия, направленные на его прекращение.

5. Аппарат эмоций как внутренние системные часы нервной системы

При принятии решений в текущей ситуации нервной системе необходимо просмотреть свою базу знаний и найти оптимальное решение. Однако просмотр базы знаний требует определенного времени, которым нервная система может не располагать в текущих условиях. Определение времени, отведенного нервной системе на принятие решений в текущей ситуации, есть также функция аппарата эмоций. Внутреннее время нервной системы выражается в количестве $k(t)$ образов, которое можно успеть проанализировать в базе знаний при принятии решений в текущих условиях. Речь здесь идет об «образах результатов» совершения действий, т.е. о тех образах, распознавание которых управляющая система может вызвать своими действиями. Это число определяется некоторой функцией

$$k(t) = f(S(t), dS/dt)$$

от значений общей текущей эмоциональной оценки и ее производной (Рис. 2). Эта функция монотонно возрастает от некоторых предельно допустимых минимальных значений как $S_{min}(t)$, так и $(dS/dt)_{min}$. Т.е., чем хуже ситуация или чем быстрее она ухудшается, тем скорее надо принимать решение. При минимальных граничных значениях функция равна нулю, здесь нервная система «падает в обморок», так как не в состоянии успеть принять решение.

При больших значениях аргументов функция f имеет плато или спад, что индивидуально для индивида, и, возможно, в целом определяет тип его темперамента (сангвиник, холерик и т.д.).

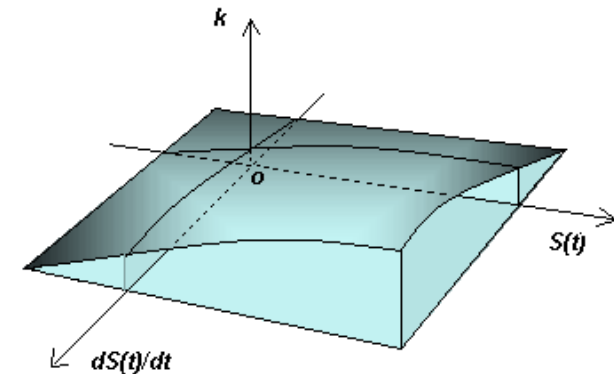


Рис. 2. Примерный вид зависимости времени, отведенного на принятие решений (глубины просмотра «Базы знаний»), от значения общего текущего эмоционального состояния и его производной.

Очевидно также, что те $k(t)$ образов, которые управляющая система может проанализировать при принятии решений в текущих условиях, должны быть наиболее важными образами, т.е. образами, с максимальными, по модулю, оценками. Множество образов всегда упорядочивается по убыванию модуля их эмоциональной оценки, и в каждый текущий момент отсекаются для принятия решений только $k(t)$ первых из них, и это число не очень велико. Из психологии известно «правило 7 +/- 2», отмечающее, что, например, человек может контролировать одновременно всего лишь около семи параметров. Можно предполагать, что живой организм, вынужденный, помимо всего прочего, поддерживать свой гомеостазис, в большинстве случаев борется именно с образами, имеющими большие отрицательные оценки, связанные с вызванными возмущающими факторами отклонениями от нормы основных параметров, определяющих нормальное функционирование организма. В этом отношении управляющая система проявляет себя как реактивная система, реагирующая на возмущающие факторы. Однако, хотя и сравнительно редко, но управляющая система распознает условия, благоприятные для получения некоторого большого выигрыша – возможности совершить некое действие, которое приведет к распознаванию образа с большой положительной оценкой. Тогда управляющая система может выбрать именно это действие, пренебрегая даже необходимостью поддержания гомеостазиса своего организма. Например, это проявляется в ситуациях, когда живое существо идет на некий опасный для него риск в надежде получить крупный выигрыш, и этот риск пропорционален

возможному выигрышу. В этих случаях управляющая нервная система проявляет себя не как реактивная, но как активная система, она совершает действия, не вызванные непосредственными воздействиями среды, но идущими от ее сложившейся информационной модели мира.

Итак, в каждый момент функция f аппарата эмоций подсказывает нервной системе, сколько образов с максимальными по модулю оценками она может успеть просмотреть в базе знаний с целью прогнозирования их поведения при альтернативных вариантах действий, и определяет лучшее из возможных действий с учетом последствий только по этим образам. Остальные образы, на учет которых при прогнозировании не хватило времени, получают в результате выбранного действия случайные изменения. Здесь следует обратить внимание на то, что «результат», который достигается вследствие совершения управляющей системой выбранного ею действия, не является детерминированным по нескольким причинам, а именно: а) потому что управляющая система еще не знает всех последствий того или иного своего действия, у нее еще не сформированы соответствующие образы, не хватает знаний; б) управляющая система просто не успевает в текущих условиях учесть даже те последствия (образы), которые она знает, не хватает времени; в) а также по другим причинам. Отсюда следует, в частности, что от величины выработанной управляющей системой эмоциональной оценки образа зависит, попадет ли этот образ в число тех, которые будут «приняты во внимание» при принятии решения, или вовсе не попадет. Например, у человека эмоциональные оценки образов закладываются во многом в процессе его обучения и воспитания, т.е. от системы привитых человеку эмоциональных оценок образов в корне зависит его поведение, приоритеты, которые он выбирает при принятии решений.

6. Аппарат эмоций, как канал передачи качественной информации

Наконец, следует упомянуть о важной роли участия аппарата эмоций в общении индивидов. При необходимости передачи знаний от одного индивида к другому, животному приходится передавать информацию как об образах некоторых объектов, так и о своих эмоциональных оценках этих образов. Поскольку даже в элементарном передаваемом знании должна содержаться информация и об образах условий, и об образах действий, и об образах ожидаемых результатов действия, то нужно передавать и сведения об эмоциональных оценках каждого из таких компонент. Каналом для передачи такой многообразной эмоциональной информации служит мимика животного, управляемая аппаратом эмоций, с помощью которой передается и распознается эмоциональная информация об условиях, действиях и об ожидаемых результатах действия. Здесь требуется обращение к специальным знаниям, чтобы понять, какие именно части тела, лица, и с помощью каких мимических выражений и жестов передают информацию о качественных оценках, являющихся актуальными в текущий момент. Очевидно, что чем развитее

управляющая система животного, чем больше образов формируется ею и принимается в расчет при принятии решений, чем более социально животное, тем более развитым должен быть этот «мимический» канал для передачи эмоциональной компоненты информации при общении между индивидами.

7. Использование аппарата эмоций в прикладных управляющих системах

Описанные здесь принципы работы аппарата эмоций частично использовались нами при построении прикладных систем на основе метода ААУ [4].

В системах адаптивного управления угловым движением космического аппарата [4,5,6], автомобильной подвески, мобильного робота [7,8], системы управления моделью сердечно-сосудистой системы [9] важную роль играла подсистема эмоций. Использовались эмоциональные оценки, описанные выше в пунктах (1), (2), (4). В дальнейшем планируется реализовать и остальные возможности и функции аппарата эмоций.

В заключение отметим, что общепринятое на сегодняшний день понятие, что эмоции животных определяются какой-либо актуальной потребностью и оценкой вероятности ее удовлетворения [10], нам представляются мало конструктивными.

Литература

1. Анохин П.К. Очерки по физиологии функциональных систем. – М.: Медицина, 1975. – С. 447.
2. Zhdanov A.A., A.N. Vinokurov, Emotions Simulation in Methodology of Autonomous Adaptive Control. 1999 -14th IEEE International Symposium on Intelligent Control /Intelligent Systems and Semiotics ISIC/ISAS'99. Special session Emotions and Intelligent Systems. September 15-17, 1999, Cambridge, Massachusetts, USA. Paper 99-0021-6.
3. Жданов А.А. Об одном имитационном подходе к адаптивному управлению. Сборник "Вопросы кибернетики" №2. Научный совет по комплексной проблеме "Кибернетика" РАН. М., 1996, С. 171- 206.
4. Жданов А. А., Метод автономного адаптивного управления // Известия Академии Наук. Теория и системы управления, 1999, № 5, с. 127-134.
5. Жданов А.А., Арсеньев С.В., Половников В.А. Об одной методологии автономного адаптивного управления. Труды института системного прог-раммирования РАН. 1999. Том 1. М.: Биоинформсервис, 2000.- С. 66-83. (англ. том. Zhdanov A.A., S.V. Arsenjev, V.A. Polovnikov, On autonomous adaptive control methodology.// Proceedings of the Russian Academy of Sciences Institute for System Programming. N 1, 1999, pp. 55-70).
6. Жданов А.А. Земских Л.В. Беляев Б.Б. Система стабилизации углового движения космического аппарата на основе нейноподобной системы автономного адаптивного управления. // Космические Исследования, М. 2004 (принята редакцией).

7. Жданов А.А., Крыжановский М.В. Преображенский Н.Б. Нейронная адаптивная система управления. Труды международной конференции "Интеллектуальные и многопроцессорные системы" IMS'2002., С. 115-118. Казивели.
8. Жданов А.А., Крыжановский М.В., Преображенский Н.Б.. Бионическая интеллектуальная автономная адаптивная система управления мобильным роботом // Мехатроника, 2004 (принята редакцией).
9. Zhdanov Alexander, Maxim Karavaev and Helen Maklakova, Claire Medigue, Michel Sorine. Simulation of control mechanisms in the cardio-vascular system. French-Russian A.M. Liapunov Institute for Applied Mathematics and Computer Science. Transactions. Vol. 4. Pp. 233-245. Moscow. 2003.
10. Психофизиология. Под. Ред Ю.И. Александрова. Изд-е 2-е. М., СПб и др.: «Питер», 2003, с. 267.