Williams, J.L., Hooey, B.L. & Foyle, D.C. (2006). 4-D taxi clearances: pilots' usage of time- and speed-based formats. Proceedings of the AIAA Modeling and Simulation Technologies Conference, Paper AIAA-2006-6611.

4-D Taxi Clearances:

Pilots' Usage of Time- and Speed-based Formats

Jennifer L. Williams[*] and Becky L. Hooey[†]
San Jose State University at NASA Ames Research Center, Moffett Field, CA, 94035

David C. Foyle[‡]
NASA Ames Research Center, Moffett Field, CA, 94035

Automated surface management systems are being developed that utilize dynamic algorithms to calculate the most efficient movement of all surface traffic in order to increase the efficiency with which airport surfaces are utilized. If these systems are to be implemented, pilots will be required to comply with 4-D taxi clearances, in which a pilot is required to be at a specific location at a specific time. This pilot-in-the-loop simulation study is an investigation of how to present the information necessary for pilots to comply with such 4-D taxi clearances. This study is aimed at determining the fundamental abilities that pilots have in complying with clearances given as time-based versus speed-based commands, without the use of an advanced display. Using a medium-fidelity surface operations simulator at NASA Ames Research Center, with Boeing 737 modeled dynamics, 18 commercial airline Captains each completed 45 taxi clearance scenarios. Each taxi clearance scenario required pilots to follow a cleared taxi route and to make an active runway crossing at a specific future time by following either a time-based or speed-based clearance. Taxi clearances were presented using: 1) a Speed format that displayed current ground speed and a commanded average ground speed; 2) a Time format that displayed elapsed time and a commanded time of arrival; or, 3) a Speed/time format that displayed all of the information from the Speed and Time formats simultaneously. Pilots' time-of-arrival (TOA) absolute error, TOA error bias, and velocity standard deviation (SD) were recorded for each trial. Overall, the results suggest that having both speed and time information provides more accurate and less variable 4-D clearance compliance. Additionally, the results indicate that utilizing shorter taxi clearances can maximize TOA accuracy. This baseline study indicates that the development of flight deck displays for 4-D taxi clearances may likely require the presentation of both time-based information and speed-based information for accurate and efficient use by the flight crew.

I. Introduction

Delays caused by airport surface congestion, such as those incurred by aircraft waiting to cross active runways, result in the largest delay cost in the US airspace system.[1] These delays result from a competition for limited resources, such as runways and taxiways. Researchers have pointed out that reducing airport surface congestion requires not only increasing the amount of usable airport surfaces (i.e., adding runways and taxiways), but also increasing the efficiency with which existing runways and taxiways are utilized through more precise taxiing and

[*] Research Associate, San Jose State University at NASA Ames Research Center, Human Systems Integration Division, MS 262-4, AIAA graduate student membership.
[†] Senior Research Associate, San Jose State University at NASA Ames Research Center, Human Systems Integration Division, MS 262-4.
[‡] Senior Research Scientist, Human Systems Integration Division, MS 262-4.

improved runway coordination.[2] This may be accomplished with the development of surface management systems (SMS) that provide information to air traffic controllers to help them direct traffic in a more efficient and safe manner than that of current-day operations.[1] Although most of the SMSs under development provide only advisory information, automated systems are being developed that utilize dynamic algorithms to calculate the most efficient movement of all surface traffic based on changing conditions.[3,4]

In addition to physical information about surface traffic that these systems can utilize, such as location, heading, and speed; time can be included as an additional component of movement in order to coordinate traffic. This fourth dimension (time) could allow the system to predict future traffic locations and enable four-dimensional (4-D) coordination of surface traffic as aircraft move forward in time. This information can be presented to air traffic controllers and pilots in order to coordinate runway crossings, and it could potentially lead to shorter and fewer taxi delays. If these systems are to be implemented, pilots will be required to comply with 4-D taxi clearances provided by the automation. It is expected that 4-D taxi clearances will keep traffic flowing smoothly by requiring pilots to reach a particular runway crossing or taxiway intersection by a specified time in order to ensure that they do not have to stop to wait for cross traffic. Research has shown that aircraft take twice as long to cross active runways when starting from a standstill compared to crossing without having to stop.[2] If pilots can reach a runway crossing within a window of time that allows them to cross without stopping, they can potentially achieve shorter taxi times, increase fuel efficiency, and increase traffic throughput of the airport.[2] Future implementation of automated 4-D taxi clearances would likely require advanced displays to support pilots during taxi operations. The purpose of the present study was to determine the basic information required for those displays and the fundamental abilities that pilots have in complying with 4-D taxi clearances.

Although 4-D taxi clearances introduce the dimension of time into surface operations, speed is also an important component as a dimension of movement. Thus, this study compared taxi commands presented in speed and time formats because both speed and time were considered relevant pieces of information for complying with 4-D taxi clearances. The general task that pilots performed was to taxi to an active runway crossing by a given time, or at an average speed, with as little error as possible.

The use of 4-D taxi clearances is a novel concept, and as such very little literature exists to guide the development of 4-D taxi systems, or predict whether pilots will be able to comply with 4-D clearances. However, literature in other domains can be used to guide this research effort. Studies of the perception of speed indicate that in general, people tend to underestimate the speed at which they are traveling.[5,6] This phenomenon is more pronounced at lower speeds than higher speeds. In other words, drivers typically estimate that they are going slower than they actually are, and this error tends to be larger at lower speeds, but improves at higher speeds. Despite the existence of various theories about the underlying processes of time estimation, there appears to be a general finding that prospective time estimations tend to be negatively related to task demands.[7,8] The more difficult or demanding a task is, the shorter an individual's time estimation of that task will be. This trend is more pronounced for longer durations.[9] These two findings are at odds with one another in the context of a production task. If pilots do underestimate their average speed they will travel too quickly and arrive too soon, causing their average speed to be faster than the commanded speed. However, if pilots underestimate time, they will travel too slowly and arrive later than the commanded time of arrival (TOA).

Providing both speed and time information is another consideration for 4-D taxi systems. On the one hand, it would provide the most information for taxiing, yet on the other hand, more information is not always better due to confusion that can result from display clutter or conflicts between competing stimuli.[10,11] However, if the different pieces of information are congruent and complementary to each other, they can facilitate more accurate performance, known as a redundancy gain.[11] Examples of redundancy gain include presenting instructions with both pictures and text, or giving a warning signal both verbally and spatially. Research has shown that people use a subjective form of the equation for speed (velocity = distance / time) to relate estimates of speed, time, and distance.[12] By providing pilots with speed and time information, two pieces of that equation are available to be compared and checked against each other. Thus, the speed/time format, in which they receive both speed and time information, should produce the most accurate compliance with taxi commands due to the redundancy of the information, compared with the speed or time formats individually.

A. Present Study

The main research question of the present study was to determine whether pilots are able to comply more accurately with 4-D commands given in a speed, time, or speed/time format. The variability of their ground speed was also of interest in order to determine how consistent pilots' speed was within the routes. In order for 4-D taxi clearances to increase the efficiency of taxi operations, not only is temporal compliance at the final destination (a runway crossing) important, but consistency throughout a taxi clearance is also necessary for the coordination of numerous aircraft and for an automated system to be able to accurately predict where various aircraft will be in the future. Additionally, how pilots used the displays, as measured by the proportion of time that they looked at the different displays, was of interest in order to assess when they use different types of information (speed vs. time) and how much they refer to the information. This is important for determining if providing speed and/or time information causes attention tunneling, which occurs when pilots focus on the display too much, at the expense of objects in the environment, thus reducing their situation awareness.

Because time information is considered to be less salient than visual speed cues, it was expected that performance with the speed format would be more accurate than with the time format. However, because the speed/time format could produce a redundancy gain, it was expected that performance with the speed/time format would be more accurate than with either the speed or time format. Additionally, it was expected that, as in other studies, participants would underestimate their speeds in the speed format, causing them to travel at an average speed that is faster than the commanded speed. Alternatively, in the time format, participants' time estimations would be significantly shorter than the commanded times, resulting in arrival times that are later than the commanded arrival times.

Accuracy was expected to decrease as distance increased because error would accumulate as pilots progressed through the route, making short routes the most accurate and long routes the least accurate, with the increase in error due to pilots arriving earlier for longer routes. It was also expected that there would be more accurate compliance and later arrival times as commanded speed increased. However, because the commanded speed is an average across the trial, pilots might have difficulty safely compensating for slower spool-up and turn speeds during short routes with faster than average speeds, thus causing them to taxi at an average speed slower than the commanded speed. Thus, a pattern of less error as commanded speeds increased, for medium and long routes, but not for short routes was expected.

Additionally, it was expected that the formats would produce different patterns across the duration of the taxi route. Because the speed format did not provide inherent performance feedback, pilots were not cued to correct their error at the end of the route. However, it was expected that for the time and speed/time formats, error would accumulate during the beginning and middle of the route and then, when time information became more relevant near the end of the route, pilots would use the time information to correct their error. These different patterns across the taxi route for the formats were expected with medium and long routes, but not with short routes because they will not provide enough distance/time to accumulate substantial error.

Standard Deviation (SD) of speed is a measure that can be used to assess speed consistency. Because the speed and speed/time formats provided performance/control information that could be used to regulate speed maintenance, variability for the time format was expected to be greater than the speed and speed/time formats, which should not differ. Pilots' speed variability with the three formats was also expected to produce different patterns during the routes. With the speed format it was expected that variability would be large near the beginning of the route and decrease and stabilize throughout the middle and end of the route. In the time format, the pattern should show that variability remains at a consistently low level until the end of the route where it would increase due to error correction. The speed/time format was expected to produce a combination of the patterns expected from the speed and time formats. Specifically, the variability would be high at the beginning and end of the routes. However, because the short routes might not provide enough time and/or distance during which pilots could simply maintain their speed instead of actively attempting to correct their error, these format patterns were not expected with the short routes.

In order to gain insight into how pilots used the different formats, a head-mounted eye tracker was used to collect information about what percentage of the trial pilots looked at the display information. In general, pilots were expected to look at display information more when correcting error than when simply maintaining their performance. Because the time format provided conformance monitoring information as opposed to control information, it was expected that pilots would look at the time display less than the speed display. Alternatively, the speed/time format provided a combination of status and control information from the speed and time formats, making it likely that pilots would look at the speed/time display more than the time or speed displays. The patterns of display usage during the routes were expected to differ between formats. With the speed format, pilots should look at the display more during the beginning of the route, whereas with the time format they should look at the display more near the end of the route. Again, the speed/time format should show a combination of the speed and time patterns, with more use near the beginning and end of the route, because it is during those phases of the routes that the displays are most salient for error correction.

II. Methods

B. Participants

Eighteen instrument-rated, current commercial airline captains or first officers, who had flown a minimum of 500 hours as captains within the last 4 years, participated in the experiment. The mean age of the pilots was 49 years (SD = 8 years) and the mean number of flight hours as a commercial captain was 8,383 hrs (SD = 8,031 hrs) with a range of 500 to 30,000 hrs. All participants were male, and had normal vision or vision corrected to normal with soft contact lenses. Pilots were compensated for their time.

C. Apparatus

The experiment was conducted at the NASA Ames Research Center in a medium-fidelity surface operations simulator with Boeing 737 modeled dynamics. The simulated environment modeled Dallas Fort-Worth International Airport at 1,200 Runway Visual Range. The main out-the-window scene was projected at approximately 3,500 Lumens by an LCD projector onto a 1.83 m by 2.44 m screen that was approximately 2.44 m away from the eye point of the participant (see Fig. 1). Peripheral views were displayed on two 41 cm (H, 31.89 deg visual angle) by 30 cm (V, 24.19 deg) monitors located approximately 71 cm to the left and right of the participant, respectively. A static, north-up airport diagram and text taxi clearance were displayed on a 33 cm (H, 18.46 deg visual angle) by 24 cm (V, 13.54 deg visual angle) monitor located approximately 102 cm in front of the pilot, below the out-the-window scene (see Fig. 2). Pilots controlled the simulated aircraft using a tiller, throttle, and rudder toe-brakes. While taxiing, pilots wore an Applied Science Laboratory head-mounted eye tracker, Model 501, attached to an adjustable headband. The eye tracker and headband combined weighed 227 g.

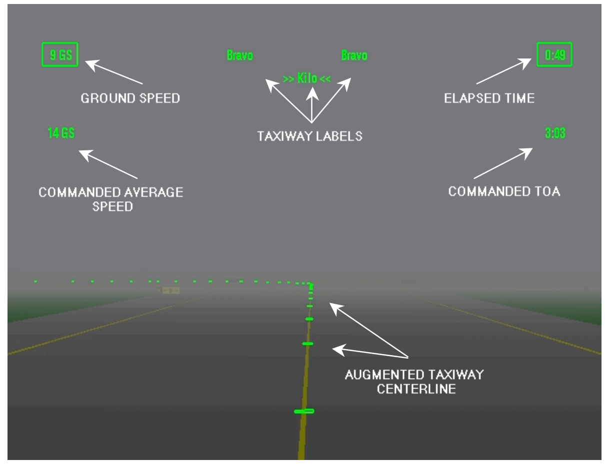

Figure 1. Actual out-the-window view of the speed/time display (not to scale). Note: Symbology was displayed in green.

At the beginning of the experiment, pilots filled out a demographics questionnaire. At the end of each of the three blocks of experimental trials, participants completed a written questionnaire assessing their subjective measures of workload and of their own performance for each block. The questionnaire for the final block also included questions assessing their preferences for the command formats.

D. Displays

The current instantaneous ground speed indicator, elapsed time, commanded taxi speed, and commanded TOA were superimposed graphically over the out-the-window screen simulating a Head-Up Display (HUD). When the command was in the Speed format, the current ground speed was displayed inside a line box in the upper left corner of the main screen and the commanded speed was displayed 15 cm below it (see Fig. 1). The text of the ground speed measured 5 cm (H, 1.15 deg) by 3 cm (V, 0.57 deg), and the commanded speed measured 5 cm (H, 1.16 deg) by 3 cm (V, 0.58 deg). For the Time format, elapsed time was displayed inside a line box in the upper right corner of the out the window screen and the commanded time was displayed 15 cm below it. The text of the elapsed time measured 5 cm (H, 1.14 deg) by 3 cm (V, 0.58 deg) and the commanded time measured 5 cm (H, 1.15 deg) by 3 cm (V, 0.57 deg). When the command was in the Speed/Time format, current

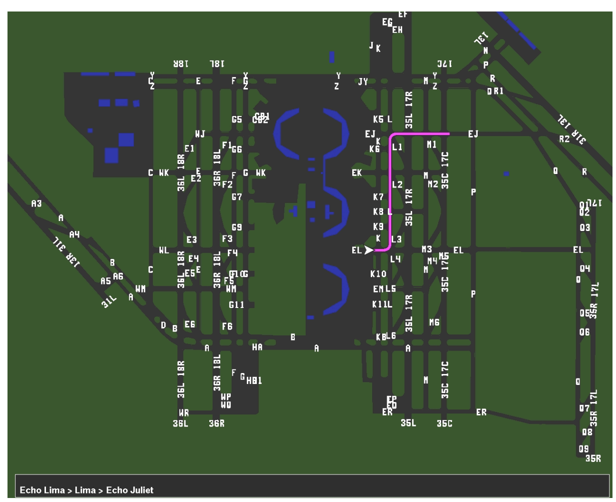

Figure 2. Electronic map of Dallas-Fort Worth Airport with the taxi clearance highlighted in magenta and the ownship icon at the beginning of the taxi clearance. The text clearance is displayed below the map.

ground speed, commanded ground speed, elapsed time, and commanded time were displayed in their respective locations simultaneously (see Fig. 1).

During all trials, the simulated HUD displayed a “text triad” in the upper center of the out-the-window screen (see Fig. 1) that was 46 cm (H, 10.71 deg) by 10 cm (V, 2.39 deg). The text triad indicated the current taxiway on which the aircraft was traveling and the upcoming taxiways on the left and right respectively. In order to aid with navigation in the absence of a first officer, the HUD also displayed a highlighted centerline superimposed over the out-the-window scene along the cleared taxi route. This “scene-linked” centerline symbology conformed to the out-the-window scene as though it was part of the simulated environment.[13]

An electronic, north-up diagram of the airport and a text taxi clearance were also present during all of the trials. At the beginning of each trial, while the pilot received and read back the taxi clearance, the map displayed an ownship icon at the starting point of the clearance and a magenta line highlighting the cleared taxi route (see Fig. 2). After the pilot repeated the clearance back to ground control (the experimenter), the trial started and the ownship icon and highlighted route disappeared in order to maintain a map similar to those used in current day operations and to avoid drawing the pilot’s attention away from the out-the-window view. The text taxi clearance remained at the bottom of the diagram throughout the trials.

E. Experimental Design

This experiment was a 3 (format) x 3 (distance) x 4 (4-D command), within subjects design. The three levels of command format were speed, time, and speed/time. For the speed format, pilots were instructed to taxi to a specific runway crossing at a commanded average speed. With the time format, pilots were instructed to taxi to a specific runway crossing by a commanded TOA. There was also a speed/time format in which they were given both a commanded speed and its corresponding commanded TOA and pilots were instructed to comply with both of them. The 4-D commands were based on combinations of four commanded speeds (in kts) and three route distances. In order to calculate commanded TOAs for the time and speed/time formats, the given route distance was divided by the commanded speed (converted to equivalent ft per min). Thus, when given a commanded TOA, it was based on an implicit speed/distance combination. The 4 commanded speeds for the 3 route distances were equivalent to 12 commanded TOAs, listed in Table 1. For presentation and discussion purposes, data are presented in terms of the four commanded speeds for all conditions (time, speed, and speed/time formats according to Table 1).

Table 1. Commanded Time of Arrival (min:s) as a Function of Route Distance and Commanded Speed

	Commanded speed			
Route distance	10 kts	14 kts	18 kts	22 kts
3,000 ft	2:58	2:07	1:39	1:21
6,000 ft	5:55	4:14	3:17	2:41
12,000 ft	11:51	8:28	6:35	5:23

Note: Actual distances were approximate and within 5% of the target distance.

Additional analyses were conducted with a 3 (format) x 3 (distance) x 4 (4-D command) x 4 (quartile), within subjects design. The four levels of quartile represent the 4 equal distance quarters of the route, and were based on data markers positioned proportionally to the distance of each route. The data markers were not visible to the participants and were used only for data analysis. The experiment was comprised of three blocks of 15 trials each. Each block consisted of one of the three command formats. The presentation order of the blocks was counterbalanced across subjects and the 4-D commands and route distances were randomized within each block and matched for each format.

F. Procedures

Before beginning the first experimental block, participants completed three initial phases: familiarization, training, and introduction. The participants first taxied through three familiarization trials in order to become familiar with the simulator and the airport. During the familiarization trials the pilots were given a taxi clearance and a ground speed indicator and were told to taxi at a variety of speeds at which they were comfortable, while staying on route. The training trials ensured that participants were calibrated to the simulator and were able to estimate speed based on the cues from the simulator. The participants completed two to three training trials during which they were asked to taxi at specific speeds (10, 20, and 30 kts) without using a ground speed indicator. Once they indicated that they thought they had reached the target speed, they were shown their actual speed on the main screen. They repeated this task until they had successfully taxied within 20% of the target speed two times. Finally, in order to ensure that they understood the different conditions that they would see during the experimental trials, the participants taxied through one introductory trial for each command format, consisting of the same displays and procedures as the upcoming experimental trials.

Following a short break, the experimenter gave the participants detailed instructions for the experimental trials and fit the eye tracker for calibration. The experimenter told the participant what his first command format was and he then completed 3 practice trials and 12 experimental trials for that format. For each trial, the participants received a verbal taxi clearance from ground control via an intercom, a text taxi clearance at the bottom of the airport diagram, and a taxi command in the speed, time, or speed/time format. The experimenter, acting as ground control, read the taxi clearance to the participants and confirmed that their read back of the clearance was correct. Once they received this information, a tone was sounded and the trial began. The pilots' task was to maintain the commanded average speed, or arrive at the destination at the commanded time, as closely as possible without arriving early or late. Trials contained a small amount of realistic traffic around the airport, but there were no potential incursions or close calls with any of the traffic present. After each trial, pilots were shown their actual average speed, or TOA, as well as the commanded speed, commanded TOA, or both (depending on the condition).

The participants completed one block, consisting of 15 trials of the first command format, in approximately 90 min. After they finished the first block they completed a questionnaire and then took a 45-min lunch break. After the break, the practice, experimental, and questionnaire process was repeated for the two remaining blocks until they had completed the experiment. The participants took another 10-min break between the second and third blocks. After completing all of the blocks, the participants were debriefed. Pilots participated individually and each pilot completed the entire experiment in one day, which lasted approximately 8 hrs including instruction, breaks, and lunch.

III. Results

A three-way, repeated measures analysis of variance (ANOVA) with format (speed, time, and speed/time), distance (3,000; 6,000; and 12,000 ft), and 4-D command (10, 14, 18, and 22 kts) was performed on TOA absolute error, TOA error bias, SD of speed, and percentage of dwell time, in order to assess the accuracy, consistency, and strategy with which pilots complied with taxi commands presented in different formats. Additional 4-way, repeated measures ANOVAs were also performed on the same dependent variables with format (speed, time, and speed/time), distance (3,000; 6,000; and 12,000 ft), 4-D command (10, 14, 18, and 22 kts), and quartile (1, 2, 3, 4), in order to investigate pilots' performance during different phases of the routes. Because the quartiles were based on distance (proportionate to the distance of the route), variation in pilots' taxi rates during the trials resulted in different quartile durations and unequal data sampling between quartiles. Consequently, only the results of interactions among quartile, format, distance, and 4-D command were analyzed because the calculation of main effects would be inaccurate. When analyzing performance by quartiles, TOA accuracy must be interpreted carefully, because despite being instructed to attempt to be consistent throughout the taxi routes, pilots were not explicitly instructed to get to certain points of the routes at specific times. Although the main task was to reach the end of the route at a specific time (commanded via time or speed, or both), it is useful to assess pilots' time or speed profile across the route. That is, independent of the actual route completion time, for example, did pilots reach the mid-point of the route in half the total commanded time? The dependent variables were calculated in two different ways for the quartile analyses. For TOA absolute error and TOA error bias, the commanded TOAs for the quartiles were calculated as a proportion of the overall commanded TOA (e.g., For an overall commanded TOA of 6 min the commanded TOA at Quartile 2 would be 3 min), and that was subtracted from the actual TOA at each quartile. SD of speed and percentage of dwell time were calculated as independent segments of the routes (e.g., Quartile 2 was calculated from the first data marker up to the second data marker). In order to obtain pilot feedback and guidance for future research, subjective data were collected.

Of the 18 participants, 2 pilots made a wrong turn and went off of their cleared route during one trial. These trials were stopped and were retested with new clearances with equivalent turns, formats, distances, and 4-D commands. The pilots completed the substitution trial at the end of the block during which the substitution occurred. A trial was also replaced for a third participant due to technical difficulties with the simulator. This trial was also retested with an equivalent clearance.

G. TOA Absolute Error

TOA absolute error was calculated as the absolute value of the difference between the commanded TOA and pilots' actual TOA at the designated runway crossing at the end of each route. This measure of error magnitude was used to assess the accuracy of compliance with the 4-D commands. In order to test the hypotheses that pilots would taxi most accurately with the speed/time format and that their accuracy would improve as route distances decreased and 4-D commanded speeds increased, a 3 (format) x 3 (distance) x 4 (4-D command) repeated measures ANOVA was conducted on TOA absolute error. Overall, with a Huynh-Feldt correction of sphericity, the analysis revealed that there was a main effect of format, $F(1,21) = 5.95$, $p < .05$, and a significant interaction between distance and command, $F(3,54) = 4.00$, $p < .01$. Pair-wise comparisons of the formats, with a Bonferroni correction, indicated that the TOA absolute error for the speed/time format (M = 8.87 s) was significantly lower than TOA absolute error for the speed format (M = 15.54 s), $t(17) = 3.72$, $p < .01$; and the time format (M = 20.28 s), $t(17) = 3.43$, $p < .01$. Pilots' TOA compliance was more accurate with the speed/time format than with either the speed or time formats (see Fig. 3).

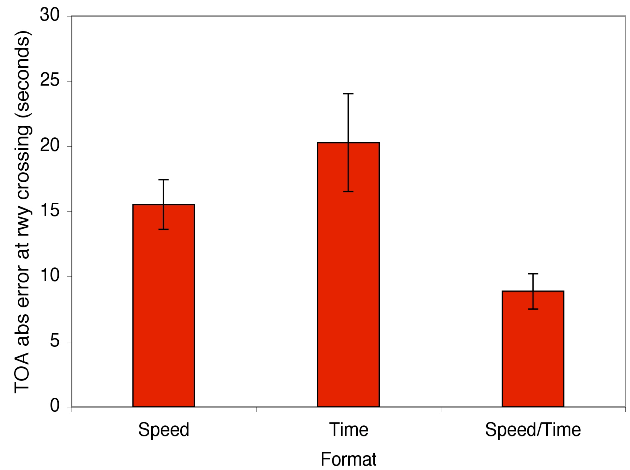

Figure 3. Mean time of arrival (TOA) absolute error for display formats. Vertical lines indicate +/- 1 standard error of the mean.

Simple effects analyses for the interaction between distance and 4-D command revealed a significant effect of distance for the commanded speed (CS) 10, $F(2,53) = 5.81$, $p < .01$; the CS 14, $F(2,53) = 11.08$, $p < .01$; the CS 18, $F(2,53) = 4.56$, $p < .05$; and the CS 22, $F(2,53) = 4.79$, $p < .05$. There was also a significant effect of 4-D command for short, $F(3,71) = 6.76$, $p < .01$, medium, $F(3,71) = 11.30$, $p < .01$; and long, $F(3,71) = 7.57$, $p < .01$, routes. A trend analysis revealed a significant linear by quadratic interaction between distance and 4-D command, $F(1,17) = 6.35$, $p < .05$. As shown in Fig. 4, TOA absolute error increased as routes lengthened and this increase in error was larger as 4-D commanded speed was slower (i.e. the slope of this trend became more negative as 4-D commanded speed decreased). These results indicate that pilots' taxied more accurately with shorter routes and faster commanded speeds.

A 3 (format) x 3 (distance) x 4 (4-D command) x 4 (quartile) repeated measures ANOVA was conducted on TOA absolute error. There was a significant interaction between format and quartile, with a Huyhn-Feldt correction, $F(3,47) = 7.30$, $p < .01$. Based on predictions about the pattern of accuracy within trials, simple effects analyses were performed on quartile at each level of format. The analyses revealed that TOA absolute error differed between quartiles with the speed, $F(3,71) = 8.18$, $p < .01$; time, $F(3,71) = 4.22$, $p < .01$; and speed/time formats, $F(3,71) = 5.23$, $p < .01$. In order to determine if the shapes of the quartile trends differed between formats, separate trend analyses were conducted on each level of format. These analyses revealed that there was a significant linear trend across quartiles for the speed format, $t(68) = 4.95$, $p < .001$; indicating that TOA absolute error increased as quartile increased; and a significant quadratic trend for the time and speed/time formats, $t(68) = -3.39$, $p < .01$, and $t(68) = -3.24$, $p < .01$, respectively. This indicates that TOA absolute error increased from early to later quartiles, but that increase diminished for later quartiles (see Fig. 5). Because the time and speed/time formats had quadratic trends, a 2 (format) x 3 (distance) x 4 (command) x 4 (quartile) repeated measures ANOVA was performed on TOA absolute error, without the speed format, in order to determine if the quadratic trends differed. There was still a significant interaction between format and quartile, $F(3,51) = 3.44$, $p < .05$; indicating that the quadratic trend was larger for the time format than for the speed/time format.

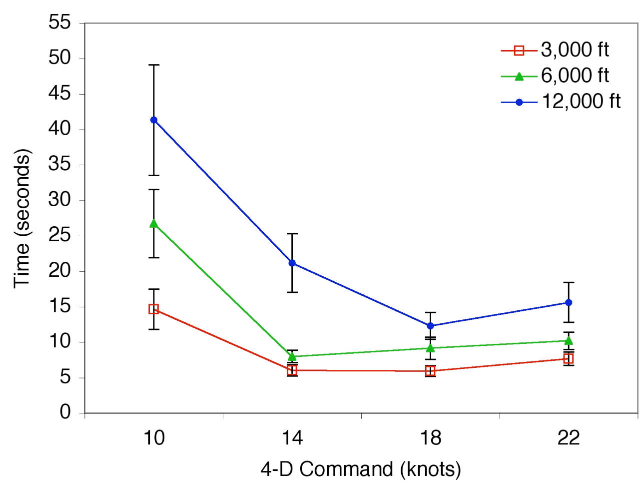

Figure 4. Mean time of arrival (TOA) absolute error as a function of the interaction between distance and 4-D command. 4-D command is plotted in terms of speed, but the 4-D commands were also presented as TOAs depending on the display format. Vertical lines indicate +/- 1 standard error of the mean.

It was also of interest to examine the simple effects of format, which revealed that TOA absolute error differed between formats for Quartile 1, F(2,53) = 29.89, p < .001; Quartile 2, F(2,53) = 55.83, p < .001; Quartile 3, F(2,53) = 26.33, p < .001; and Quartile 4 F(2,53) = 5.04, p < .01. Despite the main effect of format showing no differences between the time and speed formats at the ends of the routes (see Fig. 3 or Quartile 4 in Fig. 5), TOA absolute error was higher for the time format than for the speed format during Quartile 1, t(17) = 5.78, p < .001; during Quartile 2, t(17) = 8.52, p < .001; and during Quartile 3, t(17) = 5.71, p < .001: but the speed and speed/time formats did not differ. This analysis reveals that the speed and speed/time formats were actually more comparable, with regards to TOA accuracy during the first three quartiles, than the speed and time formats were. These results are also supported by the subjective data, which indicate that pilots rated the speed and speed/time formats as easier to use than the time format.

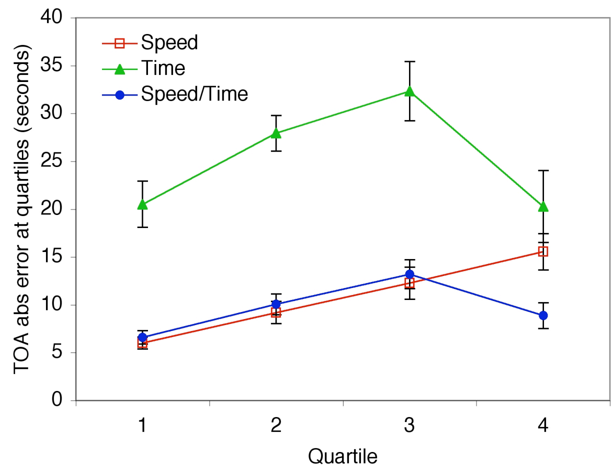

Figure 5. Mean time of arrival (TOA) absolute error as a function of the interaction between format and quartile. Note that the data for Quartile 4 are equivalent to those in Figure 3. Vertical lines indicate +/- 1 standard error of the mean.

H. TOA Error Bias

TOA error bias was measured as the signed difference between the commanded TOA and the pilots' actual TOA. This was calculated by subtracting the commanded TOA from the actual TOA. Thus, negative values indicate an early TOA and positive values indicate a late TOA. This measure of the error direction was analyzed in order to investigate how early or late pilots' compliance was and also indicated if they were taxiing, on average, faster or slower than the 4-D commanded speed. In order to test the hypotheses that pilots would arrive early with the speed format and late with the time format, a 3 (format) x 3 (distance) x 4 (4-D command) repeated measures ANOVA was conducted on TOA error bias. There was no main effect of format, nor did it interact with any other variable. Overall, with a Huynh-Feldt correction of sphericity, the analyses revealed that there was a significant 2-way interaction between distance and 4-D command, F(5,78) = 7.22, p < .001. Simple effects analyses of the interaction between distance and 4-D command revealed that TOA error bias for 4-D commands differed for short, F(3,71) = 13.19, p < .001; medium, F(3,71) = 22.43, p < .001; and long, F(3,71) = 14.26, p < .001, routes. Simple effects analyses also revealed that TOA error bias differed between route distances for CS 10, F(2,53) = 4.65, p < .05; CS 14, F(2,53) = 8.33, p < .01; and CS 18, F(2,53) = 7.23, p < .01; but not for CS 22. A trend analysis revealed a significant linear by linear interaction, F(1,17) = 18.97, p < .001. As seen in Fig. 6, with the exception of CS 22, pilots arrived at the runway crossings earlier as distance increased, and across distances pilots arrived earlier as 4-D commanded speed decreased.

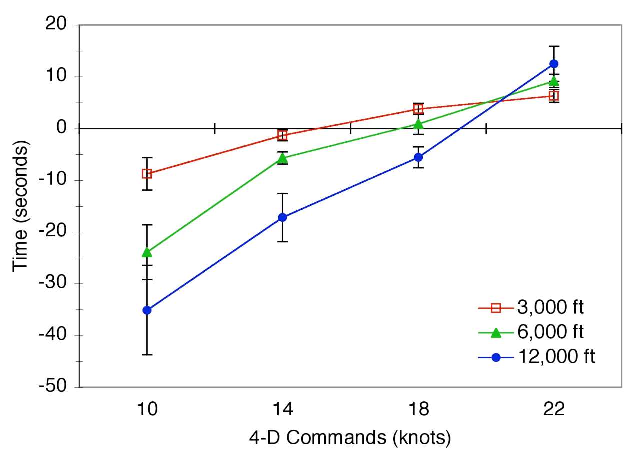

Figure 6. Mean time of arrival (TOA) error bias as a function of the interaction between distance and 4-D command. 4-D command is plotted in terms of speed, but the 4-D commands were also presented as TOAs depending on the display format. Vertical lines indicate +/- 1 standard error of the mean.

I. SD of Speed

The SD of speed was calculated as the SD of the ground speed of each trial. This variable indicated how consistent pilots' speed maintenance was around their average speed, regardless of the accuracy of their compliance. In order to test the hypotheses that pilots' speed maintenance would be more consistent with the speed and speed/time formats and with slower 4-D commands and longer route distances, a 3 (format) x 3 (distance) x 4 (4-D command) repeated measures ANOVA was conducted on the SD of speed. Overall, with a Huyhn-Feldt correction of sphericity, the analysis revealed that there was a significant main effect of format, $F(2,33) = 92.60$, $p < .001$. Pair-wise comparisons of format, with a Bonferroni correction, indicated that the SD of speed in the time format (M = 7.23 kts) was significantly higher than in the speed format (M = 4.27 kts), $t(17) = -10.96$, $p < .001$; and in the speed/time format (M = 4.74 kts), $t(17) = -10.75$, $p < .001$; but the speed and speed/time formats did not differ.

A 3 (format) x 3 (distance) x 4 (4-D command) x 4 (quartile) repeated measures ANOVA was conducted on SD of speed. Whereas there was no expected interaction among format, distance, and quartile, results did reveal a significant interaction between format and quartile, $F(3,59) = 8.54$, $p < .001$. Based on predictions about the pattern of accuracy within trials, simple effects analyses were performed on quartiles at each level of format. The analyses revealed that SD of speed differed among quartiles with the speed, $F(3,71) = 98.48$, $p < .001$; time, $F(3,71) = 50.73$, $p < .001$; and speed/time formats, $F(3,71) = 74.08$, $p < .001$. Post hoc trend analyses were conducted at each level of format in order to determine what the shapes of the quartile patterns were. These analyses revealed that there was a significant cubic trend for the speed format, $F(1,68) = 18.93$, $p < .001$; for the time format, $F(1,68) = 11.11$, $p < .01$; and for the speed/time format, $F(1,68) = 15.50$, $p < .001$.

In order to determine if these cubic trends differed from each other, repeated measures ANOVAs were performed with two combinations of two of the three formats. A 2 (format) x 3 (distance) x 4 (4-D command) x 4 (quartile) repeated measures ANOVA was performed on SD of speed, with the time and speed/time formats only. This analysis revealed a significant interaction between format and quartile, $F(3,51) = 4.85$, $p < .01$; indicating that the cubic trend of quartile was larger for the speed/time format. As shown in Fig. 7, for time and speed/time, SD of speed decreased between Quartile 1 and Quartile 2, and then increased as quartile increased; but the initial decrease in SD of speed was larger for the speed/time format than for the time format. Another 2 (format) x 3 (distance) x 4 (4-D command) x 4 (quartile) repeated measures ANOVA was performed on SD of speed, with the speed/time and speed formats only. This analysis also revealed a significant interaction between format and quartile, $F(2,38) = 8.06$, $p < .01$; indicating that the cubic trend of quartile was again larger for the speed/time format. However, in this case, the initial decrease in SD of speed is equivalent, whereas the subsequent increase of SD of speed as quartile increased was larger for the speed/time format than for the speed format.

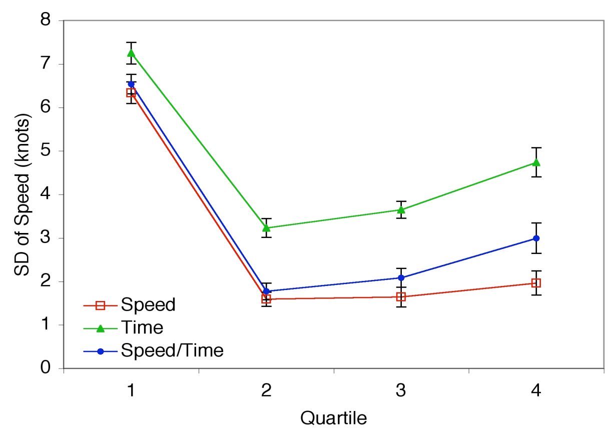

Figure 7. Mean standard deviation of speed as a function of the interaction between format and quartile. Vertical lines indicate +/- 1 standard error of the mean.

J. Percent Dwell Time

Percentage of dwell time was calculated as the proportion of dwell time that pilots were looking at various pieces of display information. Total dwell time was calculated as the cumulative time during which the eye tracker detected a minimum pupil diameter of 30 mm while the eye was fixated in one location for a minimum of 0.25 s Due to technical difficulties, no valid eye tracking data was collected for one participant. Consequently, analyses of dwell time information consisted of valid eye tracking data from 17 of the 18 participants. Data for 10 individual trials were replaced with the mean of the appropriate conditions because the eye calibration was lost due to shifting of the headband or technical difficulties with the eye tracker. These trials were distributed among five participants. Of the 10 trials, 6 trials were from the speed format, 1 trial was from the time format, and 3 trials were from the speed/time format. Additionally, data from an entire block (the time format) from one subject were lost due to technical difficulties. These data were replaced with means from the appropriate conditions.

In order to test the hypotheses that pilots would look at the speed/time display the most, a 3 (format) x 3 (distance) x 4 (4-D command) repeated measures ANOVA was conducted on the percentage of dwell time on display information. Overall, the analysis revealed that there was a significant main effect of format, $F(2,32) = 14.88$, $p < .001$. Pair-wise comparisons of format, with a Bonferroni correction, indicated that the percentage of dwell time was significantly lower for the time format (M = 9.6%) than for the speed (M = 19.8%), $t(16) = 5.10$, $p < .001$; and the speed/time (M = 19.9%) formats, $t(16) = 4.48$, $p < .01$ (see Fig. 8). Pilots' subjective judgments that they referred to the speed information more than time information in both the speed and speed/time formats support this finding.

It was also of interest to compare the percentage of dwell time on speed and time information within the speed/time format in order to determine how pilots used the information when presented together. Pair-wise comparisons revealed that percentage of dwell time on speed information in the speed/time format ($M = 12.8\%$) was significantly lower than the speed format ($M = 19.8\%$), $t(16) = 3.72$, $p < .05$. The percentage of dwell time on time information ($M = 7.4\%$) was lower than the percentage of time on speed information ($M = 12.8\%$) in the speed/time format, $t(16) = 2.69$, $p < .05$. However, percentage of dwell time did not differ between time information in the speed/time format and the time format, as seen in Fig. 9.

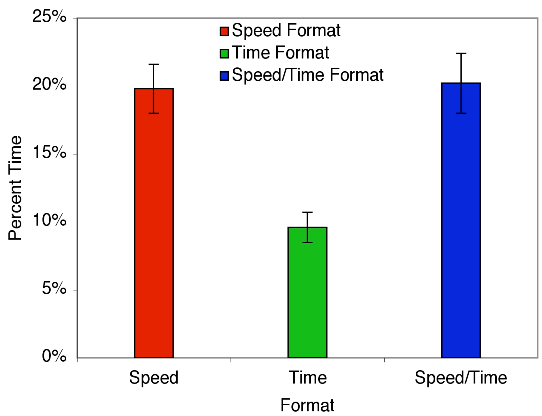

Figure 8. Mean percentage of eye dwell time for formats. Vertical lines indicate +/- 1 standard error of the mean.

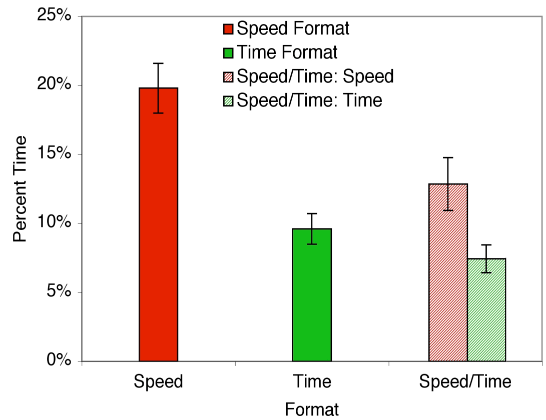

Figure 9. Comparison of the mean percentage of eye dwell time, between the speed and time formats and the speed and time information within the speed/time format. Vertical lines indicate +/- 1 standard error of the mean.

In order to investigate the differences between the quartile patterns for percent dwell time, a 3 (format) x 3 (distance) x 4 (4-D command) x 4 (quartile) repeated measures ANOVA was conducted on percentage of dwell time, which revealed a significant interaction between format and quartile, $F(3,46) = 23.80$, $p < .01$. Trend analyses revealed a significant quadratic trend across quartiles for the speed format, $t(64) = 3.33$, $p < .01$; the time format, $t(64) = 2.93$, $p < .01$; and the speed/time format, $t(64) = 3.22$, $p < .01$. As predicted, pilots looked at the speed format display most at the beginning of the route and looked at the time format display most near the end of the route, whereas with the speed/time format the quartile pattern appears to be a combination of the speed and time format patterns (see Fig. 10).

Again, it was of interest to examine the quartile patterns for speed and time information within the speed/time format. In order to investigate these differences within the trials, a 2 (format) x 3 (distance) x 4 (4-D command) x 4 (quartile) repeated measures ANOVA was conducted on percentage of dwell time, which compared the speed format with the speed information in the speed/time format only. There was a significant interaction between format and quartile, with a Huyhn-Feldt correction, $F(2,24) = 4.52$, $p < .05$. Trend analyses revealed a significant linear trend for speed information in the speed/time format, $t(64) = 2.85$, $p < .01$; and a significant quadratic trend for the speed format, $t(64) = 3.33$, $p < .01$. Whereas percentage of dwell time in the speed format decreased at the beginning of the trial and then leveled off, in the speed/time format the percentage of dwell time for the speed information declined at the beginning of the routes and continued to decline slightly near the end of the routes as well, as seen in Fig. 11.

A 2 (format) x 3 (distance) x 4 (4-D command) x 4 (quartile) repeated measures ANOVA was also conducted on percentage of dwell time, which compared the time format with the time information in the speed/time format only. There was no significant interaction between format and quartile, and the trend analyses revealed a significant quadratic trend for time information in the speed/time format and in the time format, $t(64) = 3.98$, $p < .001$; and $t(64) = 2.93$, $p < .01$, respectively. Regardless of format, when using time information, pilots' percentage of dwell time increased as quartile increased, with a large increase near the end of the routes (see Fig. 11).

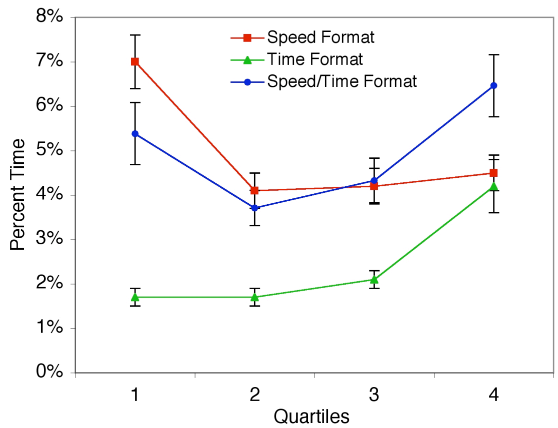

Figure 10. Mean percentage of eye dwell time as a function of the interaction between format and quartile. Vertical lines indicate +/- 1 standard error of the mean.

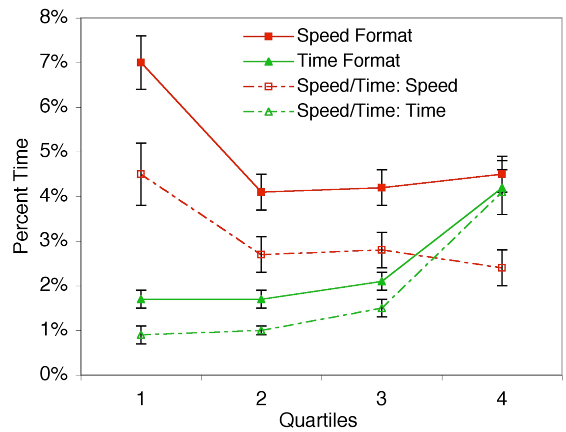

Figure 11. Comparison of the mean percentage of eye dwell time, among the speed format, the time format, and the speed information and time information within the speed/time format as a function of the interaction with quartile. Vertical lines indicate +/- 1 standard error of the mean.

K. Summary

Overall, results indicate that pilots' TOA absolute error was the lowest with the speed/time format, but when analyzed throughout the trials, performance with the speed format was equivalent to the speed/time format during the first three quartiles and did not diverge until the ends of the routes. Pilots also taxied more accurately with shorter routes and faster 4-D commanded speeds. TOA error bias data revealed that pilots arrived equally early regardless of format, but arrived earlier with slower 4-D commanded speeds and longer distances. The results also revealed that pilots taxied the least consistently with the time format and as 4-D commanded speed increased and as distance decreased. Finally, percentage of dwell time data indicate that pilots looked at the speed and speed/time displays more than the time display. Additionally, quartile analyses revealed that pilots' patterns of display usage when given both speed and time information were equivalent to the patterns of usage with the individual speed and time formats in that they looked at the speed information primarily at the beginning of the route and looked at the time information primarily near the end of the route.

IV. Discussion

L. Compliance Accuracy

As expected, pilots' compliance to 4-D taxi commands was most accurate when taxiing with the speed/time format, however, there was no difference between the speed format and the time format. Pilots did accumulate error throughout the speed format trials, whereas with the time and speed/time formats pilots accumulated error and then corrected that error, typically by slowing down, near the end of the trials (see Fig. 5). Despite the finding that there was no difference between pilots' accuracy with the speed and time formats at the end of the routes, the speed format was more accurate than the time format during the first three quartiles. Pilots also rated the speed format as easier to use than the time format. It was only at the end of the trials that the combination of the error correction that occurred with the time format and the error accumulation associated with the speed format led to equivalent error with the two formats. The equivalent compliance accuracy between the speed and speed/time formats during the first three quartiles indicates that even when presented with the speed/time format, pilots likely used the speed information as a guidance indicator during those taxi phases and incorporated the time information for conformance monitoring primarily near the end of the route. As seen in Fig. 5, pilots' compliance with the speed/time format demonstrated high accuracy as in the speed format and error correction near the end of the route as in the time format. This indicates that presenting both speed, as a guidance indicator, and time information, for conformance monitoring, did produce a redundancy gain when considered across the entire route. Incorporating both speed and time information in more advanced future displays could potentially facilitate general accuracy throughout the clearance and focused error correction near critical locations or destinations.

Pilots' compliance accuracy declined as distance increased, and that decline was more pronounced as 4-D commanded speed decreased. When taxiing short routes, pilots were most accurate and did not experience the more drastic declines in accuracy when given slower 4-D commanded speeds, as they did with longer routes. This can be explained by previous research, which indicates that people estimate slower speeds less accurately than faster speeds, and that people estimate long durations less accurately than short durations.[7,9] As hypothesized, shorter routes provided less distance and time during which estimation error could accumulate. Such an interaction between route distance and commanded speed could guide the definition of performance limitations in the presence of certain clearance characteristics. Dynamic algorithms could incorporate this interaction to set limitations on one variable when the other variable falls outside certain parameters.

In addition to investigating the degree to which pilots were accurate in complying with 4-D commands, determining the bias of that compliance provided a more complete picture of their performance. Results for TOA error bias did not support the hypothesis that pilots would tend to arrive late with the time format and early with the speed format. Pilots arrived equally early with all three formats. The pilots did, however, arrive earlier as distance increased, and as 4-D commanded speed decreased. The results also confirmed the prediction that distance would interact with 4-D command and pilots would arrive earlier as 4-D commanded speed decreased, but that pattern would not be as pronounced for short routes. It appears that pilots had a tendency to arrive early, possibly because arriving early could be easily corrected by slowing down, but arriving late could be impossible to correct if the error was detected too late. However, it is also possible that the simulated environment was not as realistic as taxiing in the real world, causing pilots to taxi at speeds faster than they would feel comfortable with in a real aircraft or causing them to disregard the 4-D command because they found it excessively slow. Additionally, whereas pilots arrived earlier with longer distances for the slower 4-D commanded speeds, they arrived equally late regardless of distance for the fastest commanded speed (22 kts). This pattern might indicate that taxiing at an average of 22 kts was too difficult and pilots had to attempt to catch up when complying with that 4-D command. If pilots taxied the simulated aircraft faster than they would a real aircraft, it is likely that they might find it even more difficult to taxi at an average of 22 kts in a real world situation.

M. Consistency

As predicted, pilots' speed maintenance was less consistent with the time format than with the speed or speed/time formats. The presence of control information in the speed and speed/time formats enabled pilots to taxi within a smaller range of speeds around their average speed. With the time format pilots did not have any feedback or explicit guidance regarding control information, resulting in less consistent speed. Speed variance was large at the beginning of the routes regardless of format because the initial speed/engine spool up required pilots to overcompensate in order to reach their 4-D command. There was a larger increase in consistency after the spool up with the speed and speed/time formats than with the time format. Additionally, with the time and speed/time formats, pilots taxied less consistently towards the end of the routes when they corrected their error; with the speed format their speed maintenance remained consistent because there was no time information to cue them to correct their error. Thus, as predicted, pilots' speed maintenance was least consistent at the beginning of the routes. Contrary to predictions, pilots' speed maintenance in the time format was also highly variable at the beginning of the routes, but did increase again at the end of the routes when time became more salient for error correction. As predicted, pilots' speed maintenance with the speed/time format appeared to be a combination of the speed and time trends, with the least consistency near the beginnings and ends of the routes. Given that the increase in speed variability near the beginning and end of the routes with the speed/time format was due to error correction, a predictive algorithm could incorporate that information into a profile of taxi phases and increase the error buffer in those areas accordingly.

N. Strategy

As expected, pilots looked at the time display less than the speed and speed/time displays, however, their use of the speed/time display and the speed display was roughly equivalent. Although pilots did not use the speed/time format more than the speed format, they did use each of these two formats twice as much as the time format. Subjective data indicate that at least half of the pilots stated that they found themselves looking at the speed information when they should have paying attention to the outside world during both the speed and speed/time formats. Given that the displays were simplistic text displays, this problem could potentially be alleviated through the use of a more advanced display designed to minimize attentional capture, such as scene-linked symbology.[12] However, more research is warranted in order to determine if pilots experience cognitive tunneling when provided with speed information during taxi operations. As expected, within the speed format, pilots looked at the display the most at the beginning of the trial while correcting their initial spool-up error. Within the time format, pilots looked at the display consistently during the first three quartiles, but used the display most near the end of the routes while they corrected their error before the runway crossing. With the speed/time format, pilots demonstrated the predicted combination of usage from the speed and time formats in that they looked at the display frequently at the beginning and at the end of the routes when they were correcting error.

In order to confirm that the increases in display usage were due to an actual combination of the speed and time patterns, analyses within the speed/time format examined the patterns in more detail. Pilots used the speed and time information proportionally within the speed/time format as they did in the speed and time formats. The data clearly revealed that when both the speed and time information were displayed simultaneously, pilots used the speed information most at the beginning of the routes and used the time information most at the end of the routes, just as they did in the speed and time formats. The pattern for the use of speed information in the speed/time format was slightly different than in the speed format, in that usage of speed information in the speed/time format decreased near the end of the trial. However, the pattern for time information was the same within the speed/time format as in the time format. Both pieces of information appear to be useful for complying with 4-D commands and produce a redundancy gain by presenting relevant information in complementary formats that are useful during different taxi phases. Future displays could incorporate these patterns by providing speed information throughout a 4-D clearance, but only providing time information during relevant phases of taxi (right before a checkpoint or destination).

V. Conclusions

In order to taxi precisely, pilots must have relevant display information to help them comply with 4-D taxi commands. Combining speed and time clearances provides redundant information about movement and enables pilots to switch between relevant information during appropriate phases of taxiing. Utilizing the high taxi accuracy associated with speed information and the error correction facilitated by time information could enable pilots to safely comply with coordinated runway crossings. The consistent speed maintenance also associated with speed information could enable SMSs to accurately predict and coordinate traffic flow for surface operations. In addition to providing both speed and time information, shorter taxi clearances also promote more precise taxiing when paired with average taxi speeds and durations.

More research is required before this study can be generalized to real world situations. Given that the simulation provided no proprioceptive feedback and a smaller than average amount of airport traffic, pilots might have taxied faster and less cautiously than they would at a real airport (although previous research in this simulator has been validated against full-mission and flight test data). Additionally, the display was intentionally simplistic in order to capture the baseline nature of the information. The display was also presented on a simulated HUD, and might not be as salient or accessible when displayed differently (e.g., head down). However, these results could be useful in guiding the design of future advanced 4-D taxi displays. Designers might consider including explicit speed information for control and time information for implicit feedback. Additionally, when designing within the limitations of pilots' compliance with 4-D commands, clearances should likely remain shorter (or be issued in multiple short segments) in order to avoid error accumulation. If incorporated appropriately, such displays could become an invaluable tool for efficient taxi operations.

Implications of results for system design:

- Providing speed and time information produces more accurate and consistent compliance.
- Shorter routes produce more accurate compliance.
- Faster 4-D commands produce more accurate compliance.
- Speed information is relevant during initial taxi phases for consistent speed control.
- Time information is relevant during critical checkpoints for precise error correction.

References

[1]Glass, B. J., & Gawdiak, Y., "Integrated navigation applications in airport surface traffic management," *Saint Petersburg International Conference on Integrated Navigation Systems,* Vol. 4, May 1997, pp. 105-113.

[2]Cheng, V. H. L., Sharma, V., and Foyle, D. C., "A study of aircraft taxi performance for enhancing airport surface traffic control," *IEEE Transactions on Intelligent Transportation Systems,* Vol. 2, No. 2, 2001, pp. 39 - 54.

[3]Hecker, P., Doehler, H. U., Korn, B., and Ludwig, T., "Extending enhanced vision capabilities by integration of advanced surface movement guidance and control systems," (A-SMGCS*). Proceedings of SPIE, the International Society for Optical Engineering*, Vol. 4363, August 2001, pp. 58-66.

[4]Cheng, V.H.L., and Foyle, D.C., "Automation tools for enhancing ground-operation situation awareness and flow efficiency," *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, 2002-4856, Monterey, CA, 2002.

[5]Denton, G. G., "A subjective scale of speed when driving a motor vehicle," *Ergonomics*, Vol. 9, No. 3, 1966, pp. 203-210.

[6]Recarte, M. A., and Nunes, L. M., "Perception of speed in an automobile: Estimation and production," *Journal of Experimental Psychology: Applied,* Vol. 2, No. 4, 1996, pp. 291-304.

[7]Zakay, D., Nitzan, D., and Glicksohn, J., "The influence of task difficulty and external tempo on subjective time estimation," *Perception & Psychophysics*, Vol. 34, No. 5, 1983, pp. 451-456.

[8]Zakay, D., "On prospective time estimation, temporal relevance and temporal uncertainty," *Time, Action, and Cognition,* edited by Macar, F., Pouthas, V., and Friedman, W. J., Kluwer Academic Publishers, Netherlands, 1992, pp. .

[9]Brown, S. W., "Time perception and attention: The effects of prospective versus retrospective paradigms and task demands on perceived duration," *Perception & Psychophysics*, Vol. 38, No. 2, 1985, pp. 115-124.

Wickens, C. D., and Hollands, J. G., *Engineering Psychology and Human Performance*, 3rd ed., Prentice-Hall Inc., New Jersey, 2000, pp. 106-107,220,299.

[10]Selcon, S. J., "Using redundancy in the design of time-critical warnings: A theory-driven approach," *Human factors in auditory warnings,* Ashgate, Vermont, 1999, pp. 201-224.

[11]Awe, C. A. and Johnson, W. W., "The effect of scene content on speed, time, and distance perception," *International Symposium on Aviation Psychology*, Vol. 7, April 1993, pp. 281-285.

[12]Foyle, D. C., McCann, R. S., and Shelden, S. G., "Attentional issues with superimposed symbology: Formats for scene-linked displays," *Proceedings of the Eighth International Symposium on Aviation Psychology*, edited by R. S. Jensen, & L. A. Rakovan, Ohio State University, Ohio, 1995, pp. 98-103.