2007

Effects of Message Complexity and Message Length on the Production of Pilot Readback Errors

O. Veronika Prinzo Ph.D.

EFFECTS OF MESSAGE COMPLEXITY AND MESSAGE LENGTH ON THE PRODUCTION OF PILOT READBACK ERRORS

O. Veronika Prinzo, Ph.D.
Federal Aviation Administration, Civil Aerospace Medical Institute, AAM510
6500 S. MacArthur, Oklahoma City OK 73169

Field data and laboratory studies conducted in the 1990s reported that the rate of pilot readback errors and communication problems increased as controller transmissions became more complex. This resulted in the recommendation that controllers send shorter messages to reduce the memory load imposed on pilots by complex messages. More than 10 years have passed since a comprehensive analysis quantified the types and frequency of readback errors and communication problems that occur in the operational environment. Hence, a content analysis was performed on 52 hours of pilot and controller messages that were transmitted from 5 of the busiest terminal radar approach control (TRACON) facilities in the contiguous United States between October 2003 and February 2004. Of importance was the finding that the number of pilot readback errors increased as the complexity and number of ATs in ATC messages increased — especially when pilots were performing approach tasks as compared with departure tasks. To limit the occurrence of communication problems and misunderstandings, controllers should be encouraged to transmit shorter and less complex messages.

Introduction

Since the implementation of changes in airport and aircraft security, the number of scheduled flights and passenger volume are at pre-9/11 levels and air traffic control (ATC) communications during peak traffic periods are at saturation points. During these times, pilots often compete for access to the same radio frequency to establish contact, receive clearances, make requests, etc. These communication bottlenecks can add to airport congestion, delays, and may increase the potential for communication problems.

Sometimes controllers adopt the strategy of sending longer, more complex messages to reduce the number of times they are on frequency, while including all the information required by FAA policy/regulations. As well-intended as the strategy is, the upper limit of verbal working memory (VWM) limits the amount of information that is successfully processed, stored, recognized and recalled (Miller, 1956).

With the onset of an ATC message, the sounds at the beginning of the message stream enter into a pilot's limited-capacity VWM, where they are processed and temporally stored as phonological representations. That is, acoustically relevant sounds are extracted and encoded into consonant-vowel-consonant clusters that form syllables that are assembled to create words, phrases, clauses, and other constituents. Successfully recoding sensory information into progressively larger chunks requires automatic recoding; otherwise, newly transmitted inputs are sacrificed while attempting to retain the name of the last group (e.g., interference effects). If not actively rehearsed they begin to decay in about 2 s (i.e., forgetting occurs, Baddeley, Thomson, & Buchanan, 1975) or are overwritten by incoming information. In fact, Baddeley et al. proposed a linear relationship between the number of words correctly recalled and speech rate. Using mathematical modeling, Schweicker and Boruff (1986) found that 95% of the variance in memory span[1] performance for words, digits, and colors was related to the number of items spoken in 2 s.

These findings, classic to cognitive psychology and psycholinguistics, have been applied to aviation. In particular, field and simulation findings (see Prinzo & Britton, 1993 for a review of the literature; Cardosi Brett, & Han, 1996; and Morrow & Prinzo, 1999) led to the recommendation that controllers should transmit more messages that were less complex, rather than fewer but more complex messages. The rationale was that less complex messages (fewer topics and less information) should not tax pilots' memory to the same extent as longer, more complex ones. Their recommendation, if made policy and implemented, should lead to fewer communication problems.

FAA Order 7110.65, The Handbook of Air Traffic Control (FAA, 2004) prescribes that controllers use a rigid set of words/phrases that narrows the definition and meaning of communication elements.[2] Some of these words and phrases serve as anchors that make the communication element

1 Memory span refers to the number of items (usually words or digits) that a person can hold in working memory. The average span for normal adults is 7.

2 As noted in Prinzo (1996), communication elements are the fundamental unit of meaningful verbal language. Within aviation communications, a communication element is identified according to its purpose, operation, or action and is restricted with regard to its AT (Prinzo, Britton, & Hendrix 1995).

more precise in its interpretation. For example, the significance of "3-5-0" is ambiguous until an anchor word appears with it — "3-5-0" can easily be interpreted as a heading, altitude, or speed. Thus, degrees are associated with heading, descend/climb/maintain with altitude, and knots with speed. Anchors assist in the interpretation of communication elements. As is often the case, ATC messages contain multiple communication elements, and message complexity is the sum of the complexity values (CVs) assigned to each one.

Wasow (1997) argues that an utterance's complexity can be derived from its grammatical weight — the amount of information expressed in its constituents as measured by the number of words, syntactic nodes, or phrasal nodes in the constituent. In the Prinzo, Hendrix and Hendrix (2006) scoring scheme, the added complexity imposed by communication elements that contain more information is reflected by assigning them larger values. For example, altitude instructions such as "three thousand five hundred," "one-zero thousand" and "four thousand" are likely to impose quantitatively different loads on VWM (e.g., articulatory loop Baddeley, 2000).

It has been 10 yrs since a comprehensive analysis was conducted to quantify the types and frequencies of readback errors that occur in the operational environment. The results presented here provide a current description and summary of readback errors that occurred during normal (TRACON) operations.

Method

Materials

Audiotapes. Digital audio tapes (DAT) of facility-recorded communications were provided by the five busiest TRACON facilities in the contiguous United States. In this report 28 hr 13 min 23 s of approach and 23 hr 56 min 32 s of departure communications were analyzed.

A Guide to the Computation of Level of Complexity. Presented in Table 1 are excerpts taken from the Instruction Complexity Guide. The guide was developed to increase the reliability and consistency of tabulating complexity for typical ATC phraseology usage. Anchors appear capitalized, are fixed in their meaning, and designate the to-be-performed action. Italicized words (in parenthesis) are qualifiers that vary with geographical location and aircraft position.

To determine complexity value, anchors, qualifiers, and excessive verbiage are assigned a value indicative of new information or importance towards understanding an instruction, traffic advisory, and altimeter setting advisory. In most cases, each anchor is counted as one element of complexity. There are several exceptions, however. Some communication elements contain multiple anchors, as is the case "turn left/right heading (degrees)." The anchor "Turn left/right" provides the direction of the turn, while "heading" indicates the aircraft's bearing.

Table 1. Excerpt from the for Instruction/Clearance Complexity Guide

AVIATION TOPIC	**COMPLEXITY VALUE**	**PHRASEOLOGY**
		3=*(altitude) two digits* +THOUSAND 2=*(altitude) one digit* + THOUSAND 3=*(altitude) two digits* + HUNDRED 2=*(altitude) one digit* + HUNDRED *2=(altitude) two digits* *1=(altitude) one digit*
	6	DESCEND/CLIMB & MAINTAIN *(altitude)* THOUSAND *(altitude)* HUNDRED *three* *five*
Altitude	5	DESCEND/CLIMB & MAINTAIN *(altitude)* THOUSAND *one zero*
	4	DESCEND/CLIMB & MAINTAIN *(altitude)* THOUSAND *four*
	*4-8	CONTINUE CLIMB/DESCENT TO *(altitude)*
	*4-8	AMEND YOUR ALTITUDE DESCEND/CLIMB AND MAINTAIN *(altitude)*
	*3-7	AMEND YOUR ALTITUDE MAINTAIN *(altitude)*
	*3-8	DESCEND/CLIMB TO *(altitude)*
	*2-6	MAINTAIN *(altitude)*
	*1-2	(*altitude,* omitted "THOUSAND" "HUNDRED")

Also, qualifiers such as the numbers that comprise an altitude must be evaluated according to the phraseology used by the speaker. That is, the number "three thousand five hundred" was assigned a value of 4 (a value of one for each number and a value of one for each anchor) since it would be more demanding than either one-zero thousand (value = 3) or four thousand (value = 2). Finally, one element of complexity is added for communication elements that contain excessive verbiage. Excessive verbiage is determined by comparing the utterance of the speaker against the phraseology designated in *FAA Order 7110.65*. If a pilot attempted a verbatim readback of a controller's transmission, then the coding procedures were applied that were used to evaluate the controllers' transmissions.

A Guide to the Classification of Pilot Readback Errors. As used here, a readback error is unsuccessful attempt by a pilot to read back correctly the information contained in the communication elements in an ATC message. The guide provided examples for each AT readback error type. As seen in Table 2, the column to the left displays the types of readback errors according to a particular type of aviation topic (AT). For example, ATC might transmit the following message to AAL10, "American Ten turn left heading two one zero." If the pilot reads back either "three one zero" or "six zero," it would be coded as a substitution error since the numbers in the original heading instruction included neither a three nor a six. If the pilot readback "American Ten turn right heading two one zero" it would be coded as an incorrect direction of turn.

Table 2. Readback Error Guide

Classification of Readback Errors
Readback Errors Type (Heading) 1 = Substitution of message numbers 2 = Transposition of message numbers 3 = One type of information read back as another type 4 = Incorrect direction of turn 5 = Omission of one or more numbers 6 = Not assigned 7 = Omission of anchor word(s) 8 = Substitution of anchor word(s)

Procedure

One set of audiocassette tapes was dubbed from each DAT and provided to the transcribers who used them to generate the verbatim transcripts. Each message was preceded by its onset and offset time. Once transcribe the ATC subject matter expert parsed each transmission into communication elements, labeled then by speech act category and AT using the procedures developed by Prinzo et al. (1995).

Then the complexity guide was used to assign a CV according to the phraseology used by the controller for that communication element. Also, each ATC transmission was paired with the pilot's response to that message and the pilot readback guide was used to evaluate the accuracy of the readback. The number of errors present was recorded (e.g., a zero indicated no error while a value of 3 indicated 3 errors).

Results

The results presented here examined the prevalence of pilot readback errors as a function of ATC message complexity and message length (as determined by counting the number of ATs in the transmission) — excluding Address/Addressee and Courtesies. Only controllers' messages that contained instruction (e.g., heading, heading modification, altitude, altitude restriction, speed, approach, departure, radio frequency, route, position, or transponder ATs) or advisory (traffic, altimeter portion of a weather advisory) speech acts were selected for the computation of message complexity. Of the 14,673 controller transmissions 11,159 met the selection criteria and were paired with its readback.

There were 723 individual readback errors present in 688 pilot transmissions — approximately 6% of the pilots' readbacks contained a readback error. Pearson correlations revealed that readback errors increased significantly as the complexity, $r(11159)=.196$ and message length (i.e., number of ATs), $r(11159)=.180$ in a controller's message increased, $p<.05$.

Message Complexity. Each ATC message was classified as either low ($\leq$ 09) or high ($\geq$10) complexity. Each pilot transmission had a readback value, and the average of those values was computed for each aircraft. An ATC Sector (Approach, Departure) by Message Complexity (Low, High) Analysis of Variance (ANOVA) was conducted on pilot readback performance. The results, evaluated using a criterion level set to $p \leq .05$, revealed that pilots produced more errors while in an approach (Mean = .126 SD = .304) compared with a departure (Mean = .038 SD = .153) sector, [$F(1,3700)$ = 129.00]. Also, more complex ATC messages had a higher incidence of being read back incorrectly (Mean = .172 SD = .375) than messages that were less complex (Mean = .038 SD = .117), [F1,3700) =

154.39]. However, these statistically significant main effects must be qualified by the presence of a statistically significant ATC Sector by Message Complexity interaction, [F(1,3700) = 97.18] that is presented in Figure 1.

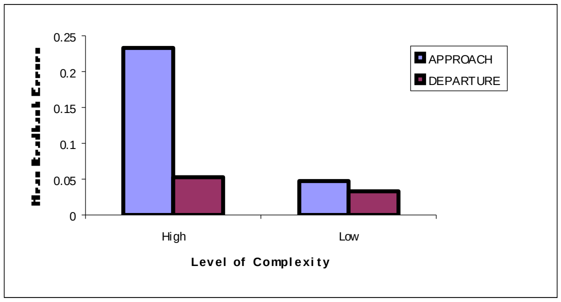

Figure 1. Mean Pilot Readback Errors Presented by ATC Sector and Message Complexity

The Tukey Honestly Significant Difference (HSD) statistic revealed that pilots experienced more difficulty reading back approach control high-complexity messages than reading back departure control high-complexity messages or low-complexity messages from either approach or departure control.

Message Length. Very short messages containing only one AT occurred for 54.2% of the transmissions, and they resulted in 3.84% readback errors (232/6049). Messages with 4 ATs appeared in 5.2% of the transmissions, producing 25.69% readback errors. Once again, pilot mean readback performance scores were computed for each aircraft call sign. The results of the ATC Sector (Approach, Departure) by Message Length (1AT, 2AT, 3AT, 4AT) ANOVA revealed that more readback errors occurred when pilots were in the approach (Mean = .113 SD = .307), as compared with the departure (Mean = .0343 SD = .157) sectors, [F(1,5599) = 78.48]. As expected, the number of readback errors varied with the number of ATs, [F(3,5599) = 21.62]. Tukey HSD comparisons revealed that the fewest readback errors occurred when ATC messages contained one AT (Mean = .036 SD = .139). There was no reliable difference between messages with 2 or 3 ATs (2AT = .062 SD = .214; 3AT = .082 SD = .258). However, messages with 4 ATs contained the most readback errors (Mean = .30 SD = .513). These main effects are qualified by a statistically significant ATC sector by message length interaction.

Figure 2 shows that as approach control messages increased from one AT to between 2 and 3 topics and 4 ATs, the mean number of pilot readback errors increased accordingly. The effect of message length is apparent only for approach control. There was no discernible difference between readback performance for approach and departure sectors for one AT.

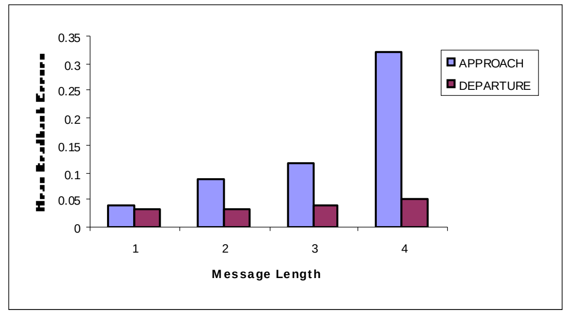

Figure 2. Mean Pilot Readback Errors Presented by ATC Sector and Message Length

Readback Errors and AT. Approximately 33% of the 723 identified readback errors involved speed instructions. Like the Cardosi et al. findings, there were proportionally more heading (22.68%) than radio frequency (17.98%) errors and proportionally fewer readback errors that involved altitude (19.50%) instructions. Route/position, approach/departure, altimeter, and transponder instructions captured the remaining 6.77% readback errors.

Although interesting in demonstrating the overall composition of readback errors, they fail to take into account the frequency of delivery of those instructions by controllers. There may be more opportunities to incorrectly read back a speed instruction simply because controllers issue them more often. Therefore, another analysis was performed that compared the number of readback errors of a particular AT (e.g., speed) to the total number readbacks of that AT. When the number of readback errors is examined in conjunction with the number of actual pilot readbacks produced, then reading back the content of an altitude restriction (18.57%) seems to posit greater difficultly than reading back the elements comprising a heading (3.93 %) instruction, as well as any of the other ATs. In fact, there were 7.68 times more attempts at reading back headings than altitude restrictions.

Readback errors fell within three major classifications — omission (63.76%), substitution (33.61%), and transposition (2.63%). The distribution of error classes differed across AT. For instance, of the 18.95% omission of anchor word(s), 12.45% involved heading (e.g., “eight zero”); almost half (11.20% of the 24.62%) of omission of number element(s) concerned speed (e.g., “eighty on the speed,” “eighty knots”); and over 67% of the omission of point/fix related to speed (e.g., in response to "… maintain speed one eight zero to depot," the pilot readback "I'll keep one eighty speed").

Substitution of anchor word(s) and substitution of number element(s) represented nearly 75% of the 7 types of substitution errors. Substitution of anchor word(s) was more likely to involve altitude restrictions and speed assignments than headings or approach clearances. Similarly, substitution of number element(s) was more likely to involve radio frequency, followed by heading and speed instructions. The combination of altitude instructions with altitude restrictions accounted for about 18% of the readback errors involving number substitution(s).

Transposition readback errors involved reordering the number element(s) or point/fix. About 95% of the transposition errors involved reversing the order of one point/fix with another.

Coincident Factors to Miscommunications. In this final analysis, transmissions that contained one or more faulty readbacks were examined for the presence of factors that might be correlated with, or have contributed to, its occurrence. Coincident factors included clipped/abbreviated transmissions, nonstandard phraseology, pilot expectation, language barriers, and transmission overlap (stepped-on, blocked transmissions).

There were 207 pilot readbacks that began with an abbreviated speech act (e.g., "Thirty heading," "Eighty speed" "One zero four thousand") that may have resulted from poor microphone technique, poor phraseology, or differences in aircraft radio transceivers. Also, once the pilot began a readback, nonstandard phraseology was another factor associated with 91 transmissions with readback errors. There was a tendency among some pilots to truncate or otherwise abbreviate the numerical values in speed, heading, or altitude assignments. Some pilots used the "point" designation associated with radio frequencies when reading back altitudes and speeds or substituted "decimal" for the word "point" when reading back a radio frequency. Also, several pilots flying for foreign air carriers displayed some problems in English proficiency and language production. Finally, pilot expectation (n = 16) played a coincidental role in pilot readback errors and was associated with the pilot of one aircraft reading back the contents of a message meant for the pilot of a different aircraft.

Discussion

The results presented in this report provide a description and summary of the controller-pilot communication process that occurred during day-to-day operations in the TRACON environment. It provides additional evidence that readback errors increased with increases in message complexity and message length. Importantly, pilots experienced the most difficulty reading back ATC messages with more than one AT and ATC messages with a CV of 10 or greater when flying the approach.

The type of readback error produced seemed to be related to the type of information read back. For example, pilots were more likely to omit an anchor word or phrase when reading back a heading and either exclude a number or leave out the point/fix in a speed instruction. They were more likely to substitute an anchor word(s) when reading back either an altitude restriction or speed assignment than a heading or approach clearance. When instructed to switch frequencies, change to a new heading or alter the aircraft's speed, pilots were likely to substitute numbers. Finally, a majority of the transposition errors involved reversing the order of one point/fix with that of another within the same message.

Whether unintentional or purposeful, many pilots also made number/letter substitutions. Some pilots truncated or otherwise abbreviated the numerical values in speed ("Twenty five knots"), heading ("one four" for a heading of one four zero), or altitude assignments ("Down to five hundred"). It is possible that some of the abbreviations were due to delivery technique or equipment use, while others may reflect a heightened workload.

Other forms of nonstandard phraseology were also associated with readback errors. It may be that some of the phraseology used (or heard) by pilots during international flights is making its way into the NAS. Some pilots used the "point" designation associated with radio frequencies when reading back altitudes (e.g., "Three point five" instead of "Three thousand five hundred") and speeds (e.g., "Two point seven on the speed" for "two hundred and seventy knots") or substituted "decimal" for the word "point" when reading back a radio frequency. Also, several pilots flying for foreign air carriers displayed some problems in English proficiency and language production — for example, reading back a speed instruction as "two zero hundred" instead of "two hundred knots," or responding to "maintain visual from traffic" as "Maintain visual approach."

In summary, a comparison between the voice communications analyzed by Cardosi et al. with those presented in this report revealed differences in message complexity and readback/hearback error rates. When the data were compared, the findings show more than 50% of controllers' messages are fairly short but information-rich. Both the Cardosi et al. 1996 report and this report show that aircraft headings and radio frequency changes still are the most frequently occurring readback errors.

Similarly, the differences in the degree of faulty pilot readbacks errors may be partially due because of the approach used to evaluate the message content. *FAA Order 7110.65* was applied by Prinzo et al. (2006), whereas Cardosi et al. do not describe their evaluation criteria. A liberal criterion reveals only a minimal increase in pilot readback errors (up 0.3%) between the two reports. Both reports show that aircraft headings and radio frequency changes still are the most frequently occurring readback errors.

Communicating for safety is the primary objective of the phraseology developed for and provided in *FAA Order 7110.65*. With increased international travel and the gradual migration of other phraseologies into the NAS, pilots and controllers must remain vigilant in the accurate production and recitation of ATC clearances, instructions, advisories, reports, requests, and other communications.

References

Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417-23.

Baddeley, A.D., Thomson, N., and Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 575-589.

Cardosi, K. (1993). *An analysis of en route controller-pilot voice communications*. DOT Report no. DOT/FAA/RD-93/11. Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.

Cardosi, K., Brett, B., and Han, S. (1996). *An analysis of TRACON (Terminal Radar Approach Control) controller-pilot voice communications*. DOT Report no. DOT/FAA/AR-96/66. Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.

Federal Aviation Administration. (2004). FAA Order 7110.65P Air Traffic Control. www.faa.gov/ATPUBS/ATC/. Accessed 4 November, 2005.

Miller, G.A. (1956). The magical number seven plus or minus two. Some limits on our capacity for processing information. *Psychological Review, 63,* 81-97.

Morrow, D. and Rodvold, M. (1993). *The influence of ATC message length and timing on pilot communication.* NASA Contract Report 177621. Moffett Field, CA: NASA Ames Research Center.

Morrow, D., Rodvold, M., McGann, A., and Mackintosh, M. (1994). *Collaborative strategies in air-ground communication.* Proceedings of Aerotech '94 Conference, pp.119-24, Technical Paper #942138, Los Angeles, CA: Society of Automotive Engineers.

Morrow, D. and Prinzo, O.V. (1999). *Improving pilot/ATC voice communication in general aviation*. Report no. DOT/FAA/AM-99/21. Washington, DC: Federal Aviation Administration.

Prinzo, O.V. (1996). *An analysis of approach control/pilot voice communications.* Report no. DOT/FAA/AM-96/26. Washington, DC: Federal Aviation Administration.

Prinzo, O.V. and Britton, T.W. (1993). *ATC/pilot voice communications – A survey of the literature*. Report no. DOT/FAA/AM-03/20. Washington, DC: Federal Aviation Administration.

Prinzo, O.V., Britton, T.W., and Hendrix, A.M. (1995). *Development of a coding form for approach control/pilot voice communications.* Report no. DOT/FAA/AM-95/15. Washington, DC: Federal Aviation Administration.

Prinzo, O.V., Hendrix, A.M., and Hendrix, R. (2006). *The outcome of ATC message complexity on pilot readback performance*. Report no. DOT/FAA/AM-06/25. Washington, DC: Federal Aviation Administration.

Schweicker, R. and Boruff, B. (1986). Short-term memory capacity: Magic number or magic spell? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 419-25.

Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change, 9,* 81-105.