Perceptions of Temporal Synchrony in Multimodal Displays

Wayne Giang, Ehsan Masnavi, and Catherine M. Burns

Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada

Current multimodal interfaces make use of several intra-modal perceptual judgements that help users "directly perceive" information. These judgements help users organize and group information with little cognitive effort. Cross-modal perceptual relationships are much less commonly used in multimodal interfaces, but could also provide processing advantages for grouping and understanding data across different modalities. In this paper we examine whether individuals are able to directly perceive cross-modal auditory and tactile temporal rate synchrony events. If direct perception is possible, then we would expect that individuals would be able to correctly make these judgements with very little cognitive effort. Our results indicate that individuals have difficulty identifying when the temporal rates of auditory and tactile stimuli in a monitoring task are synchronous. Changes in workload, manipulated using a secondary visual task, resulted in changes in performance in the temporal synchrony task. We concluded that temporal rate synchrony is not a perceptual relationship that allows for direct perception, but further investigation of cross-modal perceptual relationships is required.

INTRODUCTION

The concept of affordances, as described by Gibson (1979), refers to a relationship between an environment and an actor whom perceives and interacts with it. An affordance provides the actor with a set of actions or capabilities which are unique to that actor. For example, a cat door affords passage to cats due to its nature as an opening and its size, but it does not offer the same affordance to a human. When affordances exist in the environment and they are coupled with "information in the environment that uniquely specifies that affordance" (McGrenere & Ho, 2000, p. 189), direct perception is possible. Direct perception allows an actor to bypass mediation and internal processing and proceed directly with an action. Thus, if a cat is able to directly perceive the door due to its salient colour and shape, then it would be able to enter the house without overtaxing its feline brain.

In the realm of interface design and human factors engineering, direct perception is desirable because it offers processing advantages over information processing strategies that involve higher degrees of cognitive control (Vicente & Rasmussen, 1992). Ecological Interface Design (EID) is an interface design methodology that makes use of the concepts of affordances and direct perception for analyzing and designing interfaces for complex systems. Using the EID framework, interface designers aim to create links between the affordances of the system (i.e., what the interface can change and manipulate in the larger system) and the perceptual interface elements that operators in complex systems interact with. The linkage is created by representing system constraints and relationships as perceptual judgements. As a result of this, operators are able to directly perceive and interact with the larger system through the perceptual forms in the interface. The EID technique has been used to great effect in visual interfaces where complex cognitive interpretations of data have been reduced to simple comparisons of visual orientation, shape, and size (see Burns & Hajdukiewicz, 2004 for examples).

The use of EID for non-visual displays is still not commonplace. However, with the development of new interface technologies, such as touch-screens and tactile displays, future interfaces will most likely present and receive information to users through multiple sensory modalities. A number of researchers have already addressed why multimodal interfaces would improve operator performance. In a review on multimodal information presentation, Sarter (2006) states that information synergy, redundancy, and increased bandwidth are all benefits of multimodal displays. By using additional sensory modalities, the "design-space" for interface designers is increased, and new combinations of information presentation can be achieved.

Another often cited advantage of multimodal displays is the access to separate pools of attentional resource as described by Multiple Resource Theory (Wickens & McCarley, 2008). Separating information presentation into different sensory modalities could allow for better workload management and concurrent processing of different streams of information. While more recent research has shown that the different sensory modalities are not as independent as stated by MRT, even these cross-modal linkages can be leveraged to create better alarms and alerts (Ferris & Sarter, 2008; Spence, 2010). Finally, different sensory modalities have characteristics that best align themselves to different types of abstract data (Nesbitt, 2004). For example, vision is often used to show and compare spatial information, while temporal information is often mapped into auditory stimuli.

Taken together, these benefits for using multimodal information presentation have motivated interface designers to move beyond the visual affordances and to explore perceptual affordances in other modalities such as audition and touch. While this research is still limited, EID has been used in the design of auditory sonifications (Sanderson, Anderson, & Watson, 2000) and haptic interfaces (Arrabito, Ho, Au, Keillor, Rutley, Lambert, & Hou, 2009). Past EID interfaces have explored the use of new perceptual judgements in their respective modalities that could be used by interface designers to facilitate direct perception.

However, one area that is still unexplored is the types of cross-modal relationships that could be used by interface designers. Multimodal interfaces exist as entities beyond their individual modalities, yet many current interfaces have limited interaction between the different sensory modalities used. The affordances and perceptual tools currently available are largely intra-modal, which limits how interface designers are able to organize information that is presented in different modalities. Without these perceptual links, information in each modality must first be understood at a higher level of cognition before it is compared or grouped with information in different modality. Thus, to effectively extend EID to multimodal interfaces, we must have a better understanding of the types of cross-modal perceptual relationships individuals are able to make and whether these judgements can be used to create cross-modal affordances.

There are a number of possible cross-modal perceptual relationships that could facilitate direct perception of cross-modal affordances such as information grouping, comparison, or equivalence. The most common type of cross-modal relationship comes in the form of cross-modal matching tasks. These tasks require the observer to judge whether a stimulus in another modality is equal, along some dimension, to the one being perceived. The judgement of temporal synchrony is one example of a cross-modal matching task which may prove to be suitable for direct perception. Time is a common dimension that is shared between all modalities, and it is also one of the major contributors to multisensory integration (Stein & Meredith, 1990). It is also one of the most ubiquitous cross-modal matches that individuals make. From ages as early as 2 months, individuals are able to identify when stimuli in different modalities share temporal characteristics (Lewkowicz, 2000).

In this paper we examine whether or not judgements of temporal rate synchrony could be used to design interfaces that facilitate direct perception. If temporal rate synchrony could be used to show cross-modal affordances representing information grouping or equivalence, then we would expect that individuals would be able to correctly make these judgements with very little cognitive effort. Thus, performance in a temporal rate synchrony task would be highly resistant to changes in workload if individuals are directly perceiving the synchrony events rather than perceiving the information in each modality and then comparing their synchrony at a higher cognitive level. We examine this by having participants make judgements about the synchrony of auditory and tactile stimuli in a monitoring task, while changing workload through the use of a secondary visual task. We hypothesize that performance in the cross-modal synchrony task will be similar under both high and low workload conditions.

METHOD

Participants

A total of twenty-seven undergraduate and graduate students were recruited for this study of which twenty-four sets of data were used for further analysis. All participants had normal or corrected-to-normal vision and normal or corrected-to-normal hearing. In addition, each of the participants in the study was right-handed. One participant's data was lost due to a software malfunction. Two participants' data were removed due to poor performance in temporal synchrony task. All three participants were replaced and their data were excluded from the subsequent analysis. All participants were compensated for their time.

Apparatus and Stimuli

Auditory and tactile stimuli of varying temporal rhythms were used as the stimuli for the primary experimental task. The auditory stimuli consisted of 200 Hz pure tones played from a set of headphones which were repeated at 5 different temporal rates. The tactile stimuli consisted of a set of EAI C2 tactors secured onto the outsides of the wrists of the participants which vibrated at 250 Hz and were also repeated at 5 different temporal rates at moderate intensities.

The 5 different levels of the temporal rate condition were represented by auditory and tactile signals which were repeated at 5 different rates. Each signal was broken into 2 second units, within which the rate of the signal was varied by changing the number of "beats" (when the auditory or tactile signal is turned on). Each beat had a duration of 0.3 seconds. The fastest rate contained 5 beats within the 2 second unit, with separations of 0.4 seconds between the onsets of each beat. The other rates consisted of 4, 3, 2, and 1 beat(s) distributed evenly within the 2 second unit.

The individual temporal rate units were combined into longer auditory and tactile "streams" which were used to represent more complex monitoring tasks that would be found in multimodal interfaces and human supervisory control situations. These streams were generated into scenarios which contained "temporal rate synchrony events" when the rates in both the auditory and tactile streams were the same, as shown in Figure 1. Three 10 minute scenarios were generated, each containing 10 temporal rate synchrony events for each of the 5 rates. The synchrony events accounted for roughly 16.67% of the scenario. No synchrony events occurred immediately after another synchrony event. A fourth 2 minute scenario was generated for training purposes.

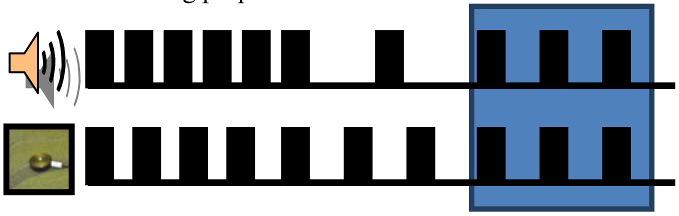

Figure 1: A temporal rate synchrony event embedded within an auditory and tactile stream.

A secondary visual task was also used within this experiment. The visual task consisted of monitoring a display adapted from Sethumadhavan (2009). A series of numbers between 100 and 199 appeared on the computer monitor at either 2 second intervals for the high workload condition, or 6 second intervals for the low workload condition.

Participants were seated in front of a 22 inch liquid crystal display monitor that displayed information pertaining to the temporal synchrony task, and a laptop with a 14 inch display which displayed the visual task. Participants responded to the tasks using two keyboards with clearly marked buttons for responses. One button, used for the temporal synchrony task, was controlled using their left hand, and the other button, for the visual task, was controlled using their right hand. The experimental software was created using the open-source PsychoPy framework (Pierce, 2007).

Design and Procedure

The experiment was a 5 (temporal rate type) x 3 (visual task workload: high vs. low vs. no visual task) within subjects design, with workload as a blocked variable and temporal rate type presented within each block. In this paper, only the workload factor was analyzed. The workload condition was manipulated using a secondary visual task. The dependent measures were the percentage of hits, misses, false alarms, correct rejections and response times for both the temporal synchrony task and the visual task. In addition, participants were asked to fill out questionnaires about their perceived performance, workload, strategies, and how they perceived the stimuli.

The experiment was divided into three blocks, one for each of the workload conditions. The order of the workload blocks and the temporal synchrony scenarios was counter balanced to control for learning effects. However, the first session completed was always the no visual task condition. This was done to ensure that participants were comfortable with the temporal synchrony task. Each workload condition was run as a separate 10 minute block and participants were given a break between each of the blocks to reduce the risk of fatigue and the effect of adaptation to the tactile stimuli.

Each participant was given a training session to familiarize the participant with the experiment stimuli and the experimental tasks. The training session consisted of a tactile and auditory familiarization activity and a practice block of the temporal synchrony task without the visual task, and then the visual task without the temporal synchrony task. If the participant felt that they needed additional practice, they were allowed to repeat any of the training activities.

After completing the practice session, the participants began the experimental tasks. In the temporal synchrony task, participants monitored both the auditory and tactile streams for occurrence of cross-modal temporal rate synchronies. When the participant identified one of these synchrony events, they responded by hitting a button with their left hand. In the visual task, participants were asked to monitor the magnitude of a number visually displayed on the screen. Participants performed this visual task in addition to monitoring for temporal rate synchronies. The participant was required to respond to the visual task whenever the number displayed was below or equal to 130 or above or equal to 170. They accomplished this by hitting a button with their right hand. Participants were asked to fill out the questionnaire at the end of each experimental block and they were given a short break before continuing with the experiment.

RESULTS

During the experimental tasks, each stimuli presentation (auditory-tactile pair and visual event) was logged. The participant's response and its correctness, based on the current stimuli presented, was also recorded. However, many participants reported that they responded after the stimuli were presented. Subsequent analysis revealed that many of the responses during the temporal synchrony task actually occurred after the two second stimuli-presentation window. An adjustment was applied to the temporal synchrony task results; responses that occurred during the first 0.75 seconds of a stimuli presentation are attributed to the previous stimuli presentation. The delays in responses may be attributed to the lack of distinguishable breaks between the presentations of different rate "units". No adjustment was made for the visual task. Post-hoc comparisons were done using a Bonferroni correction and the p-values reported for the post-hoc tests are taken from SPSS's Bonferroni adjusted p-values. All error bars represent 95% confidence intervals.

Temporal Synchrony Task

Overall, performance was relatively low for the temporal synchrony task. Across all conditions, the mean hit rate (the number of correct detections of temporal rate synchronies divided by the total number of targets present in the scenario) was 0.441 with a standard deviation of 0.175. A one-way repeated measures ANOVA was conducted with workload (high vs. low vs. no visual task) as the independent factor. This revealed that the hit rate differed significantly between the different levels of workload, $F(2,46)=9.074$, $p<.001$ (Figure 2). Post-hoc tests showed that the no visual task workload condition ($M=0.488$, $SD=0.168$) produced higher hit rates than the high workload condition ($M=0.395$, $SD=0.183$), $p=.002$. The low workload condition ($M=0.441$, $SD=0.169$) did not differ significantly from the no visual task ($p=.144$, *ns*) condition, and was only marginally different from the high workload ($p=.069$). It is important to note that the majority of the stimuli that were presented to the participants were of non-synchrony events, and only ~17% of the events encountered were synchronous.

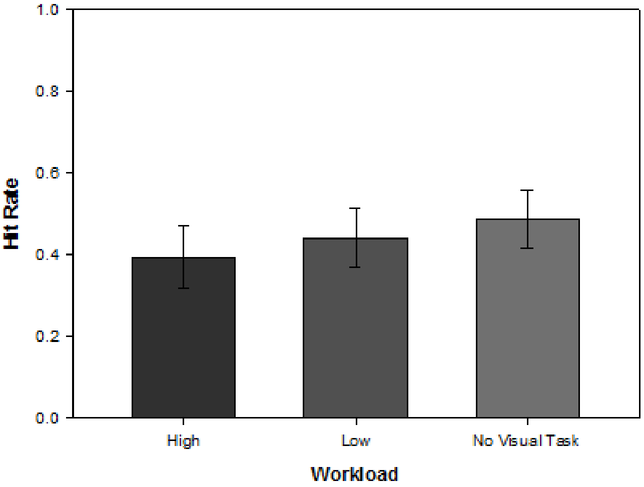

Figure 2: Hit Rate for Temporal Synchrony Task

A similar analysis was conducted for false alarm rate (the number of false positives divided by the total number of non-synchronous stimuli presentations). The one-way repeated measures ANOVA did not reveal any significant differences between the three workload conditions, $F(2,46)=2.338$, $p>.05$. The mean false alarm rate over all the conditions was 0.061 with a standard deviation of 0.043.

Signal detection theory was used to further analyze the results of the temporal synchrony task. Signal detection indices for sensitivity (d') and criterion (c) were calculated for each workload condition (high vs. low vs. no visual task). In the temporal synchrony task, sensitivity referred to the ability for the participant discriminate between stimuli with rate synchrony and stimuli that were not synchronous. When the participants had a hit rate or false alarm rate of 1 or 0 a correction of either 1-1/(2N) or 1/(2N) was used, where N was either the total number of temporal synchrony events or total number of non-synchrony events (Macmillan & Creelman, 1991).

A one-way repeated measures ANOVA showed that participant's sensitivity (d') differed between workload conditions, *F*(2,46)= 7.913, *p*=.001 (Figure 3). Post-hoc comparisons revealed that participant's performed with lower sensitivity in the high workload condition (*M*=1.295, *SD*=0.747) when compared to the no visual task condition (*M*=1.617, *SD*=0.662), *p*=.009, and the low workload condition (*M*=1.555, *SD*=0.662), *p*=.013. No other comparisons were significant.

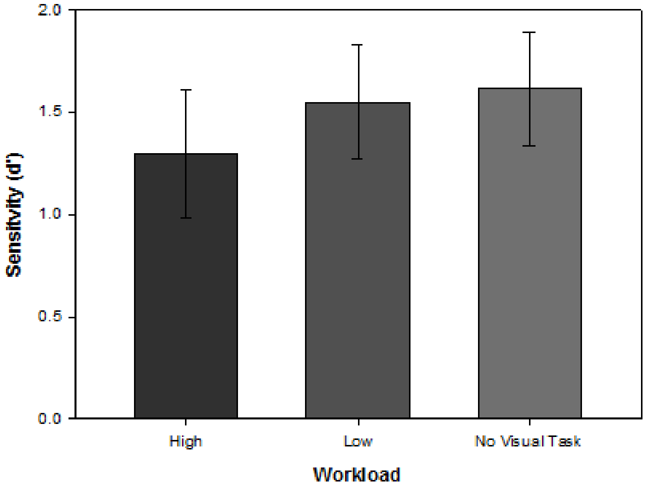

Figure 3: Sensitivity (d') for Temporal Synchrony Task

As was done for sensitivity, a one-way repeated measures ANOVA was conducted for the criterion values. The results showed that the participants' decision criterion differed between workload conditions, *F*(2,46)= 5.451, *p*= .008. The criterion values for each condition were all less than 0, which meant that participants adopted risky decision biases. However, post-hoc comparisons showed that participants' responses during the high workload condition (*M*= -.952, *SD*=0.278) used a much riskier decision criterion than the responses during the no visual task condition (*M*= -0.8412, *SD*=0.241). The low workload condition (*M*= -0.946, *SD*=0.305) did not differ from the high workload condition (*p*=1.00, *ns*) and was only marginally different than the no visual task conditions (*p*=.065).

Visual Task

Overall performance was much higher in the visual task for both hit rate (*M*=0.927, *SD*=0.109) and false alarm rate (*M*=0.057, *SD*=0.035). A paired samples t-test for hit rate revealed that the hit rate for the high workload condition (*M*=0.887, *SD*=0.125) was significantly lower than the hit rate for the low workload condition (*M*=0.966, *SD*=0.072), *t*(23)= -6.125, *p*<.001. A second paired samples t-test for false alarm rate also found a significant difference between the high (*M*=0.0712, *SD*=0.028) and low (*M*=0.0432, *SD*=0.036) workload conditions, *t*(23)=3.761, *p*=.001.

Signal detection analysis was also used to analyze the visual task by calculating sensitivity (d') and criterion c. A paired samples t-test was conducted to examine the effects of workload on sensitivity (d'). As to be expected, the test revealed that the high workload condition (*M*=2.820, *SD*=0.553) reduced the ability of participants to detect the critical visual stimuli (numbers below or equal to 130 or above or equal to 170) when compared to the low workload condition (*M*=3.796, *SD*=0.599), *t*(23)=-10.555, *p*<.001. Similarly, a paired samples t-test on the effects of workload on criterion c showed that participants adopted a riskier response bias in the high workload condition (*M*=-0.085, *SD*=0.233) than the low workload condition (*M*=0.097, *SD*=0.270), *t*(23)=-3.917, *p*=.001.

Questionnaire

At the conclusion of each experimental session, participants were asked to complete a questionnaire on the difficulty of the experimental tasks and their strategies; the questionnaire contained other questions, but they were not analyzed for this paper. Each question was answered using a 7-point scale, and answers were coded from 0 to 6.

Participants were asked to judge the difficulty of both the temporal synchrony task and the visual task for each of the workload conditions (high vs. low vs. no visual task) using a scale from "*extremely easy*" (coded as 0) and "*extremely difficult*" (coded as 6). The results of the Friedman test suggested that the high (median=4), low (median=4), and no visual task (median=4) workload conditions did not significantly change the participant's judgements of the difficulty of the temporal synchrony task, $\chi^2(2)=3.354$, *p*=.187. However, for the visual task, a Wilcoxon signed-ranked test revealed that participants found the high workload condition (median=2) more difficult than the low workload condition (median=1), *Z*= -3.132, *p*=.002. This showed that the workload manipulation using the visual task worked in its intended direction.

In the written portion of the questionnaire, many of the participants mentioned that they found the temporal synchrony task very difficult in all conditions, and stated that it required a large degree of concentration, attention, and effort. With the addition of the visual task, most participants reported that they made use of a task switching strategy where they would attempt to finish off one task (such as the visual task) before switching their attention to the other task. This was much easier to accomplish in the low workload condition than in the high workload condition where both the visual and tactile-auditory stimuli switched at two second intervals.

DISCUSSION

The results of the temporal synchrony task suggest that participants were affected by the secondary visual task. During the high workload condition, participants' performance was much lower than the no visual task condition for both hit rate and sensitivity. However, the low workload condition did not result in a lower hit rate when compared to the no visual task condition. In addition, participants in the low workload condition were much more sensitive to the temporal synchrony events than in the high workload condition and the differences between the low and high workload conditions for hit rate were approaching significance. If participants were able to directly perceive the temporal synchrony events, then we would expect no differences in performance in the different workload conditions. However, the current data does not support this hypothesis, and suggests that the participant's ability to perceive and respond to the temporal synchrony events were dependent on their current level of work.

In the visual task, under the high workload condition, participants had a lower percentage of correctly detected targets. Participants were also less able to detect the critical visual events, and they were riskier with their target designations, allowing for more false-alarms. This suggests that participants may have been partially compensating for the higher workload by spending less time on the visual task. This may have mitigated some of the detrimental effects of workload on the temporal synchrony task. Thus, the actual differences between performance on high and low workload conditions in the temporal synchrony task may be even larger if the amount of effort spent on the visual task was kept constant.

Given these findings, temporal rate synchrony, as it was tested in this study, does not appear to be a perceptual relationship that can lead to direct perception. Contrary to our original hypothesis, participants showed higher levels of performance in high workload task when compared to the low workload and no visual task conditions. There are a number of reasons why this may have occurred. Firstly, the length of the 2 second perceptual "units" may have made this task more difficult by forcing the participants to make use of working memory. By simplifying the detection task to a matching auditory and tactile onset and duration, participants may respond in a manner that was more indicative of direct perception.

Secondly, Multiple Resource Theory suggests that the same pool of resources is used for perceptual and cognitive tasks (Wickens & McCarley, 2008). Thus, the temporal synchrony task, which was presumed to be a highly perceptual task, and the visual task, which required participants to use working memory to make judgements about numbers, would draw from the same pool of resources even though the information was presented in different modalities. The fact that the temporal synchrony task may also have drawn heavily on working memory only increases the amount of interference between the two tasks.

Thirdly, both tasks required manual responses from the participants. Even if judgements of temporal synchrony lead to direct perception, participants may have experienced interference between the two tasks at the response selection stage which draws from the same pool of resources (Wickens & McCarley, 2008). Thompson, Tear, and Sanderson (2010) examined differences between responding using mental count (a larger load on working memory) and a physical clicker (greater motor demand) in a study on multisensory integration while walking. Their results suggested that participants did worse on their primary multisensory task when using the clicker than when responding using a mental count, which was contrary to their original hypothesis. One possible explanation was that using the physical clicker might have interacted with a secondary button-press task, and increased the workload of the tasks overall. A similar effect may have forced participants to direct attention away from the temporal synchrony task more often in the high workload task due to an increased number of manual responses required.

Overall, it is evident that participants were unable to perform both the temporal synchrony task and the visual task concurrently and a bottleneck existed which prevented the participants from directly perceiving the temporal synchrony stimuli from this study. Reducing the amount of working memory required for the temporal synchrony task by simplifying it to detections of synchrony onset and duration may still prove that individuals are able to intuitively parse and group stimuli in different modalities together with low cognitive load. Despite the findings of this study, the search for cross-modal relationships will still be important for multimodal interface designers. Interface designers must be able to leverage pre-existing relationships between our senses in order to represent the multitude of abstract data that is often used in complex systems. Finding ways of intuitively linking information across modalities will help contribute to multimodal interfaces that are holistic and easier to understand.

ACKNOWLEDGEMENTS

This work was funded by NSERC.

REFERENCES

Arrabito, G.R., Ho, G., Au, H., Keillor, J.M., Rutley, M., Lambert, A., & Hou, M. (2009). Proposed Techniques for Extending Ecological Interface Design to Tactile Displays: Using Tactile Cues to Enhance UAV Interface Design. In *Proceedings of the 53rd Annual Meetings of the Human Factors and Ergonomics Society*. pp. 76-80.

Burns, C. M., & Hajdukiewicz, J. (2004). *Ecological Interface Design*. Boca Raton, FL: CRC Press.

Creelman, C.D., & Macmillan, N.A. (1991). *Detection Theory: A User's Guide (2nd ed.)*. New York: Cambridge University Press.

Ferris, T., Sarter, N. (2008). Cross-modal links among vision, audition, and touch in complex environments. *Human Factors*, 50(1), pp. 17-26.

Gibson, J. J. (1979) *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lewkowicz, D.J. (2000). The Development of Intersensory Temporal Perception: An Epigenetic Systems/Limitations View. *Psychological Bulletin*. 126(2). pp. 281-308.

McGrenere, J., Ho, W. (2000). Affordances: Clarifying and Evolving a Concept. Proceedings of GI 2000, pp. 179-186.

Nesbitt, K.V. (2004). MS-Taxonomy: A Conceptual Framework for Designing Multi-sensory Displays. In *Proceedings of the Information Visualization, Eighth International Conference*. pp. 665-670.

Peirce, J. W. (2008). PsychoPy – Psychophysics software in Python. *J Neurosci Methods*, 162(1-2), pp. 8-13.

Sanderson, P., Anderson, J., & Watson, M. (2000). Extending ecological interface design to auditory displays. In *Proceedings of OzCHI2000*. pp. 259-266.

Sarter, N. B. (2006). Multimodal Information Presentation: Design Guidance and Research Challenges, *International Journal of Industrial Ergonomics*, 36, pp. 439-445.

Sethumadhavan, A. (2009). Effects of Levels of Automation on Air Traffic Controller Situation Awareness and Performance. (Doctoral dissertation). Retrieved from http://etd.lib.ttu.edu/theses/available/etd-12112008-095502/unrestricted/Sethumadhavan_Arathi_diss.pdf

Spence, C. (2010) Crossmodal Spatial Attention. Annals of the New York Academy of Sciences, 1191(1), pp. 182-200.

Stein, B.E., Meredith, M.A. (1993). *The Merging of the Senses*. Cambridge: MA: The MIT Press.

Thompson, M., Tear, M., & Sanderson, P. (2010). Multisensory integration with a head-mounted display: Role of mental and manual load. Human Factors: The Journal of Human Factors and Ergonomics Society, 52(1), 92-104.

Vicente, K., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. IEEE Transactions on Systems Man and Cybernetics, 22(4), 589-606.

Wickens, C.D. & McCarley, J. S. (2008). *Applied Attention Theory*. Boca Raton, FL: CRC Press.