AVIATION ENGLISH INTELLIGIBILITY

Julia Trippe
Eric Pederson
University of Oregon
Eugene, Oregon

Aviation English (AE) is the "primary dialect" of international aviation. Demonstrably, AE and Conversational English (CE) are distinct varieties of English. Past research shows that AE is spoken more rapidly, with less inflectional variation and different rhythm than CE. Differences are strong enough that AE and CE may not be mutually intelligible. However, flight students are not trained in AE production and perception prior to flight training. This study examines the intelligibility relationship between AE and CE by comparing native English speaking non-pilots and native English speaking pilots responding to actual air traffic controller transmissions. A difference between these groups was predicted, given their comparative AE familiarity. However, the difference in AE intelligibility proved to be stronger than expected. Additionally, results from licensed pilots indicate that AE learning continues with flight experience, suggesting there may not be adequate training prior to reliance on AE in flight.

Aviation English is the mandatory language for pilots and air traffic controllers (ATC) at international airports, if they do not share a first language. Proficiency in Aviation English (AE) and conversational English (CE) are required by the International Civil Aviation Organisation (ICAO), yet little is known about how AE and CE interact in language learning and usage. These requirements as well as AE training conventions are based on the assumption that CE proficiency aids in AE proficiency, although this may not be the case (see discussion of "plain English" in Background section below). Indeed, past research shows that AE is different from CE in ways that may affect intelligibility (Trippe & Baese-Berk, submitted). The current study examines AE intelligibility differences between native English speaking pilots and native English speaking non-pilots. If AE is not intelligible to CE speakers without aviation experience, CE proficiency cannot be sufficient to predict AE proficiency. The goal of this research is to further establish the intelligibility relationship between AE and CE and influence development of effective AE training to improve international flight safety.

Aviation English Description

AE is a variety of radiotelephony developed to convey critical information between pilots and Air Traffic Control (ATC). Although AE includes both standard phraseology and "plain English", in this study the term Aviation English (AE) is used to denote standard phraseology and "plain English" is referred to as such. Ambiguity in AE is avoided by fixing a single meaning to each word and phrase. Words whose pronunciation may cause confusion are assigned distinct pronunciations. For example, AE require that *five* and *nine* be pronounced *fife* and *niner*. Additionally, word and phrase inventories of AE are restricted. Articles, prepositions and possessives are not used except to resolve ambiguity. Any ambiguous word is given a single meaning or substituted with another word. AE standard phrases use lexical topic identifiers and specific number expressions to signify aviation topics. For example, *wind three fife zero at one two,* or *turn right heading three fife zero* both use single digits to express direction, but each phrase has a lexical identifier denoting the aviation topic addressed (i.e. *wind* v. *heading*). Lexical and grammatical differences as well as environmental factors (i.e. multiple speakers, no face-to-face contact, signal static and reduced frequency range), lead to differences in the sound profiles of AE and CE. AE is faster than CE, with fewer intonational cues and a different rhythmic signature than CE (Trippe & Baese-Berk, submitted). These differences could make AE unintelligible to CE speakers.

Aviation English Regulation

High loss-of-life accidents caused in part by communication problems (Cookson, 2011) compelled ICAO to require AE proficiency in international airspace as of 2011. However, this requirement has yet to be thoroughly operationalized. While ICAO has published general proficiency-rating guidelines, there is no agreed upon standard protocol by which to attain or prove proficiency. Dozens of tests have been developed internationally and several are in use, although ICAO recognizes only one (English Language Proficiency for Aeronautical Communication). The new requirements also pertain to CE proficiency (ICAO, 2004), although the vast majority of pilot-ATC communication is in AE, which was designed to convey all typical transactions. When AE is not sufficient to convey messages, ICAO regulations stipulate the use of "plain English". Generally this caveat applies to unusual or emergency situations. Although the implementation of AE recognizes the need to keep communications succinct and unambiguous, it is impossible to control for these needs in "plain English", because the parameters of "plain English" are not defined. Native English speakers often speak quickly and colloquially during times of duress. Although such interactions usually aid in clarification of complex situations between native English speakers, these communications may not be comprehensible to non-native English speaking interlocutors (Kim & Elder, 2009). Additionally, second language English users have more difficulty conversing in CE under conditions of stress or high cognitive load that typically trigger "plain English" use in native speakers (Farris, Trofimovich, Segalowitz & Gatbonton, 2008). Further, the requirement to use "plain English" is confounded by the fact that there exists no consistent guidance as to what is meant by *plain* English. The regulatory intent is clear: this English variety should be readily understandable to one's interlocutor. Unfortunately, it is impossible to ascertain what level of English proficiency, or indeed what model of Standard English, one's interlocutor has. In fact, language experts recommend "plain English" be avoided as much as possible in aviation communications (Day, 2004: Moder, 2012). AE fluency reduces repetitions, delays, and misunderstandings.

As the international flying community becomes more diverse, pilots will operate in airspace and on crews with individuals from different language backgrounds, increasing the potential for misunderstanding and miscommunication (Kim & Billington, 2016). In this environment, it is critical to utilize AE standard phraseology, to reduce the potential for confusion as much as possible. Rather than relying on "plain English", consideration should be given to expanding AE so that unusual situations may be addressed using this clear and constrained format and lexicon.

Aviation English Testing and Training

Testing. Newly developed AE testing protocols differ greatly. However, a common element of AE tests is a face-to-face interview with a language evaluation specialist wherein the pilot must discuss unusual situations that may arise while flying, to determine if they have a working knowledge of aviation terminology and can convey ideas in CE. Interviews are typically conducted by English-language teaching specialists who are not aviation professionals nor fluent in AE. This type of testing does not evaluate AE speech used in most pilot-ATC interactions. In fact, listening and responding to actual ATC transmissions may not be included in the pilot's proficiency test, although this is the vast majority of pilot communications (Alderson, 2009). Additionally, when ATC speech is used in testing, it is created for that purpose and is often slower, without static, accents and multiple speakers that occur in actual transmissions. Therefore, passing an AE proficiency test does not guarantee a pilot's ability to fulfill their job requirements. In their study of non-native English speaker ATC oral proficiency, Moder and Halleck (2009) found that there was no consistent relationship between AE and CE scores. Additionally, Kim & Elder (2009) asserted that CE-focused testing protocols unfairly benefit native English speakers, who are assumed to be fluent in AE, but often do not comply with AE standard phraseology.

Training. The standard for AE training has long been that radiotelephony is learned simultaneously with flight training. It is assumed that pilots will learn through immersion: monitoring and interacting with ATC. Anecdotally, it is common knowledge that student pilots are as anxious about talking on the radio as they are about flying the plane. However, the AE immersion strategy has been adopted as the model for non-native speakers training in native English speaking countries, which is where a great deal of international commercial flight training takes place. Although many flight-training programs for non-native English speakers include language training, AE courses are designed by English language teaching experts in consultation with aviation professionals, focusing on face-to-face communication in CE with emphasis on aviation terminology. AE instructors are generally not fluent in AE. This learning environment does not reflect pilots' experience or needs. In actual flight conditions, pilots must interpret messages through static and reduced frequency range, without seeing their interlocutor. If AE is as different from CE as prior research indicates (Trippe & Baese-Berk, submitted), training in CE with non-AE speakers will not enhance AE skills as much as dedicated AE training will.

Aviation English Intelligibility

To further understand the intelligibility relationship between AE and CE, it must be determined if native CE users can understand AE and vice versa. The current study addresses the first of these proposals, examining the differences between native English speakers with and without AE experience, perceiving actual ATC transmissions. If AE is intelligible to CE users, then teaching and testing CE for aviators is practical. If AE is not intelligible to CE users, teaching and testing of CE for aviators may be a misuse of time and energy.

Method

Participants

Two groups of native English speaking participants were recruited for the study. The non-pilot population was made up of 26 (17 female) University of Oregon undergraduates, mean age 20.69 (SD = 3.03). The pilot population was made up of 23 licensed pilots (4 female) from Lane Aviation Academy and Hillsboro Aero Academy in Oregon, mean age 28.30 (SD = 7.77). The pilot group consisted of licensed pilots ranging in age from 19 to 55 (*median* = 26) with flight hours from 67 to 7000 (*median* = 350), including 4 to 2500 hours under Instrument Flight Rules (*median* = 56 hours).

Procedure

Participants performed three verbal repetition tasks, starting with a 15-minute verbal working memory task, followed by a five-minute Standard American English intelligibility task to establish CE competency. A 15-minute AE intelligibility task concluded each trial. Tasks were self-paced and computer-administered using Psychopy (Peirce, 2007) software. Participants completed language background questionnaires reporting other language and/or professional radio experience. Working memory (WM) was evaluated using the Word Auditory Recognition and Recall Measure (WARRM) (Smith, et al., 2016) which required participants to repeat Standard English monosyllabic audio stimuli with intervening unrelated cognitive tasks. WM was scored on a scale from 2 to 6 points, reflecting the number of words the participant was able to remember consistently. This score was then multiplied by 16.67 to make the highest possible score 100, to be comparable with percentage scores for the other tasks.

The second task was a CE intelligibility task in which participants repeated ten sentences from the Harvard Sentence recordings (Open Speech Repository, 2016) which were approximately fifth grade reading level, phonetically balanced for Standard American English, and from seven to ten words long. Responses were tape-recorded for later analysis. Score for the CE task was the percentage of words correctly reproduced of the 83 possible words in the ten CE sentences combined.

The third verbal repetition task was an AE intelligibility task in which participants repeated 84 ATC transmissions selected from the Air Traffic Control Complete corpus (Godfrey, 1994), based on number of topics and terminology. Since past studies indicate that subjects show a sharp decrease in navigational performance for transmissions with more than three propositions (Farris & Barshi, 2013), selected transmissions were limited to two topics (i.e. *[traffic no factor] [turn right heading two zero zero]*). Half of the selected ATC transmissions had one aviation topic and half had two. Equal numbers of transmissions were chosen from 22 (3 female) apparently native American English ATC. Responses were tape-recorded for later analysis. Stimuli were pseudo-randomized so that every dozen transmissions included an equal number of one- and two-topic tokens. AE task transcription was done by two trained lab technicians and the first author. Inter-coder reliability tests resulted in 98% agreement. Words were correct if they were in order relative to other words in the transmission (see Table 1).

Table 1.
Sample Points Awarded for Participant Response

Original transmission	TURN	RIGHT	…	HEADING	…	TWO	FOUR	ZERO	(6 words)
Response	…	right	turn	heading	zero	two	…	zero	
Points	0	1	0	1	0	1	0	1	4
Percentage									66.67

Results

Statistical Analysis

Verbal repetition task scores by group. Pilot group average for the CE task (*M* = 95.55, *SD* = 3.55) did not differ significantly from non-pilots' CE scores (*M* = 97.00, *SD* = 3.11) (t (44.12) = 1.52, p = 0.14). Nor did pilots' WM task scores (*M* = 77.30, *SD* = 13.60) from non-pilots' WM scores (*M* = 71.82, *SD* = 16.56) (t (46.77) = -1.27, p = 0.21). However, Average pilots' AE task scores (*M* = 87.97, *SD* = 18.22) were significantly higher than non-pilots' (*M* = 57.27, *SD* = 26.18) (t (46.69) = -15.81, p < .001). The only apparent learning effect in the data was for non-pilots showed a learning effect between the first to the second set of AE transmissions (see Table 2).

Table 2.
Mean Aviation English Percentage Correct Over Testing Period by Group

	Testing Period						
Group	AE1	AE2	AE3	AE4	AE5	AE6	AE7
Non-Pilots	46.44^{a}	51.28^{ab}	53.86^{b}	55.71^{b}	55.36^{b}	54.83^{ab}	55.60^{b}
Pilots	82.42^{c}	83.46^{c}	84.13^{c}	84.00^{c}	85.86^{c}	86.13^{c}	87.99^{c}

Note: Values with different superscripts are significantly different p < .05

Factors predicting AE intelligibility. A linear mixed effects regression was performed using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2014) on all responses in the data. Fixed effects included CE score, WM score, group, radio experience, language experience, age, sex, number of words per transmission, number of topics per transmission and interactions with group for each fixed effect. Number of topics and number of words were collinear (R^2 = 0.54). However, given the significant interaction of number of topics by group, both number of words and number of topics were retained as factors in the model (see Table 3).

Table 3.
Linear Mixed Effects Analysis of AE Intelligibility Scores (Random effects: Subject, Transmission and Order)

Predictor	*Coefficient*	*Std. Error*	$\chi^2(1)$	*p-value*
Intercept	43.77	25.91		
CE Score	0.42	0.27	2.42	0.120
WM Score	0.11	0.06	3.21	0.073
Pilot Group	18.07	2.38	290.99	< .001***
Number of Words	-2.20	0.55	23.54	< .001***
Number of Topics	-10.74	3.54	1.66	0.197
PilotGroup*Words	-0.84	0.23	13.02	< .001***
PilotGroup*Topics	13.34	1.50	78.61	< .001***

*Note. Significance codes: .001 '***', .01 '**', .05 '*'. Non-pilot group coded as default treatment.*

Model fit determination using piecewiseSEM package in R (Lefcheck, 2015), gave a marginal (fixed effects) R^2 value of 0.46 and conditional (including random effects) R^2 value of 0.66. Regression results indicate that pilots had significantly higher AE intelligibility scores than non-pilots. Additionally, non-pilots' scores decreased with increases in number of words and number of topics, whereas pilots' scores decreased slightly more with number of words and increased with number of topics (see Figure 1). Model statistics indicated multicollinearity between CE (R^2 = .12) and WM (R^2 = 0.11). However, their addition to the model significantly increased model fit from R^2 = 0.45 to 0.46 ($\chi^2(2)$ = 7.71, p = 0.021).

Flight experience effect on AE scores. A separate regression was done on pilot group AE scores, to determine flight experience effect on AE score (see Table 4).

Table 4.
Linear Mixed Effects Analysis of Pilot AE Intelligibility Scores (Random Effects: Subject and Transmission)

Predictor	*Coefficient*	*Std. Error*	$\chi^2(1)$	*p-value*
Intercept	104.13	3.25		
Number of Words	-2.74	0.28	95.30	< .001 ***
ln(IFR)	1.88	0.44	18.69	< .001 ***

*Note. Significance codes: .001 '***', .01 '**', .05 '*'*

The full model included the above factors in addition to total flight hours (TT) and Instrument Flight Rules hours (IFR). Substitution of ln(IFR) for IFR increased model fit by R^2 of .01, therefore it was included in the model. The model resulted in a marginal R^2 value of 0.27 and conditional R^2 value of 0.50. Pilots' AE scores were significantly predicted by number of words in the transmission and by flight experience.

Types of AE Errors

One randomly selected participant's responses from each group were preliminarily coded for descriptive analysis. Errors were coded as transpositions, substitutions, number-number substitutions, omissions, and readback omissions (reflecting standard AE terminology). About half of the transmission data consisted of repetitive phrases and 42.76% were numbers. Both pilots and non-pilots transposed, or produced wrong, numbers (see Table 5). Otherwise, observation of the data, as well as analysis of these two participants, indicates that pilots typically produced errors of omission, while non-pilots' errors were more often substitution. For example, responding to the transmission *Turn right heading two seven zero,* a pilot produced ____ *right two seven zero*, whereas a non-pilot responded: *Turn right hitting two seven zero.* Non-pilots' also included substituted numbers for non-number words.

Table 5.
Breakdown of AE Errors by One Participant From Each Group

Group	total errors	transposition	substitution	wrong number (substitution)	omission	readback omission
non-pilot	96	10 (10%)	42 (44%)	13(14%)	31 (32%)	na
pilot	76	4 (5%)	10 (13%)	11 (14%)	25 (33%)	26 (34%)

Discussion and Conclusion

Results of this study indicate that AE is not intelligible to non-pilot native English speakers beyond a low threshold (53%) and acoustic learning of AE with no feedback peaks early at a level far below ceiling (~ 55%). Examining the data, we can theorize as to why CE proficiency does not imply AE intelligibility. Firstly, regression results indicate that, whereas number of words in a transmission is the primary factor in determining AE difficulty for pilot and non-pilot groups, this effect was mitigated for pilots by number of aviation topics in the transmission. This finding is consistent with the observation that expert language-users chunk information to efficiently interpret language streams. Data examination also indicates that, since AE topic identifiers are frequent and predictable, they are rapidly produced and monotone, making them less intelligible to naïve listeners. Therefore, non-pilots substituted novel terms for topic identifiers (i.e. *try to maintain* for *climb and maintain*). On the other hand, pilot AE errors reflected patterns of standard pilot-ATC communication, in which pilots repeat only critical elements of transmissions. Therefore, although instructed to repeat the entire transmission, pilots often omitted words that could be implied, (i.e. *runway, heading, turn, left, right, of, and, to, the, at*).

The logarithmic relationship of pilot flight experience with AE scores suggests that the AE learning curve is steep for low-time pilots and shallows out with experience. During flight, a small percentage of time is in direct communication with ATC and a higher percentage of time is in passive exposure. A training program in which pilots are exposed to recorded ATC transmissions including periods of active response and periods of passive listening would expose students to both flight language experiences. This type of training protocol would enable pilots to dedicate their attention in a low-stress, focused, efficient language-learning environment, rather than struggling to allocate cognitive resources during flight training, allowing them to acquire AE proficiency in far less time than it takes in flight. Although native English speakers may not be able to learn AE without feedback, AE language itself is formulaic, employing a constrained lexicon and restricted phrase inventory. This makes AE ideal for teaching, particularly when taking into account the chunking methodology that lends itself to pilot comprehension. If focus in training is on topic identifiers, novices may quickly learn how to recognize these rapidly produced language chunks.

Conclusion

This study seeks to improve international pilot language training by showing the need for pilots to learn the language they use every day on the job. Previous studies have shown that AE's rhythm and intonation are different from CE. The current study shows that AE is scarcely intelligible to CE speakers. Therefore the assumption that CE proficiency enhances AE proficiency is in doubt. The most efficient way of teaching AE is to focus on the AE language that pilots actually hear: including static, fast speech, real accents and a reduced frequency range. Because of the emphasis on CE in training, pilots may not be getting enough AE training before relying on it in flight. A small amount of classroom and/or online training focusing on familiarization with the limited inventory of AE words and phrases, as well as exposure to the rhythm and intonation of real ATC transmissions could enable pilots to effectively and confidently communicate in AE as soon as they get off the ground.

References

Alderson, J. C. (2009). Air Safety, Language Assessment Policy, and Policy Implementation: The Case of Aviation English. *Annual Review of Applied Linguistics, 29*, 168-187. doi:10.1017/s0267190509090138

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. <doi:10.18637/jss.v067.i01>.

Cookson, S. (2011). Zagreb and Tenerife: Airline accidents involving linguistic factors. *Australian Review of Applied Linguistics, 32*(3).

Day, B. (2004). Heightened awareness of communication pitfalls can benefit safety. *ICAO Journal*.

Farris, M. C., & Barshi, I. (2013). *Misunderstandings in ATC Communication: Language, Cognition, and Experimental Methodology*: Ashgate Publishing, Ltd.

Farris, C., Trofimovich, P., Segalowitz, N., & Gatbonton, E. (2008). Air traffic communication in a second language: Implications of cognitive factors for training and assessment. *TESOL Quarterly, 42*(3), 397-410.

ICAO, (2004). *Manual on the implementation of ICAO language proficiency requirements*. (Doc 9835 AN/453). Retrieved from http://caa.gateway.bg/upload/docs/9835_1_ed.pdf)

Kim, H., & Billington, R. (2016). Pronunciation and Comprehension in English as a Lingua Franca Communication: Effect of L1 Influence in International Aviation Communication. *Applied Linguistics*, amv075.

Kim, H., & Elder, C. (2009). Understanding aviation English as a lingua franca: Perceptions of Korean aviation personnel. *Australian Review of Applied Linguistics, 32*(3), 23-23.

Lefcheck, Jonathan S. (2015) piecewiseSEM: 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

Moder, C. L. (2012). Aviation English *The Handbook of English for Specific Purposes* (pp. 227-242): John Wiley & Sons, Ltd.

Moder, C. L., & Halleck, G. B. (2009). Planes, politics and oral proficiency. *Australian Review of Applied Linguistics, 32*(3), 25-1.

Peirce, J. W. (2007). *PsychoPy-Psychophysics software in Python*. Journal of Neuroscience Methods, *162* (1-2):8-13

Open speech repository (2016). Retrieved from http://www.voiptroubleshooter.com/open_speech/american.html

Godfrey, J. J. (1994). *Air Traffic Control Complete,* (LDC94S14A). CD-ROM. Linguistic Data Consortium, Philadelphia, USA.

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.