To appear in the *Proceedings of the 46th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, Human Factors and Ergonomics Society, 2002.

THE IMPACT OF COMMUNICATION DELAYS ON AIR TRAFFIC CONTROLLERS' VECTORING PERFORMANCE

Esa M. Rantanen, Jason S. McCarley, and Xidong Xu
University of Illinois at Urbana-Champaign

An experiment was conducted to investigate the impact of communication delays in the pilot—controller communication loop on air traffic controllers' performance and workload. Four levels of constant systemic audio delay (AD), 150, 250 ms, 350 ms, and 1,000 ms, and two levels of variable pilot delay (PD), zero delay and realistic delay, were employed. Vectoring accuracy and the controller's final turn initiation served as dependent variables. Subjective workload was measured by the NASA-TLX workload index. Eight subjects proficient in the task participated in the experiment. Random PD had significant effects on both vectoring accuracy and final turn initiation; accuracy was reduced and initiation times were earlier when PD was added. However, data showed no effect of AD on vectoring accuracy, and no evidence of compensatory strategies in response to increasing AD levels. Results suggest that variability in the subjects' turn initiation effectively masked the impact of even the longest AD of 1,000 ms on vectoring accuracy. The NASA-TLX showed no effect of AD on workload in either PD condition.

INTRODUCTION

An area particularly sensitive to air traffic congestion is that of radiotelephony communication between pilots and controllers. Means of communication between pilots and controllers is one of the fundamental principles of air traffic control (ATC) (Hopkin, 1995), yet air-ground communication is in many respects the weak link of the system, with many accidents attributed to improper or misunderstood communications (Morrow, Lee, & Rodvold, 1993; Nolan, 1999). Voice communication between pilots and controllers is also susceptible to factors such as noise and language problems that degrade the system's reliability, as well as to the influence controllers' and pilots' expectations, biases, and other cognitive factors. Consequently, the language, phraseology, and procedures used in air-ground communication are extensively standardized and regulated (e.g., FAA, 2000). Such measures, unfortunately, result in a cumbersome system with substantial channel occupancy times associated with even the simplest messages. As a result, frequency congestion has become a factor that constrains the capacity of the National Airspace System (NAS).

Given the dynamic nature of ATC, the temporal aspects of the controllers' mental picture are apparent, as are the temporal demands of their tasks: Controllers need to anticipate aircraft trajectories and pilot intentions well into the future, plan their actions, and then execute the planned actions at a proper time and in an appropriate sequence. Time lags, which are universally detrimental to human performance in control tasks—resulting in increased human workload and system errors (Wickens & Hollands, 2000)—could have a significant negative impact on controllers' work. Controllers direct traffic by issuing specific instructions and commands to pilots who then, after an unavoidable time delay, execute the instructed maneuvers. If the delays between command issuance and execution are very long or highly variable, then accurate prediction of response times and consequences may become difficult and subsequently increase controllers' mental workload and the probability of errors.

Two specific types of time delays between the issuance of controllers' instructions and their actual execution in the cockpit can be identified. The first, here termed audio delay (AD), is defined as the time elapsed from the depression of the controller's microphone's transmit key to the moment his or her voice is heard in the aircraft cockpit. It should be noted, however, that this definition does not account for the situation where there is an additional delay between the microphone key press and when the controller begins to speak. As defined here, AD is the product of the technology used in voice transmissions (hence regarded as systemic), and is characteristically constant (or near-constant) and very short (< 1 s). The second type of delay, here termed pilot delay (PD), is defined as the time elapsed from the moment a pilot hears the controller's instruction until the moment she either keys the microphone to respond vocally or manipulates the aircraft controls in response. Time lags of this type, being the product of human (i.e., pilot) performance limits, are highly variable and sometimes relatively long (several seconds) (Cardosi, 1993).

To effectively investigate the temporal aspects of air traffic controllers' tasks and the impact of communication delays on their performance, it is important to identify situations where these are most critical to some measurable outcome. The relevance of time in ATC becomes evident when prompt compliance with a controller's instruction is necessary for an aircraft to maintain separation from other traffic. Here, the communication is essentially one-way, and the pilot's readback is secondary in importance to execution of the instructed maneuver. Such situations frequently arise when arriving traffic is sequenced for the final approach to a given runway. This task is known as final sequencing, and it is one of the most demanding tasks in ATC. Final sequencing is commonly accomplished by radar vectoring, that is, by issuing pilots specific headings and airspeeds, which they must fly to intercept the final approach course at a given location. The final intercept point is determined by the distance from the runway threshold necessary for a stabilized approach as well as by the minimum separation from other aircraft approaching and landing on the same or closely spaced parallel runway. If the controller is late with the instruction to turn on the final the aircraft will intercept the final approach course farther from the runway than expected and will possibly end up too close to other aircraft turning onto the final behind it. If the controller is too early with the turn instruction, conversely, the aircraft will intercept the final closer to the runway than expected and possibly end up too close to the preceding aircraft on the same final (see Figure 1).

METHOD

Subjects

A total of eight subjects thoroughly proficient with the experimental task participated in the study. The mean age of the six male and two female subjects was 42.75 years with a range of 36 to 66 years. All subjects gave their informed consent to participate, in accordance with the University of Illinois policies.

Apparatus

An ATC simulator was constructed for the purposes of this experiment. The simulator consisted of a Dell Optiplex® personal computer (PC) and a 21-inch Trinitron® flat screen display. A computer program was created to display simulated STARS radar screen with runway centerline, weather returns, and aircraft symbols displayed. Realistic aircraft dynamics were used in calculation of target trajectories and positions on the simulated radarscope.

A custom-built communications box allowed the subjects to issue voice commands to the simulated targets. This system consisted of a headset and a push-to-talk switch. Pilot communications were prerecorded as wave files and played to the subject at times determined by the experimental program. This allowed for a very accurate measurement of timing of the communications between the subjects and pilots.

The Experimental Task

The subjects' task was to turn an aircraft approaching the final approach course on a perpendicular heading to intercept the final at a predetermined point. The subjects' task was limited to controlling the timing of the turn; the heading of the turn was fixed at a 30-degree intercept angle, emulating the constraints often present in the final sequencing task and emphasizing the importance of timing and the effect of time lags (see Figure 1). This task provided a good balance between experimental control and realism with respect to tasks of air traffic controllers, as well as allowed for exact definition and measurement of dependent variables.

Design

Independent variables. The effect of transmission delay was simulated by adding a specified AD to the transmission of the subjects' instruction. Four levels of AD were employed: 150 ms, 250 ms, 350 ms, and 1,000 ms. Two levels of PD were employed, zero delay and realistic delay. The zero delay condition allowed investigation of the impact of the different AD levels in isolation, without the possible masking effect of PD. The realistic delay condition allowed for investigation of effects of AD under naturalistic circumstances, embedded in PD. The realistic delays were sampled from either a Gaussian distribution with a mean of 2.5 s and a standard deviation (SD) of 2 s, or a lognormal distribution with a location and scale parameters chosen so that the sample had approximately the same mean and SD as in the Gaussian case.

Dependent variables. The dependent variables were the outcome of the subjects' actions in a final sequencing task measured as the error from the criterion intercept point at 10 nm from the runway, and the point of the subjects' instruction initiation. The former can be regarded as a "system measure," or a measure of the impact of communication delays on the ATC system performance. The latter can be regarded as a "human measure;" if the delays present in the system were perceptible to the subject he or she could have altered the point of turn initiation to achieve high intercept accuracy in spite of them, in which case the effects of AD and PD would not be evident in the intercept error measure. Workload was measured by using the NASA TLX subjective workload index. Workload was not manipulated, but the experimental scenario was designed to induce consistently high workload in all conditions.

Experimental design. This was a within-subjects factorial design with two independent variables, the audio delay (AD) and the pilot (execution) delay (PD). The experiment was thus a 2 × 4 factorial design, resulting in eight experimental blocks. Replicates were created by including 10 aircraft within each block. The order of the blocks in the experiment was completely randomized.

Procedure

The subjects were explained the purpose of the experiment, and the task was explained and demonstrated to them. The subjects then practiced the task until their performance was stable in their own judgment. After the subjects reported that they were comfortable with the task, they completed the eight experimental blocks, each consisting of 10 trials (aircraft). After each block, the subjects filled the NASA TLX workload index, administered by the same computer as the simulation.

Figure 1. A screenshot from the experiment. Because the aircraft turning radius (which is a function of bank angle and speed), baseleg location, and intercept heading were constant, the point of final intercept was determined by timing of the turn initiation instruction ("[callsign] turn left heading 210, cleared for ILS runway 18 approach") and subsequent delays in carrying out the instruction.

RESULTS

A manipulation check confirmed that independent variables were indeed present in the simulation as designed. The data were also first examined in a time series for each subject in each block. Stable performance was typically achieved from the fourth trial on; hence, only trials 4—10 were included in the subsequent analyses.

Overall, the subjects' performance in the experimental task was very good. The mean intercept error across all subjects and all trials was -0.08 nm with an SD of 0.64 and a range from -1.8 nm to 3.4 nm, where a negative error means that the plane intercepted closer to the runway than the criterion 10 nm (see Figure 2).

Data were analyzed with a general linear model (GLM) of Analysis of Variance (ANOVA), using PD and AD as within-subject factors. The level of statistical significance was set at $\alpha < .05$ in all analyses. The effect of PD was statistically significant, $F(1,438) = 16.68$, $p < .001$. The aircraft intercepted farther from the runway on the final when PD was present than when it was absent, which is consistent with the geometry of the problem. The effect of AD, however, was not significant.

Turn initiation was defined as position of the aircraft at the moment the subject keyed the microphone to begin his or her transmission to instruct the plane to turn on the final intercept heading. Because the heading was predetermined, this was the only way the subjects could adjust the intercept point on the final approach course. Overall, however, only very small effects could be observed in the turn initiation data (Figure 3). The effect of PD was statistically significant, $F(1,438) = 6.61$, $p = .01$, and indicates an appropriate response to the additional PD in the communication loop, that is, the subjects initiated the turn farther from the final as the pilots took longer to execute the turn. The effect of AD was not significant, however, $F < 1$.

The overall workload ratings on the NASA TLX subjective workload index were very low, only up to the 30s on a scale up to 100. Workload was also rated inconsistently with respect to the AD added to the PD. In many cases workload was rated lower in the PD + 1000 ms AD condition than at lower AD levels. Overall workload showed no main effect of AD, $F < 1$, but was significantly higher when PD was added to AD (mean rating = 17.3) than without PD (mean rating = 9.3), $F(1, 7) = 19.513$, $p = .003$. Data also produced a significant interaction of PD and AD, $F(3, 21) = 3.717$, $p = .027$, reflecting the fact that the difference between overall workload ratings with and without PD were greater at larger AD levels. Simple-effects tests, however, revealed no significant effect of AD either with or without PD.

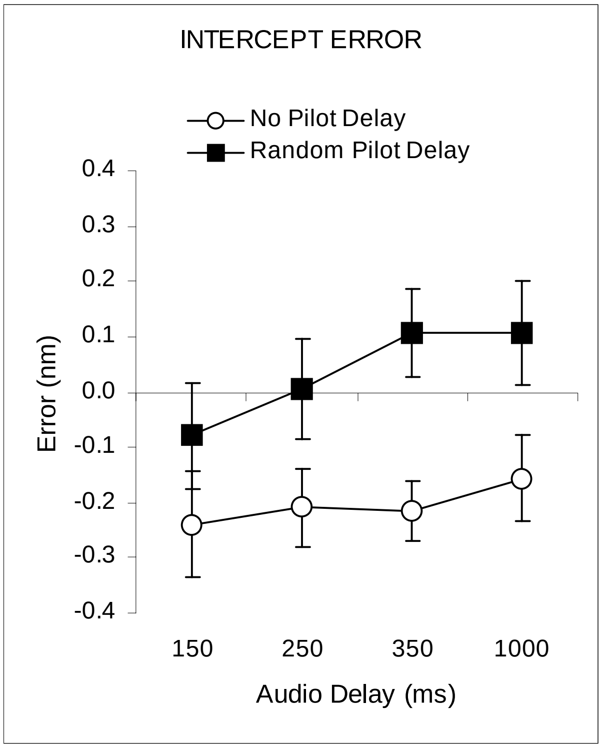

Figure 2. Intercept error, measured from the 10 nm criterion intercept point on the final approach course. A negative error means that the plane intercepted closer to the runway and a positive error farther from the runway than the criterion 10 nm. The effect of PD was significant, but not that of AD. The error bars represent ± one SE of the mean.

DISCUSSION

The subjects' impressions of the experimental task were generally very favorable. The subjects' performance was also good in all experimental conditions, with final intercepts remarkably close to the 10 nm criterion. Experimental conditions with random PD resulted in intercepts farther along the final than when PD was absent. Actually, the subjects' performance was better in the PD trials, that is, closer to the target intercept at 10 nm (see Figure 2). This might have been because the PD trials were more realistic than the trials without PD and the subjects might have had difficulty getting used to the fast responses of the pilots in the no PD trials. No one brought this up in an interview, however, so there is no evidence that the subjects were actually aware of the experimental conditions while working on the task.

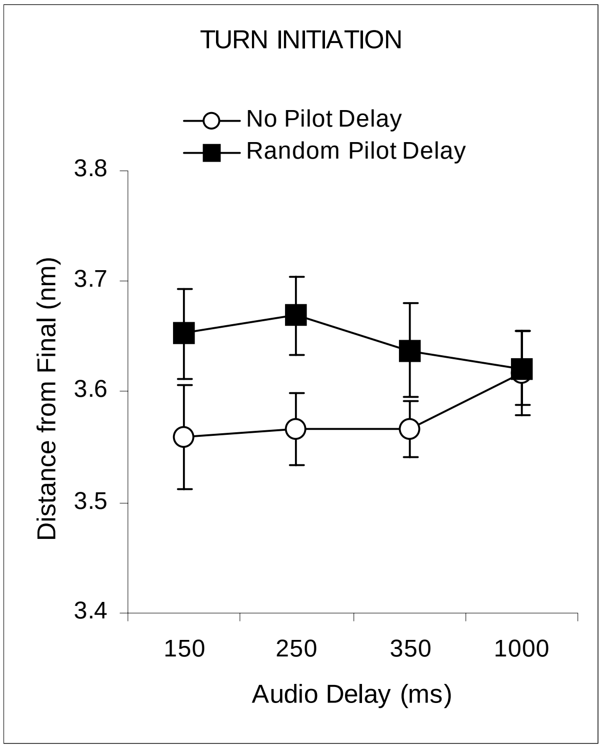

Figure 3. Turn initiation, measured as a distance from the final approach course. The effect of PD was significant, but not that of AD. The error bars represent ± one SE of the mean.

The effects of constant AD on intercept accuracy and turn initiation were generally very small, within 0.25 nm. In addition, these were realized in conditions that were very extreme: the subjects were limited to one turn to a predetermined heading and were allowed no other control techniques. A slight increase in distance from final of turn initiation can be seen in the 1,000 ms AD condition without a PD, suggesting compensatory technique. However, lack of statistical significance raises the question whether this was simply due to chance.

The overall workload ratings were very low and rated inconsistently with respect to the AD added to the PD. In many cases workload was rated lower in the PD + 1000 ms AD condition than at PD + lower AD levels. The constant AD had no effect on the NASA-TLX frustration ratings.

Finally, it is important to evaluate the results in terms of the accuracy demands of the task. Intercept errors in the most demanding PD + 1000 ms AD condition were within 0.1 nm from the target. It is doubtful that any greater accuracy would be within human capabilities even in the best of conditions.

It is most likely that the variability in PD (randomly sampled from a given distribution to simulate operational conditions) had a far greater effect on workload than a constant ADs, which were in most cases much smaller than the PDs. There was also much variability both between and within the subjects, despite the experimenters' best efforts to enforce high control in the experiment.

This research also addressed several important methodological issues. Experimental research on various aspects of ATC is very difficult due to the complexity and dynamic nature of the system and the innumerable variables present in even simplest scenarios. In addition to making ATC experiments expensive and time-consuming, these factors pose a serious threat to the validity of research. This research sought to address such problems from the onset.

First, it was imperative to operationally define the independent and dependent variables in such a manner that the successful manipulation of the former could be verified in the data and that the latter could be accurately and reliably measured. This ensured construct validity of the research. Equally important was to control extraneous variables to the highest degree possible to safeguard internal validity. Due to the typically applied nature of ATC research, external validity (i.e., generalizability of results) is also critical, albeit to a lesser extent than construct and internal validity. In other words, one cannot have external validity if construct and internal validity are compromised. To achieve reasonably high external validity the experimental task was designed to be highly realistic and representative of "worst case scenarios" controllers would encounter in the operational work. Finally, realistic tasks also ensured good face validity, that is, acceptance by the subjects participating in the research. This was important to secure subject motivation and good performance in the experimental trials.

CONCLUSION

The experiment reported here sought to examine the impact of an additional, systemic, delay in the pilot—controller communication loop. Because the delays were relatively short (150, 250, 350, and 1,000 ms), it was critical to choose an experimental task in which the delays would have a measurable impact on a clearly defined outcome. Equally important was to exert maximum control over any conceivable extraneous variables to bring out the necessarily very small size effects. The task that fulfilled these design criteria was final sequencing for an ILS approach. In addition, the task was constrained by a predetermined script to minimize between—subjects differences and performed in isolation from any concurrent tasks would normally be present in the operational ATC environment. Yet, the task was reasonable as highly constrained scenarios such as were employed here can also be found in the operational environment. In congested airspaces the controller has often only limited number of alternatives (e.g., headings or altitudes) to choose from and the success of his or her control actions will depend on their appropriate timing. In this sense, the experimental task was realistic.

The results only partially substantiate the success of the design, however. Despite the careful attention to control of extraneous variables, much uncontrollable variability was evident in the subjects' performance. The source of this variability was the subjects' rate of speech and initiation of turns for final. The implications of these findings are important: First, if the inherent variability in subjects' performance was sufficient to partially mask the effects of systemic audio delay in a controlled laboratory experiment, these effects may not be detectable at all in more realistic full-fidelity simulations or in the operational environment. Second, it is crucial to examine and understand the amount and sources of variability in operational environments when making conjectures about potential effects of new technologies. In the case of air—ground communications, the "communications landscape" in different airspaces and ATC facilities should be carefully mapped to fully predict the implications of various levels of systemic communication delays.

REFERENCES

Cardosi, K. M. (1993). Time required for transmission of time critical air traffic control messages in an en route environment. *The International Journal of Aviation Psychology*, *3*(4), 303-313

Federal Aviation Administration (2000). *Air traffic control* [7110.65M]. Washington, DC: Author

Hopkin, V. D. (1995). *Human factor in air traffic control* Bristol, PA: Taylor & Francis.

Morrow, D., Lee, A., & Rodvold, M. (1993). Analysis of routine controller-pilot communications. *Intrernational Journal of Aviation Psychology*, *3*(4), 285-302.

Nolan, M. S. (1999). *Fundamentals of air traffic control* (3rd ed.). Pacific Grove, CA: Brooks/Cole Publishing Company.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.