VISUAL VS AUDITORY COUPLING IN DYADS UNDER DIFFERENT TASK DIFFICULTY

A Thesis
Presented to
The Academic Faculty

by

Adam F. Werner

In Partial Fulfillment
of the Requirements for the Degree
MASTERS OF SCIENCE in the
SCHOOL OF PSYCHOLOGY

Georgia Institute of Technology
MAY 2019

VISUAL VS AUDITORY COUPLING IN DYADS UNDER DIFFERENT TASK DIFFICULTY

Approved by:

Dr. Jamie Gorman, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Bruce Walker
School of Psychology
Georgia Institute of Technology

Dr. Christopher Wiese
School of Psychology
Georgia Institute of Technology

Date Approved: [April 15, 2019]

To my wife, Katherine. Thank you for making my dreams possible.

ACKNOWLEDGEMENTS

I would like to thank our research assistants, Lauren, Anya, and David. I would also like to thank my advisor, Jamie, for all the help in developing and writing this thesis. Thank you for letting me conduct research that is very interesting to me and extremely relevant to my experiences as an Army Aviator. Lastly, I want to recognize and thank the Soldiers of Alpha Company, 3-1 Task Force Gunfighter for your inspiration in conducting this experiment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Due to lack of visual or auditory perceptual information, many tasks require interpersonal coordination and teaming. Dyadic verbal and/or auditory communication typically results in the two people becoming informationally coupled. Previous research suggests that coupling between two individuals can take place auditorily or visually during intentional and unintentional tasks (i.e., Richardson, Marsh, & Schmidt, 2005; Gorman, Amazeen, Crites, & Gibson, 2017). This experiment examined coupling by using a two-person remote navigation task where one participant blindly drove a remote-controlled car while another participant provided auditory, visual, or a combination of both informational cues (bimodal) to navigate the driver. Under these three perceptual-motor coupling conditions, participants' performance was evaluated using easy, moderate, and hard task difficulty conditions. I predicted that the visual coupling condition would have higher performance measures overall, and the bimodal (combination of auditory and visual cues) coupling condition would have higher performance as difficulty increased. Results indicated that visual coupling performs best overall. When auditory coupling is used (auditory and bimodal conditions), medium difficulty had worse performance compared to hard difficulty, an unexpected result. This result can be attributed to the frequency at which teams verbally communicate. Though intuitive, the faster teams speak, the better they perform. Applications within team coordination and potential theories that could explain cue rate results and poorer performance at medium compared to hard difficulty is discussed.

Keywords: Auditory coupling, visual coupling, gestures, team communication

CHAPTER 1. INTRODUCTION

At altitudes above 10,000 feet Mean Sea Level, military helicopters in Afghanistan must be able to insert troops and cargo into landing zones often only suitable for partial landings of the helicopter (i.e., only one of three wheels of the helicopter touching the ground). Known as pinnacle landings, pilots must make precise control inputs with limited fields of view of the landing zone (at the end of the landing, the pilot cannot see the landing zone). Crew chiefs, who sit behind the pilots' station and can see the side and bottom of the helicopter (specifically, the wheels in this example), provide additional information verbally to the pilots on the aircraft position in relation to the landing zone. At critical moments of the landing, when no other visual cues of the landing zone are available, the pilot will rely solely on what the crew chief says to complete the landing. This is an example of what a crew chief would say during a one-wheel pinnacle landing: "Continue forward for three, two, one. Hold hover. Continue down. Right wheel is down in three, two, one. Right wheel is down, hold position" (TC 3-04.33, *UH60 Aircrew Training Manual*, 2017). The pilot simply reacts to these auditory cues by adjusting the flight controls to move the helicopter. Without the auditory input from the crew chief, the pilot cannot safely land the helicopter (S. R. Baker & J. Grace, personal communication, October 10, 2018).[1] This is an example of perceptual-motor coupling (auditory coupling) between the pilot and the crew chief.

---

[1] Chief Warrant Officers 5 S. R. Baker and J. Grace are distinguished U.S. Army helicopter pilots and are expert instructor pilots in the CH47 and UH60 helicopters, respectively. I consulted both individuals when writing this paragraph.

There are several other examples of perceptual coupling. Pilots of large airplanes cannot safely taxi the plane into a parking spot at an airport without the help of a ground guide who uses hand and arm signals, a form of visual gesture, to direct the pilots. Perhaps the most relatable example is backing up a car at a crowded mall where the driver cannot see well behind his or her car and needs assistance from another person to back up safely. The person helping could use a gesture, auditory directions, or both to help the driver safely back up. The commonality in all these instances is the lack of perceptual information available to the person who has ultimate control over the task (i.e., pilot, driver), and interpersonal coupling is required. Without the help of another person (i.e., crew chief, ground guide) to give additional visual, auditory, or a combination of visual and auditory cues, then the task cannot be performed effectively.

The goal of this project was to study coupling (or interpersonal coordination) between a person who controls the “motor” inputs of a task and a person who fills in perceptual or auditory (semantic) details to complete the task. Furthermore, I sought to examine forms of coupling that occur between two people when they have a combination of the visual and auditory modes to examine how task difficulty affects each mode of coupling and when coupling modes are combined.

1.1 Coupling

In dynamic environments, synchronizing behavior between people has two requirements: at least two systems moving in relationship to each other and coupling between systems (Strogatz, 2004). For coupling to occur between two people, there must be a coupling medium (e.g., visual, auditory). Previous research suggests that interpersonal coupling can occur either through either visual or auditory modes

(Richardson, Marsh, & Schmidt, 2005; Gorman & Crites, 2015; Gibson, Gorman, & Hessler, 2016).

1.1.1 Visual Coupling

I define visual coupling as two people using different visual inputs to coordinate task performance with each other. Numerous studies have demonstrated that two people can coordinate their movements visually, whether it is something as complex as double-dutch jump roping or as simple as mirroring each other's finger movements (Gorman et al., 2017). Research shows that visual coupling can occur under intentional and unintentional coordination. Schmidt and O'Brien (1997) showed that unintentional interpersonal coordination occurs when participants see each other by demonstrating that two individuals will synchronize their pendulum movements by merely looking at each other. Research also suggests that when participants intentionally coordinate specific movements a "pattern of synchrony" results when visual coupling is increased (Schmidt, Bienvenu, Fitzpatrick, & Amazeen, 1998; Richardson et al., 2005).

Furthermore, these studies show that visual inputs provide enough information for two people to coordinate and synchronize their actions (Richardson et al., 2005). Ouiller et al. (2008) demonstrated that visual coupling between two people was enough to cause spontaneous 1:1 in-phase synchronization, or perfect mirroring, between the two individuals. However, Gorman et al. (2017) found that visual coupling alone is less reliable for more complex coordinated patterns (e.g., 3:1). Specifically, research in multi-frequency coordination that has participants coordinate at different levels other than 1:1 (e.g., 4:1; 5:1) shows that as coordination becomes more complicated, coordination with another individual becomes more difficult even as perceptual (visual) coupling increases (Gorman et al., 2017). As described in Gorman et al. (2017), other modes of interpersonal coupling (e.g., auditory) are required when the coordination task becomes more difficult (e.g., Double Dutch Jump Roping).

1.1.2 Auditory Coupling

With this in mind, I define auditory coupling as two people using different auditory inputs to coordinate a task with each other. Shockley, Santanna, and Fowler (2003) demonstrated that speaking is a medium by which individuals can unintentionally synchronize their postural dynamics. Multiple team experiments have shown how auditory communication is used to intentionally coordinate task performance (Gorman, Amazeen, & Cooke, 2010; Shockley et al., 2003; Richardson, Dale, & Kirkham, 2007).

1.1.3 Bimodal Coupling

In most applied settings, people combine perceptual and auditory coupling to coordinate task performance. In this experiment, I define visual plus auditory coupling (referred to as the "bimodal condition") as two people using simultaneous visual and auditory communication to coordinate interpersonally with each other. Richardson et al., (2005) examined the use of a combination of visual and auditory cues and how it compared to either visual or auditory coupling alone in an unintentional coordination task. The results of the study suggested that auditory interaction had no effect or did not increase the level of unintentional synchrony between the two participants, while also showing that visual interaction contributed to increased synchronization between participants. Another interesting finding in the study was that when they combined auditory and visual conditions, there was a slight decrease between the visual and the bimodal condition in the level of unintentional coordination between the participants. It should be noted, however, that the current study is slightly different, as it will focus on intentional coordination between participants.

Gorman et al. (2017) observed during their study of perceptual coordination amongst double-dutch jump rope teams that as the frequency of coupling increased from 1:1 to 2:1 and higher up to 4:1, that teams spontaneously changed their coupling medium. During 1:1 jump roping, the teams relied on visual perceptual cues to coordinate their actions. As jump rope patterns became more complicated, the teams coupled not only visually, but also by counting out-loud their cadence (auditory coupling). This visual and auditory communication enabled them to perform more complicated jump rope patterns. Thus, I predict that auditory coordination may be more important as the difficulty of the coordination task increases.

1.1.4 Gestures: A form of visual and auditory coupling

Visual coupling in this experiment focused on hand gestures. There is a large volume of research concerning spatial representation that demonstrates the importance of communication using hand gestures and auditory communication for spatial perception. Specifically, hand gestures play a significant role in a person's ability to communicate spatial information (Alibali, 2005). Often, speakers use gestures when they are trying to communicate spatial information to a listener. For example, when speakers are asked to verbalize information on neutral topics, such as what they did that day or auditory topics describing a book, they use significantly fewer gestures than when they describe spatial topics such as describing a route (Lavergne and Kimura, 1987).

Further research demonstrates how gestures combined with audio presentation can contribute to more effective communication with another person (in other words, a combination of visual and auditory information contributes to more successful outcomes). McNeil, Alibali and Evans (2000) demonstrated that children were more accurate in a block selection task when the information on which block to select was conveyed using gestures that were redundant with accompanying speech. Another study demonstrated that participants answered questions about clips from a cartoon story more accurately when they were presented with audio and video compared to audio-only clips. More interestingly, participants communicated object size and relative position more accurately in an audio plus video condition compared to an audio-only condition (Beattie & Shovelton, 1999).

Additionally, several studies indicate that gestures play a role in a speaker's ability to formulate spoken words (Alibali, 2005). For example, Emmorey and Casey (2001) conducted a study examining speakers who had to direct other participants to place puzzle pieces, which needed to be rotated, in a puzzle. Speakers sometimes used gestures showing the motion of the puzzle piece that needed to be rotated. These gestures, however, took place during pauses in speech that could suggest the gesture helps the speaker formulate how to verbalize a task to be accomplished (Alibali, 2005). Research also shows that speakers have a more difficult time delivering spatial information when they are not allowed to gesture compared to when they can use gestures. When speakers are prohibited from making gestures, they speak more slowly, increasing the difficulty of communicating spatial information (Rauscher, Krauss and Chen, 1996).

From the listener's perspective, research suggests that the ability to comprehend information is affected by the use of hand gestures so significantly that listeners' can identify specific information from a gesture alone (Driskell & Radtke, 2003). If the gesture is beneficial to communication, because it communicates both spatial and motor ideas, then one should expect to find more substantial coordination effects when gestures are present (Hostetter, 2011).

1.2 Current Study

The current study examined 24 teams with 8 teams per coupling condition. The task was to, as a dyad, move a remote-controlled car from a start point to a target area between obstacles as fast and accurately as possible. Within each dyad, participants were randomly assigned to the role of "driver" or the role of "spotter." The driver was responsible for the control inputs of the car but could not see the car. The spotter was responsible for viewing the car and giving auditory, visual, or a combination of cues to the driver on how to manipulate the controls of the car to accomplish the task. Difficulty varied between easy, medium, and hard conditions in which the target area was bigger or smaller based on level of difficulty for the trial. Each team conducted 12 trials (4 trials for each difficulty). Trial time, path variance (root mean square error; RMSE), and commands (gestures or auditory commands) per second were measured. Trial time was the primary performance measure.

I hypothesized that the visual condition would have the lowest overall mean trial time followed by the bimodal condition. The visual condition would also have the highest speed-accuracy trade-off demonstrated by higher RMSE. Previous research shows that visual coupling is stronger in dyads (i.e., Richardson et al., 2005, Gibson, Gorman & Hessler, 2016), and I predicted this strength will result in faster mean trial time. Additionally, the bimodal condition would have a faster mean trial time as the task difficulty increases and the lowest task error (RMSE). The bimodal condition affords the team more communication information that will result in decreased task error and better performance. Gorman et al., (2017) showed that as task difficulty increased, a combination of auditory and visual coupling helped double-dutch jump rope teams perform more complex tasks. Gesture research also indicates that the driver will be able to interpret the spotter's cues more easily (Driskell & Radtke, 2003; McNeil, Alibali & Evans, 2000). Lastly, I have an exploratory hypothesis based on gesture literature that when auditory coupling occurs (auditory and bimodal conditions), increased auditory cue rate will result in better performance as indicated by lower trial time and lower RMSE. I predict this will occur for visual cue rate as well but will be more significant for the auditory and bimodal coupling conditions.

CHAPTER 2. METHOD

2.1 Participants

48 participants (8 dyads per between-subjects condition) were recruited from the Georgia Institute of Technology School of Psychology participant pool. Based on a priori power analysis, in which I expected a medium effect between conditions (f=.2), power equal to .80, and alpha equal to .05, the total sample size should be 24 total dyads in order to produce reliable, task-based differences between the different modes of communication. Participants had to have 20/20 or correctable to 20/20 vision and English speaking to participate. The average age for the participant was 19.73 (*SD* = 1.47), and there were 28 male and 20 female participants.

2.2 Experimental Design

To simulate a task where one person lacks visual information and needs the help of another person, I had participants perform as teams of two or dyads. Within each dyad, participants were randomly assigned to the role of "driver" or the role of "spotter." Together, they were given a task to drive a remote-controlled car into a target area. The driver was responsible for the control inputs of the car but was unable to see the actual car when accomplishing the task. The spotter was responsible for viewing the car and giving auditory, visual, or a combination of the two types of commands to the driver on how to drive the car to complete the task. During the auditory condition, the driver and spotter were not able to see each other. In the visual and bimodal condition, the driver and spotter were able to see each other, but the spotter was faced away (driver saw spotters back; *Figure* 1). This manipulation served two functions. One is that it eliminated a potential confound of the driver using the spotter's eyes to make inputs on the car. Initial testing amongst experimenters showed that the spotter's eye movement could influence the driver's inputs. The second function is that it allowed the dyad to mirror each other and the spotter to mirror the car. In other words, it eliminated the need for the spotter to perform any spatial reconstruction (translate his or her right or left to the driver's right or left).

***Figure 1*. Example of coupling conditions during a trial. The spotter was located facing the task area and the driver was located behind the spotter. The picture on the left shows the auditory condition. The picture on the right shows the visual and bimodal conditions. *Figure* 4 displays the task area.**

Each dyad was randomly assigned to a between-subjects task condition (auditory, visual, or bimodal) and completed 12 total trials of 3 different within subjects task difficulties (easy, medium, and hard). I used complete counterbalancing to vary the sequence for task difficulty to ensure any results between task difficulties were not influenced by progressive effects. The independent variables were the between subjects coupling condition (auditory, visual, bimodal) and within subjects task difficulty (easy, medium, and hard). The dependent variables, described below, were trial time, number of commands per second, and path variability from an ideal path.

***Figure 2*. Example of the target area where the car must drive through to complete the task. The width between the two posts will change based on difficulty.**

To vary task difficulty, I manipulated the target size area while keeping the distance to the target equal for all conditions. The time required to move to a target area depends on the distance to the target and the width of the target (Fitts, 1954; Fitts & Peterson, 1964). The target area will vary in size to simulate hard, moderate, and easy tasks. Jones, Johnson, and Schmidlin (2011) found that for robot operators to successfully pass through obstacles, they needed an area the size of the robot's width plus 22% (SD = 15%) of the robot's width in order to complete the task. Based on these results, I used 22% width of the car for hard difficulty (10.85 centimeters). Specifically, I added one standard deviation to the medium difficulty and two standard deviations for the easy difficulty (37% width of the car (12.18 centimeters) for medium and 52% width of the car (13.51 centimeters) for easy). To complete the task (*Figure 5*), the dyads had to conduct two half turns of the car which ensures the dyads had to communicate several times. Dyads were told to complete the task as fast and accurately as possible. No time limit was applied.

I also utilized standardized forms of communication for the auditory and visual signals to the driver in order to reduce variability and increase predictability between the driver and the spotter. The spotter used hand and arm signals that were designed based on the Federal Aviation Association and U.S. Army regulations (FAA-H-8083-3B, 2016; FM 21-60, *Visual Signals*, 1987). Lastly, participants received training on their respective assignment as the driver or the spotter. Each became familiar with the standardized cues he or she would give or receive. Once initial training was complete, each participant completed a series of tasks designed to ensure each participant was at an expert level of performance prior to beginning trials. An experimenter acted as the participant's teammate during training.

2.3 Apparatus

The driver remotely operated a remote-controlled car via a controller utilizing both hands. A ten-camera Vicon Motion Capture System recorded the movement of the car from the time the participants started the task until the vehicle reached the target area. A camcorder also recorded the spotter to determine the number of commands completed.

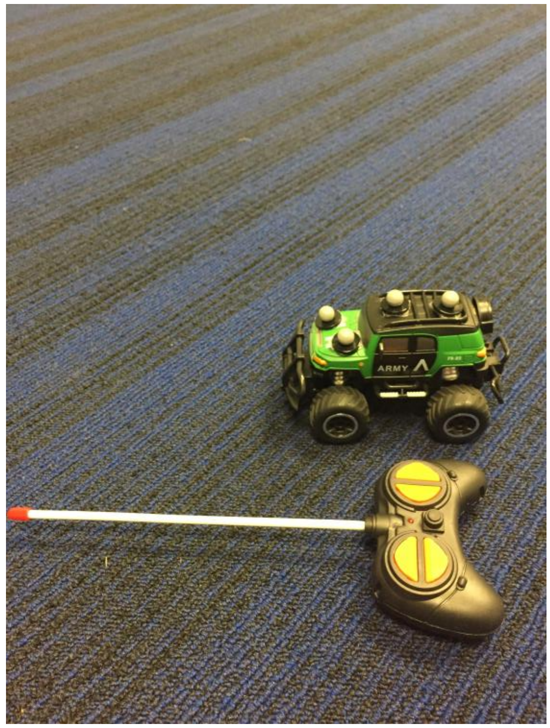

***Figure 3*. Remote control car and controller.**

2.4 Measures

The primary measure for this experiment was trial time. Trial time is the time from when the car first moves from the start point until the mid-point of the car crosses through the target area.

I also measured the number of commands (visual and/or auditory) to complete the task, which I used to determine the commands per second for each trial (i.e., divide number of commands by trial time). A singular command was the Spotter either gesturing or verbalizing a movement to the Driver. For example, “move right” or “forward right” would count as one command in each instance (FAA-H-8083-3B, 2016; FM 21-60, *Visual Signals*, 1987). These commands are the primary focus for cue analysis.

Another way to interpret cues was to assess distinct cues. These cues occur when the spotter specifies a change in the car’s path. Visually, this occurs naturally and thus, the number of visual commands would remain the same for this measure. However, auditory cue counts would change. Using this approach, the auditory cues “Forward, Forward, Forward” would count as one cue. The cues “Forward, Forward Left, Forward” would count as three cues with the “Left” showing a distinct change in the cars path. Though not the primary focus for this experiment, this approach is a potential way to analyze auditory cues and visual cues together in this task. Appendix A has a complete listing of all possible visual and auditory commands.

***Figure 4*. Example of a visual cue. The auditory cue example would be “move forward.” A complete list of cues is located in Appendix A.**

I also measured the variability of the car's path to the target area. Using similar procedures as Gorman and Crites (2013), I calculated the overall root mean squared error (RMSE) from an ideal car path to provide a single, summative measure of variability per trial. Less variability showed less error with the task. This also allowed for any detection of any speed-accuracy trade-off between conditions and difficulty.

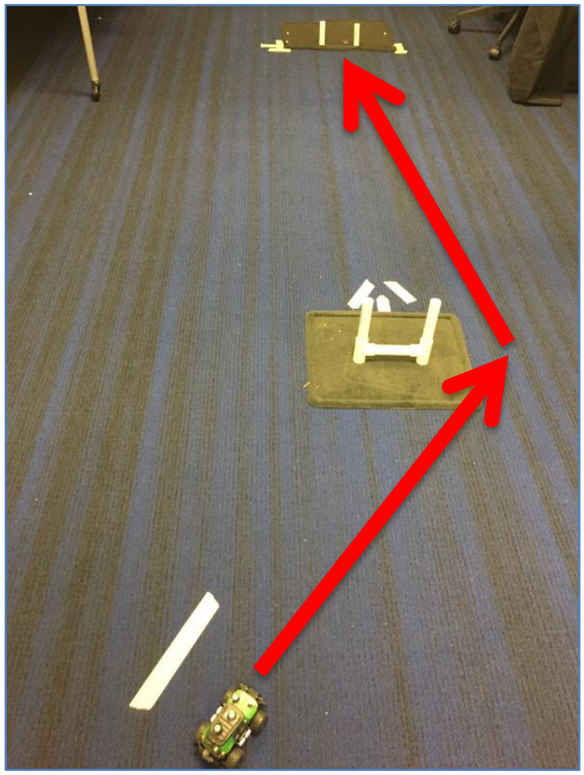

***Figure 5.* Path of car. To complete the task, the car must move in front of the obstacle and then turn towards the target area. The ideal path is designated by the red arrows and is two separate straight lines. I compared these two lines to the actual path of the car to determine RMSE.**

Lastly, participants completed a NASA-TLX survey (Hart, 2006) and mental rotation task (Shepard & Metzler, 1971) at the end of each experiment to assess workload and spatial ability. Though I do not have a specific hypothesis concerning these measures, they could provide another insight to the nature of the task overall.

CHAPTER 3. RESULTS

3.1 Trial Time

A 3x3 mixed ANOVA (Difficulty x Condition) was conducted to determine if there were significant differences between main effects and if there were interactive effects. The main effect of difficulty was significant, $F(2, 186) = 13.258$, $p < .001$, $MSE = 190.832$, $\eta_p^2 = .125$. Pairwise comparisons with a Bonferroni correction showed that significant differences existed between all difficulties: easy ($M = 29.91$, $SD = 18.13$), medium ($M = 40.17$, $SD = 17.73$) and hard ($M = 34.72$, $SD = 18.13$). Easy was significantly less than medium ($p < .001$) and hard ($p = .022$). Trial time for the medium difficulty was unexpectedly higher than the hard difficulty ($p = .043$). Pilot results indicated a similar trend though I did expect with increased sample size that medium difficulty would have lower mean trial time than hard difficulty.

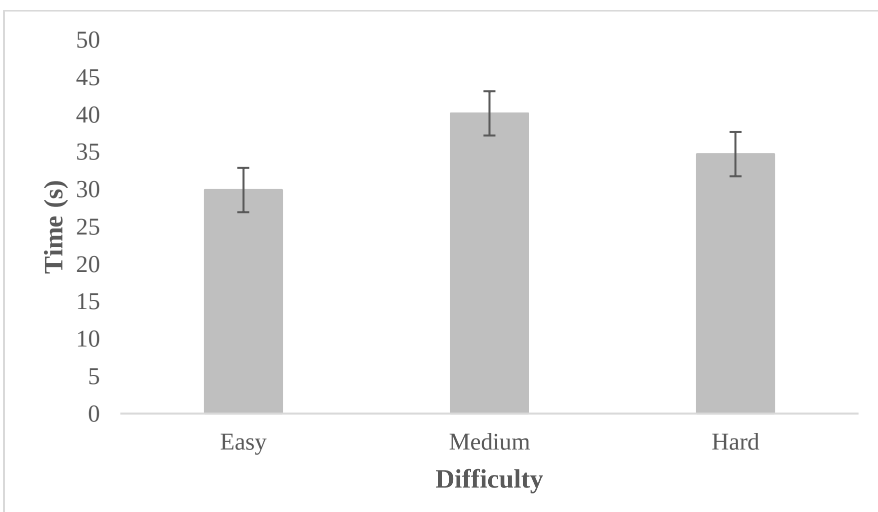

***Figure 6*. Average trial time by difficulty. Error bars represent standard errors.**

The main effect of condition was also significant, $F(2, 93) = 6.418$, $p < .001$, $MSE = 472.679$, $\eta_p^2 = .121$. Multiple comparisons, using Tukey's HSD, showed that the auditory condition (*M* = 41.20, *SD* = 22.37) had significantly longer mean trial time than the visual (*M* = 30.34, *SD* = 13.83, *p* = .002) and bimodal conditions (*M* = 33.26, *SD* = 14.77, *p* = .035). The visual and bimodal conditions did not have significant differences (*p* = .622).

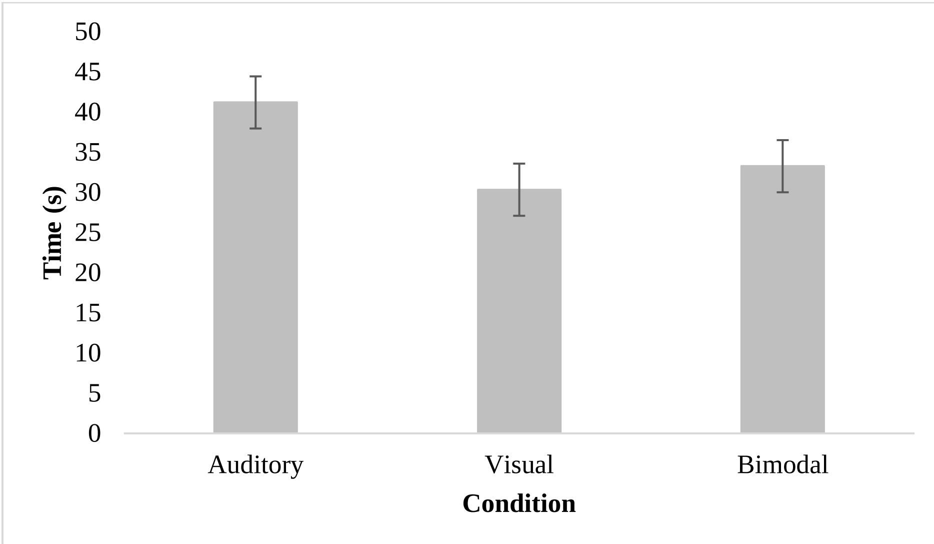

***Figure* 7. Trial time by condition. Error bars represent standard errors.**

Furthermore, there was a significant interaction between difficulty and condition, $F(4,186) = 2.479$, $p = .046$, $MSE = 190.832$, $\eta^2 = .051$. To determine the nature of the interaction, all pairwise comparisons, using a Bonferroni correction ($\alpha_{fw} = .05$), were conducted.

Table 1. ***Mean Trial Time for each Condition within Difficulty.***

	Easy		*Medium*		*Hard*	
Perceptual Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual	26.43	12.96	32.16	21.07	32.42	13.44
Auditory	35.91	12.22	45.19	21.64	42.5	15.63
Bimodal	27.38	15.66	43.16	23.95	29.24	9.48

Multiple comparisons show that the visual condition is faster than the auditory (p = .008) and bimodal conditions (*p* = .033) at medium difficulty. At hard difficulty, the bimodal condition has a faster mean trial time than the verbal condition, *p* = .009. At easy difficulty, there is no significant difference between visual and auditory conditions, *p* = .066, but the same trend exists (visual condition performs faster than verbal condition). Lastly, the bimodal condition is significantly slower at medium difficulty than at the hard (*p* = .001) and easy (*p* < .001) difficulties.

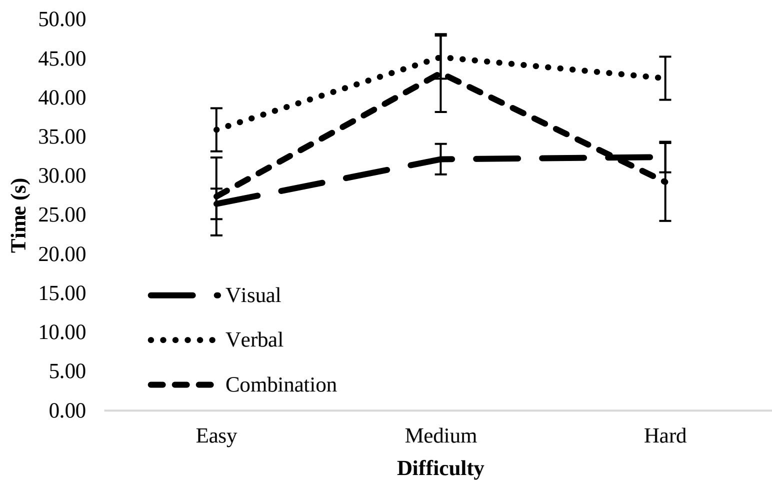

***Figure 8*. Mean trial time for condition by difficulty. Error bars represent standard error.**

At medium and hard difficulties, the visual condition does perform better than auditory. The bimodal condition acts differently through the difficulties. At medium, it performs worse than visual and equal to auditory. At hard, it performs better than auditory and equal to visual. This result was not predicted.

3.2 Commands per Second

Because of the nature of the cues given, this experiment's primary focus was not to compare visual to auditory cues. The nature of the task almost certainly results in significantly more auditory cues occurring than visual cues. Therefore, the results for cue rate have three different sections. One section that compares the auditory and bimodal conditions auditory cue rate. The second section compares the visual and bimodal conditions visual cue rate. The last section reports results for distinct cue rates, which compares all conditions.

Inter-rater reliability was assessed using inter-class correlation (ICC) for visual and auditory cues to ensure rater reliability. For visual cues, average ICC was .993. For auditory cues, average ICC was .994. These values indicate that the raters were highly reliable (Koo & Li, 2016).

3.2.1 Auditory Cue Rate

A 3x2 mixed ANOVA (Difficulty x Condition) using auditory cue rate as the dependent variable was conducted to determine if there were significant differences between main effects and if there were interactive effects. Maucley's Test of Sphericity was significant therefore an epsilon correction, Huynh-Feldt, was used for the within subjects effect. The main effect of difficulty was significant, $F(1.841, 114.145) = 5.531$, $p = .006$, $MSE = 114.145$, $\eta_p^2 = .082$. Pairwise comparisons with a Bonferroni correction showed that medium difficulty ($M = .992$, $SD = .334$) had a lower auditory cue rate than easy ($M = 1.104$, $SD = .353$), and hard ($M = 1.139$, $SD = .410$) difficulties, $p = .011$ and $p = .016$, respectively.

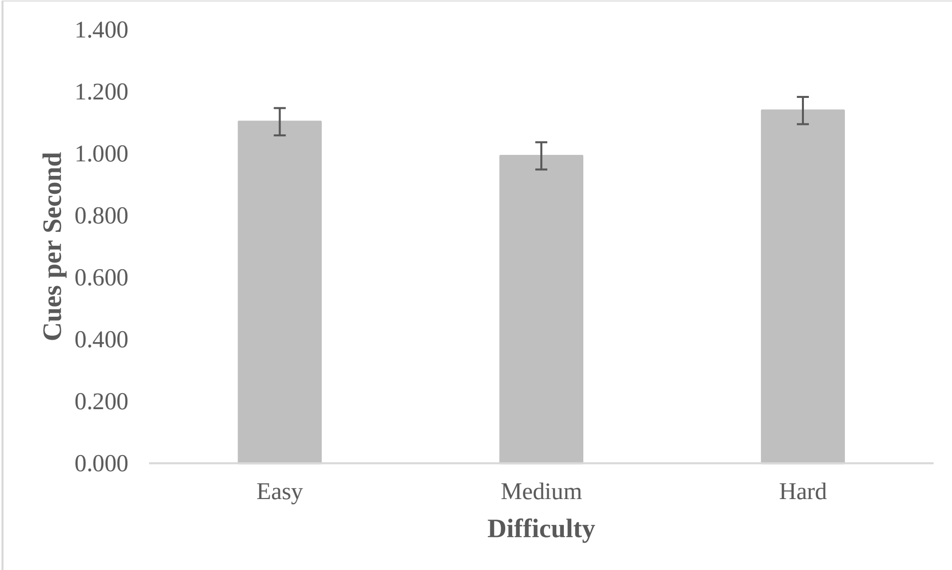

***Figure* 9. Auditory cues per second by difficulty. Error bars represent standard error.**

The main effect of condition was not significant, $F(1, 62) = .227$, $p = .635$, $MSE = 529.551$, $\eta_p^2 = .004$. Furthermore, the interaction between difficulty and condition was not significant, $F(1.841, 114.145) = .186$, $p = .813$, $MSE = 114.145$, $\eta_p^2 = .003$. A post hoc comparison between medium ($M = .992$, $SD = .334$) and hard ($M = .992$, $SD = .334$) difficulties for the bimodal condition yields significant results, p = .044. This indicates that dyads communicated auditorily less frequently during the medium difficulty compared to the hard difficult.

Overall, the auditory cue rate had a significant negative correlation to time (Table 2) showing that as auditory cue rate increases, trial time decreases. R^2 values also indicate that auditory cue rate does account for a moderate amount of the variance in task performance (trial time). Additionally, the bimodal condition appears to have higher amounts of variation accounted for by auditory cue rate than does the auditory condition.

Table 2. Pearson's *r and R^2 of Trial Time and Auditory Cue Rate.*

	Auditory Cues per Second		
Perceptual Condition	*df*	*r*	R^2
Auditory	94	-.483*	.233
Bimodal	94	-.568*	.323
Auditory and Bimodal	190	-.481*	.231

Note. *Correlation between time and Cues per second significant p < .001.

3.2.2 Visual Cue Rate

A 3x2 mixed ANOVA (Difficulty x Condition) using visual cue rate as the dependent variable showed that the main effect of difficulty was not significant, $F(2, 124) = .111$, $p = .895$, $MSE = .030$, $\eta_p^2 = .002$. The interaction between difficulty and condition also was not significant, $F(2, 124) = .434$, $p = .649$, $MSE = .030$, $\eta_p^2 = .007$. The main effect of condition was significant, $F(2, 62) = 6.893$, $p = .011$, $MSE = .121$, $\eta_p^2 = .100$. This indicates that the visual condition ($M = .657$, $SD = .309$) had a higher visual cue rate than the bimodal condition ($M = .525$, $SD = .152$).

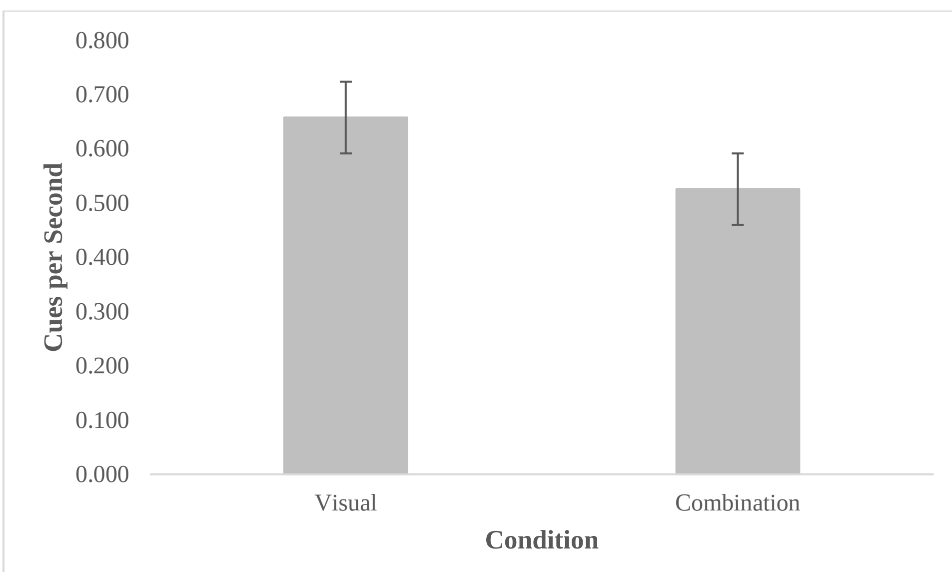

***Figure 10*. Visual Cues per second by condition. Error bars represent standard error.**

Similar to the results of auditory cue rates, the visual cue rate had a negative correlation to trial time for the visual condition and when the visual and bimodal conditions are combined. As visual cue rate increases, trial time decreases. The bimodal condition's visual cue rate does not show a significant correlation to time. Also, visual cue rate does account for a moderate amount of the variance in trial time, though less so than seen in auditory cue rates. Additionally, the bimodal condition has a lower amount of variation accounted for by visual cue rate than the visual condition. This result, taken with the auditory cue rate, indicates that for the bimodal condition, there is more variation accounted for by auditory cue rates than visual cue rates.

Table 3. Pearson's *r and* R^2 *of Trial Time and Visual Cue Rate*

	Visual Cues per Second		
Perceptual Condition	*df*	*r*	R^2
Visual	94	-0.372*	0.138
Bimodal	94	-0.139	0.019
Visual and Bimodal	190	-0.289*	0.084

Note. *Correlation between time and Cues per second significant p < .001.

3.2.3 Distinct Cue Rate

A 3x3 mixed ANOVA (Difficulty x Condition) using visual cue rate as the dependent variable was conducted to determine if there were significant differences between main effects and if there were interactive effects. The main effect of difficulty was not significant, $F(2, 124) = .111$, $p = .895$, $MSE = .030$, $\eta_p^2 = .002$. The interaction between difficulty and condition also was not significant, $F(2, 186) = .580$, $p = .678$, $MSE = .022$, $\eta_p^2 = .012$. The main effect of condition was significant, $F(2, 93) = 9.755$, $p < .001$, $MSE = .118$, $\eta_p^2 = .173$. Multiple comparisons using Tukey's HSD indicate that the visual condition ($M = .657$, $SD = .309$) had a higher distinct cue rate than the bimodal condition ($M = .493$, $SD = .173$) and auditory ($M = .450$, $SD = .186$) conditions, $p = .004$ and $p < .001$ respectively.

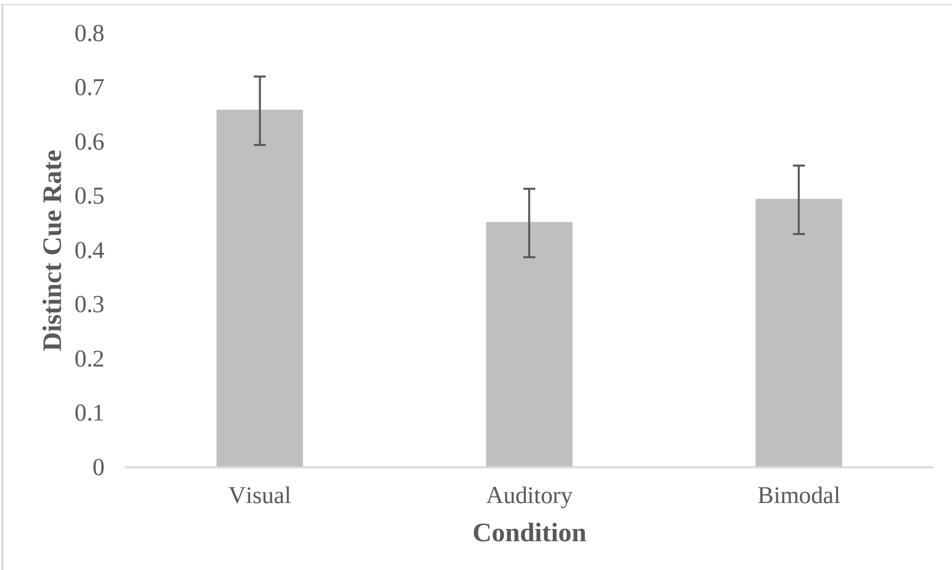

***Figure 11*. Visual Cues per Second by Condition. Error bars represent standard error.**

The distinct cue rate had a significant negative correlation to trial time overall and for each individual condition. Similar to auditory and visual cue rates, these correlations for distinct cue rates suggest that as cue rates increase, trial time decreases. Furthermore, R^2 values indicate that less variation is accounted for by distinct cue rates compared to auditory cue rates. On the other hand, R^2 values indicate that distinct cue rates account for similar amounts of variation in trial time as visual cue rates, but less than auditory cue rates.

Table 4. ***Pearson's r and R^2 of Trial Time and Distinct Cue Rate.***

	Distinct Cues per Second		
Perceptual Condition	*df*	*r*	R^2
Visual	94	-.372*	.138
Auditory	94	-.378*	.143
Bimodal	94	-.226**	.051
Overall	286	-.357*	.124

Note. *Correlation between time and Cues per second significant $p <$ .001. ** Correlation between time and Cues per second significant $p =$ .013.

3.3 RMSE

A 3x3 mixed ANOVA (Difficulty x Condition) with RMSE as the dependent variable was conducted to determine if there were significant differences between main effects and if there were interactive effects. The main effect of difficulty was significant, $F(2, 184) = 4.299$, $p = .015$, $MSE = 45626.82$, $\eta_p^2 = .04$. Pairwise comparisons with a Bonferroni correction showed that significant differences existed between the easy (M = 139.79 SD = 180.40) and medium (M = 231.10 SD = 246.50) conditions, $p = .007$. There was no difference between hard (M = 192.23 SD = 266.70) and the other difficulties. The main effect of condition was not significant, $F(2, 92) = .942$, $p = .393$, MSE = 70333.947, $\eta_p^2 = .020$.

Furthermore, there was a significant interaction between difficulty and condition, $F(4,184) = 2.479$, $p = .043$, $MSE = 472.679$, $\eta^2 = .052$. To determine the nature of the interaction, all pairwise comparisons, using a Bonferroni correction ($\alpha_{fw} = .05$), were conducted.

Table 5. *Mean RMSE for each Condition within Difficulty.*

	Easy	*Medium*	*Hard*

Perceptual Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual	158.79	224.39	177.09	195.09	138.57	129.90
Auditory	152.60	146.37	211.44	178.50	269.57	388.29
Bimodal	108.58	164.30	299.22	327.31	168.79	196.86

Comparing the conditions to each other at each difficulty yields insignificant results. Comparing the conditions individually at each difficulty has significant results for the bimodal condition. The bimodal condition's RMSE is greater at Medium than it is at Easy (p = .001), but fails to reach significance at Hard (p = .077). Though not statistically reliable, the results for medium and hard difficulties share a similar pattern to those found in trial time and cue rate for the bimodal condition.

There is also a significant positive correlation between trial time and RMSE. This indicates that as trial time increases, RMSE also increases. This contradicts our hypothesis that there would be a speed-accuracy trade off shown by a negative correlation. R^2 values also indicate similar results to those found in cue rates. RMSE accounts for a greater amount of variation in trial time during conditions that have auditory cues.

Table 6. *Pearson's* r *and R^2 of trial time and RMSE.*

		RMSE	
Perceptual Condition	*df*	*r*	R^2
Visual	93	0.411*	0.169
Auditory	94	0.664*	0.440
Bimodal	94	0.633*	0.401
Overall	285	.593*	.352

Note. *Correlation between time and RMSE significant p < .001.

3.4 NASA-TLX

A NASA-TLX survey was given after the completion of all trials to assess the participants' workload. An omnibus one-way ANOVA was conducted to compare the effect of condition with NASA-TLX and the sub-categories of NASA-TLX. No significant results existed. No significant correlations existed as well. When the data is grouped by participant role, the results for overall TLX score, performance, and effort for drivers are worth discussing. An omnibus one-way ANOVA comparing conditions indicates that performance is the only category that has statistically significant results, $F(2, 24) = 4.516$, $p = .002$, $MSE = 195.06$, $\eta_p^2 = .30$. Results were insignificant for overall TLX score ($F(2, 21) = 2.99$, $p = .072$, $MSE = 142.37$) and effort ($F(2, 21) = 2.44$, $p = .111$, $MSE = 449.35$). Multiple comparisons using Tukey's HSD indicate that for performance, the bimodal condition has a higher score ($p = .027$) than the visual condition, indicating participants thought they performed better during the bimodal condition. For overall TLX and effort scores, comparisons yielded insignificant results, $p = .060$ and $p = .093$ respectively, between the bimodal and visual conditions. Though this fails to reach statistical significance, this trend is worth reporting and suggests the bimodal condition has higher workload and requires more effort than the visual condition.

Table 7. *Mean Driver NASA-TLX Overall, Effort, and Performance Scores for Condition.*

	Overall TLX		*Effort*		*Performance*	
Perceptual Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
Visual	20.75	12.15	22.25	18.11	16.75	13.50
Auditory	26.77	12.58	35.13	26.30	20.13	8.43
Bimodal	35.27	11.02	45.63	18.13	36.38	18.22

Lastly, I analyzed the bivariate correlations between average trial time and NASA-TLX categories for each condition based on participant role to see if any category had a significant correlation to trial time. The results show that effort for drivers in the bimodal condition has a significant correlation ($r(6) = -.750$, $p = .032$) to average trial time. This indicates that as average trial time decreases, overall effort for drivers in the bimodal condition increases. No other significant correlations were found.

3.5 Spatial Ability

A mental rotation task was given after the completion of all trials to assess the spatial ability of each participant (Shepard & Metzler, 1971). The task involved correctly identifying the rotation of a two-dimensional object. Regression analysis showed that the number of correct responses ($\beta = -.419$, $p < .001$) during this task was a significant predictor of average trial time for participants who were drivers ($R^2 = .176$, $F(1, 22) = 4.693$, $p = .041$). This analysis shows that drivers with more correct responses during the mental rotation task had lower average trial times.

When analyzing the correct responses by condition, an omnibus one-way ANOVA yields insignificant results, $F(2, 21) = 2.636$, $p = .095$, $MSE = .964$. Post hoc comparisons using Tukey's HSD indicate that there are no significant differences between conditions, though the difference between the visual ($M = 7.250$, $SD = .886$) and bimodal ($M = 8.375$, $SD = .744$) conditions approaches significance, $p = .079$. Additionally, an omnibus one-way ANOVA indicates that gender did not have an effect on how many correct responses participants had during the mental rotation task, $F(1, 46) = 1.647$, $p = .206$, $MSE = 1.216$. Though these results are not statistically significant, they show a trend that warrants reporting and could impact future research.

CHAPTER 4. DISCUSSION

The overall pattern of results for this experiment further supports previous interpersonal coordination research that visual coupling is strongest compared to auditory coupling (i.e., Gibson et al., 2016). Visual coupling had the lowest trial time overall compared to auditory coupling, and this finding was consistent across all difficulties. Visual coupling did not outperform bimodal coupling overall, but it did perform better during medium task difficulties. Additionally, visual coupling did not have a significantly greater RMSE value compared to the other conditions. This indicates that visual coupling did not have a greater speed-accuracy tradeoff than other conditions, as predicted. Consequently, RMSE shows no difference between the conditions.

I did predict that the bimodal condition would outperform the other two conditions as task difficulty increased. However, the results show that the bimodal condition did not significantly outperform the visual condition at the hardest difficulty and was significantly slower than the visual condition at medium difficulty. I found that the bimodal condition was significantly slower during medium difficulty than during hard which I did not predict, though can be explained somewhat by the cue rate results— the bimodal condition's auditory cue rate is lower at medium compared to hard difficulty. Bimodal coupling did outperform auditory coupling at hard difficulty, which could possibly be attributed to simply having visual cues available. The bimodal condition also did not have a significantly lower RMSE value compared to the other conditions as I predicted.

The exploratory hypothesis that auditory cue rate and visual cue rate would impact trial time was correct and could further explain why medium difficulty performed worse than hard difficulty, a finding I did not expect. Cue rate results indicate that the more frequently the spotter communicates auditorily or visually, the faster the team was able to complete the task. Additionally, auditory cue rate does account for more variation in trial time than visual cue rate, especially for the bimodal condition. With respect to distinct cue rate, the results provide more support for the idea that when auditory coupling is used, it cannot provide distinct cues at the same rate as visual coupling. Consequently, teams need to focus on increased auditory cue rate, even though the same cue is given over and over (e.g., forward, forward, forward) to make up for the decreased coupling strength within the condition.

I suspect these results show us something unique about the nature of verbal and visual coupling. For visual coupling, I suspect that cue rate simply is not as important (accounts for small amount of variation in trial time) and visual coupling is naturally strong enough to overcome the performance decrement at Medium difficulty. When dyads use auditory coupling, especially when combined with visual coupling (bimodal), auditory cue rate plays a large role in task performance (larger amount of variation accounted for in trial time). Though intuitive, this result supports the idea that increased verbal communication, especially when combining visual and auditory cues, will result in better task performance in coupled dyads.

Wickens' Multiple Resource Theory (MRT) provides a potential explanation for the results seen in cue rates. The spotter's task was visual with a verbal or spatial output. MRT, as I interpret it, states that a visual task would be more difficult with a spatial output and easier for a verbal output (Wickens, 2002). Along these lines, MRT theory, with respect to coupling, supports why stronger correlations were observed with trial time and auditory cue rate compared to visual cue rate, though further research and analysis needs to be done to strengthen this argument.

The results of the experiment do not suggest any specific reason as to why the medium difficulty performed worse than the hard difficulty during the auditory and bimodal conditions. A potential explanation for the increased trial time at medium difficulty for auditory and bimodal conditions is based on phase transitions in dynamical systems theory. Essentially, medium difficulty is a phase transition—point of instability—between easy and hard, which results in less obvious mechanisms for synchronization. Though this experiment did not examine coupling in the sense of phase transitions, it is a potential explanation. One way to test this would be to use phase transition methodology (Kelso, 1995) to see if similar results emerge.

Another potential theory that explains the difficulty results could be the Yerkes-Dodson Law between arousal and performance (Yerkes & Dodson, 1908; Kahneman, 1973). This theory states that performance increases based on the amount of a person's stress or arousal. Once stress reaches a certain level, a person's performance would then decrease. In our experiment, I suspect that the spotter was experiencing a lack of stress to perform at the medium difficulty and therefore did not communicate as much. The spotter simply could not perceive that the task difficulty increased (from easy to medium) and thus did not perform as well. At the hard difficulty, the spotters did perceive that the task was more difficult and increased their performance, demonstrated by an increased cue rate. To test this, future research could assess the participants' workload and stress levels after each trial and also assess if the spotter was able to perceive a difference in difficulty. While the Yerkes-Dodson Law could explain our findings for auditory cue rate, it fails to explain our findings for the visual cue rate that show no differences between difficulties. If the Yerkes-Dodson Law explained our results fully, I would expect the same trend in difficulty when analyzing visual cue rates.

The spatial ability and workload measures do not explain a significant amount of my results; however, they do display interesting trends that could influence future research. Spatial ability appears to have an impact for the driver in this task—higher spatial ability scores correlated to lower average trial time. When analyzing this by condition, my results were not statistically significant, though there appears to be a trend that the bimodal condition drivers did perform better than the visual condition drivers during the mental rotation task. This, however, does not explain team performance differences between the conditions since visual coupling did have lower average trial times. One would expect if spatial ability had a large influence on performance, then the bimodal condition would have had lower average trial times. For future research, I recommend analyzing participants' spatial ability if time permits, but to prioritize other measures such as workload.

Drivers also appear to have a higher overall workload and effort level during the bimodal condition compared to the visual condition, though again, the results are marginally significant and require increased sample size to more significantly detect differences. This result is not surprising given the fact that the drivers for the bimodal condition have to respond to two different versions of cues in comparison to the other conditions. This result could explain some performance decrement for the bimodal condition, but I cannot say to what degree because I did not assess workload after each trial. Consequently, I recommend that future research analyze workload by trial, which would increase the sample size for this measure. Also, if workload measures show differences during medium difficulty compared to hard difficulty, this could provide us more insight as to why the medium difficulty performed worse than hard difficulty.

4.1 Conclusion

This study provides a unique perspective on how two people coordinate their actions to accomplish a mutual goal. There are often instances when two people have no choice in their form of communication (or coupling medium) to accomplish a task. Developing a better understanding of how different modes of coupling occur contributes to a better understanding of team cognition and team coordination.

Overall, this experiment's findings do have unique applications for real world situations. In the context of interpersonal perceptual-motor coupling (e.g., parking a large airplane), our results suggest that teams should first rely on visual over auditory communication when given a choice. In my helicopter-landing example, the current results suggest that the crew chief should speak as frequently as possible during landing when the pilot cannot see the landing zone. If both perceptual cues are available and utilized (e.g., backing up a car) then it is even more important that teams communicate frequently verbally to ensure peak performance.

In addition to these recommendations, there is the nagging question of why medium difficulty, at least as operationalized here, performed worse than hard difficulty. Assessing workload after every trial and/or collecting heart rate data could provide insight into if participants did not feel enough stress to perform at medium compared to hard difficulty. Another potential manipulation could be to assess the differences between top-down versus bottom-up knowledge, and its affects on team performance in this task (similar to Gibson et al., 2016). Manipulating if a dyad knows the trial task difficulty prior to the start of a trial compared to not knowing could provide another insight to explain the results of medium and hard difficulties. If the performance decrement does not exist when dyads know task difficulty, this could provide insights into how having more shared knowledge affects task performance in teams.

APPENDIX A. SPOTTER CUES

<table>
<tr><th>Auditory Cues</th><th>Visual Cues</th></tr>
<tr><td>“Move Forward”</td><td></td></tr>
<tr><td>“Move Backward”</td><td></td></tr>
</table>

"Move Forward and Right" or "Forward, Right"	
"Move Forward and Left" or "Forward, Left"	
"Move Backward and Right" or "Backward and Right"	

"Move Backward and Left" or "Backward and Left"	
"Stop"	

REFERENCES

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial cognition and computation*, *5*(4), 307-331.

Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of language and social psychology*, *18*(4), 438-462.

Department of the Army. (1987). *Visual Signals* (FM 21-60). Washington, DC: Headquarters TRADOC.

Department of the Army. (2017). *UH60 Series Aircrew Training Module* (TC 3-04.33). Fort Rucker, AL: Reviewed by CW4 Lamb, J. United States Army Aviation Center of Excellence.

Driskell, J. E., & Radtke, P. H. (2003). The effect of gesture on speech production and comprehension. *Human Factors*, *45*(3), 445-454.

Emmorey, K., & Casey, S. (2002). Gesture, thought, and spatial language. In *Spatial language* (pp. 87-101). Springer, Dordrecht.

Federal Aviation Administration. (2016). *Airplane Flying Handbook* (FAA-H-8083-3B). Oklahoma City, OK.

Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. *Journal of experimental psychology*, *67*(2), 103.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, *47*(6), 381.

Gipson, C. L., Gorman, J. C., & Hessler, E. E. (2016). Top-down (prior knowledge) and bottom-up (perceptual modality) influences on spontaneous interpersonal synchronization. *Nonlinear dynamics, psychology, and life sciences*, *20*(2), 193-222.

Gorman, J. C., Amazeen, P. G., Crites, M. J., & Gipson, C. L. (2017). Deviations from mirroring in interpersonal multifrequency coordination when visual information is occluded. *Experimental brain research*, *235*(4), 1209-1221.

Gorman, J. C., Cooke, N. J., & Amazeen, P. G. (2010). Training adaptive teams. *Human Factors*, *52*(2), 295-307.

Gorman, J. C., & Crites, M. J. (2013). Are two hands (from different people) better than one? Mode effects and differential transfer between manual coordination modes. *Human factors*, *55*(4), 815-829.

Gorman, J. C., & Crites, M. J. (2015). Learning to tie well with others: bimanual versus intermanual performance of a highly practiced skill. *Ergonomics*, *58*(5), 680-697.

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.

Hostetter, A. B. (2011). When do gestures communicate? A meta-analysis. *Psychological bulletin*, *137*(2), 297.

Jones, K. S., Johnson, B. R., & Schmidlin, E. A. (2011). Teleoperation through apertures: Passability versus driveability. *Journal of Cognitive Engineering and Decision Making*, *5*(1), 10-28.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.

Kelso, J. S. (1997). *Dynamic patterns: The self-organization of brain and behavior*. MIT press.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163.

Lavergne, J., & Kimura, D. (1987). Hand movement asymmetry during speech: No effect of speaking topic. *Neuropsychologia*, *25*(4), 689-693.

McNeil, N. M., Alibali, M. W., & Evans, J. L. (2000). The role of gesture in children's comprehension of spoken language: Now they need it, now they do not. *Journal of Nonverbal Behavior*, *24(2)*, 131–150.

Ouiller, O., de Guzman, G. C., Jantzen, K. J., Lagarde, J. & Kelso, J. A. S. (2008). Social coordination dynamics: Measuring human bonding. *Social Neuroscience*, *3(2)*, 178-192.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, *7*(4), 226-231.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological science*, *18*, 407-413.

Richardson, M. J., Marsh, K. L., & Schmidt, R. C. (2005). Effects of visual and Auditory interaction on unintentional interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(1), 62.

Schmidt, R. C., & O'Brien, B. (1997). Evaluating the dynamics of unintended interpersonal coordination. *Ecological Psychology*, *9*(3), 189-206.

Schmidt, R. C., Bienvenu, M., Fitzpatrick, P. A., & Amazeen, P. G. (1998). A comparison of within- and between-person coordination: Coordination breakdowns and coupling strength. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 884-900.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701-703.

Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 326-332.

Strogatz, S. H. (2004). *Sync: How Order Emerges from Chaos in the Universe, Nature, and Daily Life*. New York, NY: Hyperion.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, *3*(2), 159-177.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology*, *18*(5), 459-482.