In Proceedings of ICCM - 2007- Eighth International Conference on Cognitive Modeling. 175 - 180. Oxford, UK: Taylor & Francis/Psychology Press.

A 3-node Queuing Network Template of Cognitive and Neural Differences As Induced by Gray and White Matter Changes

Marc G. Berman (bermanm@umich.edu)
Departments of Psychology and Industrial and Operations Engineering, 530 Church Street
Ann Arbor, MI 48109-1043 USA

Yili Liu (yililiu@umich.edu) and Changxu Wu (changxuw@umich.edu)
Department of Industrial and Operations Engineering, 1205 Beal Avenue
Ann Arbor, MI 48109-2117 USA

Abstract

We present a 3-node queuing network template for simulating brain activity differences for different subject populations performing simple cognitive tasks. We hypothesize that distinct areas of cortex behave similarly to queuing network servers, whose interactions are used to simulate the interactions of different brain areas. This 3-node queuing network template accurately accounts for brain activity disparities (as found with neuroimaging techniques) for different subject populations performing verbal working memory, spatial working memory, and verbal audition tasks. Further, this 3-node queuing network template provides an account explaining the interactions between different brain areas. This account suggests that reductions in service rates (due to changes in gray matter volume or white matter anisotropy) for different brain areas alters the flow or propagation of neural activity, causing different brain activity patterns for different subject populations performing the same cognitive tasks.

Keywords: Queuing Networks; neuroimaging; working memory

Introduction

The brain is an enormously complex network of interconnected systems and sub-systems, which at this point cannot be easily understood. Most standard neuroimaging techniques tend to focus on singular brain regions that are hypothesized to be responsible for singular functions, either general functions (global approach) or specific functions (local approach; Nyberg and McIntosh, 2001). It seems more likely that behavior and thought result from the interactions of different brain regions rather than from singular brain region activations (Lashley, 1931; Bressler, 1995).

How do different brain areas interact with each other? A number of models and techniques have been proposed to examine this question. Such techniques include Partial Least Squares (PLS), Structural Equation Modeling (SEM), and Dynamic Bayesian Networks (DBN; Nyberg and McIntosh, 2001; Labatut et al., 2004). These techniques are all statistical techniques that can uncover the relationships between different brain areas. While these techniques are extremely useful, they do not explain why brain regions interact in such ways.

This paper offers a new research method based on queuing network theory to explore brain networks. The unique power of this queuing network approach for examining cortical interactions is illustrated in this paper through a simple 3-node queuing network architecture that explains differences in brain activity for different subject populations (young vs. old, literate vs. illiterate) performing the same cognitive tasks.

There are a number of major contributions of this paper. The first is to offer an alternative method to model connectivity in the brain and the subsequent interactions of different brain areas. The second major contribution is that the queuing network template provides plausible, novel, and causal explanations, which predict when certain brain areas will become active and offer explanations as to why they become active. The third contribution is the model's breadth where the same structural template may be the underlying architecture mediating task performance in a wide range of cognitive tasks. In addition, this architecture did not require many model parameter changes in order to model these different phenomena.

Queuing Networks and Psychology

Queuing Networks are a mathematical discipline that are used to simulate and model a wide array of phenomena including manufacturing processes, emergency room workload, and airport traffic. The queuing network methodology has also been applied to cognitive psychology, and was used to successfully unify various psychological models of reaction time (Liu, 1996) and multitask performance (Liu, 1997). Recently, the queuing network approach has been successfully integrated with the symbolic approach (Liu et al., 2006) for both mathematical analysis and real time simulation of human performance in a multitude of settings including in-vehicle steering (Liu et al., 2006), transcription typing (Wu & Liu, 2004), and visual search tasks (Lim & Liu, 2004). The success of the queuing network methodology in these domains is evidence of its efficacy as a model of human cognition and behavior.

Queuing Networks and Brain Activation

In this paper we attempt to model brain activation differences, as uncovered with neuroimaging, for different subject populations performing the same cognitive tasks. We will model these brain activation differences with a queuing network methodology and architecture. Rather than model changes in blood flow or volume, we attempt to model the underlying neural activity that drives the differences in blood flow as found empirically in these studies.

Positron Emission Tomography (PET) and functional Magnetic Resonance Imaging (fMRI) are two techniques used to measure blood flow changes in the brain, and both do so in different ways (Cabeza and Nyberg, 2000). fMRI measures blood flow changes via the Blood Oxygenated Level Dependent (BOLD) contrast and PET measures blood flow changes by marking blood with radioactive tracers (Cabeza and Nyberg, 2000). These hemodynamic processes are also correlated with the underlying neural activity of those brain areas exhibiting blood flow changes (Logothetis et al., 2001). This provides researchers with some confidence that while an indirect measure, these techniques can be used as to approximate underlying neural activity. In this paper we model the underlying brain activation moderating these hemodynamic changes and we do so with a queuing network framework.

3-Node Queuing Network (QN) Template

While there are many different types of queuing networks, this paper focuses on a simple 3-node queuing network with one server branching out into two parallel servers, as shown in Figure 1.

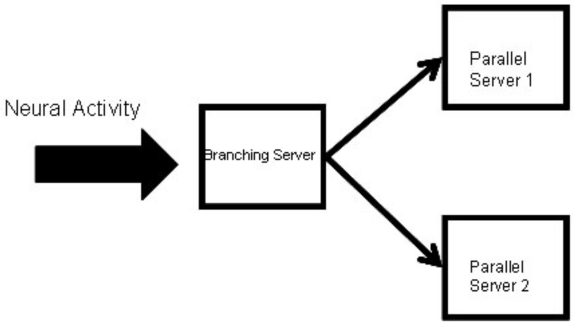

Figure 1: 3-node QN template

In all of these simulations, neural activity or neural spike trains are treated as the customers (C) in these networks that will be served by the queuing network servers in the network[1]. Each of the boxes in this network is a queuing network server (S) that provides a service to the customers that enter the network. Each of these servers will represent a unique brain area(s) that provides a unique service to the customers that enter it. Customers arrive at the branching server at some arrival rate λ customers/unit time. Once customers enter the network through the branching server, they receive some service from the branching server with a service rate of μ customers/unit time. Once customers complete service at the branching server they then travel probabilistically to either parallel server 1 or parallel server 2 for additional service. Each of these servers services customers with it's own service rate. Once customers complete service at either of those servers they have completed their full service and subsequently leave the network.

Each server also has a service capacity (the number of customers it can serve at a time) and a waiting capacity or queue capacity, which identifies how many customers can wait in front of the servers for service. The parallels between the queuing network methodology and the brain are apparent. First, it seems that different brain areas do in fact provide some unique function or service that mediate behavioral performance. Second, it is reasonable to assume that brain areas have capacity limitations in the amount of processing that they can accomplish and the speed with which they can process. Third it seems that information in the brain can be queued, as information that is not processed immediately is not immediately lost or discarded.

For these simulations we alter service rates for particular servers in the network, but we leave service capacities and queue capacities constant throughout, as it is beyond the scope of this paper to provide/hypothesize queue capacities for different brain areas. We therefore set each server's capacity to serve only one customer at a time with an infinitely large queue capacity (waiting line).

In all of these simulations we assume that each server provides a unique service and that customers need to be served by the branching server and only one of the parallel servers, but not both. Therefore, while each parallel server provides a unique service, each of the simulated tasks can be accomplished by traversing either of the parallel servers, which provide similar service.

Parameters for arrival rates, λ, were set based on the empirical parameters of the task. Service rate parameters were set based on neural evidence coming from research on aging and literacy. Zimmerman et al. (2006) have found significant gray matter volume reductions with increased age, and that these reductions have been correlated with reductions in executive functioning and working memory performance. Klingberg et al. (2000) have found that reductions in white matter anisotropy (connection fidelity), is strongly correlated with poor reading performance. We draw on these findings in setting our queuing network parameters for these simulations.

In addition one of the major assumptions of this paper is our hypothesized explanation for how neural activity flows or moves through brain neural networks. Neural activation flows in this network based on the comparative processing rates of the parallel servers, and is mediated by the following equation:

$$P_i = [\mu_i]/([\mu_i] + [\mu_j])(\text{Eq. } 1)$$

Where μ_i = the service rate for server i, and P_i is the probability of traveling to server i.

One can see that if the service rate of one parallel server in relation to the other server is much greater, then it is more

---

[1] Initial neural activation at the Branching Server is time-locked with empirical stimulus onsets.

likely that neural activity or spike trains would propagate to that server and vice versa.

There is also neural evidence that supports this routing equation. It is known that stronger synaptic connection strengths of an individual neural route, increase the probability (P_i) that neural spike trains (the customers in our network) travel through that route (Black, 1999; Chklovskiii et al., 2004; Habib, 2003). Synaptic connection weights can also be decomposed into waiting times and processing times of customers traveling a particular route (Wu, 2007). Therefore, equation 1 is supported by the aforementioned neural evidence.

Studies to Be Modeled

In this paper we model the neuroimaging results from 2 separate PET studies. These two studies were selected for two reasons. First each study found brain activity differences for different subject populations performing the same cognitive tasks. Second, in all the studies, subjects activated the same brain areas, in other words, different brain areas were not recruited for the different populations of subjects, only the distribution of the brain activation differed.

The first study, from Reuter-Lorenz et al. (2000), explored the difference in brain activity for old and young adults performing verbal and spatial Working Memory (WM) tasks. The second study from Petersson et al. (2000), investigated brain activity during a pseudoword generation task for illiterate and literate subjects. Though we have restricted the number of studies simulated in this paper, our results could conceivably be applied to many other studies of this kind.

All of our simulations were run for 30 minutes, roughly mimicking the total time of the empirical studies. In addition each simulation was run for 100 replications.

Study1: Reuter-Lorenz et al., 2000

In this study young and old subjects performed verbal and spatial working memory tasks. The authors used PET to identify the brain areas activated to perform these tasks and also explored the brain activation differences between the two groups of subjects. The major finding was that for verbal working memory young subjects showed substantial left lateralized frontal activation and for spatial working memory those subjects showed substantial right lateralized frontal activation. Older adults on the other hand showed bilateral frontal activation for both spatial and verbal working memory suggesting that older adults may be recruiting other brain areas to compensate for neural declines (Reuter-Lorenz et al., 2000). In addition, few differences were found in posterior activations for these two subject populations.

Simulation Parameters The queuing network template used to model these data can be seen in figure 2. The anterior and posterior brain areas that compose these servers can be seen in table 1.

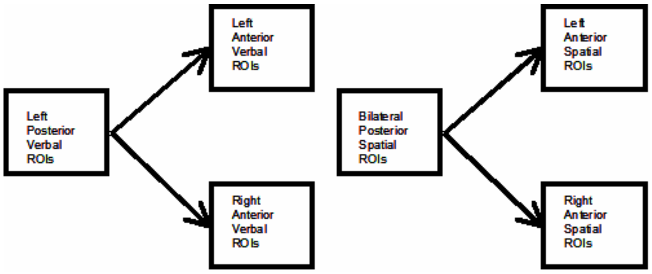

Figure 2: 3-node QN templates used to model Reuter-Lorenz et al. (2000) data

Table 1: Brain areas that compose the Queuing Network Servers as treated singularly by Reuter-Lorenz et al. (2000).

	Verbal Working Memory Task	Spatial Working Memory Task
Anterior ROIs	BA 45, 46, 10, 9 and 44 (Broca's); BA6(Supplementary Motor and premotor	BA 9, 46, 47 (DLPFC, VPFC); SMA and Premotor
Posterior ROIs	BA 40, 7 (parietal) and temporal sites BA 42 and 22	BA 40, 7 (parietal); BA 18, 19 (striate and extrastriate); BA 31 (Precuneus)

Arrival rates of the stimuli were set to be 5 seconds as this was the presentation rate of the trials to the subjects in the empirical study.

Service rate parameters were initially set to be exponentially distributed with a mean of 18 ms and have been validated by other researchers (Feyen, 2002; Wu, 2007). For young adults, service rate parameters were set in ways to show frontal lateralization. Therefore, for the verbal working memory task, the right anterior server's processing rate was treated as a free parameter and set to a value that would show lateralization (we used the same parameter value for the left anterior server for the spatial working memory task).

In addition, we feel it makes intuitive sense that left frontal areas should have disproportionately faster service rates for verbal tasks (compared to right frontal areas), and right frontal areas should have disproportionately faster service rates for spatial tasks (compared to left frontal areas), as these areas seem to be most active in the service of those respective tasks. If there were not such a difference in the processing abilities of these areas of cortex mediating performance in these tasks, we would not expect such robust lateralized activity.

For setting processing rates for older adults we used equations 2 and 3.

$$\text{Lateral frontal} = 67{,}043 - .47 * \text{Age} \quad \text{(Eq. 2)}$$

Note: the units are in mm^3

$$\mu_{subject}{}^{-1} = \mu_{control}{}^{-1} + \left(\frac{graymatter_{young} - graymatter_{old}}{graymatter_{young}} \right) * \delta$$

(Eq. 3)[2].

Equation 2, was provided by Zimmerman et al. (2006) and explains how gray matter volume in lateral frontal areas decreases with increased age. Equation 3, describes how gray matter volume changes for older adults translates into slower processing rates in lateral frontal areas. Here we are assuming that gray matter reductions reduce the abilities of those cortical areas to process information, and that this reduction in processing is additive to initial processing values. Tables 2 and list the parameters used to simulate the results from Reuter-Lorenz et al. (2000). Note: the mean age for young adults was 24, and for older adults was 69.

In addition, one may note that for older adults the service rates change for frontal areas, but not for posterior areas. With aging, there is more gray matter loss in frontal areas, compared to posterior areas, which can explain more deficits in planning, organizing and performing other executive functions with age (Zimmerman et al., 2006).

Table 2. Processing rates for old and young in the verbal WM task

Population	Older Adults	Younger Adults
Arrival rate (λ)[3]	1 every 5 sec	1 every 5 sec
Service rate Left Posterior Regions	Exponential mean 18 ms per neural spike train[4]	Exponential mean 18 ms per neural spike train
Service rate Right Anterior Regions	Exponential Mean 86 ms per neural spike train	Exponential mean 54 ms per neural spike train
Service rate Left Anterior Regions	Exponential Mean 50 ms per neural spike train	Exponential mean 18 ms per neural spike train

Table 3. Processing rates for old and young adults in the spatial WM task

Population	Older Adults	Younger Adults
Arrival rate (λ)	1 every 5 sec	1 every 5 sec
Service rate Left Posterior Regions	Exponential mean 18 ms per neural spike train	Exponential mean 18 ms per neural spike train
Service rate Right Anterior Regions	Exponential Mean 50 ms per neural spike train	Exponential mean 18 ms per neural spike train
Service rate Left Anterior Regions	Exponential Mean 86 ms per neural spike train	Exponential mean 54 ms per neural spike train

From tables 2 and 3 one can see how the initial imbalance between left and right service rates for the verbal and spatial tasks would cause more neural spike train activity to propagate to left anterior areas for the verbal task, and right anterior areas for the spatial task (see equation 1). Again, for older adults, service rate parameters in anterior areas were set based on equations 2 and 3.

Simulation Results Figure 3 displays the simulation results and the empirical results from the Reuter-Lorenz et al. (2000) study. The fits of our simulation results have an $R^2 = 0.64$ for the verbal working memory task, and an $R^2 = 0.72$ for the spatial working memory task.

The dependent variable that Reuter-Lorenz et al. (2000) report is the % change in brain activation for experimental working memory trials compared to control trials. For the experimental trials there was a higher working memory load compared to the control trials (roughly 4 times that of controls). Therefore in our simulations we altered processing by a scalar value (4) to reflect the changes in task demands from control trials to experimental trials. We report the changes in server utilization from control trials to experimental trials.

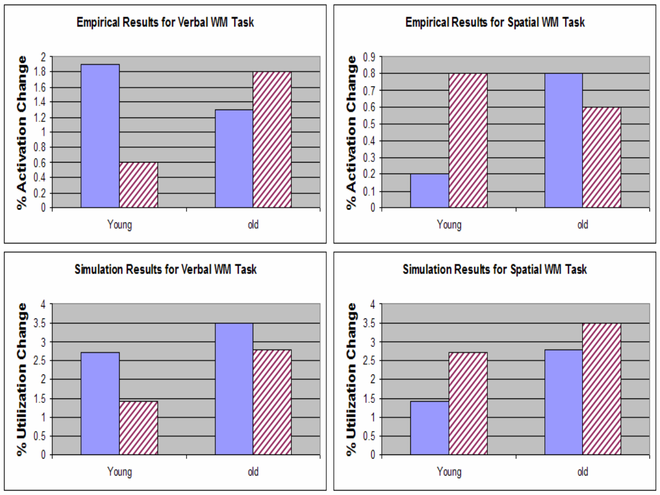

Figure 3: Empirical Results and simulation results for Verbal WM Task and spatial WM task from Reuter-Lorenz et al. (2000). Blue solid bars are left anterior areas, and magenta dashed bars represent right anterior areas. Top row shows the empirical results and the bottom row the simulation results[5]

From figure 3, one can see that our simulations do capture the empirical results well, especially the overall pattern of less lateralization with increased age. This reduced lateralization was due to processing declines mediated by gray matter loss, which reduced the ratio in processing rates of one parallel server relative to the other.

---

[2] δ is a scaling parameter was set to 100 for simulation 1 and set to 18 for simulation 2.

[3] The arrival rates were based on empirical stimulus presentation rates

[4] See Liu, Feyen and Tsimhoni (2006)

[5] The apparent reversal in lateralization for older adults in both the verbal and spatial working memory tasks was not significant

Study2: Petersson et al., 2000

In this study literate and illiterate participants performed a task where they needed to repeat verbally auditorily presented words and pseudowords. It was found that literate and illiterate subjects had similar behavioral performance in repeating words, but illiterate subjects were impaired in repeating pseudowords. It was also found that the neural networks supporting pseudoword repetition were different for the two groups, suggesting that learning to read causes functional changes in brain circuitry.

Here we concentrate on path weight differences (as found with Structural Equation Modeling; SEM) between inferior Parietal Cortex (iPC; BA 7/40) with Broca's area (BA 44) and iPC with prefrontal cortex (PFC; BA 45/46). The authors found that the path weight between iPC and Broca's was higher for literate subjects (by .18), while the path weight between iPC and PFC was higher for illiterate subjects (by .26). These path weight changes may reflect more efficient phonological loop processing for literate participants, and subsequently more reliance on executive processes for illiterate subjects to perform the pseudoword repetition task. Note: we report correlations rather than path weights, but the path weights were based on the correlation matrix of the empirical study.

Table 4. Processing rates for Literate and Illiterate Subjects

Population	Literate Subjects	Illiterate Subjects
Arrival rate (1/lambda)	U(6, 1) sec	U(6, 1) sec
Service rate iPC	Exponential mean 18ms per spike train	Exponential mean 18ms per spike train
Service rate Brocas	Exponential mean 18ms per spike train	Exponential mean 29 ms per spike train
Service rate PFC	Exponential mean 27 ms per spike train	Exponential mean 27 ms per spike train

Simulation Parameters Table 4 lists the parameters that were used to simulate the data from Petersson et al., (2000). Arrival rates were set based on empirical parameters, where stimuli were presented every 6 seconds. However, we needed to include some variance so that we could calculate the correlation of neural activations in our queuing network servers.

We set parameters for literate subjects in a similar manner to that of our simulations of Reuter-Lorenz et al. (2000). We treated service rate in the PFC as a free parameter as younger adults are biased to utilize the route connecting iPC and Brocas over iPC and PFC.

For setting parameters for illiterate subjects we depended on differences in white matter anisotropy. While there may be gray matter volume differences between literate and illiterate subjects, we were guided by white matter anisotropy (connection integrity) differences between good and poor readers (Klingberg et al., 2000).

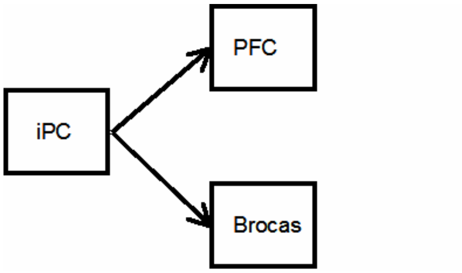

Figure 4. 3-Node queuing network used to simulate Petersson et al. (2000) data

Klingberg et al. (2000) found that white matter anisotropy in a volume connecting parietal and temporal cortices in the left hemisphere was significantly reduced in poor readers compared to normal readers, and that anisotropy in this region was significantly correlated with reading performance. We used this significant reduction in anisotropy connecting temporal and parietal cortex to change the processing rate of Broca's area for illiterate subjects as this white matter region would be connecting iPC with Broca's. We assumed that this reduction in anisotropy in this region would alter the relationship between iPC and Broca's area for illiterate subjects.

Klingberg et al. (2000) found that anisotropy in this white matter region was correlated with reading performance $r = 0.84$. Using behavioral data from the Petersson et al. (2000) study (via Castro-Caldas et al., 1998) and Klingberg et al.'s (2000) regression equation we calculated the anisotropy values for literate and illiterate subjects. These calculations yielded a 60% decrease in anisotropy for illiterate subjects compared to literate. Using Equation 3., we substituted white matter changes between literate and illiterate subjects (instead of gray matter) to obtain the service rate for Broca's area for illiterate subjects.

Simulation Results The simulation results for this task are summarized in Table 5. We obtained an $R^2 = .96$ for these simulation data (pitting our simulated correlation values against Petersson et al. (2000) path weights). One can see the changes in correlated activity match the pattern of differences in path weights exhibited from the Petersson et al. (2000) study where literate participants exhibit increased correlations between iPC and Broca's compared to illiterate subjects, and illiterate participants exhibit increased correlations between iPC and PFC compared to literate participants. We did not find the same magnitude increase in correlated activity between iPC and PFC as was found empirically, however, we did find the same overall pattern. These results indicate good coherence between our simulation and the empirical findings.

Table 5. Simulation and Empirical Results from Petersson et al. (2000)

	Empirical	Simulation
Increase in iPC and Brocas Relation for Literate compared to Illiterate	+. 18	+. 10
Decrease in iPC and PFC Relation for Literate compared to illiterate	-.26	-.10

Conclusion

In sum, our 3-node queuing network templates were able to successfully model the activity of brain networks for different populations of subjects performing the same cognitive tasks. We drew on neuroscience evidence in selecting parameters and explained changes in brain networks as being caused by relative differences in service rates, which alter neural activation propagation. We hope that with the queuing network architecture we will be able to understand more complicated brain networks and make new predictions for the behavior of brain networks that underlie human cognition.

Acknowledgments

This research was supported in part by an NSF graduate fellowship to MGB and an NSF grant to YL.

References

Black, I.B. (1999). Trophic regulation of synaptic plasticity. *Journal of Neurobiology*, 41 (1), 108-118.

Bressler, S. L. (1995). Large-scale cortical networks and cognition. *Brain Research Reviews.* Vol. 20, 288-304.

Cabeza R. and Nyberg L.(2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *J of Cog Neuro.* 12 (1): 1-47

Castro-Caldas A., Petersson KM, Reis A, Stone-Elander S, Ingvar M. (1998). The illiterate brain - Learning to read and write during childhood influences the functional organization of the adult brain. *BRAIN* 121(6): 1053-1063

Chklovskii, D.B., Mel, B.W., and Svoboda, K. (2004). Cortical rewiring and information storage. *Nature*, 431 (7010), 782-788.

Feyen R. (2002). Modeling Human Performance Using the Queuing Network - Model Human Processor (QN-MHP). Unpublished Dissertation, University of Michigan, Ann Arbor, Michigan.

Habib, M. (2003). Rewiring the dyslexic brain. *Trends in Cognitive Sciences*, 7 (8), 330-333.

Klingberg, T., Hedehus, M., Temple, E., Salz, T., Gabrieli, J. D. E., Moseley, M. E., et al. (2000). Microstructure of temporo-parietal white matter as a basis for reading ability: Evidence from diffusion tensor magnetic resonance imaging. *Neuron, 25*(2), 493-500.

Labatut, V., Pastor, J., Ruff, S., Demonet, J., Celsis, P. (2004). Cerebral modeling and dynamic Bayesian networks. *Artificial Intelligence in Medicine.* Vol. 30, 119-139.

Lashley, K.S. (1931). Mass Action in Cerebral Function. *Science.* Vol. 73(1888), 245-254.

Lim, J., and Liu, Y. (2004). A Queueing Network Model for Visual Search and Menu Selection. *Proceedings of the 48th Annual Conference of the HFES.*

Liu, Y. L. (1996). Queueing network modeling of elementary mental processes. *Psychological review, 103*(1), 116-136.

Liu, Y. L. (1997). Queueing network modeling of human performance of concurrent spatial and verbal tasks. *Ieee Transactions on Systems Man and Cybernetics Part A-Systems and Humans, 27*(2), 195-207.

Liu, Y., Feyen, R., & Tsimhoni, O. (2006). Queueing Network-Model Human Processor (QN-MHP): A Computational Architecture for Multi-Task Performance in Human-Machine Systems. *ACM Transactions on Computer-Human Interaction.*

Logothetis, NK; Pauls, J; Augath, M; Trinath, T; Oeltermann, A. 2001. Neurophysiological investigation of the basis of the fMRI signal. *NATURE* 412 (6843): 150-157.

Nyberg, L., & McIntosh, A. R. “Functional Neuroimaging: Network Analyses.” *Handbook of Functional Neuroimaging of Cognition.* Eds. Roberto Cabeza and Alan Kingstone. A Bradford Book: MIT Press, 2001. 49-72.

Pastor, J., Ruff, S., Demonet, J., Celsis, P. (2004). Cerebral modeling and dynamic Bayesian networks. *Artificial Intelligence in Medicine.* Vol. 30, 119-139.

Petersson, K. M., Reis, A., Askelof, S., Castro-Caldas, A., & Ingvar, M. (2000). Language processing modulated by literacy: A network analysis of verbal repetition in literate and illiterate subjects. *Journal of cognitive neuroscience, 12*(3), 364-382.

Reuter-Lorenz, P. A., Jonides, J., Smith, E. E., Hartley, A., Miller, A., Marshuetz, C., et al. (2000). Age differences in the frontal lateralization of verbal and spatial working memory revealed by PET. *Journal of cognitive neuroscience, 12*(1), 174-187.

Wu C. (2007). Queueing Network Modeling of Human Performance and Mental Workload in Perceptual-Motor Tasks. Unpublished Dissertation, University of Michigan, Ann Arbor, Michigan.

Wu, C., & Liu, Y. (2004). Modeling Human Transcription Typing with QN-MHP (Queueing Network - Model Human Processor). *Proceedings of the 48th Annual Conference of the HFES.*

Zimmerman, M. E., Brickman, A. M., Pau, R. H., Grieve, S. M., Tate, D. F., et al. (2006). The relationship between frontal gray matter volume and cognition varies across the healthy adult lifespan. *American Journal of Geriatric Psychiatry, 14*(10), 823-833.