Chapter 2
The Modular Organization of the Mind

The human mind is what emerges from the actions of a number of largely independent cognitive modules integrated by a central control system. Figure 1.5 from the previous chapter showed the organization of some of these information-processing modules. The basic purpose of this chapter is to elaborate on this modular organization and the reasons for it. The first section will give a brief overview of the functional needs of the human mind and physical constraints of the human brain that force a modular solution. The second section will describe the modular structure in ACT-R and how that relates to issues about modularity more generally in cognitive science. The final three sections will illustrate this modular organization in three empirical studies – a study of driving behavior, a study of perfect time sharing, and an fMRI imaging study.

Function and Structure

Humans and other creatures live in a complex world where multiple simultaneous demands are placed on them. They have multiple resources to meet these demands, but there are also severe limitations on how these resources can be deployed. Roughly, these demands and resources can be divided into perceptual, motor, and central:

1. **Perceptual.** We need to be capable of detecting important information coming in via multiple sensory media. For instance, when driving we need to be able to process a car changing into our lane and to respond to a horn warning us of danger.

2. **Motor.** We need to be able to take appropriate actions, and we often have to perform multiple actions simultaneously, such as using our feet to locomote while we use our hands to manipulate.

3. **Central**. Our actions and our thoughts must be coordinated to achieve our needs. Many activities are more than just a sum of their parts -- preparing a successful meal, delivering a successful lecture, etc. Success depends on not only what actions are taken, but on the order in which they are performed.

With respect to perceptual and motor abilities, we are not much different than other primates, but we are quite different when it comes to central control. Our ability to organize novel combinations of behavior has given our species its unique ability to acquire high proficiency at a wide range of skills for which we were not specifically prepared by evolution. Examples of such skills include driving a car in modern society, navigating in traditional Polynesian society, doing mathematics in our scientific practice, and writing a research monograph. Each of these competences requires that we exercise an internal control over our behavior. The ability for such inner control is much more advanced in our species, a point that will be considered at length in Chapter 5.

Driving a car is probably an example that most readers of this book can relate to. Indeed, driving a car while holding a conversation is probably the paradigm case of multi-tasking in the modern world. Consider the demands being placed on a driver:

Perceptual. The visual system is being presented with a complex array of rapidly changing information that needs to be monitored for important events while still processing the visual cues that guide basic driving. The auditory system has to process the speech but also respond to critical sounds (e.g., a horn honking) from the outside.

Motor. The hands are occupied with steering and shifting; the feet with using the accelerator, break, and clutch; the vocal system with speech.

Central. Driving requires integrating low-level steering adjustments with high-level decisions such as how fast to progress and when to change lanes. In addition, the driver has to worry about avoiding traffic tickets and getting off the highway at the right exit. If the driver is engaged in a conversation, this task also has to be coordinated with the rest.[1]

Ignoring the demands of speech, the basic perceptual and motor demands of driving are something other primates are capable of. However, while chimpanzees might drive vehicles as part of a circus act, we would never let them on the road, because they are incapable of the central control that real driving requires.

[1] Modern cars with their radios, cell phones, temperature controls, GPS systems, and other devices give us many other things to engage our perceptual, motor, and central systems.

Structural Constraints of the Brain

These functional demands would not be nearly so interesting were it not for the fact that there are limitations on the brain's ability to achieve them. What is really interesting is not the mere fact of limitations, but the nature of these limitations. The standard contrasts between brain and computer are instructive here: the computer is fast and serial and the brain is slow and parallel.[2] Feldman and Ballard (1982) proposed a "100 step" constraint on action -- many things that we do (such as decisions made while driving) take place in well under a second, during which our neural system can progress through no more than 100 states. In that same time, modern computers can execute trillions of instructions. On the other hand, we have tens of billions of neurons performing these computations.

There are two basic forms of limitation on parallelism. One is a matter of number: While we have perhaps 100 billion neurons, we do not have 100 trillion. Each neuron comes with a cost in space and metabolic support, and evolution has found fit to pack only so many into the human brain. Thus, some computations are just not feasible, and it does not take long to discover these limitations as a driver of a car. The second limitation is one of communication: Only so many neurons can be packed into any small space, and the further apart these neurons are, the longer it takes them to communicate and the more structure has to be used to string the "wires" between the locations. Cherniak (1990) argues that simple calculations of space imply that most connections must be local. Therefore, the brain has opted for the sensible design of placing neurons performing related computations close together.

[2] This contrast can be overstated. Modern computers do many things in parallel and strive to achieve capacity increases by parallelism. Similarly, the brain is not above trying to achieve speed to increase processing power as witnessed by the major investment it makes in myelination to increase the speed of critical information transfer.

The visual system nicely illustrates the power and limitations of neural information processing. A large portion of the primate brain is given over to processing the visual array; over 30 such separate regions have been identified (Van Essen and DeYoe, 1995). At any point in time, the input is being processed in the brain from our entire visual field. However, there is also a strategic allocation of resources by our visual system. Our eyes have foveae; the information coming in from each fovea gets a disproportionate investment throughout the nervous system. We do not give the same investment to achieving high acuity throughout the visual field because it would be just too expensive – our brains would have to be an order of magnitude larger if they devoted as equal processing to all sections of the visual field. Every time we move our foveae to a new location, we are choosing to dedicate our most powerful processing resources to what we deem most important. The eyes of drivers are normally a dead giveaway as to what they are concerned with at any instant. For many purposes, it almost seems that the only thing we are aware of is what is close to our foveae. However, we only have to have an object fly at us from our periphery (we will duck) to be assured that some parts of our nervous system are still processing the rest of the visual field.

The visual system also illustrates the push to have related information processing close together. While visual information processing is occurring in many areas of the brain, each of the 30+ visual areas is specialized to process a different sort of information about the visual signal. Some regions do basic extraction of visual information, some are involved in object recognition (the ventral stream), and others are involved in extracting action affordances (the dorsal stream). Also, many of the early visual areas are topographically organized; cells that process similar areas of the visual field are placed close together.

The need for local, parallel computing in the brain is the fundamental structural fact driving the modular organization of the human mind. Essentially, parts of the brain are given over to performing related computations so that it is easier for the necessary neurons to interact. The extreme of localization is the cortical minicolumn -- the parcellation of the brain into small units of about 100 neurons that have a very restricted mission. For instance, cortical columns in the primary visual cortex are specialized to process information about one orientation, from one location, in one eye.

The need to do as much computing locally as possible is not enough by itself to require a modular organization. One could imagine a brain organization where function gradually shifted over the brain without any discrete clumping into modules. In part, this is the organization of the brain, It as manifested at a smaller scale in the topological organization of columns within areas such as primary visual cortex – nearby columns process nearby regions of the visual field. At a larger scale, different cortical areas that do similar processing are often adjacent, such as primary visual cortex and secondary visual cortex. However, there is also a clear clumping of areas in the cortex such that there are discrete changes in cytoarchitectural features defining distinct regions that seem to correspond to distinct functions. These regional differences are reinforced by the fact that different fiber tracts project to these regions. Maybe the emergence of modules in the mind is just a consequence of accidental variations in cortical anatomy. However, at the end of this section we will note Herb Simon’s argument that such hierarchical organizations, including those in the brain, are not accidents, but rather are requirements for functioning systems.

This claim that neural computation is localized goes against a long-standing tradition to regard information about a particular function as spread out over the entire brain. Lesion data going back to Lashley (1950) have been cited in favor of this conclusion – in one of his studies he found that he could remove any part of a rat's cortex and it could still run a maze that it had learned. Lashley concluded, " the memory trace is located in all parts of the functional area; that various parts are equipotential for its maintenance and activation." (p. 469). Brain imaging data have also been cited in favor of this equipotential viewpoint – in performing any task, numerous disparate regions of the brain show activity. However, interpreting such results as evidence for equipotentiality fails to appreciate the fact that different regions can perform different functions, all required for successful execution of the task. For instance, in the model for the algebra task (Figure 1.7), there were visual, procedural, declarative, goal, imaginal, and manual modules simultaneously active in different brain regions. Also, the modern interpretation of Lashley's results is that the rat had multiple representations of the maze (for instance, spatial and motor[3]) and that damage to any area of the cortex only impacted one representation.

An excellent example of the complexity of brain representation concerns the phenomenon of blindsight (Weiskrantz, 1986). Damage to the primary visual cortex leaves patients in a state where they cannot identify an object that is presented to them and they claim not to know it is there. However, they can still point to the object when asked. Thus, damage to a specific area results in a specific deficit, but similar information processing is occurring elsewhere and can serve for certain tasks.

[3] See discussion of place and response learning in Chapter 4.

Coordination: The Basal Ganglia

While different regions of the brain do their own processing, they have to at least sometimes coordinate to achieve a functional system. Consider working out a solution to an algebra problem. At the mechanical level, the eyes must move appropriately to guide the hand. Memory must be interrogated for arithmetic facts relevant to the numbers that are seen in the problem. If one tries a particular type of solution (e.g., factoring rather than the quadratic formula), control must be exercised to keep the actions moving towards that solution rather than the other. To achieve this coordination, tracks of brain fibers connect multiple cortical regions. Particularly important are paths of communication that go between cortical regions via subcortical regions. The connections through the basal ganglia have attracted a lot of attention from various researchers (see Figure 2.1).

The basal ganglia are a connected set of subcortical structures. Most of the cortex sends projections to the caudate and putamen which are collectively referred to as the striatum. Various researchers (e.g., Amos 2000; Frank, Loughry, and O'Reilly 2001; Houk and Wise, 1995; Wise, Murray, and Gerfen, 1996) have proposed that the striatum performs a pattern-recognition function – essentially recognizing patterns of activation distributed over the cortex. This portion of the basal ganglia projects to a number of small regions known collectively as the pallidum[4]. The projections to the pallidum are substantially inhibitory, and these regions in turn inhibit cells in the thalamus, which project to select actions in the frontal cortex. Graybiel and Kimura (1995) have suggested that this arrangement creates a "winner-lose-all" system such that active striatal projections strongly inhibit only the pallidum neurons representing the selected

[4] GPe and GPi in Figure 2.1. Loosely, one can also include SNr and STN from Figure 2.1. STN participates in a more complex loop than the one described here. The numbers given in Figure 2.1 are the approximate number of cells in the human brain. Note the great reduction in the number of cells in going from the striatum to the pallidum.

action, which then no longer inhibit the thalamus from producing the action. This is a mechanism by which the winning procedure (in ACT-R, a production) is determined. According to Middleton and Strick (2000), at least five non-motor regions of the frontal cortex receive projections from the thalamus and are controlled by this basal ganglia loop.[5] These regions play a major role in controlling behavior.

There are three key features in this characterization of the basal ganglia loop:

1. It allows information from disparate regions of the brain to converge in making a decision.
2. It requires a great compression of information from what is happening in these individual regions because the number of receiving neurons is so much smaller.
3. Processing that involves this multi-synapse loop is necessarily much slower than processing that can occur in a single brain region.

The existence of structures, such as the basal ganglia, that have these properties is almost a necessity given the need for coordination of information and the limitations on the human nervous system. While the basal ganglia have been targeted as a promising site to be performing this coordination function, there is no reason to suppose this is the only such system. Also, while the basal ganglia are a paradigm case of a brain region that coordinates communication among multiple cortical regions, one cortical region can communicate directly with another without any coordination with other sites. A good example is the frontal eye field, which plays an important role in voluntary eye movements. It is directly connected to posterior visual areas.

[5] These regions in turn have projections to posterior cortex and can influence processing there. There is also evidence of projections directly from the basal ganglia to the temporal lobe (Middleton and Strick, 1996).

Summary

In conclusion, to achieve the rapid processing required for functionality of the mind, different information-processing functions are computed as much as possible by different independent modules associated with different brain regions. However, the need for coordination requires communication among these modules. A particularly prominent sort of coordination is where multiple modules communicate with a single coordinating module, such as the procedural module associated with the basal ganglia.

While this section has been about the brain and how structural and functional considerations force it to a modular structure, it is worth stopping for a moment to recognize that this is an instance of a much more general phenomenon. Simon (1962) noted that nearly all complex systems whose design is driven to achieve a function have a hierarchical organization of nearly decomposable subsystems. In this he included such artifacts as books and computers, biological entities, social organizations, and cognitive activities including problem solving and language. He argues that a hierarchical structure facilitates the evolution and reproduction of such systems. Specifically, it is possible to tinker with one subsystem without disrupting another subsystem. In such hierarchies, the parts (modules) tend to occupy a small portion of contiguous space or time. This physical compactness promotes high interaction within a unit of the hierarchy, and the distances between the units results in much lower bandwidth interactions among the units.

Modular Architecture

ACT-R's Modules

The overall structure of ACT-R is illustrated in Figure 2.2. That figure is an elaboration of the earlier Figure 1.5 and illustrates the eight modules that are standard as part of the ACT-R 6.0 simulation system[6]. There are two perceptual modules: a visual module and an aural module. There are two response modules: a manual module and a vocal module. The other four modules are responsible for different aspects of central processing. There is an imaginal module that holds a current mental representation of the problem. For instance, in the context of solving an equation such as $3x - 7 = 5$, it might hold a representation of an intermediate equation such as $3x = 12$. There is a declarative module that retrieves critical information from memory such as that $7 + 5 = 12$. There is a goal module that keeps track of one's current intentions in solving the problem – for instance, one might intend to factor a quadratic equation. Finally, there is a procedural module that embodies various rules for behavior, such as the rules for solving equations. A later section of this chapter will review the current proposal for associating these modules with specific brain regions.

Each of these modules is capable of massively parallel computation to achieve its objectives. For instance, the visual module can process the entire visual field; and the declarative module can search through the large database of memories. However, when it comes to communication among the modules, there are serial bottlenecks. The only way these modules can communicate is through buffers associated with each module. Only a little information can be put into a buffer associated with the module – a single object perceived, a single problem state represented, a single control state maintained, a single fact retrieved, or a single program for hand movement.

[6] There is no implication that these are the only modules of the mind; these are just the ones currently implemented in ACT-R.

Each buffer can only hold a chunk (a structured unit bundling a small amount of information; see Figure 1.9 for an example of a chunk).

Communication among these modules is achieved via a procedural module (associated with the basal ganglia). The procedural module can respond to information in the buffers of other modules and put information into these buffers. The response tendencies of the central procedural module are represented by production rules such as the one illustrated in Figure 1.10. A significant architectural constraint in ACT-R is that only a single production rule can execute at a time. Moreover, it takes 50 ms. for a production rule to fire – which I think of as the time needed for the multi-synaptic loop through the basal ganglia to complete. Since communication among the modules must progress through the procedural module, it becomes the overall central bottleneck in information processing.

While we have extended the term "module" to the procedural system, it is worth noting that there are ways in which it is not like the other modules. In particular, it does not have a buffer associated with it in which it can deposit structures. It really is just a system of mapping cortical buffers to other cortical buffers and is not really an object in itself. Related to this difference is that we have associated the procedural module with the basal ganglia, which are not cortical structures.

It is interesting to consider what about the architecture in Figure 2.2 might be uniquely human. Elsewhere (Anderson, 2005a) I have argued that the goal module has unique properties in the human that enable humans to achieve a distance from their immediate circumstances that other primates cannot. This enables human means-ends analysis, as described in this quote from Newell and Simon motivating their early General Problem Solver (GPS) theory:

> “I want to take my son to nursery school. What’s the difference between what I have and what I want? One of distance. What changes distance? My automobile. My automobile won’t work. What is needed to make it work? A new battery. What has new batteries? An auto repair shop. I want the repair shop to put in a new battery; but the shop doesn’t know I need one. What is the difficulty? One of communication. What allows communication? A telephone . . . and so on. This kind of analysis—classifying things in terms of the functions they serve and oscillating among ends, functions required, and means that perform them—forms the basic system of GPS” (Newell and Simon, 1972; p. 416).

The key to means-ends problem solving is the ability to disengage from what one wants (the end) to focus on something else (the means). Papineau (2001) has argued that means-ends analysis is a unique human capability. Means-ends reasoning underlies human tool making. Benjamin Franklin claimed that tool making was the distinguishing human trait. While there have been modest demonstrations of tool making in other primates, Franklin was right that this is a capacity qualitatively different in the human species. The goal module is what enables this capacity. The fifth chapter of this book will delve more into what underlies the remarkable cognitive plasticity of the human species.

Before progressing to examples that illustrate the module system, it is useful to consider the relationship of this proposal to other ideas in cognitive science. The ACT-R architecture can be viewed as a summary of an emerging consensus in the field. However, there still are significant controversies entangled in the consensus; these tend to obscure that consensus. I will consider two issues of controversy that are particularly relevant to the module system. One concerns the ideas set forth by the philosopher Jerry Fodor (1983) and the lively discussion that has followed. The other concerns long-standing and evolving issues in experimental psychology regarding capacity limits on cognitive processing.

Fodor Modules

Fodor proposed that a fragment of human cognition was achieved by what he called modules. He thought a modular structure best characterized certain input systems such as vision. He listed no less than nine properties he associated with input modules, six of which are reviewed below:

1. Domain Specificity: Fodor argued that a module processes only a restricted set of stimuli. This is transparently true for ACT-R's visual and aural modules, but formally it is equally true for all the modules. Just as the visual module only processes visual input, the declarative module only process memories and the goal module only process control states.

2. Mandatory Operation: Fodor thought that when certain input arrived, the modules had to act, and how they acted could not be modified. Thus, we cannot help seeing and hearing the world in a certain way. This is equally true of all modules: memory cannot help how it responds to a retrieval request, nor can the goal module help how it responds to request for a control change, nor can the procedural module help how it responds to a certain pattern in the buffers of other modules, etc. All this comes down to saying that these are mechanical systems that function according to specific laws (the next two chapters will elaborate on the laws describing the declarative and procedural systems). However, as part of Fodor's claim, he held that these modules were not affected by the system's beliefs. Thus, for instance, we see a visual illusion even if we know that it is an illusion. The degree to which there is top-down influence on perception is hotly debated, and there are many demonstrations of contextual effects on perception. However, it should be noted that nothing in the ACT-R architecture prohibits such influence. The input to a module can include higher-level beliefs as well as sensory information.

3. Information Encapsulation. Fodor emphasized that the information that modules process is internal, and that they do not need to make requests of other systems for information. This is again something largely true about ACT-R modules. Almost all information processing is within module, although modules do offer a narrow band (their buffers) for trading information with other modules. Thus, we have "near-encapsulation" reflecting the predominance of short- over long-distance connections in the brain.

4. Fast Operation: Fodor argued that, as a consequence of information encapsulations, modular processes are the fastest cognitive processes. This is again true of ACT-R modules. However, it does take some time for the modules to do their thing, and the overall speed of cognition will be determined by these module times (e.g., see Figure 2.8).

5. Shallow Outputs. Fodor described the outputs of modules as "shallow." Here he was very much influenced by his focus on input systems. He viewed these systems as just reporting the "facts" – for example, "there is a red square" rather than "there is a position on a checkerboard."

While not all of the ACT-R modules report simple perceptual results, the restriction of buffer contents to single chunks makes the output of the modules very limited.

6. Fixed Neural Architecture. Fodor argued that there are dedicated neural structures associated with these modules, again an attribute of ACT-R modules. This tends to come with strong nativist claims, such as that the basic functioning of the architecture is prespecified and not something that can be learned. At some level, ACT-R agrees that the functioning of a module is not learned, but certainly the contents of modules – such as what is in the declarative module – are influenced by experience. What are prespecified in the declarative module are the principles for encoding and retrieving experiences, not the experiences themselves. Given evidence about widespread neural plasticity, all modules are similarly capable of adjusting their behavior with experience, even if they cannot change their basic principles of processing. However, this book will only consider learning processes in the procedural and declarative modules. Learning in other modules, such as perceptual learning in the visual system, tends to be a much slower process and has not yet been modeled within ACT-R.[7]

Each of the claims by Fodor could be questioned, even with respect to input modules. However, as they are stated above, they do not seem much like the stuff of controversy. They come close to summarize emerging consensus. Nonetheless, Fodor's claims have been associated with substantial controversy because of further claims that he and others have made about modules. It is worth reviewing three of these extensions:

[7] However, we have begun some promising interactions with the Leabra (O'Reilly and Munakata, 2000) research group to enhance the ACT-R visual module with the kinds of slow learning that characterize their visual modules – e.g., Taatgen, Juvina, Herd, and Jilk (2006).

1. Language. Fodor proposed that a dedicated input module processes the syntax of language. This proposal, combined with claims about information encapsulation and the innate basis of syntax, has generated considerable debate in cognitive science. While there is not a linguistic module among the eight modules in Figure 2.2, ACT-R is quite agnostic on the issue of whether there are special language modules, or what they might do.[8] The modules in Figure 2.2 are by no means a complete account of the human mind, and it remains to be determined what is missing.

2. Content-Specialized Modules. Others have proposed modules with rather specialized content, such as for primitive numeric judgment (Dehaene, Spelke, Stanescu, Rinel, and Tsivkin, 1999), recognition of faces (Kanwisher, McDermott, and Chun, 1997), and detection of cheaters in social situations (Cosmides and Tooby, 2000). This is sometimes called the "Swiss Army Knife" model of cognition, in which there is a blade (module) for every purpose (Duchaine, Cosmides, and Tooby, 2001). Fodor does not particularly endorse these sorts of modules, and he is especially dismissive of the proposal for a cheater detection module (Fodor, 2000). In any case, ACT-R is agnostic about such proposals, just as it is about proposals for language modules.

3. Central Cognition. Fodor seems to want to restrict such modules to input (and perhaps output) systems. He does not think there are central modules. For instance, he says there appears to be no brain center associated with modus ponens (Fodor 2000, pp. 60-62). Fodor's claims about central cognition extend beyond just denying that there are central modules; he argues that central cognition cannot be understood computationally. It is here that ACT-R and

[8] I have not always been so agnostic. Anderson (1983) argued that language processing did not involve special modules and depended instead on modules used in general cognitive processes. I was also dismissive of evidence involving localization of language.

Fodor part ways. Evidence for central modules and a computational explanation of higher-level cognition comes from the whole body of work that has been done by the ACT-R community.

Consider Fodor's remark about no central module for modus ponens. This reflects his predisposition to see central cognition as logical processing. This is not the conception of cognition in ACT-R, nor indeed in any of the cognitive science tradition emanating from Newell and Simon (1972). They treated reasoning as a special case of problem solving and were already beginning to treat problem solving as handled within a production system. Interestingly, while Fodor is right about there being no brain center for modus ponens, the basal ganglia do appear to implement a production system, and production rules provide much of the power of modus ponens.

Fodor's reason for doubting that cognition can be modeled computationally comes from his concern with the frame problem. The frame problem was started in artificial intelligence as a technical concern with how to update knowledge in logical systems (McCarthy and Hayes, 1969), and workable solutions have been developed.   However, philosophers such as Fodor have focused on bigger epistemological issues that are only somewhat related. In Fodor's mind, the real issue is information encapsulation, and he argues that the knowledge that humans bring to bear on a task cannot be bounded. The most important intellectual discoveries require bringing together disparate pieces of knowledge.   He thinks that this exceeds the capacity of any computational system, but he does not really specify any specific case of knowledge integration that is beyond a computational system. Fodor talks about analogy generally as being beyond the bounds of computational systems, and yet there are successful computational models of analogy making (e.g., Gentner, Holyoak, and Kokinov, 2001, Salvucci and Anderson, 2001). The last task described in Chapter 5 was deliberately selected because it required people to put knowledge together in novel ways. I hoped this would bring ACT-R face-to-face with Fodor's problem. While it did pose some challenges to the architecture, it turned out to be quite amenable to computational modeling that was faithful to human behavior. Moreover, it was quite capable of being modeled within a modular architecture. In summary, Fodor's worries seem not to have been realized in a documented instance of human cognition.

Modules and Capacity Limits

The basic motivation for a modular structure is to get the best performance possible given the limitations of brain processing. Experimental psychology has long been concerned with sorting out the behavioral manifestations of the limitations on human information processing. In experimental psychology, this often takes the form of trying to identify which processes occur in parallel and which processes occur serially. While it is not logically necessary (Townsend and Wenger, 2004), the assumption typically is that parallel processes are not capacity limited and serial processes are. This is approximately the way it works out in ACT-R. Because different modules can function autonomously and in parallel, they allow the system to process multiple things in parallel. Because the computations within a module can also progress in parallel, they often can avoid capacity limitations, although within a module there is potential for interference among different simultaneous processes (the next chapter will discuss interference in memory). The hardest limitations occur in the communication between modules. Because so little can be placed in a buffer, it is hard to communicate information rapidly from one module to another.

For instance, the visual system can hold up processing as it attends to different objects, one at a time, serially putting them in its buffer until the desired object is found.

Another serial limitation arises from the necessity for all modules to communicate through the production system. Since the production system can execute only a single rule at a time, it becomes the central bottleneck (Pashler, 1998) in the overall processing. Therefore, cognition can be slowed when there are simultaneous, different demands for processing the information in the buffers of the modules. The idea that such a central bottleneck exists reflects another emerging near-consensus in cognitive psychology; ACT-R gives an architectural expression to this consensus[9].

However, the idea of a central bottleneck in information processing does come in for repeated challenges. Curiously, one of the prominent recent challenges comes from the Meyer and Kieras (1997) EPIC architecture-- curious because EPIC is a production-system architecture that has been influenced by the ACT architecture and in turn has strongly influenced the current ACT-R architecture in its modular design. Indeed, substantial aspects of the simulations of the ACT-R perceptual and motor modules are taken directly from EPIC. Meyer and Kieras also made clear to us that models of human cognition would never be adequate if they continued to focus solely on the ethereal intellect and did not acknowledge cognition's perceptual-motor grounding. The limitations of the perceptual-motor components have substantial impact on many higher-level cognitive processes.

[9] Interestingly, if we accept that this central bottleneck resides in the loop through the basal ganglia, we can even see in Figure 2.1 where this bottleneck becomes most narrow: The cells that make up the pallidum, which we associate with production rule selection, are really quite few.

One category of limitation that ACT-R takes from EPIC is that each of these peripheral modules suffers its own serial-like bottleneck. Different perceptual modules can attend to one thing at a time, and motor modules can program one thing at a time. This limitation is realized in ACT-R through the existence of limited-capacity buffers associated with the modules. These buffers are the means of communication among modules. The only thing that the central production system can detect is what is put into the buffers. Thus, while the visual system may be processing many things at once, the rest of the system can only respond to that little bit put into the visual buffer. Similarly, the non-peripheral modules can only communicate through small buffers: only a single thing can be retrieved from memory, a single situation imagined, etc.

The major point of disagreement with EPIC is whether the central production system also has a central bottleneck limitation. ACT-R's production system is limited in that only one production rule can fire at a time. In contrast, unbounded, many production rules can fire simultaneously in EPIC. Curiously, while the theories differ in terms of the number of rules that can fire at once, both theories agree that it takes 50 ms for a production rule to fire. Indeed, this seems an emerging point of consensus among many production system architectures (Anderson, John, Just, Carpenter, Kieras, and Meyer, 1995).

There are functional reasons for limiting production rule firing to a single rule at a time. This avoids the problem of multiple rules making contradictory demands on the same module. Thus, for instance, ACT-R does not have to worry that different production rules will fire that ask for contradictory changes to an imagined problem representation. In EPIC, where such contradictory rules can fire in parallel, explicit coordinating production rules are needed to avoid such conflicts or deal with them when they arise. While EPIC models come with special handcrafted control rules that seem to work, these rules are task-specific and would need to be learned for each new task. It is unclear what can guide the system to form the right control rules; certainly people do not get explicit instruction at this level of detail. No learning mechanism has been proposed in the EPIC framework. In contrast, as will be discussed in Chapter 4, production learning has been a recent success story in ACT-R. Production learning in ACT-R involves learning new productions from old; this is easier to do if the learning mechanisms do not have to deal with simultaneous productions firing.

While such functional issues are critical, most of the attention of the field has been on empirical evidence for a central bottleneck. This involves dual-task experiments where participants are asked to carry out two tasks in parallel. The second example in this chapter will involve such an experimental task. The first example, which comes next, will also involve dual tasking, but it is a task where the real emphasis is on functionality rather than the experimental details that separate parallel from serial production rule firing.

Driving: Modules in Action

Before getting into examples that involve detailed experimental analyses of the behavior of specific modules, it is important to start with an example that illustrates the functionality of the overall architecture. Driving is such an example; as already noted, it requires a driver to do many things at once. The most critical task is controlling the vehicle – exercising lateral control (steering) to keep the vehicle correctly in the lane and longitudinal control (acceleration and braking) to maintain a safe speed and distance from the car ahead. While it is most critical to monitor the vehicle immediately in front, a good driver should also monitor for other vehicles and objects, such as cars in an adjacent lane. Dario Salvucci (e.g., Salvucci, 2005; Salvucci, 2006; Salvucci, Boer, and Liu, 2001) has developed a driving model in ACT-R that incorporates the two subtasks of control and monitoring:

Control: The basic ideas for the vehicle control in Salvucci's model can be found in many mathematical models of driving (e.g., Donges, 1978). The input for control involves keeping track of two points – a near point directly in front of the vehicle that indicates where one is in the lane, and a far point (either the vanishing point of the road, the car ahead, or the tangent point of the road). The output is an adjustment to the lateral and longitudinal control parameters. The control model runs in a tight loop in which each of these points is noted and the control parameters of the driving are updated. At a minimum, this loop consists of three productions (each 50 msec long) – one production to note the near point, one to note the far point, and one to update the control parameters. This system requires the visual system to update the location of these points and the motor system to translate the control information into motor commands. Once given the control parameters, motor modules can make their adjustments autonomously of the central production system, and therefore the loop does not have to wait upon the completion of these motor actions. In summary, the control model calls upon the visual module, the procedural module, the manual module, and an implicit pedal module.

Monitoring: The monitoring model selects which lane to encode and whether to encode information in front or behind (in the rear mirror). Whenever it identifies a new vehicle, it notes its lane and position. This information is held internally and used to help guide such decisions as whether to change lane. The minimum loop for this is a production cycle of two rules: one chooses where to monitor; the second determines whether there is an object in that position. Thus, the monitoring model calls upon the visual module, the procedural module, and the declarative module.

One of Salvucci's contributions, over and above the driving model itself, is the proposal of a scheme for interleaving the two subtasks. After each iteration of the control cycle, the model determines how stable the driving situation is. If it is not stable, the control cycle repeats; if it is stable, the control cycle times out for roughly 500 msec (there is a noisy timing process) and monitoring takes over. Salvucci's model cannot wait much longer than 500 msec without suffering serious control problems. Such a system naturally devotes more attention to control in difficult driving situations (e.g., a lane change) and more attention to monitoring in easy situations (staying in a lane on a straight highway).

Salvucci is able to use distribution of eye movements to track this shift between the two tasks. His eye movement theory (EMMA -- Salvucci, 2001) assumes that the eyes follow shifts of attention in order to achieve higher resolution. However, eye movements are slow and stochastic and do not provide total tracking – for instance, there are not 50 msec saccades from near to far point for each control cycle. Still, the overall correspondence between eye movements predicted by his driving model and the data is quite impressive – Figure 2.3 shows the match between human and model proportions of gazes to different parts of the visual array.

The more telling analysis by Salvucci concerned the switch between control regions and monitoring regions. For this purpose, forward gazes to the same lane were classified as control gazes and all other gazes where classified as monitoring gazes. Figure 2.4 shows the probability of switching from monitoring to control (part a) and from control to monitoring (part b) as a function of time performing that activity (monitoring or control). As the model predicts, these two distributions are quite different. The switching from monitoring shows a peak at about .5 seconds, while the probability of switch from control is concentrated at short intervals. This difference is a consequence of the need to switch back to control after half a second. The length of time spent on control depends on road conditions and the stability of the vehicle.

One of the interesting features of driving is that we interleave it with many activities. When Salvucci's model turns to such interleaved activities as tuning a radio or dialing a cell phone, the demands of the secondary task largely push out situational monitoring, resulting in an alternation between the secondary task and the critical control task that guarantees the car stays on the road and avoids accidents. However, the switch of control between the two tasks is basically the same, driven by the same need to return to control at approximately half-second intervals to maintain safe driving. One obtains distributions of fixations between control and one of these secondary tasks much like the distribution in Figure 2.4 between control and monitoring.

Figure 2.5 comes from a study of cell-phone use. It compares the time to key each digit in a 10-digit number when not driving (baseline) versus when driving. There are long breaks before the beginning and before each group in the American 3-3-4 grouping of the telephone digits. In Salvucci's model, these are the points where the system is retrieving the groups of numbers. Salvucci's model will switch back to control while these retrievals are progressing. The retrievals may complete while the model still is in the control subtask, and then the model cannot get back immediately to keying. Thus, the effect of driving is to slow down the keying of digits at these points. The times to key the other digits are not affected.

Figure 2.6 shows the effect of dialing a cell phone number on two measures of safe driving, lateral deviation and speed deviation. As can be seen, the model correctly predicts that dialing a phone number has an effect on lateral deviation and that there is not an effect on speed deviation. Apparently, only lateral deviation is impacted by the relatively minor increase in delay in returning to control. The effects are quite modest. While the model overpredicts the effect of dialing on lateral deviation, all lateral deviation measures in Figure 2.6 are well within the safe zone. Examples like that in Figure 2.6 illustrate the relatively small cost that can occur when we insert a new task into the performance of a skill. This is possible because of the autonomous processing of the modules -- for instance, the keying of a chunk from the telephone number can continue while the participant has returned to monitoring and control.

In conclusion, Salvucci's driving model illustrates that real-world interleaving of task demands can be achieved within a modular structure such as that in ACT-R. The goal module handles the switches among the tasks. Salvucci has been concerned with extending the goal module so that it can handle arbitrary combinations of tasks.

The purpose of this driving example was to get the big picture across. There are lots of important details to driving, for which the reader is encouraged to go to the Salvucci sources. The next example bores down into some of the details of how individual steps of cognition interleave in a dual task, but in a much simpler task where these details will not be too many.

Dual Tasking: Modular Parallelism and Seriality

The second example goes to the Psychology laboratory to look in more detail at the kinds of temporal organization of the modules that underlay the driving example. There are two types of parallelism and two types of seriality associated with all of the ACT-R modules:

Within-Module Parallelism. Within each module, massively parallel computation is happening. For instance, the whole visual field is being simultaneously processed; a retrieval request involves a simultaneous search through multiple memories; the procedural component must simultaneously test all production rules looking for a match; motor programming requires the simultaneous execution and monitoring of multiple muscles, etc.

Within-Module Seriality. The need for communication and coordination poses serial bottlenecks within each module. For instance, a single visual object is attended, a single memory is retrieved, a single production rule is selected to fire, a single molar action is chosen to be performed. In the case of perceptual modules (and declarative memory is like a perceptual module that perceives the past), all the parallel computation must settle in a choice.

Between-Module Parallelism. Computation in one module can proceed in parallel with computation in another module. Thus, there is the potential for the parallel threads. In driving, for instance, vision is progressing in parallel with motor, which is progressing in parallel with central activities such as retrieval of a phone number.

Between-Module Seriality. However, in many cases one module must wait on another because it depends on the information from that module. Thus, for instance, we cannot dial a phone number before we retrieve it.

The ACT-R and EPIC conceptions of this situation are identical except for the central bottleneck. Because only one production can fire at a time (within-module seriality), communication among other modules can be held up (between-module seriality). ACT-R's position is more uniform in that it claims every module has a bottleneck. EPIC, on the other hand, claims no central bottleneck. This has put EPIC in opposition to central bottleneck theories such as that of Pashler and has created a fair amount of controversy and interest in the literature.

Much of the evidence for a central bottleneck involves studies of what is called the psychological refractory period (PRP—see Pashler, 1994, for a review), where one is asked to do two tasks. In the typical PRP experiment, a first task is presented and then, after a short delay but usually before the first task is finished, a second task is presented. The fact that the first task is still ongoing produces some delay in the performance of the second task. This has been taken as evidence for the existence of a central bottleneck. Meyer and Kieras (1997) argued that these effects arise because participants are asked to give the output of the two tasks in the order they occur and they thus have to put tests in the execution of the second task to assure that it comes out second. From this perspective, a better paradigm would be one where participants are asked to perform two simultaneous tasks as fast as they can with no constraint on order of response. Usually, the result in such experiments is considerable interference between the two tasks, but combinations of tasks can be found where, with enough practice, near perfect time-sharing occurs and the two tasks are performed together nearly as fast as alone. One such example of near perfect time-sharing was demonstrated in a series of experiments that began with Schumacher et al. (2001) and were continued by Hazeltine et al. (2002). Meyer et al. (2001) argued that these experiments provide evidence of an EPIC-like theory with unlimited central processing rather than an ACT-R-like theory with a central bottleneck.

In the basic experiment used in these studies, participants responded to the presentation of a circle and a tone. The circle appeared in one of three horizontal locations, and participants made a spatially compatible response with their right hand, pressing index, middle, or ring finger to left, middle, or right locations. The 150 ms tones were either 220 Hz, 880 Hz, or 3520 Hz, and participants responded “one,” “two,” or “three.” In the single-task condition, participants did just the visual-manual task or just the aural-vocal task. In the dual-task condition, both stimuli were presented simultaneously and participants were asked to do both tasks simultaneously. Over many days of practice, participants come to respond virtually as quickly to each task in the dual-task condition as in the single-task conditions. Thus, participants were able to perform two tasks at once with virtually no cost.

Anderson, Taatgen, and Byrne (2005) did an extensive analysis of the version of this paradigm that was reported by Hazeltine et al. Figure 2.7 displays the performance of the participants and the performance of an ACT-R model for the task. The figure shows the time to perform the aural-vocal task and the time to perform the visual-manual task, separately plots the time to do each task in isolation and in conjunction with the other task, amd plots data from three points in the task performance: the first two sessions, two sessions late in the experiment, and two later sessions where some additional transfer tasks were inserted. Even on the first two sessions, there is a relatively modest dual cost of about 50 msec, but this reduces to about 10 msec by the end of the experiment. The model reproduces the overall speed increase and the reduction in the dual-task cost from about 50 msec to about 10. Compared to the data, it produces a somewhat larger dual cost in the aural-vocal task and a smaller dual-task cost in the visual-manual task. However, for current purposes, the important observation is that the model, despite its serial production-rule firing, can produce small dual-task effects that are of the same order of magnitude as seen with human participants. Hazeltine et al. ran the same participants in Figure 2.7 through a series of additional experiments and eventually got the average dual cost effect down to 3 ms (by my calculations from their data). Anderson et al. ran their model to the limit and got its dual cost down to 4 ms.

The details of the model are described in Anderson, Taatgen, and Byrne (2005); Figure 2.8 illustrates the behavior of the model early in the experiment (part a) and late (part b). That figure tracks the activity of five modules: the vocal module generating speech, the aural module processing sound, the procedural module interpreting production rules, the visual module processing vision, and the motor module controlling hand movements. In both parts a and b, the visual and aural modules are evoked once to encode information and the vocal and manual modules are evoked once to generate output. The big difference between early and late is the number of rules that have to fire – three productions that initiate processing are collapsed into one; five rules that generate the finger press and assess the outcome are collapsed into one; two productions (plus some later assessment rules not shown) are collapsed into a single rule that generates the word.[10] The learning process that allows ACT-R to compress the initial productions into just three will be discussed in Chapter 4.

For current purposes, what is of interest is the mixture of parallelism and seriality. Let us review the four types of categories described above:

Within-Module Parallelism. Each box in these figures reflects a lot of parallel activity happening within the module.

Within-Module Seriality. However, each box reflects the conclusion of this parallelism in a single action. The within-module combination of parallelism and seriality also applies to the procedural system: All the productions are tested in parallel, but only one gets to fire, creating the central bottleneck in Part a of Figure 2.8.

[10] Actually, one production rule needs to be learned for each separate stimulus-response mapping. This means three visual-manual rules and three aural-vocal rules must be learned, but only one of each will apply on a particular trial.

Between-Module Parallelism. Different modules can operate in parallel. So, for instance, in part (b) the execution of the finger press overlaps at different times with the encoding of the tone, the production that selects the word to say, and the generation of the word.

Between-Module Seriality. What is determining the ultimate timing of the response, particularly in part b, is the communication of information between modules. For instance, the finger cannot be pressed until the rule selects which finger to press, and this rule cannot fire until the location is encoded.

Figure 2.8 illustrates why near-perfect time-sharing does not occur initially but is achieved ultimately. Initially, the demand on the central production system is high, and firing of productions for one task must wait on the firing of productions for the other task. In the illustration in Figure 2.8, the execution of the aural-vocal task waits on the execution of the visual-manual task, but it could have been the other way around. After extensive practice, however, there are very few productions and the rules for each task can "fit in" while non-central aspects of other tasks are performed. In Figure 2.8b, the rule that chooses the finger can fire during the encoding of the sound and the rule that maps the sound onto the word can fire while the key is being pressed.

In the example in Figure 2.8, it is particularly convenient that it takes longer to identify the tone than to identify a location, as this creates a gap for the visual-manual production rule to fire. Note in Figure 2.8 that the aural-vocal task takes substantially longer than the visual-manual task. Much of the research in the later Hazeltine et al. report was aimed at eliminating this difference, either by making the visual-manual task more difficult or by changing the onsets of the two tasks, and near-perfect time-sharing was still observed. Also, even in the original task, while the aural-vocal task was longer on average, on some trials participants did complete it before the visual-manual task. The model predicts this because of variability in the length of the steps. The length of the boxes in Figure 2.8 just represents their average length; the variability of stages can result in a partial overlap of times when productions could fire. This is why the model predicts some small dual-task cost even when the model achieves its ultimate compact form in Figure 2.8b. This residual dual cost reflects the average amount that one production in Figure 2.8b will delay another.

With highly practiced participants, Hazeltine et al. never found dual costs greater than about 10 ms. even when they tried to manipulate the length of the visual-manual task or the relative onset of the two tasks. Somewhat surprisingly, the model does not predict dual costs greater than an average of 10 msec. One would have thought these manipulations would have created a greater degree of overlap with the central bottleneck. However, the variability in timing that produces some overlap when the average times for the stages are completely non-overlapping in Figure 2.8b creates non-overlap when the average times are maximally overlapping. Also, the maximum delay that one task can produce in another is the 50 ms for the one production rule that must fire for that task. Since only one task can be delayed on a single trial, the maximum average delay in the performance of the two tasks is only 25 msec. Thus, it is not that hard to imagine how, with variability in timing and the slack time in Figure 2.8b, one can get a stubborn small delay (less than 10 msec) that does not seem to change much.

Delays can get much more substantial in situations such as that for the beginning of the experiment in Figure 2.8, where more central processing is going on. Byrne and Anderson (2001) studied a number of complex tasks (including doing addition and multiplication simultaneously) where the time to do two tasks at once was sometimes even greater than the sum of the times to do each singly. Such tasks provide strong evidence for the ACT-R conception of matters rather than the EPIC conception. Tests of a central bottleneck are much more telling when the amount of central processing becomes substantial.

However, it would be a mistake to leave this discussion focusing on the residual differences between the ACT-R and EPIC conceptions. In fact, the views are identical on most scores, and the ACT-R conception has been strongly influenced by the EPIC position. Moreover, the EPIC and the ACT-R position at a general level have substantial overlap with many other conceptions in the field, such as that of Pashler (1998) or Card, Moran, and Newell (1983). It is also worth noting that while productions can fire in parallel in EPIC, in many of the Meyer and Kieras PRP models they enforce seriality on production firing. Thus, there is an increasing consensus on how parallel processing and serial processing combine both within modules and between modules, even though we are still working out details such as the role of a central bottleneck.

Mapping Modules onto the Brain

Localizing Eight Modules

As discussed earlier, modular organization is the solution to a set of structural and functional constraints. The mind needs to achieve certain functions, and the brain must devote local regions to achieving these functions. This implies that if these modules reflect the correct division of the functions of the mind, it should be possible to find brain regions that reflect their activity. Our lab has developed a mapping of the eight modules in Figure 2.2 onto specific brain regions,illustrated in Figure 2.9.[11] We have used in an extensive series of fMRI imaging experiments. The eight regions can be organized into four peripheral modules and four central modules. The four peripheral modules are:

1. Visual module. While large portions of the brain are devoted to processing the visual signal, we have found one region, the fusiform gyrus in the temporal lobe, that seems to best reflect the focused visual processing of attended information. Other research (Grill-Spector et al, 2004; McCandliss et al, 2003) has shown that this plays a critical role in perceptual recognition.

2. Aural module. This is associated with secondary auditory cortex but not the primary auditory cortex. As in the case of the visual module, we are tapping a region that reflects relatively advanced processing of the auditory signal rather than early processing.

3. Manual module. This is reflected in the activity of the region along the central sulcus that is devoted to representation of the hand. This includes parts of both the motor and sensory cortex.

4. Vocal module. Further down the motor strip is a region that represents the face and tongue. It also includes parts of both the motor and sensory cortex.

[11] The regions appear on the surface in Figure 2.9a but are below the actual surface of the cortex to varying degrees. Figure 2.9b is a midline illustration of the two structures truly buried deep in Figure 2.9b. The coordinates given for the brain regions in this figure are slightly different than in Figure 1.8. This reflects a correction that has been made to deal with the fact that our reference brain was acquired very slightly misaligned with the AC-PC line.

The four central regions are widely distributed throughout the brain:

5. Imaginal module. We have associated the imaginal module with a posterior region of the parietal cortex. This association is roughly consistent with the research of others who have found that this area is involved spatial processing (Dehaene, Piazza, Pinel, and Cohen, 2002; Reichle, Carpenter, and Just, 2000). However, the exact functions of different parietal regions remain a matter for study. We have found this region to be sensitive to representational changes in tasks as varied as equation solving (Anderson, 2005) and Tower of Hanoi (Anderson, Albert, and Fincham, 2005). Its response seems largely insensitive to the input modality; it seems to instead reflect the effort made in transforming a mental representation.

6. Declarative Module. We have found a region of prefrontal cortex to be sensitive to both retrieval and storage operations. Focus on this area is again consistent with a great deal of memory research (Buckner, Kelley and Peterson, 1999; Cabeza, Dolcos, Graham, and Nyberg, 2002; Fletcher and Henson, 2001; Lepage, Ghaffar, Nyberg, and Tulving, 2000; Wagner, Maril, Bjork and Schacter, 2001; Wagner, Pare-Blagoev, Clark, and Poldrack, 2001). However, the exact memory function of different prefrontal regions again seems a matter for continuing study.

7. Goal Module. We have associated the goal module that directs the internal course of cognition with a region of the anterior cingulate cortex. There is consensus that this region plays a major role in control (Botvinick et al, 2000; D'Esposito et al, 1995; Posner and Dehaene, 1994), but again there is hardly consensus on how to characterize this role.

8. Procedural Module. As noted earlier, there is a general belief that the basal ganglia play a role like that of production rules in terms of pattern recognition and selection of cognitive actions. The region of basal ganglia that we have selected is the head of the caudate, although in some studies this region has not been particularly responsive.

As the citations above indicate, there is nothing particularly novel about the association of these brain regions with these functions. What is novel is the association of these regions with parts of an integrated architecture and their use to trace out the components of that architecture. It is important to recognize that we have used the exact same predefined regions across a number of studies, including the ones described in this book. This has a number of advantages over using exploratory regions. For instance, we avoid the problem of trying to correct for the danger of getting a spurious result in all the tests that go into exploratory studies. Furthermore, the estimate of the response we get in these regions is not biased by the selection process.

Two points need to be made to qualify any simple conclusion of a 1-1 mapping of function onto structure. First, as noted earlier, the brain tends to distribute similar but distinguishable processes in different regions. For instance, over thirty regions perform visual processing. Again, multiple regions in the frontal and temporal cortices serve memory functions. Thus, there is no claim that the one region we have identified is the only region associated with a function. Second, there is no necessary reason why these brain regions should perform only a single architectural function. Nonetheless, in the range of studies that we have used in our laboratory, we have been fortunate to be able to associate the activities of these regions with just the assigned functions.

The Experiment

Having postulated eight modules and their associated brain regions, it would be appropriate to describe a single study that exercised all of these modules. We performed such an fMRI study, the details of which are reported in Anderson et al. (in press). The experiment manipulated the input module by presenting material either visually or aurally. Similarly, it manipulated the output module by having the participants respond either vocally or manually.

A rather peculiar cognitive task was chosen in order to separate the behavior of the imaginal module (parietal) from the declarative module (prefrontal). Much imaging research finds that these two regions, although widely separated, often give similar responses. This happens because memory retrieval and representational changes are naturally correlated. In order to make a retrieval request, one needs to create a representation to hold the elements of the retrieval request. Following that, the consequence of a successful retrieval is often to change the underlying representation. Consider the task of solving the equation 7x+3=38, as described in the previous chapter with respect to Figure 1.7. Representation of this equation may lead to the request for the difference between 8 and 3. Successful retrieval of 8 - 3 = 5 enables the re-representation of the equation as 7x = 35, which in turn enables another retrieval request to determine the value of 35 divided by 7. Thus, retrieval and representation operations tend to occur together and we get similar behavior in prefrontal and parietal regions, as can be seen in a comparison of Figures 1.8b and 1.8d. To break this natural correlation, one needs an artificial task where successful retrievals will not necessarily result in re-representations so that re-presentations can take place without retrieving any information to guide them.

Table 2.1 illustrates how the experiment attempted to manipulate retrieval and representation demands orthogonally. In the first phase, outside the fMRI scanner, participants memorized information that they would use in the second phase of the experiment that took place in the scanner. The material to be memorized involved associations between two-letter words and two-digit numbers, such as:

AT → 23 and BE → 24

In the second phase of the experiment, participants either heard or saw permutations of the words "Dick," "Fred," and "Tom" paired with visual presentation of the two-letter words or two-digit numbers. Table 2.1 illustrates the various conditions of the experiment. Participants were told that the two-digit codes that they had learned were instructions for transforming the three-word sequences. Thus, 23 meant that the second and third words should be switched. Applied to "Tom, Dick, Fred," it would produce "Tom, Fred, Dick." Some two-digit codes were "no-ops" such as 24 because one of the digits is greater than three, and thus in this case does not require a transformation. The difference between no-op digit pairs and ones that require an operation is referred to as the **transformation** factor in Table 2.1. Participants can either be given the digit pair directly, in which case no retrieval is required, or be given a word from which they have to retrieve the digit pair. The requirement to perform this retrieval is referred to as the **substitution** factor in Table 2.1 because it required the participant to substitute the digit for the word. The expectation was that the transformation factor would draw more upon the parietal region for manipulating problem representation and that the substitution factor would draw more upon the prefrontal region for retrieving information.

In addition to the factors represented in Table 2.1, participants could either hear the words or see them – a manipulation of an input factor. Finally, participants could either say the words or key them out (they had learned to associate Dick with the index finger, Fred with the middle finger, and Tom with the ring finger), a manipulation of an output factor.

Figure 2.10 is an attempt to display the effects of the four factors (input modality, output modality, transformation, and substitution) on the eight regions associated with the modules. It displays for each of the eight regions the F-values that come from a statistical test of the significance of each of the four factors – input modality, output modality, transformation, and substitution. Stars indicate which statistical tests are significant. The results are largely as expected: the input modality has the strongest effects on perceptual regions, the output modality has the strongest effect on motor regions, transformation has the largest effect on the parietal region, and substitution has the largest effect on the prefrontal region. As expected, both cognitive factors, transformation and substitution, have affects on the cingulate, but neither the input or output modality affects this region. The caudate is a disappointment, not responding significantly to any of the factors (it was expected to respond like the cingulate and show effects of both the substitution and transformation factors). Two of results for the auditory cortex require a little comment. Output modality has an effect on this region because participants hear themselves giving the response. The effect of transformation here is not expected and anomalous; the transformation condition has the weaker response[12].

In summary, the general pattern of results is largely consistent with the proposed associations. The failure to get the expected effects in the caudate is the most distressing. Across the experiments run in our lab, we have only sometimes found the caudate to respond to manipulations (for instance, it did in Figure 1.8e). This may be in part related to the relatively weak magnitude of response in this region.

Predicting the BOLD Response

As illustrated in the previous chapter, ACT-R does more than just specify what regions will be affected by what factors. It predicts the exact time course of the BOLD (blood oxygen level dependent) response in each of these regions. To understand these predictions, Figure 2.11 shows the detailed procedure of the experiment as it was administered in the fMRI scanner. Each trial involved 28.5 seconds in which there were nineteen 1.5-second scans of the brain. Participants either heard or read words at the rate of one each half second. Then they either had a 4-second delay or not. The purpose of the delay was to manipulate the shape of the BOLD response for purposes of testing the model.[13] Then they saw the digit or word command. They were instructed to perform the transformation mentally and to press the right thumb key when they were ready to give the answer. The time to press the thumb key is the most important

[12] Perhaps this effect is just a spurious significant result among all the tests. However, in a number of studies we have seen a tendency for decreased BOLD response in the auditory region when there is increased cognitive engagement. Perhaps it reflects the fact that subjects are distracted from the rather aversive sound of the scanner.
[13] The delay allows the BOLD response from the initial encoding to start to go back down to baseline.

behavioral measure, reflecting the time to comprehend the instruction and plan the response. When the thumb key was pressed, subjects had to key out their letters quickly (if the output modality was manual) or say them quickly (if the output modality was vocal). Then they were given feedback in the form of the correct sequence, presented at the rate of 1 word per half-second. There were large effects of about 1.5 seconds for either substitution or transformation on time for the thumb press. Thus, the cognitive factors were having large effects on the task; this helped in separating out representation and retrieval effects. Interestingly, neither input nor output modality had an effect on time to do the task, despite the large effects these factors had on the brain regions associated with the perceptual and motor modules.

Anderson et al. report an effort to fit a detailed ACT-R model to each of the regions. These fits are reproduced in Figure 2.12. The following are the key concepts for understanding the predictions of the model:

(1) The X-axis gives the time from the beginning of a trial (see Figure 2.11).

(2) The Y-axis gives the change in the BOLD signal from baseline at the beginning of the trial.

(3) When a module is engaged, it will make a metabolic demand.

(4) This metabolic demand will show up in the BOLD response as an effect smeared over time. The BOLD response reaches a peak 4 to 5 seconds after the demand.

The methodology behind producing such fits is discussed in the appendix to this chapter . The discussion here will focus on the three largest effects from Figure 2.10: input modality on the auditory region, output modality on the manual region, and substitution on the anterior cingulate. This provides a representative of an input module, an output module, and a central module.

Figure 2.12b illustrates the time course of the BOLD response in the auditory region and the fit of the model. For purposes of this display, it plots separately the data for the visual and auditory presentation. To better test the time course of the BOLD response, it also plots separately the results for the delay and no-delay conditions. The differences among the conditions are quite striking. Particularly compelling is the aural delay condition where there are separate rises for the initial presentation of the words and for the feedback. It might seem odd that the model predicts the small rises in the visual condition. However, these occur only for the participants who are seeing the material but saying the answers. The rises reflect their processing of their own speech. Another interesting feature of this region is that delay has no effect on area under the curves. Essentially, the same amount of auditory processing is being differently distributed for the delay and no-delay conditions.

Figure 2.12c shows the data for the region of the motor cortex that corresponds to the hand. This figure breaks out the data according to whether the response was manual or vocal. Again the BOLD responses for the different conditions are strikingly different. There is some response even in the vocal condition. In part this reflects the fact that all participants, including those in the vocal condition, issued their timed response as a thumb press. However, it also reflects a failure to totally separate the motor region devoted to the hand from the nearby region devoted to the face.[14] Note in this figure that the motor region begins to respond in the delay condition before the actual response. This reflects motor rehearsal by the participants to bridge

[14] The face area (Figure 2.12d) also shows some response in all conditions, but it does respond much more strongly in the vocal condition.

the delay. In contrast to the aural region, this area shows greater area under the curves in the delay condition, reflecting the rehearsal of the responses in the delay period. [15]

Figure 2.12g shows the data for the anterior cingulate broken down according to whether a substitution was required. Whenever there are multiple rules that can apply to a situation but only one or a few are appropriate, a special control state needs to be set to select the appropriate rule, and the anterior cingulate will show an increased response reflecting the setting of this control state. Thus, the model predicts the effect of substitution because it must set a special control state to wait for the result of the retrieval. Also, the BOLD response is greater (measured by area under the curve) in the delay condition because a control state must be set to bridge that delay. The model also predicts the effect of transformation on this region (not shown in Figure 2.12g) because a control state must be set to bridge the interval while the representation is being transformed. While the anterior cingulate responds to these three factors, it does not respond to either input or output modality, as Figure 2.10 illustrates. As emphasized in Anderson et al. (in press), it is a region that responds to the abstract information-processing demands of the task.

The general pattern of effects in the experiment (Figure 2.10) and the ability to explain the exact shape of the BOLD function (Figure 2.12)[16] provide strong support for an association of these regions with these modules. The associations displayed in this experiment have been replicated in many experiments in our lab and are roughly consistent with other results in the literature.

[15] This is not the first experiment to find motor rehearsal, but it was a surprise when it occurred in the first such study (Anderson, Qin, Stenger, and Carter, 2004). It led to a change in our models to include such motor rehearsal. As such it is an example of how imaging can inform the development of a model.

[16] It should also be noted that much poorer fits are obtained when one tries to fit the wrong module to one of these regions. For measures of goodness of fit see Anderson et al. (in press).

They give strong reason for believing that brain regions can be mapped onto function despite the doubts expressed by researchers such as Uttal (2001).

Overall Conclusions

With respect to Newell's question of how the human mind can occur in the physical universe, this chapter offers a general answer and some specific details. The general answer is that the mind partitions itself into specific information-processing functions and these functions are achieved in relatively localized brain regions where the processing can be done effectively. There are paths of connections among these regions that assure the coordination of these functions into a coherent sequence of activities. The specific answers are the eight modules, their associated regions, and their coordination by the central production system.

The general answer of a modular partition seems the only way to achieve the multi-purpose functionality that humans need to meet the demands of their world given the structures of their brains. However, the specific answers offered here are far from final. Certainly, these eight modules do not exhaust the functions of the mind or the regions of the brain. However, beyond the issue of completeness, one can wonder whether the proposed partitioning of function and paths of communication are correct. One movement in recent research has been to partition the brain and its functions much more finely than is done here. Such efforts include the memory functions of the prefrontal cortex (e.g., Badre et al., 2005), the control functions of the anterior cingulate (e.g., van Veen and Carter, 2005), and the representational functions of the parietal cortex (e.g., Dehaene et al., 2002). Also as noted, the loop through the basal ganglia does not begin to exhaust the paths of communication in the brain. In the hindsight of a decade or two, the partitioning offered here in this book might seem rather crude. As Newell warned we would be, we are just a little ways into an answer. Nonetheless, the structure-function associations proposed in this chapter are sufficiently similar to many other ideas in the field that it seems unlikely that further refinements will completely overturn these associations.

The successes reviewed in this chapter and elsewhere in the book strongly suggest that Fodor was wrong in his pessimism about central modules and the impossibility of a computational theory of central cognition. Figure 2.2 contained declarative, imaginal, goal, and procedural modules that are all central modules. There has been considerable success in associating these modules with specific brain areas. They seem to meet Fodor's criteria for being called a module. Moreover, they play effective computational roles in models of a wide variety of cognitive tasks.

The next two chapters in this book will look at two of the modules about which the ACT-R architecture has the most to say, the declarative and the procedural modules. The fifth chapter will be concerned with what in this architecture might be uniquely human.

Appendix: Predicting the BOLD Response

Our laboratory has developed a methodology for relating the profile of activity in modules like those in Figure 2.2 to Blood Oxygen Level Dependent (BOLD) responses from the brain regions that correspond to these modules. The fundamental idea is to use a timeline of module activity like that in Figure 1.7. Anderson et al. (in press) provide specification of the timelines behind all the predictions in Figure 2.12, but for current purposes let us just look at predictions from the timeline for the auditory cortex (Figure 2.12b) and just for the delay condition with auditory input. Figure 2.13a presents this timeline as a demand function for this condition giving the proportion of time the module is active in each 1.5 second scan. There is a peak of 100% activity when the words are presented during the second scan. The time at which the feedback is presented varies a bit and so there is a distribution of times at which module is active later to process this feedback.[17]

The basic theory we have developed of the BOLD response claims that the while a module is engaged there is an increased metabolic demand in the corresponding region producing a hemodynamic response. We have adopted the standard gamma function that other researchers (e.g., Boyton, Engle, Glover, and Heeger, 1996; Cohen, 1997; Dale and Buckner, 1997; Glover, 1999) have used for the hemodynamic response. If the module is engaged, it will produce a BOLD response *t* time units later according to the function:

$$H(t) = m\left(\frac{t}{s}\right)^{a} e^{-(t/s)}$$

[17] See Anderson et al. for a discussion of the slight negativity at the end.

where *m* governs the magnitude, *s* scales the time, and the exponent *a* determines the shape of the BOLD response such that with larger *a* the function rises and falls more steeply. Figure 2.13b illustrates the function assumed for the auditory region. As is typical of such functions, it shows a slow response that peaks about 4 to 5 seconds after the actual activity. The peak of the function is at *a**s*. The parameter *a* is 7 for this function and *s* is .63 seconds and *a**s* is 4.41 seconds, which is where the function in Figure 2.12b peaks.

The BOLD response accumulates whenever the region is engaged. Thus, if *D(t)* is a demand function giving the probability that the region is engaged at time *t*, then the cumulative BOLD response can be obtained by convolving this function with the hemodynamic function:

$$B(t) = \int_0^t D(x)H(t-x)dx$$

This is the prediction for the BOLD response that we will observe in the region associated with that demand function. Figure 2.13c shows the results, predicted BOLD response in this case. As can be seen, the observed response preserves some of the structure of the demand function in Figure 2.13a, but the convolving with the BOLD response blurs some of the temporal structure and delays the peaks.

In summary, a model for the time course of this task yields demand functions *D(t)* like that in Figure 2.13a. By convolving the demand functions with the hemodynamic function one can obtain predictions for the BOLD response in the regions associated with the modules. Anderson (2005) can be consulted for ways of assessing the match between the predictions and the data. A similar convolution methodology is frequently used in analysis programs for fMRI data where one takes the condition structure of trials in an experiment and convolves it with a hemodynamic response to produce a condition-sensitive pattern of activity. This pattern is regressed against brain activity to find which regions are sensitive to these conditions (e.g., Friston, 2003). The methodology we used is finer grained conceptually (using model behavior within a single trial) and is used for confirmatory purposes rather than exploratory purposes.

Table 2.1

Illustration of the Four Conditions of the Experiment

Associations: AT is associated to 23; BE to 24

	No Transformation	Yes Transformation
No Substitution	**Stimulus:** Tom Dick Fred **Probe:** 24 **Response:** Tom-Dick-Fred	**Stimulus:** Tom Dick Fred **Probe:** 23 **Response:** Tom-Fred-Dick
Yes Substitution	**Stimulus:** Tom Dick Fred **Probe:** BE **Response:** Tom-Dick-Fred	**Stimulus:** Tom Dick Fred **Probe:** AT **Response:** Ring-Fred-Dick

Figure Captions

Figure 2.1. Schematic diagram of the major structures of the basal ganglia and their interconnections. Abbreviations: GP, globus pallidus; GPi, internal segment of globus pallidus; GPe, external segment of globus pallidus; EP, entopeduncular nucleus; STN, subthalamic nucleus; SNr substantia nigra, pars reticulata. Numbers indicate the total number of neurones within each structure. From “Basal Ganglia: Structure and Computations,” by J. Wickens, 1997, *Network: Computationin Neural Systems*, *8*, p. R79. Copyright 1997 by IOP. Publishing Ltd.

Figure 2.2. The modules implemented in ACT-R 6.0.

Figure 2.3. Aggregate portion of gaze time for visual regions in a multi-lane highway experiment (Salvucci, 2005).

Figure 2.4. Time before switching for (a) monitoring and (b) control (Salvucci, 2005).

Figure 2.5 Task-switching points as illustrated by key delay times for (a) human drivers, and (b) model simulations. Errors bars represent standard errors. (Salvucci, 2005).

Figure 2.6 Aggregate effects of dialing on driving as measured by lateral deviation and speed deviation. Errors bars represent standard errors (Salvucci, 2005).

Figure 2.7. Learning to time share: (a) Experiment 1 from Hazeltine et al. (2002) and (b) ACT-R simulation

Figure 2.8. ACT-R module activities early in the experiment (but not as early as illustrated in Figure 2 of Anderson, Taatgen, and Byrne, 2005) and relatively late in the experiment.

Figure 2.9. An illustration of the locations of the 8 regions of interest. In part (a) are the regions close to the surface of the cortex and in part (b) are the regions deeper in the brain. The Tailarach coordinates of these right regions are given (left homologue can be obtained by switching the sign of the x coordinate). Most of the regions are cubes 5 voxels long, 5 voxels wide, and 4 voxels high. The exceptions are the procedural (caudate), which is 4 x 4 x 4 and the goal (anterior cingulate cortex), which is 5 x 3 x 4. A voxel in our research is 3.125 mm long and wide and 3.2 mm high.

Figure 2.10. A display of the F-values for the main effects of input modality, output modality, transformation, and substitution for the 8 predefined brain regions. Stars indicate significant effects. From Anderson et al (in press).

Figure 2.11. The 28.5-second structure of an fMRI trial.

Figure 2.12. Observed (dotted lines connecting points) BOLD responses and predictions (solid lines) for the eight predefined regions. The data and predictions are plotted as a function of the mean time of each scan. From Anderson et al. (in press).

(a) Effects of input modality and delay of the left fusiform gyrus.

(b) Effects of input modality and delay on the left and right auditory cortex.

(c) Effects of output modality and delay on the left motor area that is associated with the right hand.

(d) Effects of output modality and delay on the left and right motor areas that are associated with the face and tongue.

(e) Effects of transformation and delay on the left parietal region

(f) Effects of substitution and delay on the left prefrontal region.

(g) Effects of substitution and delay on the left anterior cingulate.

(h) Effects of substitution and delay on the right caudate

Figure 2.13 An illustration of the methodology behind the predictions in Figure 2.12:

(a) The aural demand function for the auditory-delay condition in Figure 2.12b.

(b) The hemodynamic function assumed.

(c) The resulting prediction of the BOLD response for the auditory region.

Figure 2.1

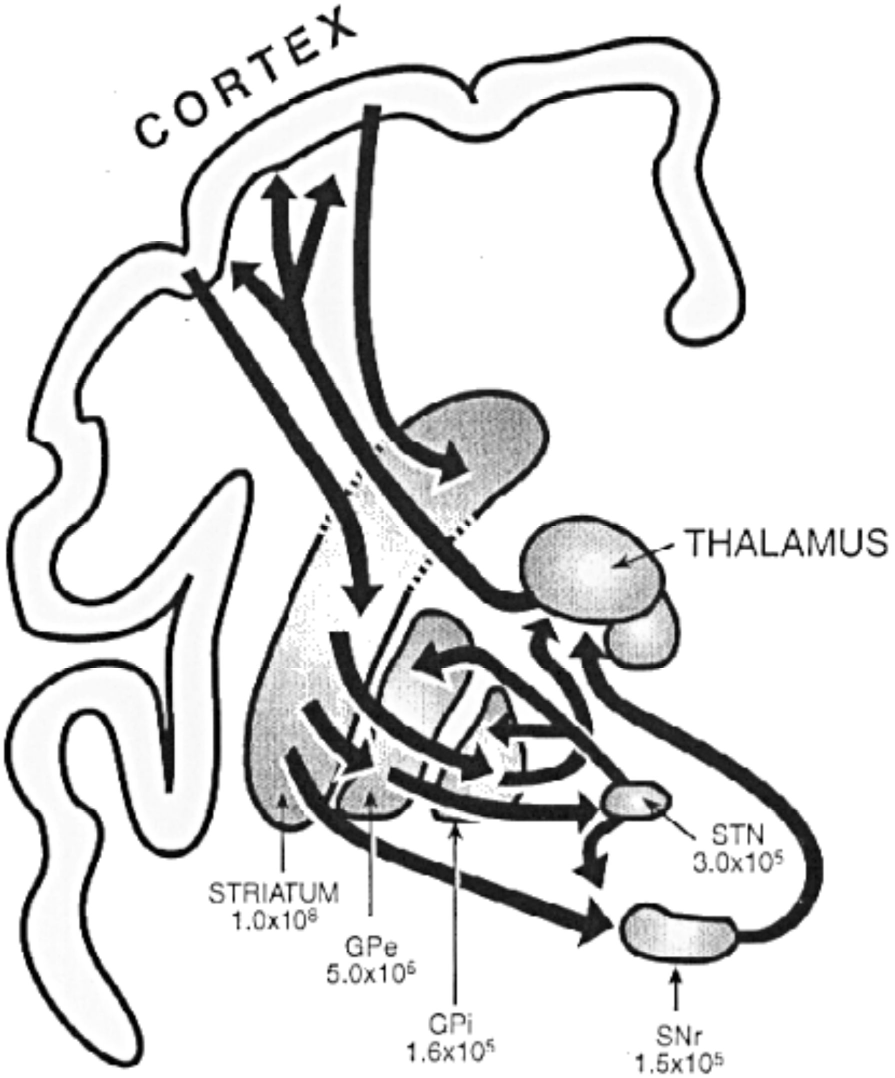

CORTEX
THALAMUS
STN
3.0x10^{5}
STRIATUM
1.0x10^{8}
GPe
5.0x10^{5}
GPi
1.6x10^{5}
SNr
1.5x10^{5}

Figure 2.2

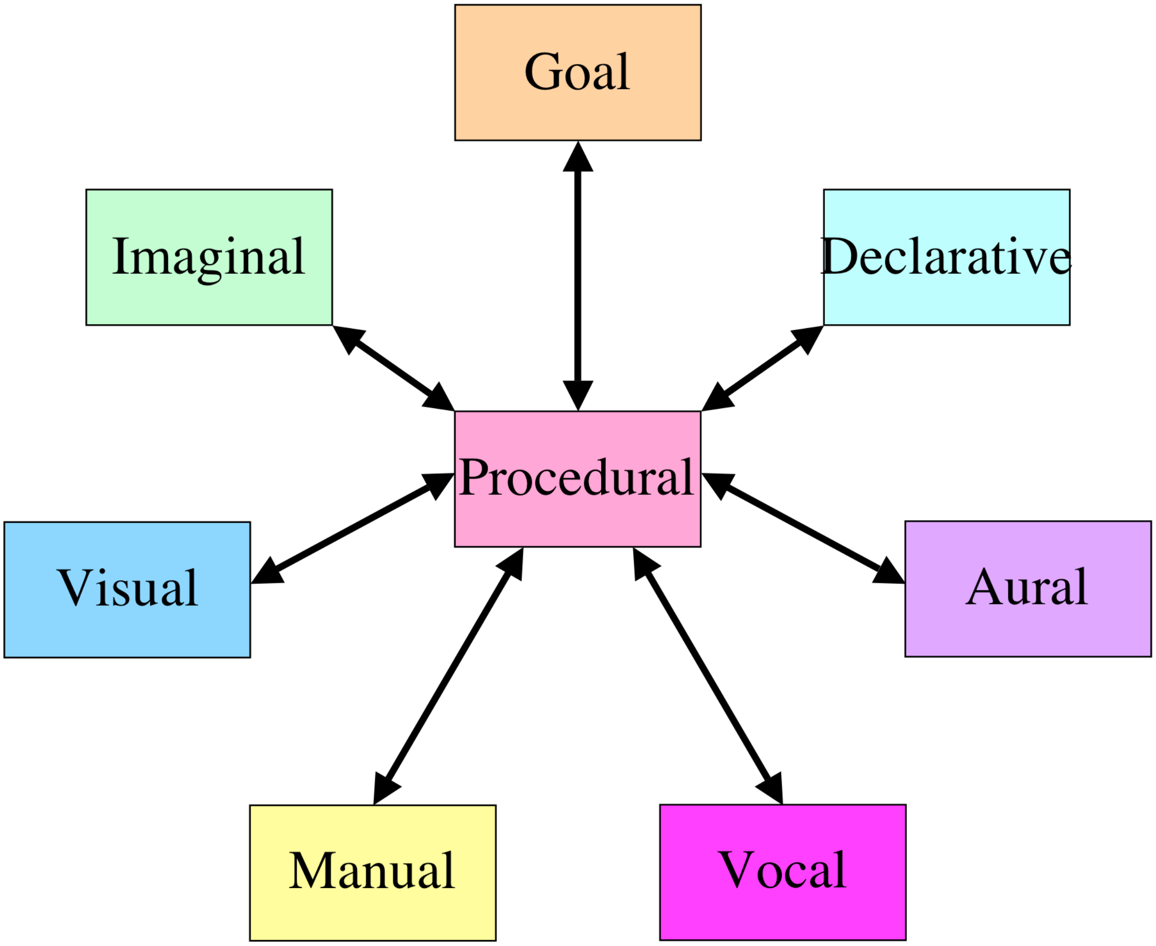

Goal
Imaginal
Declarative
Procedural
Visual
Aural
Manual
Vocal

Figure 2.3

D. D. Salvucci/Cognitive Science 29 (2005) 473

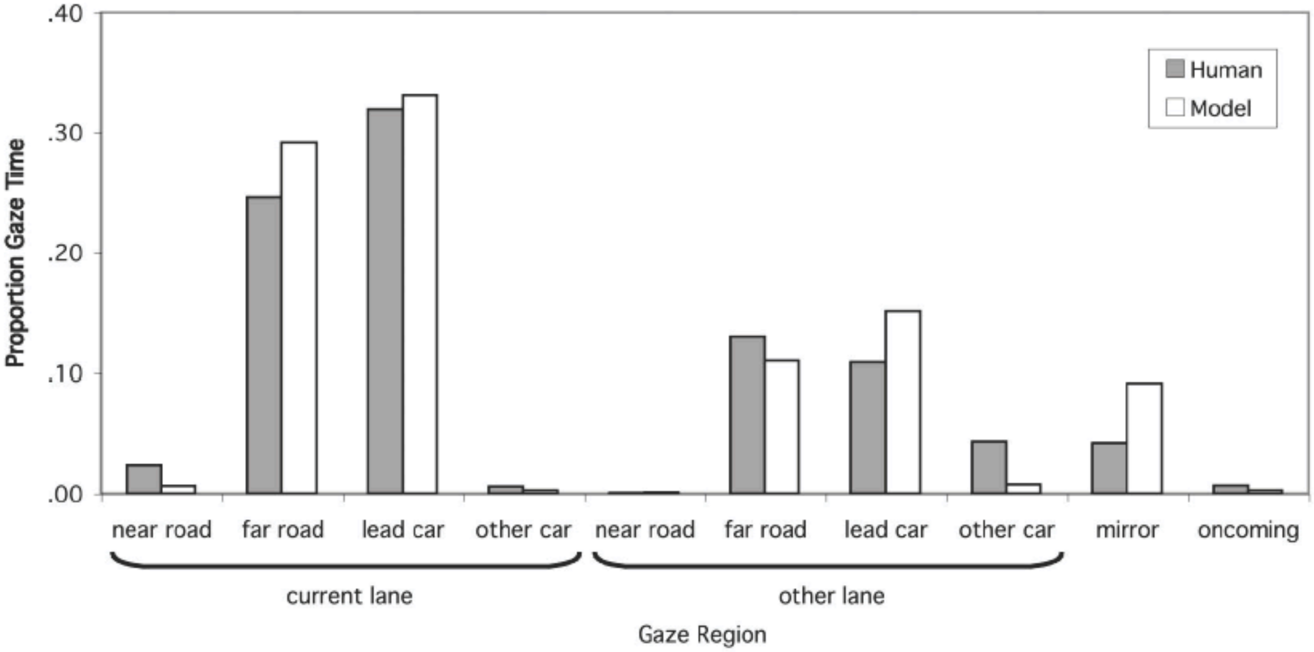


Fig. 4. Study 1: Aggregate proportion gaze times for visual regions in the multi-lane highway environment.

Figure 2.4

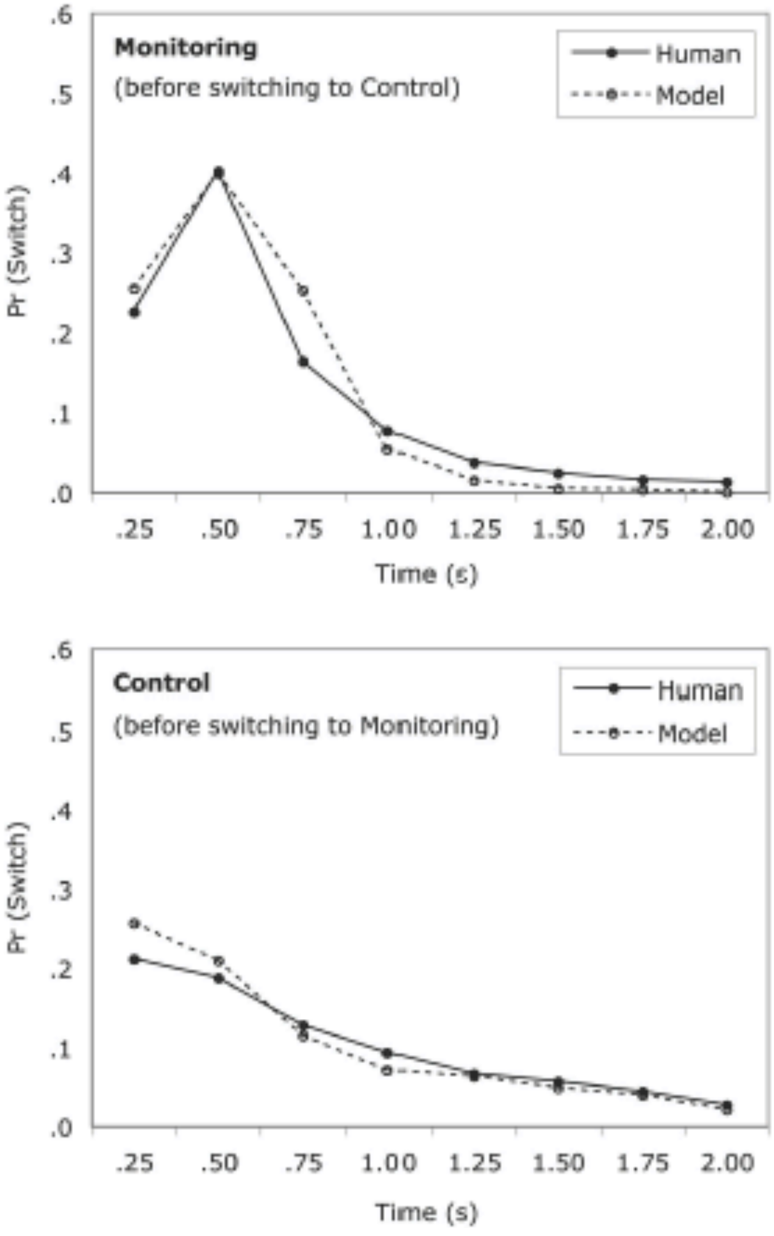

Monitoring
(before switching to Control)
Human
Model
Pr (Switch)
.6
.5
.4
.3
.2
.1
.0
.25
.50
.75
1.00
1.25
1.50
1.75
2.00
Time (s)
Control
(before switching to Monitoring)
Human
Model
Pr (Switch)
.6
.5
.4
.3
.2
.1
.0
.25
.50
.75
1.00
1.25
1.50
1.75
2.00
Time (s)

Figure 2.5

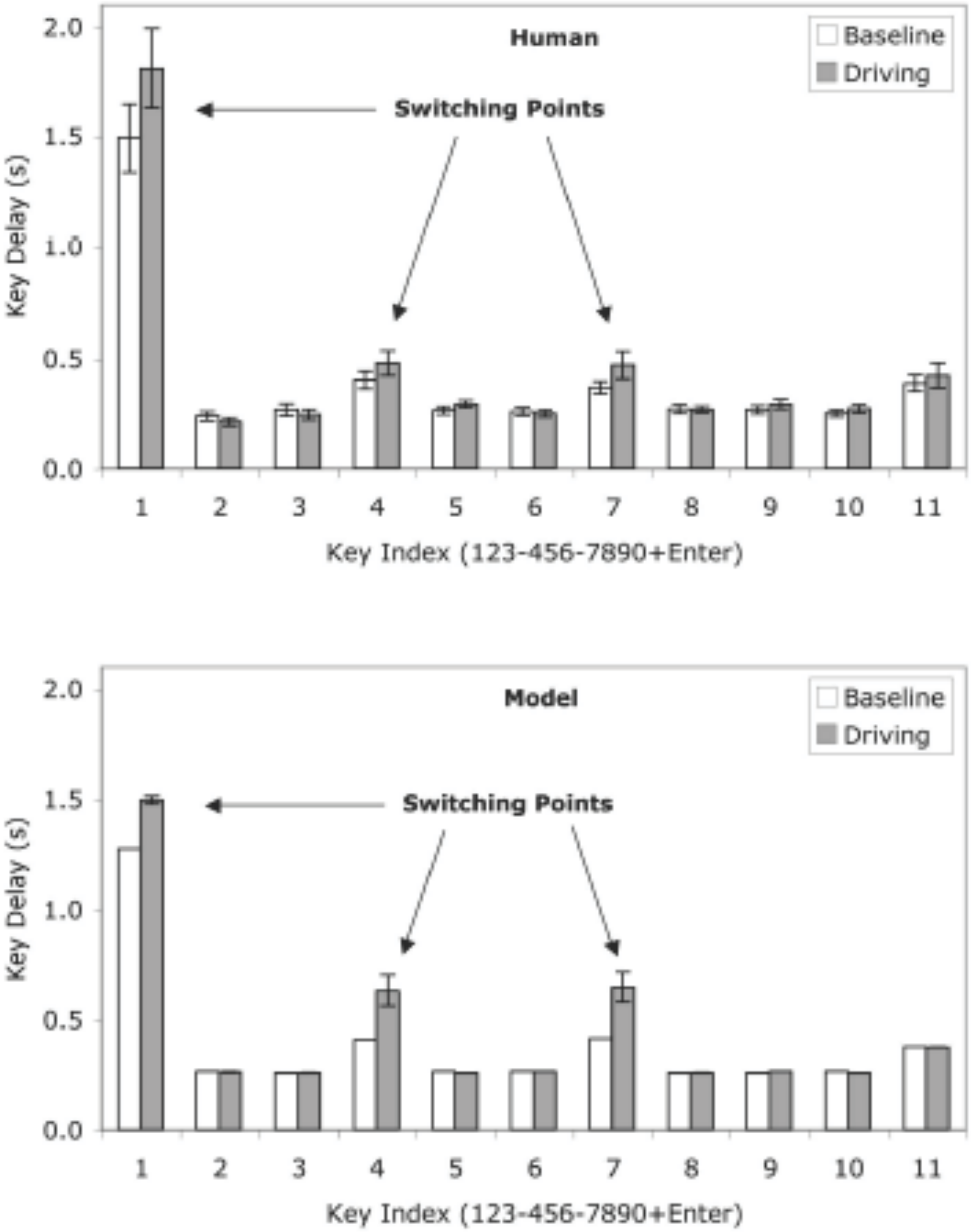

Human
□ Baseline
▨ Driving
Switching Points
Key Delay (s)
2.0
1.5
1.0
0.5
0.0
1
2
3
4
5
6
7
8
9
10
11
Key Index (123-456-7890+Enter)
Model
□ Baseline
▨ Driving
Switching Points
Key Delay (s)
2.0
1.5
1.0
0.5
0.0
1
2
3
4
5
6
7
8
9
10
11
Key Index (123-456-7890+Enter)

Figure 2.6

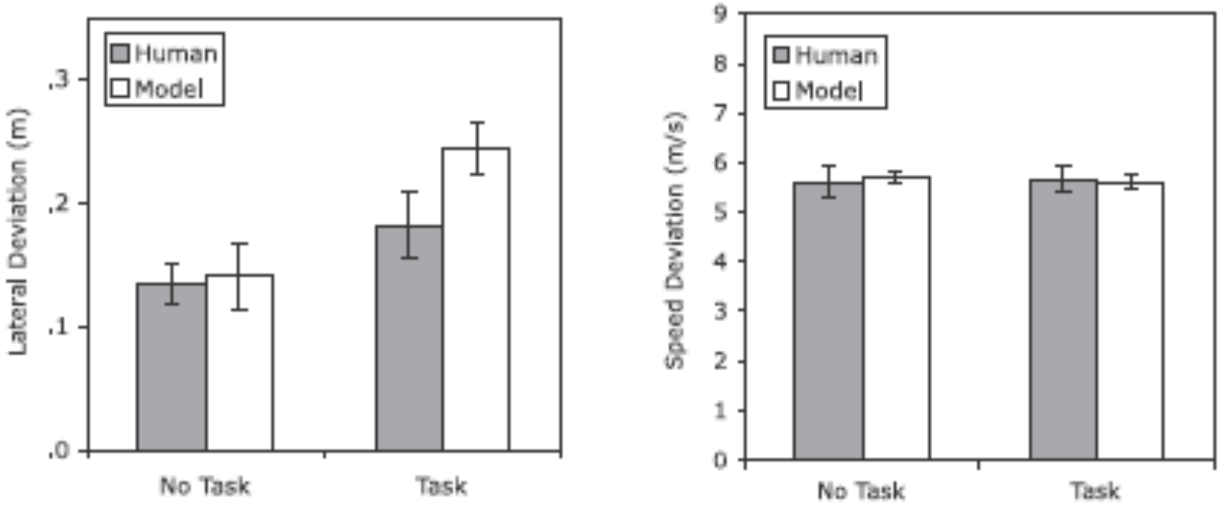
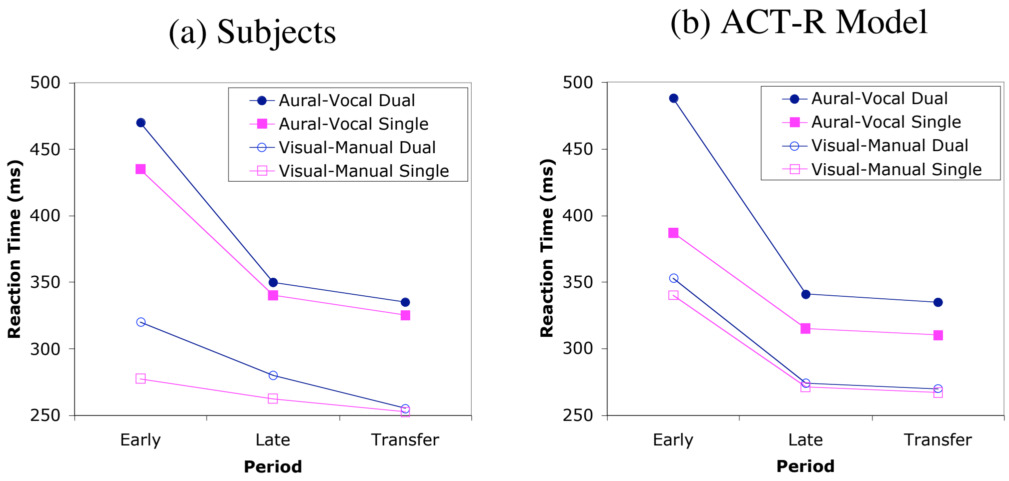

Human
Model
Lateral Deviation (m)
.0
.1
.2
.3
No Task
Task
Human
Model
Speed Deviation (m/s)
0
1
2
3
4
5
6
7
8
9
No Task
Task

Figure 2.7



(a) Subjects

(b) ACT-R Model

Figure 2.8

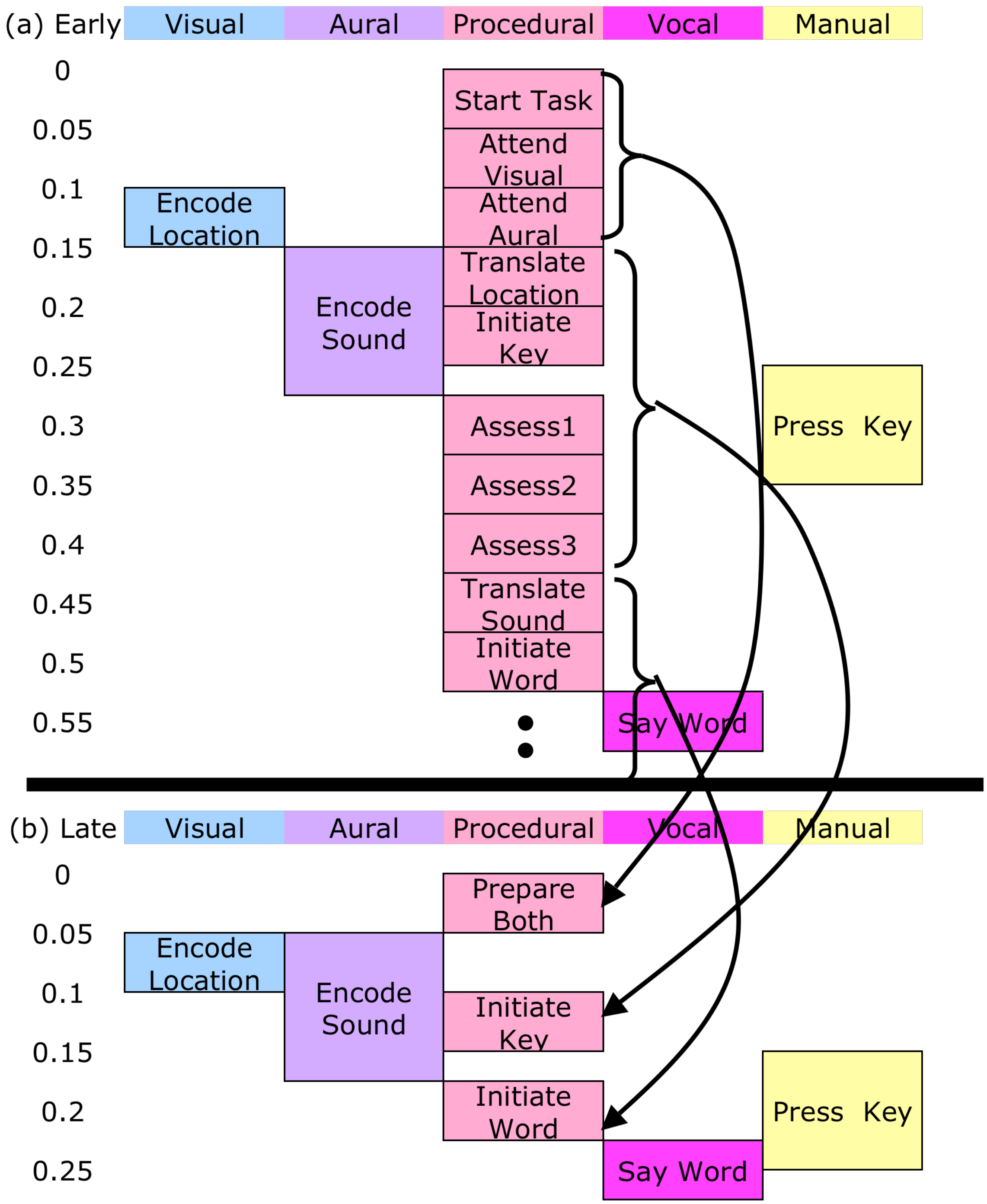
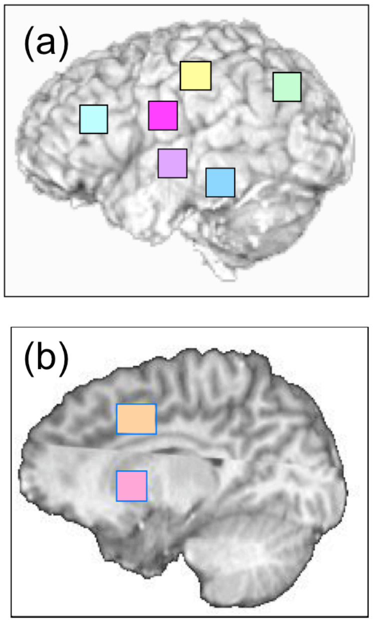

(a) Early
Visual
Aural
Procedural
Vocal
Manual
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
Start Task
Attend Visual
Attend Aural
Encode Location
Encode Sound
Translate Location
Initiate Key
Press Key
Assess1
Assess2
Assess3
Translate Sound
Initiate Word
Say Word
(b) Late
Visual
Aural
Procedural
Vocal
Manual
0
0.05
0.1
0.15
0.2
0.25
Prepare Both
Encode Location
Encode Sound
Initiate Key
Initiate Word
Press Key
Say Word

Figure 2.9



		x	y	z	Region
Manual	🟨	37	-25	47	Motor
Imaginal	🟩	23	-64	34	Parietal
Vocal	🟪	44	-12	29	Motor
Declarative	🟦	40	21	21	Prefrontal
Aural	🟪	47	-16	4	Auditory
Visual	🟦	42	-60	-8	Fusiform
Goal	🟧	5	10	38	ACC
Procedural	🟥	15	9	2	Caudate

Figure 2.10

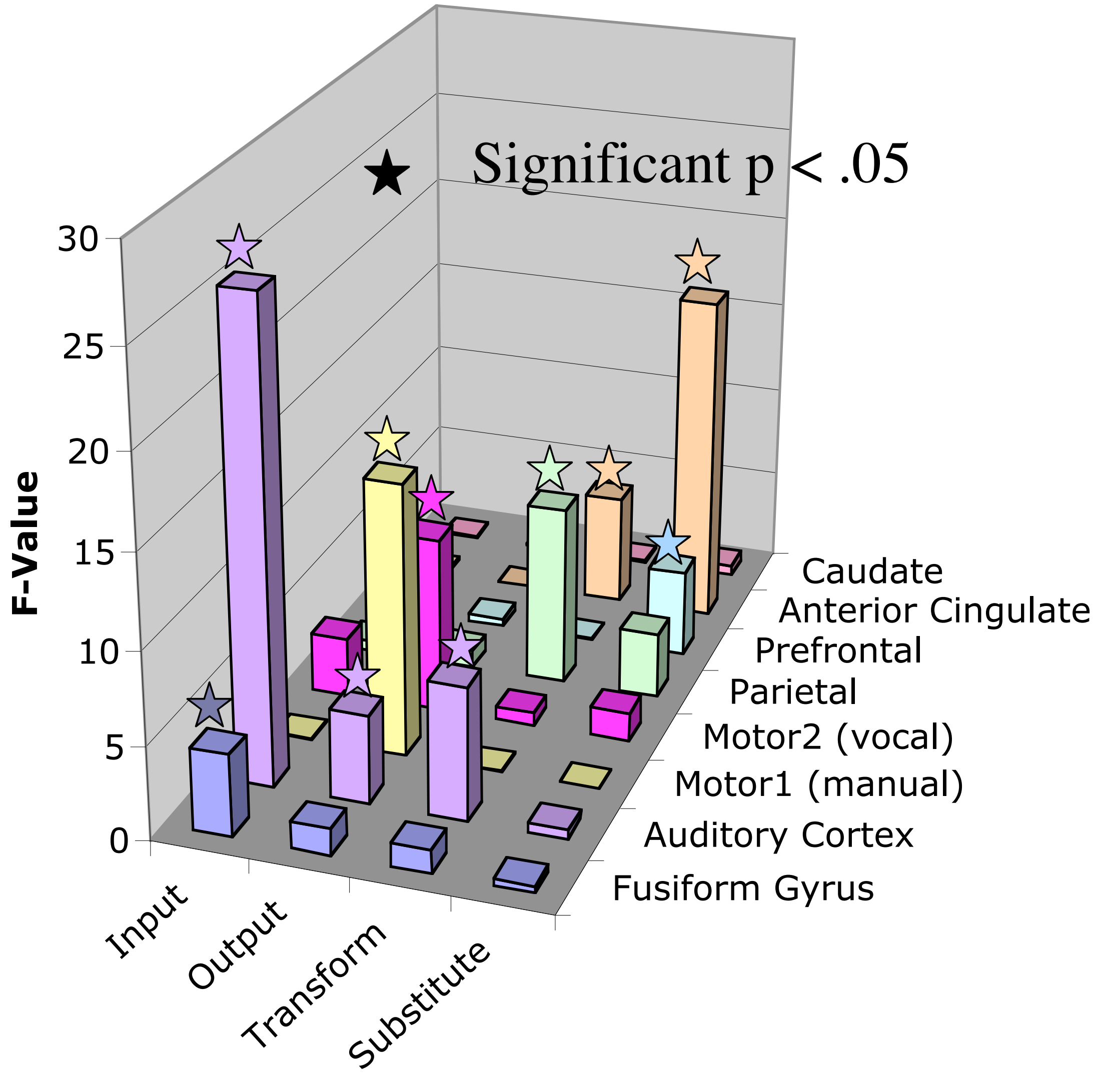

★ Significant p < .05
F-Value
30
25
20
15
10
5
0
Input
Output
Transform
Substitute
Caudate
Anterior Cingulate
Prefrontal
Parietal
Motor2 (vocal)
Motor1 (manual)
Auditory Cortex
Fusiform Gyrus

Figure 2.11

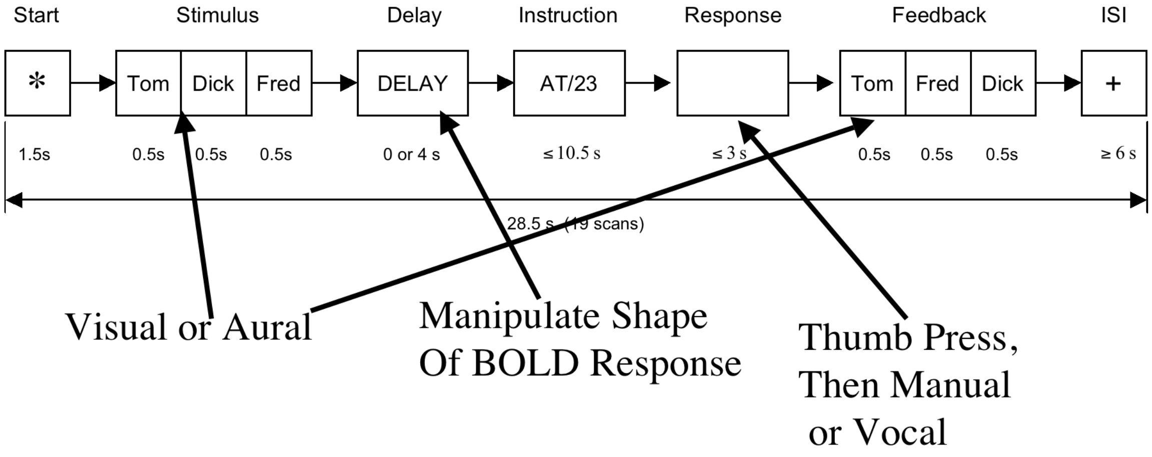

Start
Stimulus
Delay
Instruction
Response
Feedback
ISI
*
Tom
Dick
Fred
DELAY
AT/23
Tom
Fred
Dick
+
1.5s
0.5s
0.5s
0.5s
0 or 4 s
≤ 10.5 s
≤ 3 s
0.5s
0.5s
0.5s
≥ 6 s
28.5 s (19 scans)
Visual or Aural
Manipulate Shape
Of BOLD Response
Thumb Press,
Then Manual
or Vocal

Figure 2.12

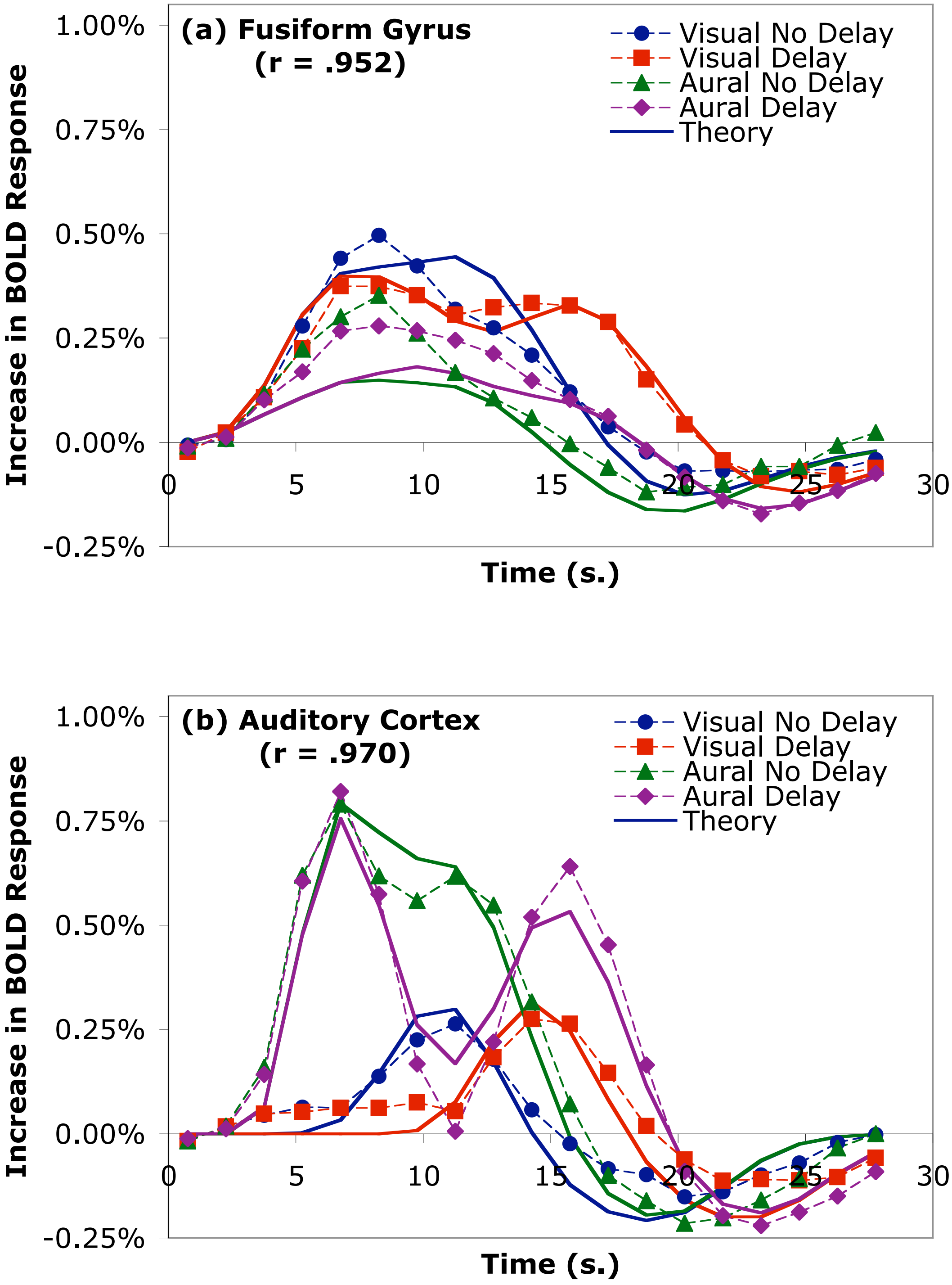
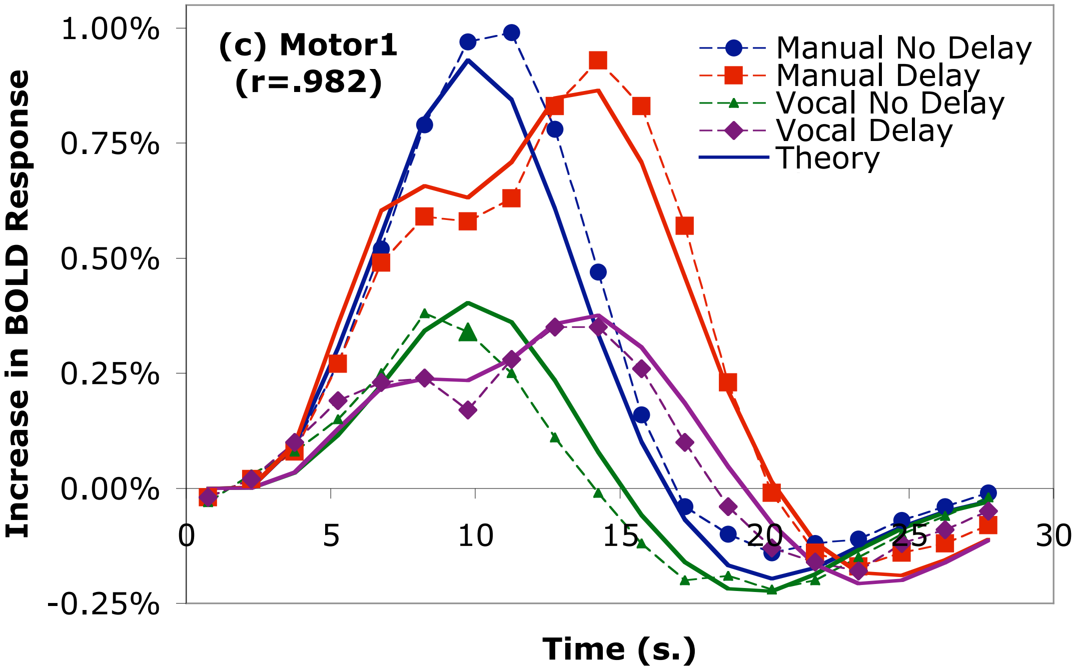
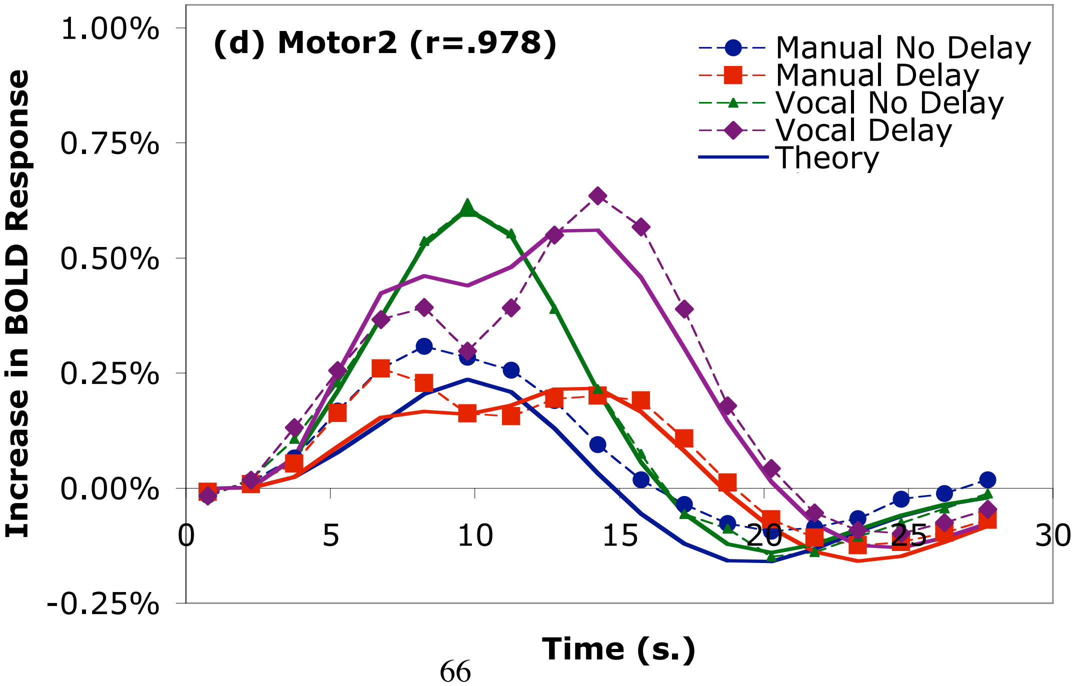

(a) Fusiform Gyrus
(r = .952)
● Visual No Delay
■ Visual Delay
▲ Aural No Delay
◆ Aural Delay
Theory
Increase in BOLD Response
1.00%
0.75%
0.50%
0.25%
0.00%
-0.25%
0
5
10
15
20
25
30
Time (s.)
(b) Auditory Cortex
(r = .970)
● Visual No Delay
■ Visual Delay
▲ Aural No Delay
◆ Aural Delay
Theory
Increase in BOLD Response
1.00%
0.75%
0.50%
0.25%
0.00%
-0.25%
0
5
10
15
20
25
30
Time (s.)

Figure 2.12 continued





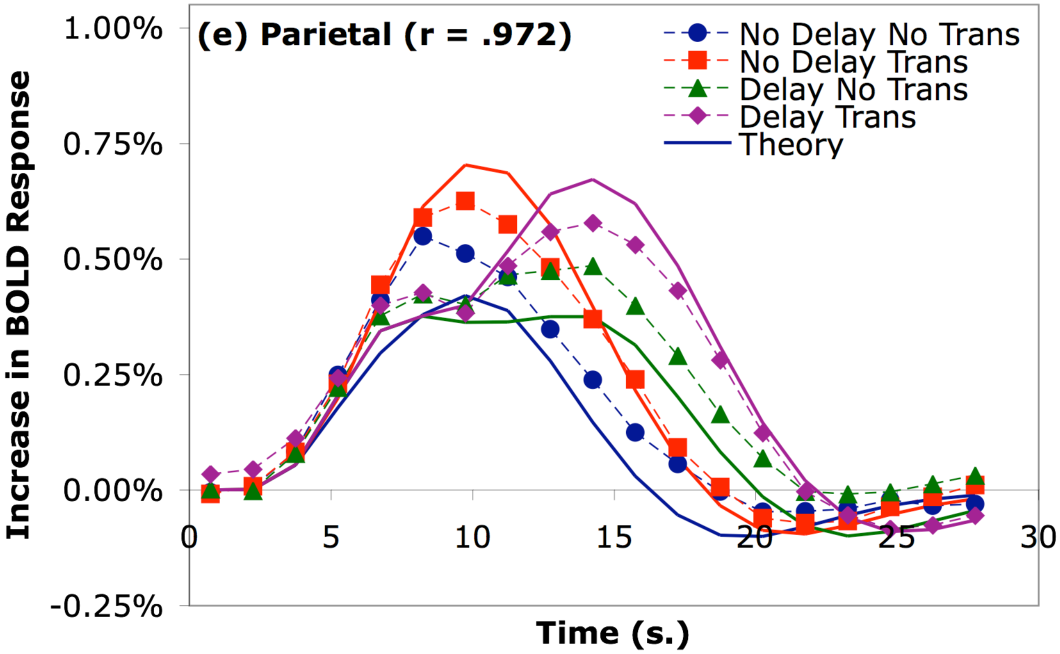


Figure 2.12 continued

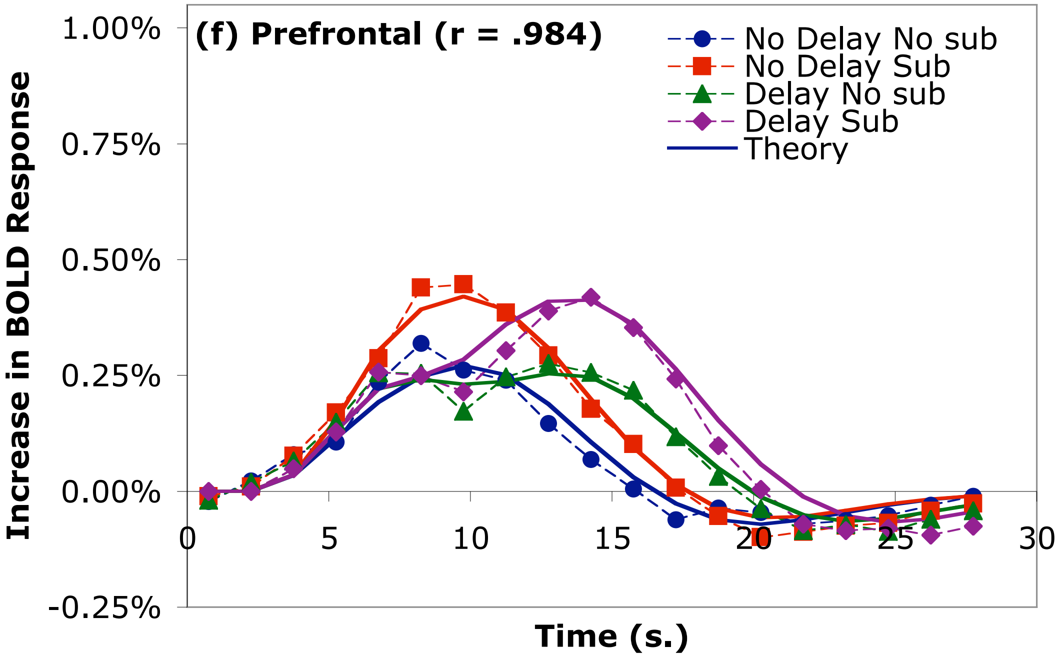


Figure 2.12 continued

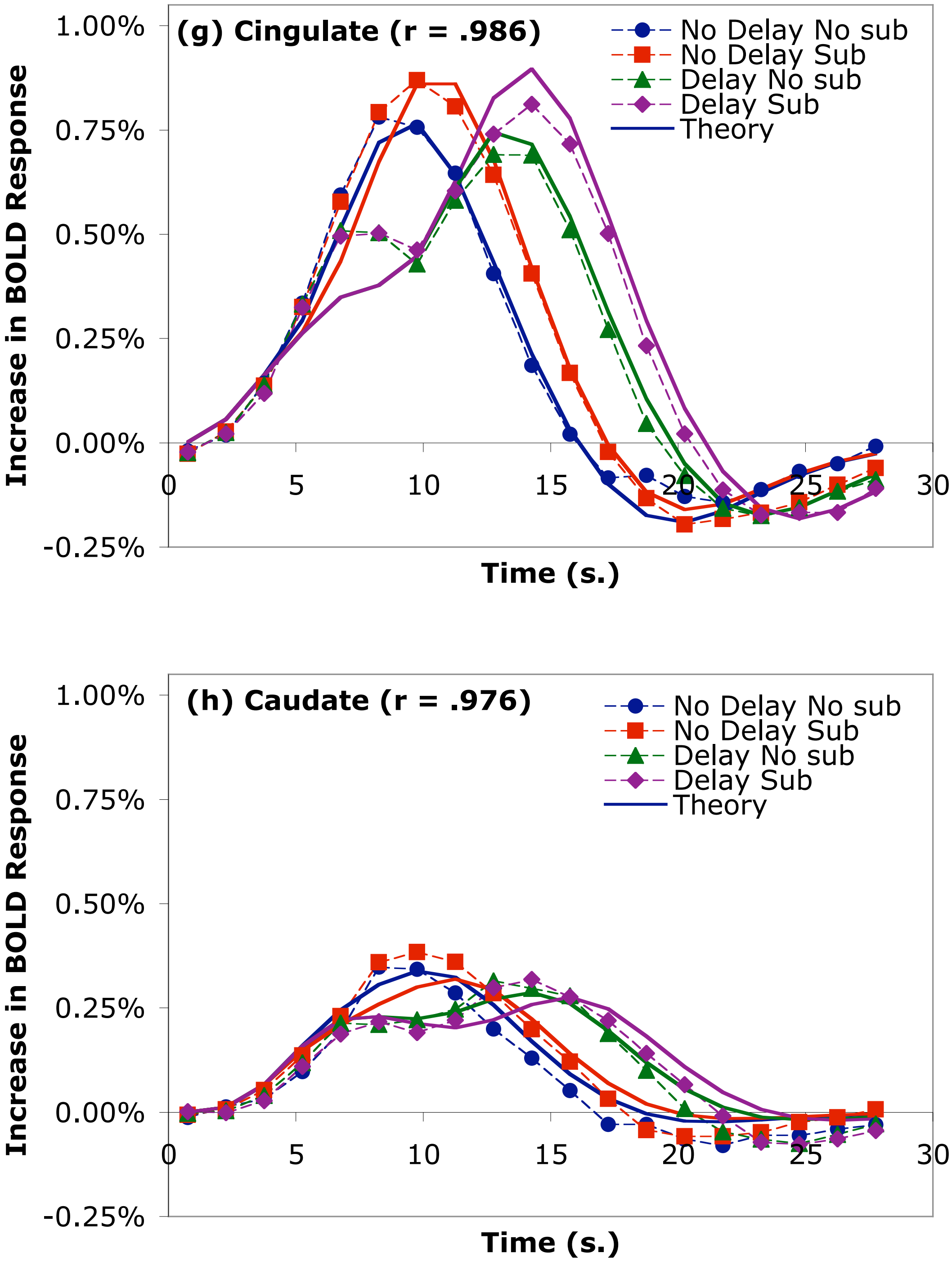
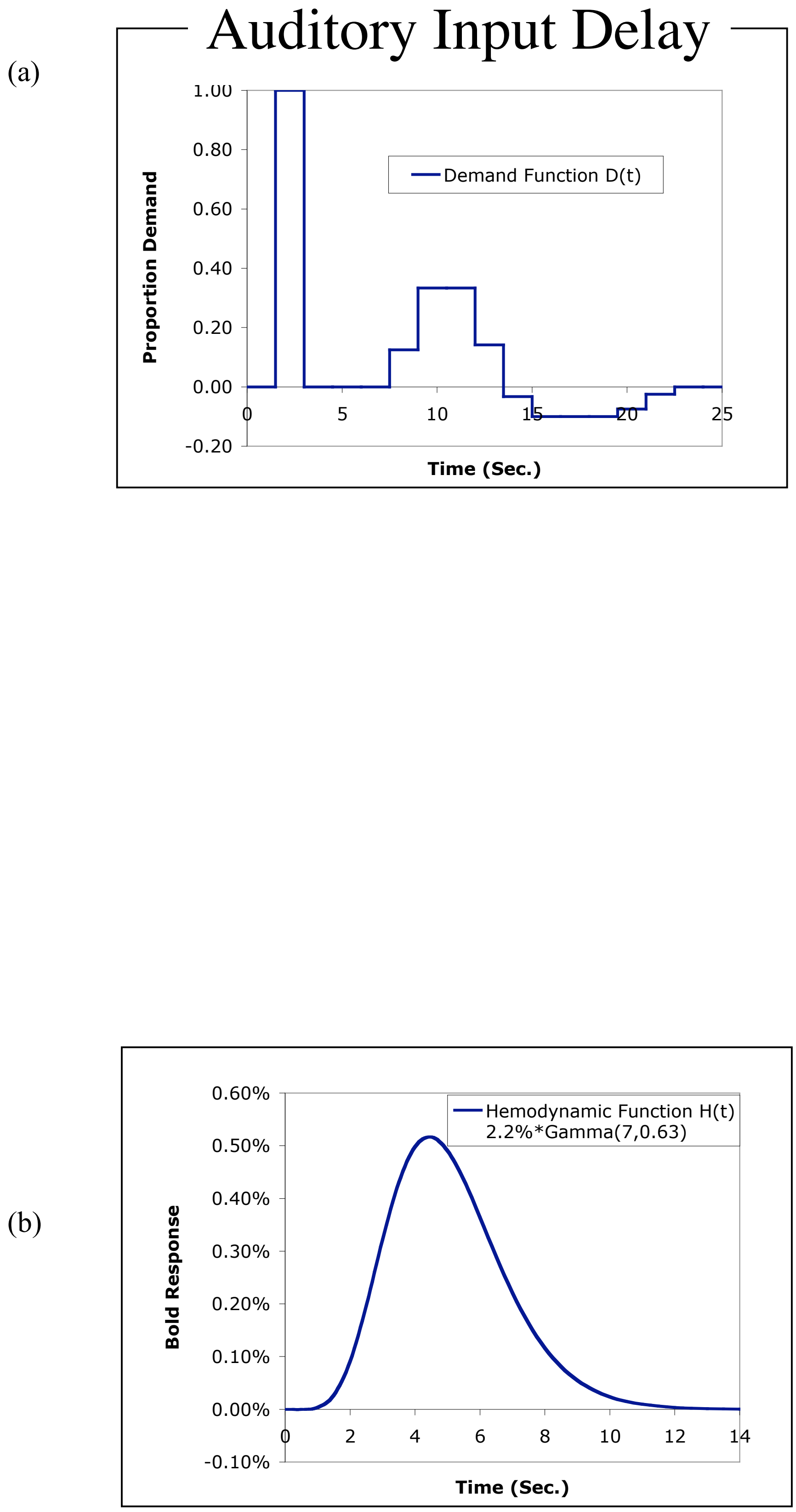
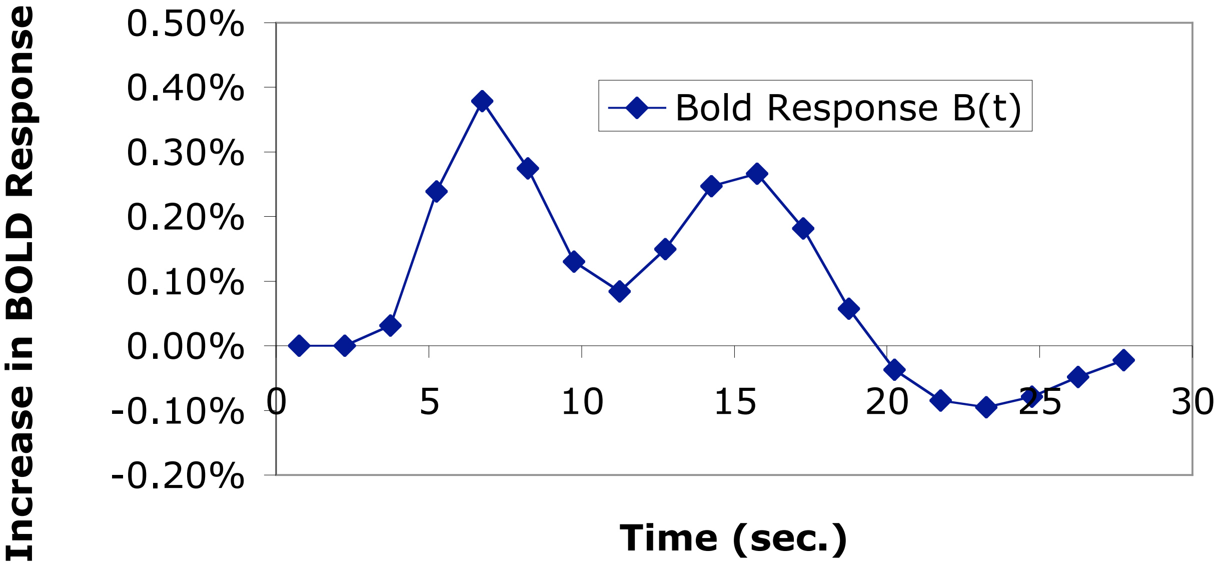


Figure 2.13


(a)
Auditory Input Delay
Demand Function D(t)
Proportion Demand
1.00
0.80
0.60
0.40
0.20
0.00
-0.20
0
5
10
15
20
25
Time (Sec.)
(b)
Hemodynamic Function H(t)
2.2%*Gamma(7,0.63)
Bold Response
0.60%
0.50%
0.40%
0.30%
0.20%
0.10%
0.00%
-0.10%
0
2
4
6
8
10
12
14
Time (Sec.)

(c)


Increase in BOLD Response
0.50%
0.40%
0.30%
0.20%
0.10%
0.00%
-0.10%
-0.20%
Bold Response B(t)
0
5
10
15
20
25
30
Time (sec.)