

Analysis of Voice Transmissions of Air Traffic Controllers in the Context of Closed Loop Communication Deviation and its Relationship to Loss of Separation

Christopher S. Lieber, Yancy Vance Paredes, Aaron Zhen Yang Teo, and Nancy J. Cooke
Arizona State University
Mesa, AZ

The present research examines a pattern-based measure of communications based on Closed Loop Communications (CLC) and non-content verbal metrics to predict Loss of Separation (LOS) in the National Airspace System (NAS). This study analyzes the transcripts from six retired Air Traffic Controllers (ATC) who participated in three simulated trials of various workloads in a TRACON arrival radar simulation. Results indicated a statistically significant model for predicting LOS based on CLC deviations (CLCD), word count in transmission, words per second, and traffic density. However, more research is required to evaluate the significance of each communication variable to predict LOS.

INTRODUCTION

The U.S. National Airspace System (NAS) consists of controlled and uncontrolled airspace immediately over the U.S., serving 45,000 flights totaling 3,000,000 passengers daily (FAA, 2021a). The U.S. FAA introduced NexGen, a suite of 40 projects, to modernize and improve the safety and efficiency of the NAS (FAA, 2020). Here, ATCs (air traffic controllers) play a vital role in ensuring air safety. Among their many tasks is maintaining safe separation between aircraft (minimum of 3 miles spacing horizontally within 40 miles of a major airport and 5 miles otherwise) (FAA, 2021b). Failure in doing so results in what is known as a loss of separation (LOS), which could lead to a serious safety violation or a fatal collision. A contributing factor of LOS is the division of attention, such as introducing secondary tasks (e.g., other aircraft under control or those needing special attention). This attention-performance relationship has been demonstrated in driving studies (Biondi et al., 2015; Kantowitz, 2000; Tijerina & Goodman, 1996). For ATCs, this could be attributed to the demanding nature of the job, particularly pertaining to workload. Although not the only factor, workload has been identified as a determining factor in causing human error (Kantowitz & Sorkin, 1983), such as communication errors that potentially result in an aircraft's altitude deviation or loss of standard separation (Cardosi et al., 1998). Workload also affects how controllers communicate.

Communication has been identified as one of the factors for the success of the Trajectory Based Operations of NexGen (FAA, 2022). ATCs communicate with pilots through radio, exchanging verbal messages, and are prone to human error. Most aviation accidents can be attributed to human error (Billings & Cheaney, 1981). An example is miscommunication which can be affected by several factors, including the pilot's workload, audio signal quality, speech accents of ATC and pilot, English proficiency, and use of standard phraseology (Molesworth & Estival, 2015). Several recommendations have been made to reduce miscommunication. For example, controllers are recommended to speak slowly and distinctly (Cardosi et al., 1998). Another recommendation is to keep transmissions short

with a maximum of four instructions per transmission to prevent misunderstandings (Barshi, 1997). Message length could be a clear source of error in miscommunication in aviation communication (Barshi, 1997; Cardosi et al., 1998). This is attributed to the cognitive theory, where the limit of short-term memory is reached. Communication is susceptible to distortion of systematic, environmental, and internal factors that affect its comprehensibility or speech intelligibility (SI). In ATC, many factors contribute to low intelligibility, such as the quality of spoken utterance, background noise, speech accent, or a fast speaking rate. Low SI of ATC has been identified as one of the major causes of aircraft accidents attributed to the high speaking rate (O'hare et al., 1994).

Both controllers and pilots are responsible for ensuring clear and comprehensible radio communication. They follow a standard format that involves a loop of confirmation or correction (i.e., closed-loop communication (CLC)) where (1) a sender transmits a message; (2) the receiver actively listens to the message; (3) the receiver repeats the message back to the sender; and (4) the sender actively listens for the readback making sure it is correct. This reduces errors and information loss (Burke et al., 2004; Lingard et al., 2008; Parush et al., 2011). Furthermore, it can alleviate workload during heavy traffic (Härgestam et al., 2013). However, the success of this strategy depends on the correct execution of the four steps. This could be due to the shared cognition (Cooke et al., 2004), which helps reduce the workload demand from both the controller and the pilot.

The purpose of this paper is to use the above-reviewed literature to analyze the potential of measuring different communication features and patterns of CLC to predict LOS. This was supported by the following research questions:

RQ1: What significant relationships exist between different non-content communication variables and LOS?

RQ2: Are there any significant relationships between communication errors and LOS?

RQ3: What relationships exist between the types of communication errors and non-content communication metrics?

METHOD

Metacraft simulated the Terminal Radar Approach Control (TRACON) in Phoenix Sky Harbor (KPHX) Quartz West and Southeast arrivals. Three 25-minute scenarios with varying workloads were designed, namely (1) baseline, (2) high workload, and (3) high workload off-nominal. Participant workload was manipulated by increasing traffic density (4-5 aircraft for baseline and 10-12 for high workload) and incorporating off-nominal events (e.g., moderate turbulence, pilot deviation NORDO (No radio aircraft), runway switch, and minimal fuel advisory). The study was conducted at Arizona State University TRACON Simulation Lab.

Participants

Six retired ATCs were recruited for this study. All participants had experience working in the civilian TRACON and had experience working in various sectors of airspace ($M=30$, $SD=10.97$). Participants were exposed to all three experimental scenarios. They were debriefed, interviewed, and compensated after the experiment.

Materials and Apparatus

Metacraft. Metacraft is a mid-fidelity radar simulation used to train ATC students. ATCs in this environment have access to digital flight strips and can use a mouse to click on aircraft to accept them. In this experiment, five computers were used, and at each computer, users are assigned their role within the simulation: ATC, ghost controller, or one of three pseudo-pilots. The participant performs the role of the TRACON. The ghost controller signs into a station and functions as a tower. The pseudo-pilots work in three stations, with each station allowing them to control Standard Terminal Arrival Route (STAR); ARLIN4, BLYTHE5, SUNNS8, and HYDRR1. Metacraft captures information on the distance between aircraft, the time an aircraft enters the airspace, and the total duration of the experiment.

Transcripts. Audio recordings from the workstations were processed using automatic speech recognition software. The experimenters manually proofread and corrected the transcripts afterward. The corresponding speaker of every transmission was identified.

Procedure

The experiment took approximately four hours (1 hour for training, orientation, and consent forms; 1 hour and 15 minutes for hands-on practice; three 25-minute sessions; and 30 minutes for debriefing). Participants were made aware of the following constraints for all scenarios:

- You must accept all handoffs from the center approach. The center will not hold.
- You will only hand off to the final approach/KPHX tower. No route modifications that result in aircraft leaving your control.

- You will not request/issue commands to land at an airport other than the field destination. No alternate airports. You may only hand off to the final approach.
- Keep aircraft in your airspace. No handoffs (except to KPHX tower) and no point outs.
- You must not declare emergencies.

In each scenario, participants interact with trained pseudo pilots trained on phraseology, proper communication procedures, and simulation controls (console commands used to manage aircraft assigned to them). Pseudo pilots control several aircraft following their assigned routes.

ANALYTICAL PROCEDURE AND RESULTS

The analytical procedure used for this experiment involved a multiple regression analysis to measure various non-content verbal features in predicting LOS. The second analysis used was a two-step hierarchical regression model which measured the effect of each communication error that was detected by CLCD and contributed to the prediction of LOS beyond density. Pearson correlation was then applied to measure the relationships between each communication variable and communication error type.

Measures

Closed-Loop Communication Deviations (CLCD). A coding scheme was used to classify any deviations from an expected pattern of CLC. Closed-loop patterns were analyzed and checked whether they conform to expected communication patterns, specifically: (1) pilot→ATC→pilot communication or (2) ATC→pilot→ATC. Any deviation from these was counted as CLCD using binary coding applied to each communication in the transcript: 0=expected pattern and 1=CLCD.

Density Category. Each scenario was categorized based on traffic density as either high (10-12 aircraft in a sector) or low (average 4-5 aircraft in a sector).

Loss of Separation (LOS). For this experiment, LOS was counted when the distance between two aircraft was less than 1000 feet vertically and 5 nm laterally.

Word Count. The total number of words for every transmission was counted using the transcript.

Words Per Second. The number of words of every transmission was divided by its time duration, which was computed using its timestamps.

CLCD classification. Transmissions that were flagged as being a CLCD were categorized and counted based on the definitions found in Table 1 as used in Lieber et al. (2021).

Table 1
Types of CLC Deviations

Type	Description
Normal	Pattern changes due to normal circumstances. This could result from a pilot changing radio channels as advised by the ATC or a pilot concludes a communication with an ATC, and a second pilot calls in.

Step Over	A pilot calls the ATC, or an ATC calls out to a pilot during another transmission. When two people are attempting to talk on the same frequency simultaneously.
Interruption	This happens when a pilot or an ATC starts a communication before the loop of the communication finishes
Repeat	A communication that was already transmitted by the sender is transmitted again by the same sender shortly following the first transmission.
Error Correction	A transmission is sent by a sender, and the sender repeats the transmission shortly afterward correcting wrong information conveyed in the original transmission.
Interruption-Neglect	Interruptions that were not addressed were counted as being neglected and coded under the category Interruption-Neglect.

Multiple Regression

A multiple regression was applied to predict LOS from density, word count, words per second, and CLCD. The model met the assumption of linearity as assessed by partial regression plots and a plot of studentized residuals against the predicted values. There was independence of residuals, Durbin-Watson statistics of 1.638. There was homoscedasticity assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. There was no evidence of multicollinearity as all tolerance values were greater than 0.1, and no values for Cook's distance above 1. Leverage values varied, the majority of cases had a leverage value less than .5, with one case having a leverage value of .869; this case was inspected and determined to not affect the regression coefficients and followed the pattern established by all of the other data points. There were no studentized deleted residuals above ± 3 standard deviations. The assumption of normality was met, as assessed by a Q-Q Plot. The multiple regression model statistically and significantly predicted LOS, $F(4, 13)=3.603$, $p<.05$, adj. $R^2=.380$. Although the model produced was statistically significant, there was no support, as shown in the coefficients in Table 2, that any of the variables were statistically significant, $p<.05$. Details on the regression coefficients and standard errors can be found in Table 2.

Table 2
Regression Coefficients to Predict LOS

Variable	<i>B</i>	β	<i>SE</i>
Constant	244.896		19.486
Density Category	62.376	.472	44.037
Word Count	-16.281	-.233	15.981
Words per second	-9.587	15.365	-1.33
CLCD	2.071	.306	2.367

Note: Averages were calculated for WPS and Word Count.

Hierarchical Regression

A hierarchical multiple regression was run to determine if the addition of each communication error detected by CLCD improved the prediction of LOS over and above density. The complete model of density, normal CLCDs, and each categorical communication error detected by CLCD (step-over, interruption, error correct, repeat, and neglect) to predict LOS was not statistically significant $R^2=.679$, $F(7, 10)=3.024$, $p=.055$; adj. $R^2=.455$. Analysis of the coefficients indicated that two of the communication errors detected by CLCD, neglect and repeat, negatively affected the prediction of LOS, details of these coefficients as well as the coefficients for each variable in the model can be found in Table 3.

Table 3
Hierarchical Regression Coefficients

Variable	<i>B</i>	β	<i>SE</i>
Constant	-7.618		37.699
Density Category	72.933	.552	41.402
Normal	3.625	.304	2.762
Step Over	20.051	.284	15.333
Interruption	8.465	.331	11.028
Interruption-Neglect	-9.451	-.268	15.275
Repeat	-6.947	-.218	9.540
Error Correction	8.621	.156	13.662

Note. No variables were statistically significant, $p>.05$.

Table 4
Hierarchical Regression Coefficients (excluding neglect and repetition)

Variable	<i>B</i>	β	<i>SE</i>
Constant	-12.779		35.368
Density Category	59.212	.448	36.514
Step Over	3.389	.284	2.572
Interruption	25.193	.357	12.814
Normal	1.565	.061	5.405
Error Correct	4.030	.073	11.969

Note. No variables were statistically significant, $p>.05$.

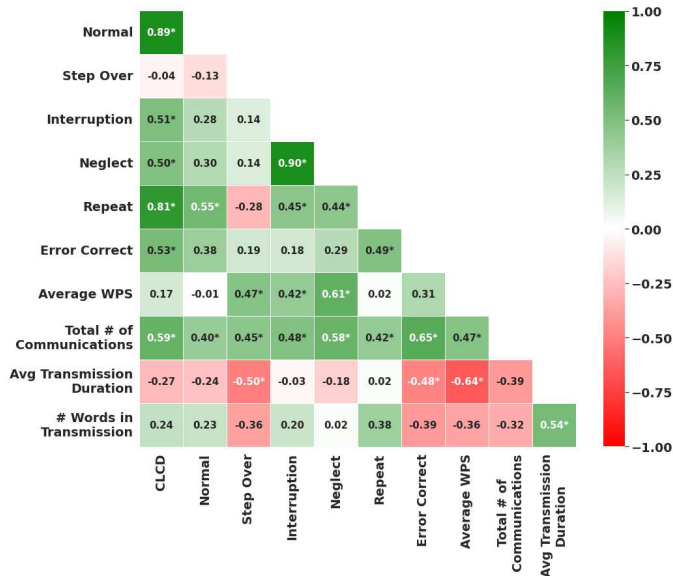
The model was reevaluated, removing neglect and repetition as predictors for LOS; this was done to explore the relationship between communication error and LOS. The hierarchical multiple regression model ran with step-over, interruption, error correction, repeat, neglect, and normal CLCD to predict LOS beyond density, leading to a statistically

significant model $R^2=.653$, $F(5, 12)= 4.57$, $p<.05$; adj. $R^2=.508$. The addition of step-over, interruption, error correction, repeat, neglect, and normal CLCD to predict LOS lead to a statistically significant increase of $R^2=.175$, $F(4, 12)= 1.514$, $p<.05$. Analysis of the coefficients found no statistically significant relationship for the input variables. Complete details of the coefficients can be found in Table 4.

Pearson Correlation

A Pearson correlation was conducted between the variables represented below in Figure 1. The purpose of this analysis was to evaluate the relationship these variables had with one another and to determine the effectiveness of the approach to measuring for CLCD's was effective in capturing problematic communication errors.

Figure 1
Correlation Matrix of Identified Variables (* $p<.05$)



DISCUSSION

The results of the multiple regression, which measured words per second, word count, density, and CLCD to predict LOS worked together to significantly predict LOS. However, none of the reported coefficients were observed to be statistically significant. Density in this model performed as expected along with CLCD. However, words per second and word count indicated a negative relationship between these variables and LOS. This could be explained by these variables' relationships to an increase in the total number of transmissions. An increase in words per second could potentially be associated with efficiency in communication, which translates to efficient airspace management. This may also infer that the communications are simple and non-demanding, indicating manageable communications.

Results of the modified hierarchical regression indicated a minor increase in the predictability of LOS. Most notably, interruptions had a relatively high value of $B= 25.193$ in a model to predict LOS. Compared to the other communication

errors detected by CLCD, interruptions had the highest effect on predicting LOS. Interruptions are detrimental to efficient communications and the exchange of information and have the potential for causing confusion.

The results of the Pearson correlation ran between the communication-related variables, categorical CLCD's, and CLCD. Most notably, there were moderate to strong positive statistically significant correlations between CLCD and each error detected by CLCD. This supports CLCD as a pattern-based approach to detecting the categorized communication errors. Another point of interest is that the data support a statistically significant negative correlation between average transmission duration and words per second, error correction, and step over. This relationship can be attributable to the requirement to restart transmissions or reduce the time spent on a communication abruptly, which has the potential for causing an operational error. Qualitative analysis of the communication transcripts indicated some circumstances where these communication errors caused confusion, inaccurate communication of information and/or the abandonment of the communication exchange altogether.

The importance of this work contributes to an ongoing conversation as to how to best implement communication-based measures into diagnostic and prognostic models used to predict and detect errors in real-time. Furthermore, this research supports the development of CLCD as a means to measure anomalous patterns of communication used for prediction and detection.

CONCLUSION

The data was inconclusive in supporting the measures of words per second, transmission duration, and words in transmission as variables to increase the predictability of LOS. However, more research is required to determine these variables' relationship to LOS and their potential for increasing accurate predictions of LOS. The results analyzed supported CLCD as a measure to detect various communication errors. Future research should evaluate the refinement of measuring for CLCD to improve its accuracy for predicting communication errors. In addition to detecting problematic errors, CLCD's capability to predict normal changes in the pattern should be analyzed with respect to other features such as workload.

LIMITATIONS AND FUTURE WORK

This analysis had a small data set of only 6 participants leading to low statistical power. Some measures were either averaged or counted for each trial. Future studies should utilize a more robust analysis that better detects changes in communication. In future analysis, the communication measures will be implemented into a time series analysis which includes various other measures such as workload, head position data, facial recognition data, and interbeat heart rate intervals, to name a few.

ACKNOWLEDGMENTS

The research reported in this paper was supported by funds from NASA University Leadership Initiative program (Contract No. NNX17AJ86A, Project Officer: Dr. Anupa Bajwa, Principal Investigator: Dr. Yongming Liu). The support is gratefully acknowledged.

REFERENCES

- Barshi, I. (1997). *Effects of linguistic properties and message length on misunderstandings in aviation communication* [Doctoral dissertation, University of Colorado at Boulder]. ProQuest Dissertations & Theses Global.
- Biondi, F., Turrill, J. M., Coleman, J. R., Cooper, J. M., & Strayer, D. L. (2015). Cognitive distraction impairs driver's anticipatory glances: An on-road study. In *Proceedings of the Eighth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design Distractive* (pp. 23–29).
- Billings, C. E., & Cheaney, E. S. (1981). *Information transfer problems in the aviation system*. National Aeronautics and Space Administration.
- Burke, C. S., Salas, E., Wilson-Donnelly, K., & Priest, H. (2004). How to turn a team of experts into an expert medical team: Guidance from the aviation and military communities. *BMJ Quality & Safety*, 13, i96–i104.
- Cardosi, K., Falzarano, P., & Han, S. (1998). *Pilot-controller communication errors: An analysis of Aviation Safety Reporting System (ASRS) reports*. Office of Aviation Research Federal Aviation Administration.
- Cooke, N. J., Salas, E., Kiekel, P. A., & Bell, B. (2004). Advances in measuring team cognition. In E. Salas & S. M. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 83–106).
- FAA. (2020). *Nextgen annual report: A report on the history, and future of national airspace system modernization*. <https://www.faa.gov/nextgen/media/NextGenAnnualReport-FiscalYear2020.pdf>
- FAA. (2021a, November 2). Air traffic by the numbers. Retrieved February 20, 2022, from https://www.faa.gov/air_traffic/by_the_numbers
- FAA. (2021b, December 2). ATC clearances and aircraft separation. Retrieved February 20, 2022, from https://www.faa.gov/air_traffic/publications/atpubs/aim_html/chap4_section_4.html
- FAA. (2022, January 5). This is NextGen. Retrieved February 20, 2022, from https://www.faa.gov/nextgen/this_is_nextgen
- Härgestam, M., Lindkvist, M., Brulin, C., Jacobsson, M., & Hultin, M. (2013). Communication in interdisciplinary teams: Exploring closed-loop communication during in situ trauma team training. *BMJ Open*, 3(10), e003525.
- Kantowitz, B. H. (2000). Attention and mental workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 3-456–3-459).
- Kantowitz, B. H., & Sorkin, R. D. (1983). *Human factors: Understanding people-system relationships*. Wiley.
- Lieber, C. S., Demir, M., Cooke, N., & Ligda, S. (2021). Deviations in closed loop communications between air traffic controllers and pilots as a predictor of loss of separation. In *Proceedings of AIAA Aviation 2021 Forum* (p. 2320).
- Ligda, S. V., Dao, A. Q. V., Vu, K. P., Strybel, T. Z., Battiste, V., & Johnson, W. W. (2010). Impact of conflict avoidance responsibility allocation on pilot workload in a distributed air traffic management system. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 55–59).
- Lingard, L., Regehr, G., Orser, B., Reznick, R., Baker, G. R., Doran, D., ... & Whyte, S. (2008). Evaluation of a preoperative checklist and team briefing among surgeons, nurses, and anesthesiologists to reduce failures in communication. *Archives of Surgery*, 143(1), 12–17.
- Molesworth, B. R. C., & Estival, D. (2015). Miscommunication in general aviation: The influence of external factors on communication errors. *Safety Science*, 73, 73–79.
- O'hare, D., Wiggins, M., Batt, R., & Morrison, D. (1994). Cognitive failure analysis for aircraft accident investigation. *Ergonomics*, 37(11), 1855–1869.
- Parush, A., Kramer, C., Foster-Hunt, T., Momtahan, K., Hunter, A., & Sohmer, B. (2011). Communication and team situation awareness in the OR: Implications for augmentative information display. *Journal of Biomedical Informatics*, 44(3), 477–485.
- Tijerina, L., & Goodman, M. J. (1996). Use of workload assessment measures and methods to assess safety-relevant impacts of in-vehicle device use among heavy vehicle drivers. In *Proceedings: International Technical Conference on the Enhanced Safety of Vehicles* (pp. 1961–1972).