5-7-2019

Aviation English Listening and Repeating Task for Native English Speaker and Non-Native English Speaker Pilots

Julia Trippe

AVIATION ENGLISH LISTENING AND REPEATING TASK FOR NATIVE ENGLISH SPEAKER AND NON-NATIVE ENGLISH SPEAKER PILOTS

Julia Trippe
Pacific Science and Engineering
San Diego, CA

Aviation English, based on a coded jargon from World War II, is a mandatory form of communication for pilots and controllers in international airspace. The International Civil Aviation Organization also requires proficiency in Conversational English, for use in non-standard communication. However, our past research indicates that Aviation English and Conversational English are distinct varieties of English, suggesting that assumptions about native English speaker proficiency and additive learning for non-native English speakers may be false. To establish how different these language varieties are, we present a study of Aviation English intelligibility for non-native and native English speaking pilots. Results suggest that non-native English speaking pilots exhibit high proficiency in Aviation English without parallel proficiency in Conversational English. Non-native English speaking Aviation English users suffer the unfair burden of having to learn and maintain proficiency in two language varieties. The impact on learning, training and testing of Aviation English is discussed.

A necessary step towards understanding the relationship between Aviation English (AE) and conversational English (CE) is to determine the extent to which these varieties of English are mutually intelligible. After determining that native English speakers (NESs) not versed in AE are scarcely able to understand AE (Trippe & Pederson, ISAP 2017), it remains to be determined if AE-using non-native English speakers (NNESs) can understand CE. The current paper addresses this proposition by examining NNES AE users. If NNES pilots are more proficient in AE than in CE, it may be reasonable to assume that CE proficiency is not necessary for AE proficiency and that initial aviation language training should focus on AE and not on CE, as is the current practice. If this is the case, then dedicated AE training would be a more efficient use of time, energy, and financial resources for pilots and air carriers: enabling students to absorb language lessons in a less stressful environment and reserving valuable aircraft time for flight training rather than language instruction. Additionally, native English speaking AE users cannot be presumed to have proficiency in this variety of English without undergoing testing.

Methods

Participants

In order to establish NNES AE proficiency as compared to their CE proficiency, groups of NNES pilots and NES pilots were given identical oral performance tasks and their results were compared. At a minimum, all participants were Federal Aviation Administration (FAA)-rated private pilots. The NNES pilot group (CP) was made up of 29 (1 female) Chinese flight students, ranging in age from 22 to 26 (M = 23.38, SD = 1.08). The NES pilot (EP) group was made up of 23 (4 female) North American flight students and instructors, ranging in age from 19 to 55 years (M = 28.30, SD = 7.77). Table 1 summarizes population descriptives: age, total flight time (TT), and instrument flight rules flight time (IFR).

Table 1.
Population Age, Total Flight Time and Instrument Flight Time by Group

	CP (n = 29)			EP (n = 23)		
	Range	*Mean*	*SD*	*Range*	*Mean*	*SD*
Age	22-26	23.38	1.08	19-55	28.30	7.77
TT	110-200	156.83	25.84	67-7000	1078.30	1767.40
IFR	10-66	37.21	14.79	4-2500	301.65	620.96

Note: CP = Chinese (NNES) Pilots; EP = native English-speaking (NES) Pilots

Procedure

Participants underwent three verbal repetition tasks, starting with a 15-minute verbal working memory (WM) task to establish baseline differences that could affect repetition of verbal elements. This was followed by a five-minute intelligibility task of Standard American English (SAE) to establish CE competency. The final task was a 15-minute AE intelligibility task to determine how well participants perceived AE transmissions. Tasks were administered by computer using Psychopy (Pierce, 2007) software and were self-paced. All participants received the tasks in the same order with no feedback for response accuracy.

Working memory Task. Participants underwent a verbal WM task to determine any correlation between WM and AE abilities. WM was evaluated using the Word Auditory Recognition and Recall Measure (WARRM) (Smith, Pichora-Fuller, & Alexander, 2016) which required participants to repeat Standard English monosyllabic audio stimuli. Possible raw scores ranged from 1.0 to 6.0, depending on number of words consistently remembered after performance of unrelated cognitive tasks. Raw score was multiplied by a factor of 16.67 to make the highest possible score 100.

CE intelligibility task. Participants were asked to repeat ten CE sentences verbatim. The sentences are approximately fifth grade reading level, ranging in length from seven to ten words. Score for the CE task was the percentage of words correctly reproduced of the 83 possible words in the combined ten CE sentences.

AE intelligibility task. Participants were asked to repeat 84 ATCO utterances verbatim. Half of the selected ATCO transmissions consisted of one aviation topic and half contained two topics. Transmissions ranged in length from two to 19 words. Stimuli was organized in eight pseudo-randomized sets in which every dozen utterances included an equal number of one- and two-topic tokens, so that analysis could explore improvement over seven sets of twelve utterances. Score for the AE task was the percentage of words correctly reproduced for each response, each set of twelve responses, and all 84 responses combined for each participant.

Results

Verbal Repetition Task Scores by Group

Native English speaking EPs performed significantly better on all of the tasks than CPs (see Figure 1). As would be expected, EPs scored higher on the CE task and on the WM task, since they are both English repetition tasks. Non-native English speaker CP CE task scores averaged 43.23 (SD = 10.45) and EP CE scores averaged 95.55 (SD = 3.55). CP WM task scores averaged 50.56 (SD = 11.06) and EP WM scores averaged 77.31 (SD = 13.60) (see Figure 1). These two task scores confirm expected differences in English proficiency.

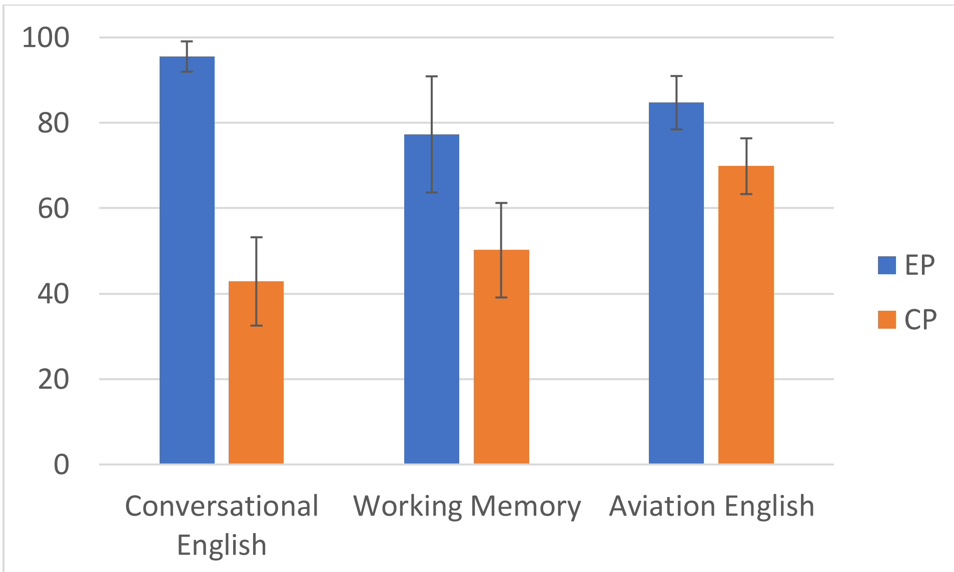

Figure 1. Average Conversational English, Working Memory and Aviation English task scores by Group: Chinese pilots (CP) and native English speaker pilots (EP). Error bars reflect Standard Deviations.

Both groups' results were examined for a learning effect. Using the Bonferroni correction in a pairwise comparison of the seven successive groups of twelve transmissions, it was determined that neither the CP nor EP group showed a learning effect.

Factors Predicting Aviation English Performance

A linear mixed effects regression was performed using nlme package (Pinheiro, Bates, DebRoy, & Sarkar, 2014) in R (R Core Team, 2014) to create the best model fit for predicting AE scores for all responses in the data. The regression included random effects of specific transmission, order presented in the task, and individual participants (subject). The full regression model included fixed effects of pilot group, CE task score, WM task score, age, sex, number of words per transmission, number of topics per transmission, total flight time (TT), Instrument Flight Rules time (IFR), and interaction with group for each of the fixed effects. The final model includes only significant factors that were not correlated (see Table 2).

Examination of correlation of fixed effects using R indicated that TT and IFR were correlated ($r^2 = 0.67$), as were CE and Group ($r^2 = 0.79$), and number of words and number of topics ($r^2 = 0.55$). Accordingly, we examined each of these factors for its contribution to the model. We selected IFR rather than TT, as the measure of AE exposure, because pilots must be in constant contact with ATCOs in the IFR environment. Number of IFR hours was log transformed for inclusion in the regression. We retained both number of words and number of topics as factors, so that possible group differences in language parsing could be discovered.

Table 2.
Linear Mixed Effects Model Summary of Native-English Speaking Pilots (EP) and Chinese Pilots (CP) AE Performance Scores

Predictor	*Coefficient*	*Std. Error*	*t value*	$\chi^2(1)$	*p-value*
Intercept	97.92	4.98	19.66		
EP Group	-3.00	2.42	-1.24	19.80	< .001
WM Score	0.14	0.06	2.36	5.56	0.018
Number of Words	-4.12	0.373	-11.06	91.14	< .001
ln(IFR hours)	1.52	0.58	2.62	6.88	0.009
EP*Number of Words	1.40	0.15	9.57	91.57	< .001

Note: Random effects of Subject, Transmission and Order were included in the model.

Regression results[1] indicate that a combination of significant factors affect AE scores for the entire population of pilots in the study. The primary effect appeared to be number of words in a transmission, which affected group outcomes differently. For every word in a transmission, CPs' AE average scores decreased by 4.12 percentage points, whereas EPs' fell by only 2.72. Model fit was further facilitated by an across-the-board 3-percent point decrement for EPs. These factors combine to indicate, for the shortest (2-word) transmissions, CP and EP scores were almost the same, but for the longest (19-word) transmissions, EPs scored much higher than CPs.

In addition to number of words affecting AE task scores, pilot exposure (as measured by the natural log of their IFR time) was a significant factor. Each unit of ln(IFR) corresponded to a 1.52 percentage point increase in AE scores. This value is greater for the EP group generally, since their mean IFR time is higher (see Table 1). Working Memory task score was also a significant predictor of AE score, probably because this score reflects Standard English proficiency. Every percentage point correct on the WM task, corresponded to a .14 percentage point increase in participants' AE scores. Once again, this effect had a greater benefit for EPs, since their WM scores were higher, on average, than the CP groups' (Figure 1).

CE Task Effect on Non-Native English Speaking Pilots' AE Scores

To determine the possible effect of CE proficiency on non-native English speaker AE scores, we did a separate regression on AE scores for the CP group. The full model for this regression included the above factors in addition to CE task score. The resultant mixed-effects regression model indicates that AE scores for Chinese pilots are significantly predicted by number of words in the transmission and CE score. Similar to the previous regression on both pilot populations, the number of words in a transmission predicted a 4.11 percentage point per word decrease in AE scores for Chinese pilots. Additionally, every percentage point in Chinese

[1] Model fit determination using piecewiseSEM package in R (Lefcheck, 2015), gave a marginal (fixed effects) R^2 value of 0.33 and conditional (including random effects) R^2 value of 0.58.

pilots' CE task score predicted a 0.37-point increase in their AE scores. Chinese pilot AE scores were not predicted by number of topics or WM score.

Discussion and Conclusion

Results of this study indicate that NNES pilots, as represented by a group of Chinese students at a US flight school, exhibit higher proficiency in standard phraseology of Aviation English (AE) than in conversational English (CE). These results suggest that CE ability does not necessarily imply AE ability. Although study results indicate that CE proficiency is correlated with AE proficiency in this population, it is not a consistent predictor. Counterexamples of this relationship are prevalent in the data. Fully 34.5% of CP participants' AE and CE task scores were negatively correlated. Although it requires further study, one possible implication of these findings is that language training specifically focused on AE is likely a more efficient way of increasing AE proficiency than CE training.

Both pilot groups exhibited familiarity with AE as indicated by the fact that there was no adjustment period / learning effect over the brief AE task duration. However, the regression model including all the pilot participants indicates flight experience predicts AE proficiency (see Table XX). This effect appears to be driven by the EP group, since it was not a significant factor in the within-group analysis of Chinese pilots. Results from our previous study on the same EP group suggest that their AE learning curve is initially steep and shallows out with flight experience, reaching asymptote at about 100 hours of IFR time. Although the small number of higher time pilots in the EP population restrains us from generalizing these findings, one conclusion that could be drawn is that, although a brief exposure to AE (during testing) may not be sufficient to increase proficiency, longer exposure does. It is impossible to test this theory on the CP group data, since the Chinese pilots in the current study were all in the early phases of their flight training and had similar, low numbers of IFR hours ($M = 37.21$, $SD = 14.79$), as compared to the native-English speaking pilots in the EP group ($M = 301.65$, $SD = 620.96$).

Regression results indicate that the primary factor in determining difficulty of repetition for both pilot groups was number of words in a transmission, especially for NNES pilots. These findings are consistent with non-native speech studies regarding the cognitive load of translation (Estival & Molesworth, 2016; Farris, 2007). Even comparing pilots with similar flight experience, we would expect NES pilots to have higher AE proficiency than their NNES counterparts, since CE and AE share vocabulary and phonotactics, requiring less translation for NESs.

This study seeks to improve international pilot language training by enhancing the industry's understanding of NNES acquisition of AE standard phraseology. Consistent with our previous study, it appears that CE proficiency does not imply AE proficiency. In the case of the current study, CPs lack of CE proficiency did not limit their AE proficiency. Rather, it appears that the determining factor in their AE abilities was exposure to actual ATCO speech during flight training. Since the rhythm and usage of AE are different from CE (Borowska, 2017; Trippe et al. 2018), language education for professional pilots should be in AE standard phraseology.

Just as for NNES pilots, a short period of AE ground training for NES pilots should enhance AE proficiency similar to the first hundred hours of AE exposure in IFR flight and would serve to prepare pilots for more fluent AE communication. Future research can determine the proper amount of time in listening and repeating actual ATCO transmissions to replicate this initial instrument flight experience. Since flight training is expensive and stressful for pilots, a language-training module for practicing pilot/ATCO communication before and during flight training would be highly beneficial.

AE standard phraseology training should be the basis for all AE communication. Conveyance of more complicated messages could be addressed by expanding AE standard phraseology to include non-routine situations. Emergency and other high-stress situations should not require CE fluency, especially since it is recognized that the cognitive load of speaking in a second language adds to the stress that may accompany such a situation.

References

Borowska, A. (2017c). Avialinguistics. *The Study of Language for Aviation Purposes, Frankfurt am Main: Peter Lang (in print)*.

Estival, D., Farris, C., & Molesworth, B. (2016). *Aviation English: A Lingua Franca for Pilots and Air Traffic Controllers*: Routledge.

Farris, C. (2007). *The Effects of Message Length, L2 Proficiency and Cognitive Workload on Performance Accuracy and Speech Production in a Simulated Pilot Navigation Task*: Concordia University (Canada).

Lefcheck, Jonathan S. (2015) piecewise SEM: Piecewise structural equation modeling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, *7*(5), 573-579. doi:10.1111/2041-210X.12512

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2014). R Core Team (2014). *nlme: linear and nonlinear mixed effects models*. R package version 3.1–117. Retrieved from https://cran.r-project.org/web/packages/nlme/index.html.

R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: http://www.R-project.org/.

Smith, S. L., Pichora-Fuller, M. K., & Alexander, G. (2016). Development of the Word Auditory Recognition and Recall Measure: A working memory test for use in rehabilitative audiology. *Ear and Hearing*, *37*(6), e360-e376.

Trippe, J., & Baese-Berk, M. (2019). A prosodic profile of American Aviation English. *English for Specific Purposes*.

Trippe, J., & Pederson, E. (2017). *Aviation English intelligibility*. Proceedings of the 19th International Symposium on Aviation Psychology, Dayton, Ohio.