AVIATION ENGLISH IS DISTINCT FROM CONVERSATIONAL ENGLISH:

EVIDENCE FROM PROSODIC ANALYSES

AND LISTENING PERFORMANCE

by

JULIA E. TRIPPE

A DISSERTATION

Presented to the Department of Linguistics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

March 2018

DISSERTATION APPROVAL PAGE

Student: Julia E. Trippe

Title: Aviation English is Distinct from Conversational English: Evidence from Prosodic Analyses and Listening Performance

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

Eric Pederson	Chair
Melissa Redford	Core Member
Melissa Baese-Berk	Core Member
Robert Mauro	Institutional Representative

and

Sara D. Hodges	Interim Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded March 2018

DISSERTATION ABSTRACT

Julia E. Trippe

Doctor of Philosophy

Department of Linguistics

March 2018

Title: Aviation English is Distinct from Conversational English: Evidence from Prosodic Analyses and Listening Performance

International aviation professionals converse in a register of English derived from postwar radiotelephony. Decades of use and regulatory pressure established Aviation English (AE) as the lingua franca for pilots and air traffic controllers. Recently, the International Civil Aviation Organization (ICAO) required aviation professionals prove AE proficiency, resulting in development of a variety of AE programs and tests derived from English language pedagogy, without accounting for unique aviation language requirements. This dissertation explores linguistic characteristics that must be accounted for in international AE programs.

Although AE standard phraseology is a limited code, regulation allows limited use of “plain language”. Unfortunately, this caveat has paved the way for native English speakers (NESs) to use colloquial English that is often opaque to non-native English speakers (NNESs). Accordingly, the ICAO further required international pilots and controllers to have conversational English (CE) proficiency.

Structural differences in AE predict an emergent prosody with rate and rhythm differences from CE. In addition to environmental differences, this distinct prosody predicts differences in production and perception of AE and CE. My dissertation examines both of these phenomena, first by evaluating prosodic differences between AE and CE; next by analyzing AE listening and repeating performance.

To compare AE and CE prosody, I examined two radio corpora: air traffic controllers and radio newscasters. From these data I quantified rhythm, intonation and speech rate differences across registers.

Using laboratory listening performance studies of pilot and non-pilot NESs and NNESs, I examined AE intelligibility differences. NNES pilots scored worse on CE tasks and better on AE tasks than NES non-pilots, indicating CE proficiency is not a predictor of AE proficiency.

Dissertation findings suggest AE language training should focus on AE and not on CE, as is current practice. Given phonological and other differences between AE and CE, enlisting all AE users to learn and adhere to AE phraseology will save time and money in training and alleviate miscommunication and confusion in flight, potentially saving lives.

This dissertation includes unpublished coauthored material.

CURRICULUM VITAE

NAME OF AUTHOR: Julia E. Trippe

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
Lane Community College, Eugene, Oregon
Embry-Riddle Aeronautical University, Daytona Beach, Florida (online)
Lane Community College, Eugene, Oregon
University of Alabama, Birmingham, Alabama
Rice University, Houston, Texas

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2018, University of Oregon
Bachelor of Arts, Linguistics, 2010, University of Oregon
Associates Degree, 1996, Lane Community College

AREAS OF SPECIAL INTEREST:

Phonetics/Phonology
Aviation Human Factors

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellowship, Linguistics, 2011-2014, 2016-2018

Aviation Human Factors Research Assistant, Decision Research, Eugene, Oregon, 2012-2016

Graduate Research Assistant, Department of Linguistics, University of Oregon, Eugene, Oregon, 2012-2014, 2016-2018

Undergraduate Research Assistant, Department of Linguistics, University of Oregon, Eugene, Oregon, 2009-2010

Teaching Assistant, Department of Linguistics, University of Oregon, Eugene, Oregon, 2011-2013

Professional Pilot: Corporate, commuter, cargo, glider, instructor, 2000-2009

GRANTS, AWARDS, AND HONORS:

Foreign Language and Area Studies (Korean) Summer Fellowship, University of Oregon, 2012

College of Arts and Sciences Everett D. Monte Scholarship, University of Oregon, 2011-12

Foreign Language and Area Studies (Korean) Fellowship, University of Oregon, 2010-11

College of Arts and Sciences Everett D. Monte Scholarship, University of Oregon, 2010-11

Edmundson-Davis Scholarship, University of Oregon, 2009-10

Osher Reentry Scholarship, University of Oregon, 2009-10

Undergraduate Research Fellowship, University of Oregon, 2009-10

Cornelia Fort Scholarship, Women in Aviation International, 2001

Amelia Earhart Scholarship, The Ninety-Nines, 2001

Gary Stanley Scholarship, Lane Community College Flight Technology Department, 1996

Wayne Shields Scholarship, Lane Community College Flight Technology Department, 1995

PUBLICATIONS:

Sherry, L., Mauro, R., & Trippe, J. (in press). Design of a primary flight display (PFD) to avoid controlled flight into stall. *Aviation Psychology and Applied Human Factors*.

Trippe, J., Pederson, E. (under review). Aviation English intelligibility. *The International Journal of Aerospace Psychology*.

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

Trippe, J. & Mauro, R. (2017). Automation surprise in commercial aviation. In Vidulich, M., Tsang, P., & Flach, J. (Eds.) *Advances in Aviation Psychology*. Ashgate, Farnham, UK.

Trippe, J. & Mauro, R. (2015). Understanding automation surprise: Analysis of ASRS reports. *Proceedings of the 18th International Symposium on Aviation Psychology*, Dayton, Ohio.

ACKNOWLEDGMENTS

I especially want to thank my advisor Eric Pederson for his patience and perseverance with me on this journey. I would also like to thank my committee members: Melissa Baese-Berk for her diligence and generosity; Robert Mauro for his friendship and guidance; Lisa Redford for the benefit of her intelligence and experience; and my first advisor, Susan Guion-Anderson, who will live in my memory as the consummate academic. Thanks are due to Charlotte Vaughn for hours of her time discussing this unusual topic. I thank my fellow grad students who contributed ideas and critiques as well as hugs, tea and beer: Ying Chen, Wook Kyung Choe, Paul Olejarczuk, Shahar Shirtz, Hideko Teruya, Prakaiwan Vajrabhaya; and the linguistics undergrads who shouldered some of the more tedious tasks: Drew McLaughlin and Lathrop Hughes for running subjects and coding data more cheerfully and reliably than I imagined possible. Special thanks to the administrators and students of Hillsboro Aero Academy and Lane Community College Flight Aviation Academy (my alma mater) for their interest in this project as well as their valuable time. And finally, thanks to the Department of Linguistics at the University of Oregon: I could not have asked for a more generous and intelligent group of people to share my graduate school experience with.

This dissertation is dedicated to my wonderful wife Ker, for sticking it out
with me while actually continuing to support my dreams of
communication and flight.

Also, to my dear parents, who instilled in me the belief that I could do
anything as well as the desire and curiosity to do it.

TABLE OF CONTENTS

Chapter Page

Chapter Page

Chapter Page

LIST OF FIGURES

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

This chapter uses portions of an article entitled A Prosodic Profile of Aviation English that is currently under revision for the journal *English for Specific Purposes,* co-authored by Melissa Baese-Berk. This chapter was written entirely by me, with my coauthor providing editorial assistance.

The great enemy of communication, we find, is the illusion of it.

--William Whyte, *Is Anybody Listening? 1950*

1.1. Introduction

The purpose of this dissertation was to examine some phonological aspects of the variety of English referred to as Aviation English (AE): the form of communication used between pilots and air traffic controllers (ATCOs) internationally. Since native English speakers (NES) and non-native English speakers (NNES) use this variety of English, my studies focused on how similar or different AE is to conversational English (CE). This relationship informs usage and practice for NES and NNES populations. Establishing the phonological relationship between AE and CE may also clarify the veracity of assumptions regarding the learnability of AE and the extension of CE proficiency to AE.

Two assertions commonly made regarding the relationship between AE and CE are that AE sounds different from CE and that novice CE speakers have a difficult time understanding AE. The purpose of this dissertation was to support these claims with evidence in order to illustrate how distinct AE is from CE. The findings illustrated here have implications for AE training of NESs and NNESs. In the discussion, I will be calling for changes in protocol, although there are more steps to attend to prior to implementing them. This exploration of a practical issue is relevant to linguists concerned with English for Specific Purposes and is theoretically grounded in language acquisition literature. Documented structural differences between these English registers have implications for rate and rhythm, as well as production and perception. These elements, in turn, have implications for the learnability, training and testing of AE.

The two experiments described in the following chapters do not directly relate to one another. However, they each contribute to answering the question of how different AE is from CE. The first experiment examines and compares the prosody of these language varieties to determine how different they are. The second evaluates AE listening and repeating tasks by different populations to determine whether people with only a CE background can understand AE.

This study juxtaposes AE speech recordings with Standard English speech recordings to describe the phonology of AE. Using this comparison, I illustrate prosodic differences between these varieties of English to help refine a description of AE. To further compare the relationship between AE and CE, I run AE intelligibility tests on populations of NES non-pilots, NES pilots and NNES pilots to determine whether language background predicts AE proficiency. These tests measure AE intelligibility via verbal repetition of NES ATCO transmissions. AE intelligibility as well as error production was compared between the groups. Results of these studies are applied to documented challenges in international AE training.

1.2. Background of the problem

Aviation English is not a conversational style, but a distinct register of English: a codified, abbreviated, jargon-filled register using numbers paired with descriptors to convey crucial information succinctly. As a form of radiotelephony, AE standard phraseology is designed to be decipherable without face-to-face contact, in a time-critical environment that includes radio static and multiple speakers sharing a single radio frequency (International Civil Aviation Organization, 2004; Melnichenko & Melnichenko, 2009; Philps, 1991). Messages must be conveyed quickly and concisely via clear, emotionless delivery (Prinzo, Lieberman, & Pickett, 1998). Radiotelephony also requires speakers to audibly occupy the radio frequency or risk being interrupted. This specialized register of English has been adopted as a lingua franca by aviation professionals around the world (Estival, Farris, & Molesworth, 2016; Hazrati, 2015; Kim & Elder, 2009). Since 1951, the International Civil Aviation Organization (ICAO) has required that ATCOs and pilots who do not share a native language speak AE with one another. If pilots and controllers share another language, they are permitted to use that. However, all flight crews in an area are privy to each other's communications (similar to a party-line phone exchange) and develop an awareness of each other's positions and intentions from the content of these transmissions (Prinzo & Campbell, 2008). If crews are not able to understand each other, situational awareness is diminished. Consequently, the practice of speaking a non-English aviation language is discouraged. Accordingly, some countries are choosing to implement English-only rules in their airspace (Dennis, 2015; Tallantyre, 2014). Because AE communication problems continued to contribute to aviation accidents, ICAO implemented a further requirement that international pilots and controllers undergo AE proficiency testing starting in 2011 (ICAO, 2004). AE communications affect the safety of some three and a half million passengers daily

(ICAO Annual Report 2014, 2015), underlining the need for description and standardization of this variety of English.

Even though AE proficiency is mandated for international pilots and ATCOs, little is known about how to evaluate or train AE users. Difficulty describing AE stems from the fact that the standard phraseology described in regulatory publications (Federal Aviation Administration, 2015; International Civil Aviation Organisation, 2010) is not the only sanctioned form of AE radiotelephony. This standard phraseology may be circumvented in non-routine situations and in…

> … cases, where phraseology provides no ready-made form of communication, pilots and controllers must resort to plain language (ICAO, 2010: 3.3.13.).

Therefore, proficiency in plain language, or plain English, when AE is in use, has become mandatory in addition to proficiency in AE standard phraseology (ICAO, 2010). Since plain language is not sufficiently defined, this stipulation amounts to a CE proficiency requirement for pilots and ATCOs working internationally.

Although proficiency in AE standard phraseology and plain English are required by the ICAO, little is known about how these varieties of English interact in language learning and use. One sociolinguistic study analyzing different usage characteristics of AE standard phraseology and plain AE (including subject-matter, situation-type, participant roles, mode, and medium of discourse) concluded that they are actually separate specialized registers of English for which the ICAO specifically outlines conditions of use (Bieswanger, 2016). Although the focus of my study was not sociolinguistic, I subscribe to this analysis. However, ICAO requirements support AE training conventions rooted in the assumption that CE proficiency aids in AE proficiency.

However, requiring proficiency in both these varieties of English may be a burden on practitioners and increase miscommunication.

1.2.1. Aviation English standard phraseology

In *Linguistic security in the syntactic structures of air traffic control English*, Philps (1991) outlines the construction of AE as a semi-artificial language dependent upon English grammar with systematic syntactic modifications. This generative study is an excellent survey of the exact phonological (i.e. pronunciation), syntactic, lexical, semantic and discourse-level modifications to CE employed to derive AE. In her discourse-focused chapter for *The Handbook of English for Specific Purposes,* Moder (2012) describes Aviation English as having "specialized vocabulary, restricted syntax and interactional characteristics" (p. 235). In another discourse analysis of AE, Hinrich (2008) notes that, "ICAO phraseology is designed to minimize syntax and intonation in order to more accurately pass information between pilots and controllers through the use of specific formula-based phrases" (p. 264).

AE standard phraseology word order and terminology are fixed, and marked by a lack of articles, auxiliary verbs, prepositions and pronouns (Hinrich, 2008; Moder, 2012; Philps, 1991). AE transmissions often consist of commands issued by an air traffic controller and acknowledged by a pilot. Indeed, these command/response communications are described as the "core role of pilot-controller communication" (ICAO, 2010, 3.4.7.). Command topics include flight path parameters (altitude, heading, airspeed), weather phenomena (wind, visibility, turbulence), location of other aircraft in the vicinity, and permission to perform particular tasks. Each transmission may include several topics. For example, here is an AE standard phraseology transmission from an ATCO to Delta flight 1019:

> Delta ten nineteen, twelve miles south of the marker heading three two zero (,) maintain four thousand till established (,) one nine zero knots to the outer marker (,) cleared I-L-S approach runway three five right (Godfrey, 1994).

In this example, proper standard phraseology is reflected in topic order as well as lexical topic identifiers and specific number expressions for each topic addressed. The aircraft is identified by a call sign made up of the carrier name and flight number read as a group, rather than as separate digits (i.e. *Delta ten nineteen*). Although AE limits articles and prepositions, to specify location, the phrase *south <u>of the</u> marker* is required to alleviate ambiguity. AE identifies compass direction commands by the word "heading" followed by three digits (i.e. *heading three two zero*). Altitude commands are identified by "(climb/descend and) maintain" with numbers given in digits except for "hundred" and "thousand" (i.e. *maintain four thousand*). Speed commands are also stated in digit format, followed by the "knots" identifier (i.e. *one nine zero knots).* To avoid ambiguity the phrase *<u>to the</u> outer marker* indicates the geographical limit of the speed requirement. Finally, clearance to fly a specific approach track must be given with the type of procedure, including the runway designation (i.e. *cleared I-L-S approach three five right*).

AE standard phraseology is designed for "maximum clarity, brevity and unambiguity [sic]" (ICAO, 2007: 3.2.2.). AE standard phraseology avoids ambiguity by fixing a single meaning to each word and phrase. Words whose pronunciation may cause confusion are assigned distinct pronunciations. For example, *five* and *nine* sound similar in noisy transmissions where initial and final sounds are obscured. Therefore, AE requires they be pronounced *fife* and *niner*, respectively. Additionally, AE standard phraseology restricts word and phrase inventories to "around 200 phraseologic [sic] English words and phrases" (Tajima, 2004, p. 458). Function words are not used except to resolve ambiguity. AE standard phraseology gives any word that is ambiguous or confusing in CE a single meaning or substitutes it with another word. For example, *descend* is juxtaposed to *climb* rather than *ascend*, since the *descend/ascend* distinction could be lost in a noisy signal. In the case of, *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 followed by a number expression made up of digits and the term *thousand* or *hundred.* These lexical and grammatical differences in addition to environmental factors (i.e. multiple speakers, signal static and reduced frequency range), lead to differences in the sound profiles of AE and CE.

Studies of miscommunications have analyzed the impact of these linguistic features on AE form and usage in accidents and incidents. See Prinzo and Britton (1993) for an extensive historical survey of 46 studies in which researchers use taxonomies, acoustic correlates, and cognitive/psycholinguistic approaches to analyze pilot-controller miscommunication. A 1997 field study of actual AE usage determined that, of ATCO speech rate, number of topics and linguistic element ellipsis, the only significant source of miscommunication was number of topics per transmission (Barshi, 1997). In another study of ATCO/pilot communications, Prinzo found that pilot readback (repeating commands from ATC by way of message confirmation) errors increased with message complexity and length (2008).

A series of experimental studies examining cognitive load effect on pilots responding to ATCO commands follows the paradigm developed by Immanuel Barshi in 1997. For these studies, non-aviators perform navigational tasks by moving tokens on a computer screen in response to verbal commands. Experimenters asked subjects to repeat the commands but measured performance by the correct movement of the tokens. Barshi and Healy (1998) used this experimental design in a study of native and non-native English speakers responding to English commands. Their results indicated that message length and not rate of speech was a causal factor in misunderstanding for NESs and NNESs. Farris (2007) continued this series of experiments by examining fluency of spoken responses to ATC-like commands for NESs and NNESs (high and low English proficiency) under cognitive load conditions. Study results indicate cumulative negative effects of cognitive load, message length and low language proficiency. Once again, message length was shown to be detrimental for all participants and low-proficiency NNESs' fluency suffered with the introduction of cognitive load.

Carrying the cognitive load studies into the realm of actual pilot communication, Estival and Molesworth (2016) undertook a simulator study of NES and NNES pilots communicating under four different cognitive load and control conditions. Results indicate that number of items was the most consistently problematic element for all pilots, although faster speech (without pauses) was more challenging for all but the more experienced NES pilots. All subjects' communication performance was worse under workload conditions and none suffered in the radio congestion condition.

Researchers have studied Aviation English communication between NESs and NNESs as a contributing cause in specific accidents, as well. Cushing (1994) examined high loss of life accidents in which ambiguity, homophony, and uncertain reference were among the sources of AE communication confusion. Cookson (2011) performed an in-depth analysis of two accidents involving AE miscommunication between NESs and NNESs, describing language as one of many causal factors including cultural, environmental and biological elements. For the period between 1971 and 2002, Jones (2003) enumerated 35 accidents and incidents, that resulted in 3,295 deaths that were caused in part by miscommunication (see Section 1.2.2. for a description of his findings).

For a more comprehensive survey of AE literature and a compelling description of the special characteristics of AE, please refer to Anna Borowska's recently published *Avialinguistics: The Study of Language for Aviation Purposes* (Borowska, 2017c).

While the syntax and vocabulary of AE standard phraseology are relatively well described, there is a lack of descriptive work on the AE sound system. Studies of pilot and controller judgments of communication challenges indicate the need for such a description. Prinzo, Campbell, Hendrix, and Hendrix (2010b, 2010c, 2010d, 2011) present results from interviews with 48 professional US pilots describing their communication challenges in international operations. Amongst others causes of AE miscommunication, pilots in this study cite accent, pronunciation, speech rate and language proficiency. In a study of over a hundred NES and NNES Australian general aviation pilots, Estival and Molesworth (2011) found that pilots reported the greatest communication challenge as understanding other pilots. The authors speculated that this could be "…due to noise in pilot-to-pilot transmission or to the poor aviation communication skills (production and/or understanding) of the pilots" (p. 372). In another survey of AE users, Kim and Elder (2009) interviewed 42 professional Korean pilots and ATCOs about their AE communication experiences. A primary complaint of both populations was that NESs deviated from AE standard phraseology and were not intelligible.

Given these broad, qualitative categorizations of miscommunications, it is difficult to determine their precise source. “Accent” alone has been used heuristically to describe a number of linguistic features, including rhythm (or timing), speech rate, and intonation (Prinzo, Campbell, Hendrix, & Hendrix, 2010a). Counter to these AE user perceptions, Barshi and Farris (2013) report that speech rate has no effect on error production in actual AE usage or comprehension for NES and NNES subjects listening to AE-like navigation instructions. A confounding factor in all the studies evaluating AE usage is the fact that AE consists of two distinct varieties of English: AE standard phraseology and plain English.

1.2.2. Plain English

Unlike AE standard phraseology, plain AE is relatively undefined. However, the intent of the regulation is clear. The ICAO stipulates that

> Plain language in aeronautical radiotelephony communications means the spontaneous, creative and non-coded use of a given natural language, although constrained by the functions and topics (aviation and non-aviation) that are required by aeronautical radiotelephony communications, as well as by specific safety-critical requirements for intelligibility, directness, appropriacy [sic], non-ambiguity and concision. (ICAO 2010: 3.3.14)

Because of the vagueness of this description, individual pilots and ATCOs must determine the construction of plain English utterances. Underlying the “plain English” exception is the assumption that this form of English will lead to more reliable communication than AE standard phraseology in non-routine circumstances. In cases where pilots and controllers share a first language, either a regional English or another language, this assumption may be accurate, and the use of a shared conversational English should typically clarify communication. However, AE is mandated for all pilots and controllers in international airspace *who do not share a first language.* Therefore, the assumption that CE will be a reliable form of communication may be inaccurate. NNESs fluent in AE standard phraseology may have less facility in the inherently more complex and nuanced productions of CE. Given that there are dozens of varieties of English, CE itself (and therefore plain English) may be quite different regionally and in different contexts, therefore it is not a reliable resource for unambiguous, safety-critical aviation communication.

In his study of English as an international aviation language, Jones (2003) goes so far as to aver that English is too ambiguous to be used in this context. He cites numerable confusing usages of English in AE standard phraseology, including homophones, synonyms, homonyms, compound terms and misnomers.

> Non-native English-speaking pilots, about half of the pilots of the world, find it very difficult to deal with the massive irregularities of English. (Jones, 2003, p. 243)

These factors along with inconsistent use of measuring units (metric v. imperial) and AE accentual differences, lead the author to conclude that English is not suitable as the basis for an international aviation language (Jones, 2003). However, convention necessitates the continued use of English in this capacity (ICAO, 2004, 2.7.6).

1.2.3. Regulating Aviation English

1.2.3.1. International Civil Aviation Organization (ICAO) Regulations

High loss-of-life accidents caused in part by communication problems (Cookson, 2011; Cushing, 1994; Jones, 2003) compelled the ICAO to require AE proficiency in international airspace as of 2011. However, this requirement has yet to be thoroughly operationalized. Although ICAO has published general proficiency-rating guidelines, there is no agreed upon standard protocol by which to attain or prove proficiency. While dozens of tests have been developed internationally in the wake of AE proficiency regulation and several are in use, ICAO recognizes only one (English Language Proficiency for Aeronautical Communication) at this time. The official guidance for AE testing consists of a general definition of mandatory proficiency levels. For instance, for AE pronunciation, operational (level 4) proficiency is defined as:

> Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding (ICAO, 2004).

The above standard also pertains to plain English language proficiency (ICAO, 2004), although the majority of pilot-ATCO communication is in AE standard phraseology, which was designed to convey all typical transactions. When AE standard phraseology is not sufficient to convey a message, ICAO regulations stipulate the use of plain English. Generally, this caveat applies to unusual or emergency situations. Although the implementation of AE standard phraseology recognizes the need to keep communications succinct and unambiguous, it is impossible to control for these needs in “plain English”, because its parameters are undefined. NESs often speak quickly and colloquially during times of stress. Although such interactions may aid in clarification of complex situations between NESs, these transmissions may not be comprehensible to NNESs (Kim & Elder, 2009). In his study of Aviation Safety Reports, Jones states:

> … when a controller's working language and native language are the same he is apt to lapse into puzzling idioms and to make assumptions about the pilot's command of English (p. 243)

Additionally, NNESs have more difficulty conversing in English under conditions of stress or high cognitive load that typically trigger plain English use in native speakers (Estival & Molesworth, 2016; Farris, Trofimovich, Segalowitz & Gatbonton, 2008).

The ICAO recommendation to use "plain language" in non-routine situations is further confounded by the fact that there exists no consistent guidance as to what is meant by *plain* AE. The regulatory intent is clear: this English variety should be readily understandable to the listener. Unfortunately, it is impossible to ascertain what level of English proficiency, or indeed what model of Standard English, a given person has. Day (2004) states that whenever AE users deviate from AE standard phraseology they stand a higher chance of being misunderstood or misinterpreted. Howard's (2008) study of actual pilot-ATCO interactions concluded that casual deviations from AE standard phraseology often lead to problems in communication. Language experts recommend that pilots and ATCOs avoid plain English as much as possible in their communications (Day, 2004: Moder, 2012). AE standard phraseology fluency reduces repetitions, delays, and misunderstandings.

As the international flying community becomes more diverse, pilots will interact with crews having 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 an undefined, limited form of CE (i.e. plain

AE), consideration should be given to expanding AE standard phraseology so that unusual situations may be addressed using this clear and constrained format and lexicon.

1.2.3.2. Federal Aviation Administration (FAA) Regulations

Although the US is a member –state of the ICAO, its internal regulations differ from those of the ICAO, as do the regulations of other member states'. It is important to note that many pilots and controllers from other countries come to the US for training. In order to attain an FAA license, the Federal Aviation Regulations (FARs) mandates that an individual, "… is able to read, speak, write, and understand the English language" (FAR§61.75). This self-reported proficiency is only tested indirectly during the flight test administered by an FAA-designated flight examiner. Until recently, examiners were very rarely taken to task for not ensuring that candidates had sufficient English skills, especially if it was known that the candidate would be going "back home" to work in aviation. However, since the change in the ICAO regulation, more examiners are being held responsible for this language assessment and some have even lost their jobs due to negligence in assessing English ability (as per a confidential personal communication).

Another difference between FAA and ICAO regulations is in the stipulation regarding plain language. In the FAA advisory publication, the Aeronautical Information Manual (AIM), it is recommended that:

> Since concise phraseology may not always be adequate, use whatever words are necessary to get your message across (FAA, 2017, 4-2-1).

Therefore, pilots trained in the US, NESs and NNESs alike, will have a propensity to use colloquial English whenever they think they need to, rather than after exhausting all possible standard phraseology application.

1.2.4. Testing and Training of Aviation English

1.2.4.1. Testing

ICAO regulations stipulate a mandatory level of competency in AE and CE (ICAO, 2001), but give vague, often contradictory, guidelines as to how to test for that proficiency (ICAO, 2004).

> … one of the difficulties in developing tests in response to the ICAO LPRs [language proficiency requirements] is that tests are to be developed in response to policy and to largely theoretical notions of language use in the aviation context, as opposed to being developed in response to empirical studies of the way language is actually used in this context (Farris, 2016, p. 84).

In fact, the guidelines emphasize the need for testing to evaluate non-routine, and therefore plain AE language (or CE), and not AE standard phraseology. Newly developed testing protocols differ greatly. However, AE tests commonly include a face-to-face interview with a language evaluation specialist. In these interviews, the pilot must discuss unusual situations that may arise in flight, 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 that are neither aviation professionals nor fluent in AE. This type of testing does not evaluate the AE speech used in the vast majority of AE user interactions. In fact, listening and responding to actual ATCO transmissions is not evaluated (Farris, 2016) and may not even be included in the pilot's language proficiency test (Alderson, 2009). Additionally, when ATCO 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. In short, passing an AE proficiency test does not guarantee a pilot's ability to fulfill their job requirements.

In their study of oral proficiency in NNES controllers, Moder and Halleck (2009) found that there was no consistent relationship between AE and CE scores. In an effort to clarify the relationship between standard phraseology and plain English proficiency, this study assessed 14 Asian ATCOs' language abilities in three areas of aviation communication: standard phraseology, and plain English in common and less-expected occurrences. On average, ATCOs scored higher on standard phraseology tests than plain English tests, however, individual scores varied greatly. Although these were practicing ATCOs, there were individuals who scored very well on the plain English tests and poorly on the standard phraseology. Others did not reach proficiency on plain English tasks but scored very high on standard phraseology. Correlation between these tasks was only .40.

Other AE language scholars assert that CE-focused testing protocols unfairly benefit NESs, who are assumed to be fluent in AE (Estival & Molesworth, 2016), but often do not comply with standard phraseology (Kim & Elder, 2009) (see Section 1.2.1 in this dissertation for a description of standard phraseology). As previously cited from the Moder and Halleck (2009) experiment, proficiency in CE does not predict AE proficiency.

1.2.4.2. Training

The standard for AE training has long been that radiotelephony is learned simultaneously with flight training. Scholars have posited that this is due to the fact that AE does not comply with a written form (Borowska, 2017c; Estival, 2016, Hansen-Schirra & Maksymski 2013). Although flight training programs generally expose student pilots to the published guidelines regarding AE usage, they dedicate little or no ground instruction to AE radiotelephony. The first time pilots hear AE is usually in flight while they are learning to master aircraft control. It is assumed that pilots will learn through immersion: monitoring and interacting with ATCOs. This AE immersion strategy has been adopted as the model for NNES training in native English-speaking countries, which is where a great deal of international commercial flight training takes place. Exact documentation regarding commercial pilot training numbers is not available, highlighting the lack of critical analysis of the current AE situation.

Many flight-training programs for NNESs include AE training courses designed by English-language teaching experts in consultation with aviation professionals. Taught by non-AE speakers, these courses mirror CE classes by focusing on face-to-face CE communication with aviation terminology. In actual flight conditions, pilots must interpret rapid AE transmissions through static and reduced frequency range, without seeing the speaker. Therefore, it would probably be beneficial to train specifically in naturally produced AE.

Scholars in ICAO member-states throughout the world report problems with new training and testing tools. On the one hand, in a study examining the training of Thai AE instructors, Suksiripakonchai (2012) describes shortcomings in the existing protocol in which teachers drawn from the aviation industry are not TESL-rated (teaching English as a second language) or given teaching guidance. On the other hand, a recent study of an AE textbook used in Saudi Arabia finds that the book does not address aviation subject matter, only conversational English (Alshabeb, Alsubaie, & Albasheer, 2017). Similar findings arose from a Russian study of AE textbooks by Melnichenko and Melnichenko (2009). These scholars state that there are “no good books”, because authors are unfamiliar with the AE environment. They assert that the focus of the texts needs to be on speaking and listening (as required by ICAO, 2010), to reflect aviation communication needs, and not on reading and writing (Alshabeb et al., 2017; Melnichenko & Melnichenko, 2009). Zolfagharian and Khalilpour (2015), in their review of an AE textbook used in Iran, also point out that the text reflects little knowledge of the AE environment, has very few spoken fluency exercises and does not include any AE standard phraseology.

There is no consistent approach to AE training and evaluation. In attempting to comply with regulations regarding testing of communication in non-routine situations, educators have focused on CE to the detriment of AE standard phraseology. The result is that pilots still learn proper AE radiotelephony on the fly. This learning through adaptation is not conducive to retention, as pointed out by Merritt and Maurino (2004).

> It is a commented fact that cosmetic behaviors crumble under stress and a reversion to native behaviors takes place. The safety concern about dealing with cross-cultural interfaces through adaptation is obvious; in stress situations adaptation may become ineffective (p. 156).

Based on linguistic theory and pedagogical research, this dissertation argues that AE radiotelephony training needs to be intentional, focused and stress-free, preceding and in preparation for flight training.

1.2.5. Voice Communication versus Text

Bypassing voice communication altogether may be another way of reducing in-flight communication problems. Since verbal communication is a source of error and potential risk in aviation, it has been suggested that transactions be undertaken in text format. I have included a brief description of this phenomenon in order to give a fuller picture of current and future AE communication. As the world’s airports get busier and the airways get more crowded, there is more and more discussion of "voice-by-exception", that is voice communication only if absolutely necessary, otherwise, electronic transfer of textual messages would be the norm. This transfer of written information is called Controller-Pilot Data-link Communications (CPDLCs), which continues to be the anticipated future state of Aviation pilot / controller communication (Cardosi, Falzarano, & Han, 1998; Prinzel, Shelton, Jones, Allamandola, Arthur, & Bailey, 2010). Another important reason given for developing this technology is to free up radio frequencies. In crowded airspace, there currently is a shortage of frequencies on which ATCOs can talk to pilots. Data-link could leave the airwaves open for verbal communication of non-routine situations, especially emergencies. A summary of CPDLC research follows.

Here are lists of some positive and negative attributes of data-link communication technology (Prinzel et al., 2010, unless otherwise cited).

Positive data-link characteristics:

- Good for large amount of data needing no immediate response
- Textual transmissions are permanent and afford referencing
- Crews can take their time reading over the clearance and responding
- Data-link may be easier than voice communications for NNESs, since it involves reading and writing instead of listening and speaking
- Reduces voice frequency congestion

Negative data-link characteristics:

- Too slow for communicating in unexpected circumstances
- Rather than decreasing workload for the crew, merely redistributes it (Kerns, 1991)

- Increases visual tasks and the chance of visual channel overload and head-down time (often cited as a cause of disorientation and/or lack of situational awareness)
- Loss of "party-line" information for the development of situational awareness
- Vocal cues to mood or subtext in off-nominal situations (i.e. intonation, stress) are not available
- Flight crews could lose verbal communication proficiency
- It takes longer to read and interpret a message and enter a response in a keyboard than to listen to and repeat a command (Lozito, Verma, Martin, Dunbar, & McGann, 2003), especially for NNESs
- Flight crews can delay responding until it is a problem for other aircraft or the ATCO (Rakas & Yang, 2007)

1.2.5.1. NASA Data-Link Study

It has been suggested that voice communication is too unpredictable to rely upon in the aviation context. Accordingly, to facilitate conveyance of detailed flight sensitive information, data-link technology was developed to convey messages between ground personnel and aircraft crews (Roy, 2004). Aviation radio frequencies and airwave space are limited. In some areas, there are no more available frequencies to be used for aviation communication. Using data-link, some time-consuming one-way transmissions could be sent via text. The ICAO has advocated for text ground-air communications in order to reduce congestion and maintain clarity. The argument for data-link made by scholars like Cushing (1994) is that the written word is more reliable and can be referenced after the fact, whereas voice communication passes instantly. Currently data-link is regularly used by ground controllers to send flight plan information prior to push back from the gate, since long flight plan clearances are burdensome and often confusing or misunderstood.

Many flight crews also use data-link to send flight details to their company ground crews for coordinating refueling and loading efforts.

Many studies have been done on the efficacy of data-link technology. Because it is not the purview of this dissertation, I describe only one representative study done for NASA by Prinzel et al. (2010), in which professional crews flew 20 simulated approach scenarios using voice, data-link or a combination of the two. Their results indicate that data-link only scenarios had the highest workload and lowest situational awareness and data-link plus voice (completely redundant) increased the workload for both crew and ATCO, although it also increased flight crew situational awareness. Data-link plus readback only in which ATCOs sent textual transmissions and crews responded with verbal readback to acknowledge receipt of directives resulted in a 50% reduction of voice traffic, allowing more time to respond to clearances and more room for non-routine communications. Situational awareness was maintained in this condition via crews hearing other pilots reading back their intentions (Prinzel et al. 2010).

Given the current state of technology, although data-link technology may be used to decrease ATCO workload and radio congestion, it will never fully replace voice communication "… due to equipage [i.e avionic equipment] and operational constraints" (Lozito et al., 2013).

ICAO language recommendations stipulate that "natural language" will always be the best way of communicating in the aviation environment (ICAO, 2004 Section 1.2.4).

> As all other options fall short [i.e. data-link, translation], natural language continues to be the most reliable and efficient form of human communication. (ICAO, 2004 Section 1.2.4)

> Attempts to delimit the scope of a language will always fail at some point, when the need to communicate a new and unexpected situation exceeds the resources of the artificially constrained language. (ICAO, 2004 Section 1.2.5)

As accurate as this statement is, it is not realistic to force all international pilots and controllers to be highly proficient CE speakers who understand all regional and colloquial forms of the language. Some linguistic constraints must be maintained.

1.2.6. Restricting Plain Aviation English

Since natural spoken language cannot be completely avoided, it may be necessary to devise a universal language to supplement AE standard phraseology, rather than rely on interlocutors' conception of "plain English". Scholars conclude that English is not a suitable source for international aviation communication (Cookson, 2011; Cushing, 1994; Jones, 2003). Some support the idea of data-link technology (Cushing, 1994), while others suggest the creation of a completely new form of verbal communication akin to Esperanto (Jones, 2003). Despite these dissenting arguments, in 2003, ICAO implemented new regulations further instantiating English as the language of aviation by requiring proof of proficiency in both AE standard phraseology and "plain language" for all international pilots and controllers. ICAO cites the history of English in aviation as well as the prevalence of global English usage as arguments for keeping English as the international aviation language (ICAO, 2004, 2.7.6). This was no doubt perceived to be the path of least resistance in a realm where AE had already been in use by interlocutors who did not share a first language for several decades, and in a global environment where world Englishes are increasingly adopted as lingua francas in every field. Unfortunately, this decision by the ICAO created an imbalance in aviation communication, whereby NNESs are at a disadvantage and often treated as less capable than NESs (Borowska, 2017a; Hansen-Schirra & Maksymski, 2013; Kim & Elder, 2009). Scholars in NNES ICAO member-states describe a situation of "linguistic dominance" fostered by the adoption of English as the official language of aviation (Borowska, 2017a; Hansen-Schirra, 2013).

> The fact that the Aeronautical English is based on natural English does not mean that native speakers of English are released from 'learning' it. (Borowska, 2017b, p.139)

The veracity of this argument has been recognized and alluded to in guidance published by the ICAO:

> [A]n ethical obligation arises on the part of native speakers of English, in particular, to increase their linguistic awareness and to take special care in the delivery of messages (ICAO, 2004, 3.2.1.)

But situations in which NESs do not comply with standard phraseology or produce confusing utterances in plain AE continue to occur, resulting in the following amendment to the 2010 ICAO regulations:

> The burden of improving radiotelephony communications should be shared by native and non-native speakers (ICAO 2010).

It is clear that there persists a lack of responsibility and linguistic awareness on the part of NESs. It has been suggested that NESs rely on their native speaker status and, as a result, often lack "interactional effectiveness" (Barshi & Farris, 2013) and that there is a

> …need for native speakers of English to develop more effective paraphrasing strategies that rely on basic phraseology in order to accommodate the abilities of aviation professionals with widely varying English proficiency levels (Moder, 2012, p. 240).

To clarify the use of a plain variety of English comprehensible to all aviation professionals, scholars have suggested that the term "plain English" be replaced with either Plain Aviation English (Bieswanger, 2016), Plain Aeronautical English (Borowska, 2017c), and the structure and content of this language be strictly specified and learned alongside AE standard phraseology.

1.3. Purpose of This Study

International aviation is a high stakes environment. Today's global economy requires ready access to every nation in the world. Any disruption or threat to this system can throw economies, businesses and personal lives into confusion, and potentially destabilize the delicate global political and economic balance. As with any dynamic human endeavor, aviation requires coordinated communication to ensure safe and efficient functioning of an ever-changing variety of human, mechanical, meteorological, geographic, and environmental factors. Unique to aviation communication is the combination of time pressure, noisy signal, lack of face-to-face contact and high loss-of-life risk. Any aviation accident causes insecurity and ripples of lack of confidence that can disrupt the global economy. Therefore, aviation safety is a critical concern. Risk of aircraft systems failure continues to decrease with modern mechanical and technological improvements. As a result, communication is emerging as one of the primary aviation safety issues (ICAO, 2010).

Attempts at regulating international aviation communication have resulted in the ICAO requiring proficiency in a common English-based language for all international airspace users who do not share a first language. However, since this language is based on English, NESs don't feel the need to comply with strict usage recommendations, effectively placing the burden on NNESs to understand CE. The resulting inequity must be rectified for the community of AE users to be able to interact responsibly. These concerns are particularly important because so many NNESs are trained in NES countries.

In order to fully appreciate the differences between AE and CE, this dissertation seeks to go beyond grammatical / lexical comparisons to create a prosodic description. Differences in language structure have an impact on the rhythm and rate of AE, as well as its production and perception. My findings may assist regulators, educators and users in understanding the impact of requiring proficiency in both of these varieties of English. This dissertation examines actual AE usage. In order to help create an accurate description of AE standard phraseology, recordings of professional ATCOs' real world interactions with pilots are used as the source of phonological data and to explore AE intelligibility for NES and NNES populations.

This dissertation has three related goals: First, to evaluate the reasonableness of assumptions underlying AE proficiency requirements. If the prosodic analysis of AE and CE demonstrate that they are measurably different from one another and if one or the other variety is more intelligible to NESs or NNESs, it is perhaps inappropriate to require both varieties of English for pilot licensure. If AE is different than CE, we cannot expect that teaching CE will result in understanding AE. Linguistic theory informs us that learners must be exposed to AE specifically. Second, to assess areas of AE difficulty for NES and NNES pilots. Examination of error production by the different populations will yield patterns of problem areas for NES and NNES pilots, as well as NES non-pilots, which can be used to extrapolate challenges for AE learners. The third goal was to clarify some strategies for enhancing AE training for all pilots. Focusing on AE challenges for

NES and NNES pilots, new training techniques can be recommended, based on linguistic theory and pedagogical studies.

1.4. Summary and Preview

Chapter 2 summarizes of the prosodic analysis of AE as compared to CE. Data from two speech corpora are divided and compared for differences in rhythm, rate and intonation. The majority of this chapter was written as an article in *English for Specific Purposes,* co-authored by Melissa Baese-Berk.

Chapter 3 introduces the AE listening and repeating study and presents the comparison of study between NES non-pilots and pilots. The majority of this chapter was written as a journal article co-authored by Eric Pederson and Melissa Baese-Berk.

Chapter 4 presents the comparison of AE listening and repeating study results between NES and NNES pilot populations.

Chapter 5 presents the comparison of AE listening and repeating study results between NES non-pilots, NES pilots and NNES pilots, as well as the error analysis for all three populations.

Chapter 6 summarizes the dissertation findings and describes their implications for the industry and future directions.

CHAPTER II

PROSODIC PROFILE OF AMERICAN AVIATION ENGLISH

This chapter uses portions of an article entitled A Prosodic Profile of Aviation English that is currently under revision for the journal *English for Specific Purposes,* co-authored by Melissa Baese-Berk. Data analysis was performed by Julia Trippe. This chapter was written entirely by Julia Trippe, with Melissa Baese-Berk providing editorial assistance.

2.1. Introduction

There are no empirical studies quantifying phonological aspects of AE. One way to describe these characteristics is by analyzing the rhythm and intonation, in other words prosody, of the language. The purpose of the current study was to describe Aviation English (AE) prosody and compare it to that of Standard English. Grammatical, lexical and environmental differences in AE suggest that its emergent prosody will be different than that of Standard English. A clear prosodic comparison of AE and Standard English will help establish teaching standards and proficiency measures for this specialized register, as well as illuminating possible problem areas for native English speakers (NESs) and non-native English speakers (NNESs).

For the purposes of this study, I have chosen to represent conversational English with Standard American English (SAE), or the form of English commonly used in writing and speaking in the United States (US). This study seeks to clarify the relationship between NES AE standard phraseology and SAE by analyzing and comparing phonological aspects of each register.

2.1.1. Intelligibility

Intelligibility is the ability to understand individual words in a speech stream. Lack of intelligibility can lead to misunderstandings, which can lead to further communication challenges, costly delays, or even fatal consequences. During the rapid expansion of commercial aviation in recent decades, several high loss-of-life accidents including Tenerife in 1977 and Avianca Flight 52 in 1991 have been attributed to miscommunication between flight crews and air traffic controllers (Alderson, 2009; Cookson, 2011; Cushing, 1994). It is possible that phonological attributes of AE itself may be the source of some confusion. Aviation radiotelephony has been described as difficult to understand because it is rapidly produced and lacks prosodic cues (i.e., intonational and rhythmic changes) that English speakers use to divide a speech stream into meaningful units (McMillan, 1998; Prinzo, Hendrix, & Hendrix, 2008). These characteristics may lead to misunderstandings, particularly for novice AE users.

Prosody (intonation and rhythm) influences language intelligibility in general (Cutler & Carter, 1987; Cutler, Mehler, Norris, & Segui, 1992; Mattys, Jusczyk, Luce, & Morgan, 1999; Tajima, Port & Dalby, 1997) and has been described as being crucial for AE intelligibility, in particular. In addition to these studies, several others point out the relationship between AE miscommunication and prosody. In a study of pilot / controller communications, Prinzo et al. (2008) found that pronunciation (including rhythm and intonation) contributed to problems in communication. In several other studies examining pilot impressions of their fellow aviators' and controllers' AE communication, unnatural or irregular prosody is reported as a source of misunderstanding for all AE users, regardless of language background (Estival & Molesworth, 2011; Farris & Barshi, 2013;

Kim & Elder, 2009; Philps, 1991; Prinzo et al., 2010b, 2010c, 2010d, 2011). As testament to the importance of prosody to AE intelligibility, the ICAO has cited pronunciation and its stipulated subcategories of intonation and rhythm, as necessary areas of AE proficiency for international pilots and controllers (ICAO, 2004). The ICAO requires that pilots and controllers maintain operational proficiency in AE. Under the area of Pronunciation, Level 4 (Operational) proficiency is defined as:

> Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding (ICAO, 2004).

This proficiency requirement was amended in 2010 (ICAO, 2010) specifically to include NESs, to combat the faulty assumption that NESs are necessarily fluent in AE (Kim & Billington, 2016).

In fact, NESs' prosody in AE has been a source of miscommunication. In a Federal Aviation Administration (FAA) report on international pilot flight language experiences, pronunciation, accent and speech rate are cited as the primary sources of misunderstanding between NES pilots and controllers (Prinzo et al., 2010d). Although AE vocabulary and grammar are based on English, they are sufficiently modified to make learning this jargon necessary, regardless of native language. AE has evolved into "… probably the world's most successful semi-artificial international language" (Robertson & Johnson, 1988). AE, as does any other language, has a distinct prosody, emergent from its particular structure and environment.

To native speakers, language is so recognizable that infants can distinguish their native prosody from that of an unfamiliar language (Nazzi, Bertoncini, & Mehler ,1998; Ramus, 2002). Not only is prosody one of the first aspects of language that humans recognize, it is one of the hardest to acquire when learning a second language. Several studies have found that even advanced and early learner NNESs differ prosodically from NESs, especially in function word reduction, stress patterns and speech rate (Aoyama & Guion, 2007, Flege, 1987; Guion, 2005; Guion, Flege, Liu & Yeni-Komshian, 2000).

2.1.2. Elements of Prosody

Prosody encompasses a wide range of factors, including intonation, rhythm, and speech rate. These elements have been reported as being noticeably different in AE and conversational English: “Lack of intonation, rhythm and pauses is typical of rapid aviation radio communication and particularly problematic because no visual cues are present.” (Estival, 2016, p. 48). Because my study describes AE speech production, which rarely includes pauses, I have chosen to limit my examination to intonation, speech rate and rhythm. These three measurable elements can be combined to create a prosodic profile of AE. Using conventional methods to measure each of these phenomena, the current study will generate a description of AE prosody that can be directly compared to SAE and other languages.

To analyze intonation, I examined its phonetic correlate, which is pitch, or the quality of speech that creates the percept of highness or lowness of voicing. In Standard English, specific pitch contours are used to convey meaning. For example, the same sequence of similarly stressed words can convey doubt, interrogation, response or incomplete response, depending on how much and in which direction pitch changes throughout an utterance, or a continuous segment of speech (Pierrehumbert, 1980). AE is heuristically described as having a flat pitch contour (Philps, 1991), lending no cues to the meaning of an utterance. However, no studies have measured this phenomenon. The current study examined the validity of this observation, by measuring the pitch range of AE utterances versus SAE utterances. If pitch does not change measurably throughout an utterance, there may be fewer cues to meaning.

Rhythm is the temporal patterning of speech sounds within a language. Language rhythm influences the recognition of spoken language, especially in noisy conditions (Mattys, White & Mellhorn, 2005). A number of factors influence linguistic rhythm, including lexical and grammatical stress patterns as well as phonotactic constraints, or the patterns of consonants and vowels allowed in a language (Dellwo, 2006; Grabe & Low, 2002; Ramus, Nespor, & Mehler, 2000). For example, English allows complex clusters of consonants throughout a word (e.g. *strictest*), whereas other languages restrict the types of clusters allowed in many positions. For instance, Mandarin only allows a vowel or a nasal consonant in word final position. The phonotactic constraints of English result in specific patterns of temporal reduction (i.e., shortened duration) in fluent speech, including consistently reduced unstressed vowel intervals (i.e. *reduce*) and typically less reduced complex consonant clusters (i.e. *wellness*) (Dellwo, 2008). This patterning yields a wide variety of syllable types and, therefore a more varied rhythm than some other languages, like Spanish, which retain vowel length and primarily allow only single consonants, creating similar syllable durations throughout an utterance, resulting in a more uniform-sounding rhythm[1] (see Section 2.1.5). Although AE and SAE share lexical

---

[1] A conventional categorization of language rhythm is that of stress- or syllable-timing based on the permissible variability of consonant and vowel durations (Grabe & Low, 2002; Ramus et al., 2000). Languages like English, with large syllable inventory consisting of full and reduced vowels as well as a wide range of consonant cluster complexity, are typically considered to be stress-timed. On the other hand, languages like Spanish, that limit consonant clusters and allow less vowel reduction, have been considered syllable-timed.

items and phonotactic constraints, the grammatical and environmental differences in AE (i.e. noisy signal, multiple speakers, rapid speech rate, no visual contact) may influence prosodic properties differently. For instance, although semantically light function words like prepositions and articles are typically unstressed and reduced in English (i.e. *the dog*), for the sake of brevity AE was designed without these terms. Because there are fewer opportunities for vowel reduction in AE, patterns of vowel interval durations may be different than in SAE. These vowel reduction differences in AE may result in rhythmic patterns that are quantitatively different than those in SAE.

Another key element of prosody is rate of speech, or articulation rate when measured without pauses. This is a measure of number of speech units produced over time and is frequently quantified as the number of syllables per second. Results from several studies support the description of AE as faster than SAE. Previous research demonstrates average Air Traffic Controller speech rates[2] of 6 and 6.6 syllables per second (Farris & Barshi, 2013 and Prinzo et al., 1998, respectively). Studies of SAE show a greater variety of average articulation rates, in part due to variation in register, with averages in several studies ranging from 4.3 to 5.8 syllables per second (Dellwo & Wagner, 2003; Morgan & Fosler-Lussier, 1998; Williams & Stevens, 1972). However, AE and SAE articulation rates have not been directly compared using consistent standards and methodology within the same study. If a rate difference between registers is supported by the current study, it could motivate further analysis of how speech rate may affect other aspects of AE prosody, including rhythm, since rhythm and rate have

---

[2] It is not clear if these were articulation or speech rates. However, since Aviation radiotelephony discourages pausing, Aviation English rates may be considered comparable to articulation rates.

been shown to interact. For example, some cross-linguistic studies of rapid speech have found that vowels are shortened more proportionally than consonants as speech rate increases (Barry, Andreeva, Russo, Dimitrova, & Kostadinova, 2003; Dellwo, 2009; Gay, 1978), therefore the rapid production of AE may affect duration patterns. These differences with SAE could make AE challenging to learn for first or second language speakers of SAE.

2.1.3. Importance of Speech Rate and Rhythm in Language Comprehension

As a primary cue to word segmentation, rhythm is central to communication in any language (Cutler & Carter, 1987; Cutler, Mehler, Norris, & Segui, 1992; Mattys, Jusczyk, Luce, & Morgan, 1999). Importantly for Aviation English, because it is no one's native language, non-native speech timing is directly correlated to lack of comprehensibility in a second language (Tajima, Port, & Dalby, 1997). Tajima et al. (1997) showed that modifying second language production by superimposing native temporal patterns substantially improves intelligibility for native listeners. Rhythm is also a cue for native speakers of the same language. In Lang's (1975) study on the intelligibility of deaf speech, normal-hearing individuals' comprehension of arrhythmic deaf speech, complete with errors, was significantly improved after superimposing it onto a hearing speaker's temporal pattern.

Speech rhythm has also been shown to be more critical for intelligibility in noisy environments, when higher-level cues are not accessible (Mattys, White, & Melhorn, 2005; Smith, Cutler, Butterfield, & Nimmo-Smith, 1989). These lower level cues are even more important for NNESs for whom contextual cues are not as familiar (Bradlow & Alexander, 2007). I would expect rhythmic cues to be more important still in the AE environment, because in addition to it being noisy and (at least) half of AE speakers being NNESs, but there is no face-to-face contact and speech is very rapid.

Indeed, studies demonstrate that rhythm mitigates the effect of rapid speech rate on comprehensibility. Studies of compressed speech indicate that, as long as the context language is of a similar rhythmic profile to the tested language, listeners adapt quickly to rapid speech and maintain perceptual learning over periods of exposure to slower speech, different speakers, and different languages. These perceptual gains were not maintained, however, if intervening languages are of a different rhythm class (Dupoux & Green, 1997; Mehler, Sebastian, Altmann, Dupoux, Christophe & Pallier, 1993; Pallier, Sebastian-Gallés, Dupoux, Christophe, & Mehler, 1998). These findings imply that as long as speakers maintain a similar rhythm to surrounding speech, they can speak rapidly and be understood. However, if a speaker exhibits an unusual speech rhythm, it could reduce the speech perception abilities of even fluent users of a language. Because studies have shown prosody is acquired through exposure to language production (Pierrehumbert, 2003; Trofimovich & Baker, 2006), the importance of exposing AE learners to natural AE productions cannot be overstated.

2.1.4. Measuring Rhythm

In order to discuss rhythm in language, it must first be defined and then quantified. As described in Section 2.1.3, natural languages have historically been categorized as syllable or stress-timed in terms of units of equal duration either being inter-stress intervals or syllables. However, there was enough counter evidence to the isochrony argument to suggest that this was not an accurate classification, even though language rhythm could be classified as having a relationship to stress/accent. Dauer's

(1983) description of rhythm suggests instead that languages be situated on a spectrum from less to more stress-based, whereupon English is strongly stress-based and Spanish or French less so.

> Even though Spanish also has stress, the syllable structure, lexical composition, processes of reduction and phonetic realization of stress militate against the impression of regularity of interstress intervals. (Dauer, 1983, p. 59)

A number of metrics used to quantify rhythm do so by juxtaposing consonant and vowel durations. Vowels are generally the salient elements of speech, with consonants creating noisy gaps between them (Ramus, et al., 2000). Together, vowels and consonants create identifiable units for segmenting the speech stream (i.e. syllables). More stress-based languages are described as those, like English, that allow vowel length reduction as well as a wide variety of consonant cluster combinations. On the other hand, less stress-based languages like Spanish do not reduce vowel length and have a restricted inventory of consonant combinations, frequently utilizing a C-V syllable structure (White & Mattys, 2007).

To quantify these differences, a number of metrics comparing properties of vowel and consonant intervals have been proposed. Vowels (and, be default, consonants) have been characterized by the ratio of vowel duration to total duration of an utterance (%V; Ramus et al., 2000), the pairwise interval difference between sequential vowel segments, normalized for speech rate (VnPVI; Grabe & Low, 2002) and overall standard deviation of vowel duration, normalized for speech rate (varcoV; Dellwo, 2006). Consonants in turn have been quantified using the raw standard deviation of consonant interval durations (deltaC; Ramus et al., 2000), the pairwise difference between sequential consonant intervals (CrPVI; Grabe & Low, 2002) and overall standard deviation of consonant duration normalized for speech rate (varcoC; Dellwo, 2006). The current study utilizes these six rhythm metrics to compare AE and SAE.

Although substantial environmental and grammatical differences between AE and SAE exist, the AE lexicon is made up of SAE words. Therefore, these registers share not only lexical items but also phonotactic constraints. English phonotactic constraints result in specific patterns of reduction, including irreducible complex consonant clusters as well as reducible vowel intervals. English rarely allows deleting or centralizing of consonants; however, unstressed vowel segments are frequently reduced and centralized.

2.1.4.1. Predictions for Rhythm Metric Comparisons

Rhythm metric comparisons between AE and SAE may be affected by vocabulary and structural differences between these registers as well as differences in speech rate. SAE is described as a stress-based language because it contains reducible vowels and complex consonant clusters. As stated above, AE also complies with these phonotactic constraints. However, the content of these language varieties differs in meaningful ways. One important difference is in the use of function words. SAE commonly reduces the duration of unstressed syllables, including function words (Johnson, 2004). AE does not utilize function words, unless necessary for disambiguation, and therefore could show more regularity in vowel duration. As a result, AE rhythm metrics could be different from SAE and could exhibit metric values more consistent with those of less stress-based languages, in which vowel intervals are more uniform. Less vowel reduction in AE could also make vowel interval durations more similar, decreasing variability, in which case, VnPVI and varcoV would be lower for AE than SAE. If this is the case, %V may also be higher in AE than SAE. However, AE allows complex consonant structures found in

SAE, therefore AE and SAE may exhibit similar consonant variability and the resulting CrPVI, varcoC and deltaC measures could be similar in AE and SAE.

In addition, if speech rate is significantly faster in AE, metrics that are not normalized for rate, like CrPVI and deltaC, may be affected. Given the shorter duration of all segments in rapid speech and, therefore, less possible variability between segments (Dellwo & Wagner, 2003), standard deviation of AE consonant intervals might be lower in faster productions. If that is the case, CrPVI and deltaC may be lower for AE than Standard English. However, due to the articulatory properties of consonants, they cannot be compressed as much as vowels in rapid speech (Bertinetto & Bertini, 2008). As speech becomes very fast, the rate of decrease in consonant variability could lessen, so that AE could exhibit higher rate-normalized consonant variability (varcoC) than SAE. By examining these six rhythm metrics together, the prosody of AE can be described and compared to the prosody of SAE. In this study, two corpora of radio speech were used to compare the speaking rate and rhythm of AE and SAE.

To contextualize the discussion of prosodic differences between AE and SAE, it is important to note that the majority of pilots trained in the US learn AE standard phraseology in the aircraft while they are learning to fly, regardless of their native language. Some flight schools offer non-native flight students a classroom course in conversational English using aviation terminology; however, this is not the same as learning to converse in standard phraseology. This is problematic because the high cognitive load required to control an aircraft detracts from available resources for absorbing and responding to language input (Robertson & Johnson, 1988). This detriment is particularly onerous for NNESs (Farris, 2007). In the current study, I hope to illustrate that there are sufficient differences between AE and SAE prosody to suggest dedicated training in this register for all pilots. As stated above, language prosody is critical for intelligibility, which is necessary to flight safety in the complex and dynamic environment of international aviation.

2.2. Method

2.2.1. Corpora

Two corpora of US NESs were used to compare AE and Standard English, so regional differences could be minimized. The Air Traffic Control Complete corpus (ATC) was the source of the AE sample (Godfrey, 1994) and the Boston University Radio Corpus (BUR) (Ostendorf, Price, & Shattuck-Hufnagel, 1995) was the source of the SAE sample. These corpora were chosen for comparison because they are examples of fluent professional speech in specific registers of English. The recordings share the medium of radio, in which speakers have no visual contact with their interlocutors. Both corpora are of predictable speech--scripted in the case of radio broadcasters, and repetitive in the case of air traffic controllers--that requires clarity and mandates conveyance of specific information. One difference between these registers of English is radio broadcasters do not have interlocutors to provide feedback, but controllers and pilots do. However, AE responses primarily consist of repetition of key elements of a transmission, with little or no commentary as to comprehension of the message. Repetition is often automatic and, unless a request for clarification is made, the controller does not know whether a pilot has received a message accurately until the aircraft changes trajectory (or not). Most importantly, for the purposes of this study, the BUR corpus represented the standard of American English familiar to first and second language English speakers, whereas the ATC corpus represented a specialized register derived from that standard.

The ATC corpus consists of recorded aviation radio transmissions from three international airports in the US (Washington National, Boston Logan, and Dallas/Fort Worth). It includes air traffic controllers interacting with native and NNES pilots using AE. From these data, only NES US air traffic controllers' standard phraseology was examined. The choice to use only controllers' speech was based on the fact that professional controllers speak AE more frequently than pilots and may have more consistent, accurate productions than pilots, and thus could be used as a basis for a language model. Due to excessive noise in the signal for the Washington National airport data (caused by data collection technique), only the data from Boston and Dallas were included in the analyses. In comparison, the BUR corpus consists of recordings of professional, NES US radio announcers from a public radio station in Boston, MA, USA. All of the speech data were analyzed using conventions established in previous studies of rhythm and speech rate, as will be discussed in the relevant subsections.

2.2.2. Procedure

A similar amount of data from each corpus was analyzed (see Table 2.1), totaling 40.12 minutes of speech from 23 air traffic controllers (ATCOs) (3 females) and 41 minutes of BUR speech from 7 radio announcers (3 females). In order to facilitate comparison, pause-free utterances of similar duration were selected from the corpora.

Table 2.1.
Utterance descriptives by Corpus

Corpus	Utterance N	Number of words		Length in Syllables		Duration	
		Mean (SD)	Total	Mean (SD)	Total	Mean (SD)	Total
BUR	703	11.52 (5.47)	8,102	18.60 (8.78)	13,077	3.57 sec (1.46)	41.00 min
ATC	757	13.90 (6.34)	10,519	21.64 (9.37)	16,244	3.12 sec (1.45)	40.12 min

Because of the rapid, turn-taking style of aviation radiotelephony, every ATC transmission is brief and without intentional pauses. Accordingly, complete turns or transmissions were selected as ATC utterances. Infrequently, ATC utterances included disfluencies in the form of hesitations, repetitions or restarts. Disfluent segments were deleted, while maintaining the natural flow of speech in the remaining utterance, following the practice in previous studies (Dellwo, 2006; Grabe & Low, 2002; White & Mattys 2007). In contrast, the BUR data consisted of error-free narrative speech streams, which included pauses for breath and phrasing. Therefore, BUR transmissions were divided at audible breath pauses (greater than 100 ms) for use as BUR utterances. This division fortuitously generated pause-free BUR utterances with similar durations to ATC utterances (see Table 2.1), allowing for more accurate cross-corpus comparison.

2.2.2.1. Intonation

The first prosodic element compared between the corpora in this study was intonation. To quantify intonation, pitch range was calculated for each utterance. It is conventional to quantify the percept of pitch using the fundamental frequency (in Hz) of a voiced interval. Using Praat speech analysis software (Boersma & Weenink, 2015), a script was devised to locate the maximum and minimum fundamental frequency values and calculate the difference for each utterance. A visual depiction of each sound file with a pitch track overlay was reviewed to find cases where Praat was unable to track pitch accurately due to background noise. Cases for which it was not possible to accurately interpret pitch contours (< 3.5% of ATC and 0% of BUR data) were excluded from the analysis.

2.2.2.2. Rhythm Metrics

Several metrics comparing properties of vowel and consonant intervals were used to quantify rhythm of the corpora (see Table 2.2). In past studies, these metrics are typically examined in pairs, in order to assess both vowel and consonant behavior in a language. The most basic of these measures are %V and ΔC (Ramus et al., 2000). %V is the percentage of total utterance duration made up by vowels, which is generally lower in languages, like English, that allow vowel reduction. ΔC is the standard deviation of consonant intervals in an utterance, which is generally higher in languages with a wide variety of consonant combinations, like English. However, this overall utterance variability does not account for systematic localized stress patterns, therefore VnPVI and CrPVI (Grabe & Low, 2002) were calculated. VnPVI is the pairwise interval difference between sequential vowel segments, normalized for articulation rate and CrPVI is the raw pairwise interval difference between sequential consonant segments. Pairwise variability, as measured by VnPVI and CrPVI, is expected to be high for languages utilizing an alternating stress pattern, like English. Lastly, to account for articulation rate differences across registers, VarcoV and VarcoC (Dellwo, 2006) were examined for the corpora. These measures represent overall standard deviation of vowel or consonant interval durations, normalized for articulation rate over the entire utterance. VarcoV and VarcoC are expected to be higher in languages with a wide variety of vowel and consonant durations, like English.

Table 2.2.
Rhythm metrics' descriptions (for formulae, see Appendix A)

Metric	**Description**
%V	Percent vowel of total duration
ΔC	Raw overall consonant variability
VnPVI	Rate-normalized pairwise vowel variability
CrPVI	Raw pairwise consonant variability
VarcoV	Rate-normalized overall vowel variability
VarcoC	Rate-normalized overall consonant variability

No single metric, or pair of metrics has proven definitive in describing rhythmic differences between languages (White & Mattys, 2007), yet they each capture an aspect of temporal patterning that contributes to the prosodic profile of a language. Since there is a suspected articulation rate difference between AE and SAE, it is important to examine rate-normalized metrics like VnPVI, VarcoV and VarcoC. Additionally, non-rate-normalized metrics like %V, ΔC and CrPVI may help us understand how vowel and consonant reduction and language rhythm change with respect to articulation rate.

2.2.3. Possible Predictors of Rhythm Metrics

As described in the last section, the rhythm metrics used in this study are based on vowel and consonant durations in each of the corpora; therefore, cross-register acoustic factors may have an effect on these metrics, making it necessary to include them in an examination of differences between AE and SAE. Possible acoustic sources of variability in the calculation of rhythm metrics include articulation rate, utterance duration and amount of final lengthening. Speech rate has been reported to interact with rhythm (Barry et al., 2003; Dellwo, 2009), perhaps as a result of articulatory constraints (Bertinetto & Bertini, 2008) or differences in vowel and consonant reduction limitations (Gay, 1978). Utterance duration may also affect the calculation of rhythm metrics, since differences in duration could imply more or fewer segments to average variability over. Another acoustic factor potentially affecting rhythm metrics is final lengthening, whereby a long final vowel segment could disproportionately affect the variability of vowel duration in an utterance (Grabe & Lowe, 2002; Ramus et al., 2000). I calculated articulation rate, utterance duration, and final lengthening measures for each utterance and undertook analyses to examine each rhythm metric including these three acoustic factors and their possible interactions with the corpora.

2.2.3.1. Automatic Speech Aligner

To calculate vowel and consonant durations, each utterance sound file was paired with an orthographic transcription and aligned using Prosodylab-Aligner (Gorman, Howell, & Wagner, 2011), which created a phoneme-level TextGrid in Praat, dividing the utterance into phonemes and pauses. Examination of the phoneme tier revealed that the aligner improperly labeled some low amplitude word transitions as “small pauses”, due to noise in the signal. Therefore, a Praat script was run on all of the utterances to split “small pauses” of less than 100 ms evenly between adjacent phonemes as an approximation of the phoneme boundary (see Duez, 1985 and Hieke, Kowal, & O'Connell, 1983 for a justification of this pause duration). Labeled pauses greater than or equal to 100 ms in the data were aligned by hand. Occasionally, there were problems with ATC utterances being too long for the aligner to parse accurately, due to the sequential structure of the automated sound analysis. In these cases, ATC sound files were divided into separate utterances between aviation topics (e.g. altitude, wind, etc.). A Praat script subsequently divided phoneme tiers into consonant and vowel segments following established conventions. Praat scripts were run over the segmented TextGrids to generate the six rhythm metrics: %V, ΔC, VnPVI, CrPVI, VarcoV and VarcoC (see descriptions in 2.2). A single data point was created for each rhythm metric for every utterance.

2.2.3.2. Articulation Rate

In the current study, articulation rate was measured in syllables per second for each pause-free utterance in both corpora. Syllables were calculated using a syllable counting script (Kendall, 2013) in R (R Core Team, 2014). Duration was calculated in seconds from onset of the initial word to offset of the final word in the utterance, after removal of disfluencies.

2.3. Results

2.3.1. Pitch Range

An analysis of variance (ANOVA) indicated that pitch range was significantly smaller in the ATC corpus (M = 58.70 Hz, SD = 21.04) than in the BUR corpus (M = 133.06 Hz, SD = 31.64), (F (1, 24) = 48.50, $p < .001$) when comparing speaker means (see Table 2.3). Female speakers tend to have a larger pitch range than males (Henton, 1995); and since there are proportionally more males than females in the ATC than the BUR corpus, the effect of speaker sex was examined between groups as well. Pitch range for females in the ATC corpus (3 speakers) was larger than for males (F (1, 17) = 32.06, $p < .001$). This was true for the BUR corpus, as well. Pitch range for BUR females (3 speakers) was larger than for males (4 speakers) (F (1, 5) = 8.55, $p = .033$) (see Table 2.3 and Figure 2.1).

Table 2.3.
Pitch range by corpus

Corpus	Females Mean (SD)	Males Mean (SD)	Total Mean (SD)
ATC	96.99 Hz (4.01)	51.60 Hz (13.48)	58.70 Hz (21.04)
BUR	159.92 Hz (14.22)	112.91 Hz (24.58)	133.06 Hz (31.64)

Note: Hz = Hertz

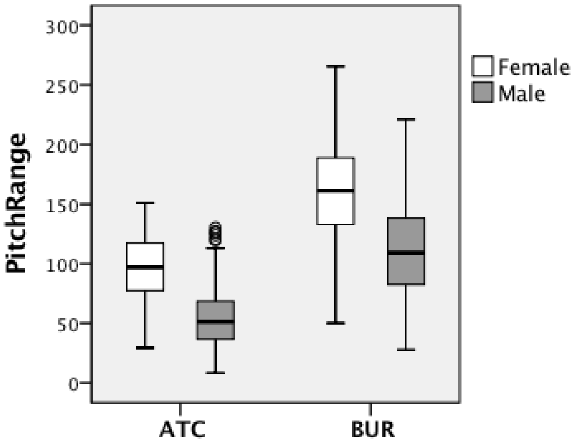

Figure 2.1. Pitch range by corpus and sex

2.3.2. Rhythm

2.3.2.1. Corpus Rhythm Metric Means Comparison

Mean values for all of the rhythm metrics described in Section 2.1 (see Table 2.2 for descriptions) proved to be significantly different between the corpora (see Table 2.4 and Figure 2.2). As predicted in Section 2.1.4.1, rate normalized vowel variability metrics (VnPVI and varcoV) were lower and overall proportion of vowel (%V) higher for AE than CE, possibly due to less frequent vowel reduction in AE. Raw variability metrics (CrPVI and deltaC) were lower for AE than CE, possibly due to faster articulation rate. However, rate normalized consonant variability (varcoC) was higher for AE than CE, possibly as a result of compression limitations at very high articulation rates.

To evaluate these rhythm differences between the corpora, mixed effects linear regression models were estimated for each of the rhythm metrics. The three acoustic measures described in Section 2.2.3, and corpus, were included as factors in the models.

Table 2.4.
Rhythm metric differences by Corpus (means compared using Welch's t-test)

Metric	Mean (SD)		*df*	*t*	*p*-value
	ATC (*n* = 757)	BUR (*n* = 703)			
%V	43.84 (4.87)	41.86 (6.44)	1304.9	6.58	< .001
VnPVI	54.58 (11.87)	63.21 (13.02)	1419	13.21	< .001
VarcoV	57.99 (11.91)	61.89 (14.00)	1383	5.71	< .001
ΔC	4.76 (1.37)	6.05 (1.61)	1385.1	16.42	< .001
CrPVI	51.61 (15.29)	67.83 (19.74)	1321.1	17.46	< .001
VarcoC	56.36 (10.49)	51.79 (10.48)	1450.1	8.32	< .001

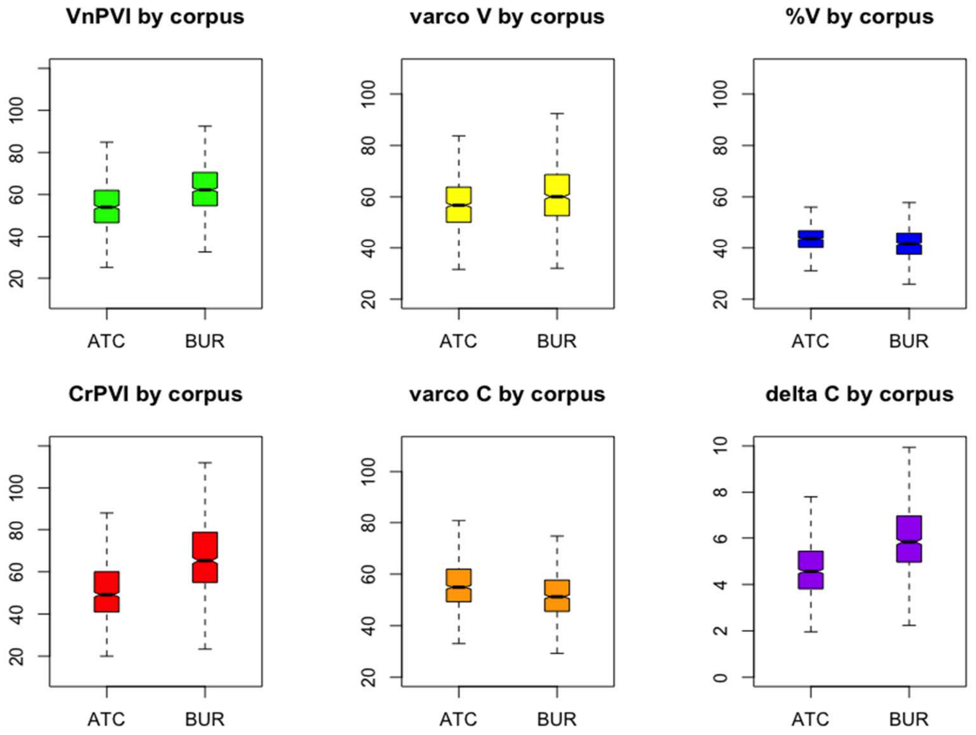

Figure 2.2. Rhythm metric means by corpus. Whiskers indicate standard deviations. Outliers are not depicted.

2.3.2.2. Rhythm Metric Acoustic Factors

Acoustic factors of articulation rate, utterance duration, and final lengthening were calculated for each utterance and proved to be significantly different between corpora (see Table 2.4). Articulation rate was significantly faster for the ATC corpus than the BUR corpus (*t* (1297) = 37.33, *p* < .001). Utterance duration was significantly shorter for ATC than for BUR (*t* (1400) = -5.49, *p* < .001). Final vowel lengthening was calculated as the ratio of the final vowel interval to the mean of all prior vowel intervals in the utterance and was lower for ATC than for BUR (*t* (1356) = -8.62, *p* < .001).

Table 2.5.
Acoustic factors by corpus

Corpus	Articulation Rate Mean (SD)	Utterance Duration Mean (SD)	Final Lengthening Ratio Mean (SD)
ATC	7.06 s/s (1.13)	3.12 sec (1.45)	1.26 (0.80)
BUR	5.22 s/s (0.72)	3.57 sec (1.66)	1.66 (0.98)

Note: s/s = syllables per second, sec = second

Although all of these factors were significantly different between corpora, their distributions overlapped (see Figure 2.2). Articulation rate had more distinct distributions and less overlap between the corpora than utterance duration or final lengthening did. Accordingly, an area of the overlapping articulation rates between the corpora was examined in addition to the full range of the data, to determine possible rhythm metric differences within the range of similar speech rates.

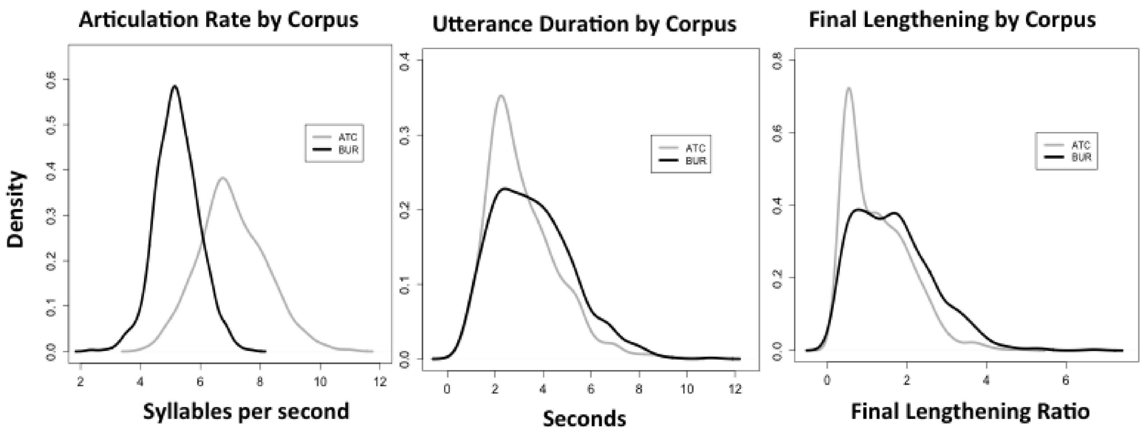

Figure 2.3. Density plots of articulation rate, utterance duration and final lengthening ratio by corpus.

2.3.2.3. Mixed Effect Regression Models of Rhythm Metrics

In order to determine if acoustic factors rather than corpus significantly predicted rhythm metric values for the entire range of data, linear mixed effects regression analyses were performed for each metric using the nlme package (Pinheiro, Bates, DebRoy, & Sarkar, 2014) in R (R Core Team, 2014). Independent variables included fixed effects of corpus, articulation rate, utterance duration, and final lengthening ratio. Interactions for corpus by each of the continuous variables of articulation rate, utterance duration and final lengthening ratio were included as fixed effects in the full models, as was the interaction of final lengthening ratio by utterance duration. The latter interaction (final lengthening ratio by utterance duration) proved not to be a significant predictor for any of the metrics and was removed from all of the models reported here. Continuous fixed effects of utterance duration and final lengthening ratio were log transformed and centered. Articulation rate was also centered. The full models included random intercepts for speakers, as well as random slopes for articulation rate-by-speaker and utterance duration-by-speaker. Mixed models included the maximal random effect structure that would allow the models to converge (specified in Table 2.5). A separate predictive model was generated for each of the six rhythm metrics. To determine significance of each predictor, model fit of the full model for each metric was compared to reduced models by subtracting one factor at a time. The models reported below include only significant predictors. In the case of a significant interaction, both factors are included in the model, regardless of individual significance. Collinearity among factors in each model was examined using variance inflation factors (VIF). No main effect had a VIF exceeding 2.5, suggesting that there was not substantial collinearity among factors included in the models, according to my personal assessment of this metric.

The articulation rate range-matched corpus subsets included data between 5.7 and 6.5 syllables per second and represented roughly 20% of each of the corpora (147 ATC and 149 BUR tokens). Sample size for this subset of data was too small to run mixed effects regressions. Therefore, means of corpus metrics were compared in this range.

2.3.3. Analysis of Rhythm Metrics

Corpus, or an interaction of corpus with an acoustic predictor, significantly improved model fit for five of the six rhythm metrics: %V, ΔC, CrPVI, VarcoV and VarcoC, but did not improve model fit for VnPVI. Models for each of the rhythm metrics, including all significant main effects and interactions, are described in Table 2.5. Notably, articulation rate improved model fit for four of the six metrics: ΔC, VnPVI, CrPVI and VarcoC, but did not improve model fit for %V or VarcoV. Below, vowel and consonant metric regression model results are described for the full range of data. The significance of main effects and interactions is reported in terms of chi-squared results comparing model fit with and without the fixed effect. Following the regression model descriptions, in order to control for speech rate effect, I compare rhythm metric values in the articulation rate range-matched corpus subsets.

All of the chosen acoustic factors had some effect on vowel and/or consonant rhythm metrics. Of these factors, articulation rate effects proved to be directionally consistent between the corpora, with corpus determining the degree of those effects. To the contrary, effects of the acoustic factors of utterance duration and final lengthening were directionally opposed for the two corpora.

2.3.3.1. Vowel Metrics: Regression Results

VnPVI results indicate pairwise vowel variability is lower for the more rapidly produced ATC corpus, meaning that adjacent vowels are more similar in duration in the ATC data than in the BUR data. Additionally, dissimilar effects of utterance duration and final lengthening on %V and VarcoV indicate that the ATC corpus and the BUR corpus exhibit different patterns of vowel durations.

Examining acoustic predictors individually, we see that articulation rate had a similar effect on vowel interval variability for the full range of data. Increases in articulation rate significantly predicted decreases in VnPVI for all the data (χ^2 (1) = 21.59, $p < .001$). It should be noted that VnPVI shows an effect of rate, regardless of the fact that the metric is normalized for articulation rate. This result is similar to previous findings (Barry et al., 2003; White & Mattys, 2007) and may be due to the fact that VnPVI is normalized for local articulation rate over each pair of vowel intervals and does not capture effects due to overall articulation rate or substantial variability in articulation rate across utterances (Barry et al., 2003).

Table 2.5.
Model summaries for linear mixed effects analysis of rhythm metrics

Predictor	*Estimate*	*Standard Error*	*t-value*
%V (Random effects: speaker intercept, rate-by-speaker, duration-by-speaker slopes)			
Intercept	43.8548	0.5385	81.44
Corpus(BUR)	-1.4188	1.2192	-1.16
Final Lengthening Ratio	-0.3703	0.2911	-1.27
Corp(BUR)xFLR	1.6599	4.4053	4.10
ΔC100 (Random effects: speaker intercept, rate-by-speaker slope)			
Intercept	5.36941	0.09744	55.10
Corpus(BUR)	-0.24733	0.18710	-1.32
Articulation Rate	-0.69995	0.05396	-12.97
Corp(BUR)xRate	-0.23289	0.10183	-2.29
Utterance Duration	0.19043	0.07017	2.71
VnPVI (Random effects: speaker intercept, rate-by-speaker, duration-by-speaker slopes)			
Intercept	57.8775	0.8410	68.82
Articulation Rate	-3.1960	0.5353	-5.97
CrPVI (Random effects: speaker intercept, rate-by-speaker slope)			
Intercept	58.4341	1.1423	51.16
Corp(BUR)	-0.9196	2.1816	-0.42
Articulation Rate	-7.1883	0.5600	-12.84
Corp(BUR)xRate	-3.7387	1.0548	-3.54
Final Lengthening Ratio	2.6234	0.8673	3.02
Corp(BUR)xFLR	-4.7837	1.2143	-3.94
VarcoV (Random effects: speaker intercept, rate-by-speaker, duration-by-speaker slopes)			
Intercept	58.1844	1.0360	56.16
Corpus(BUR)	3.7480	2.1336	1.76
Utterance Duration	3.2029	1.0377	3.09
Corp(BUR)xUttDur	-4.2399	1.4642	-2.90
Final Lengthening Ratio	-1.2979	0.7139	-1.82
Corp(BUR)xFLR	3.8587	0.9944	3.88
VarcoC (Random effects: speaker intercept, rate-by-speaker, duration-by-speaker slopes)			
Intercept	57.0397	0.7985	71.43
Corpus(BUR)	-5.5063	1.5637	-3.52
Articulation Rate	-0.7659	0.3862	-1.98

Note: Corp = Corpus, FLR = Final Lengthening Ratio, Rate = Articulation Rate, UttDur = Utterance Duration

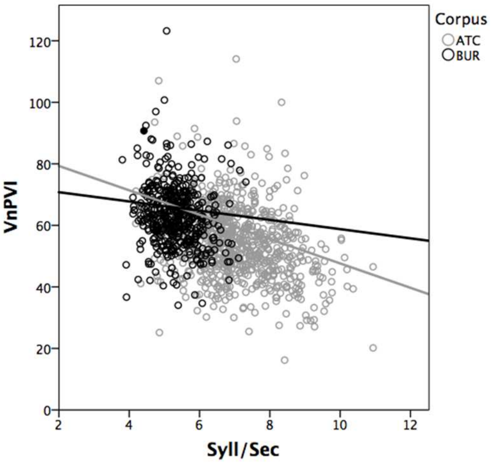

Figure 2.4. VnPVI as a function of rate only for each corpus. Individual data points and regression lines are depicted for ATC (gray) and BUR (black) corpora.

Unlike articulation rate, the acoustic predictor of utterance duration affected vowel metrics differently for each of the corpora. Increases in utterance duration significantly predicted increases in VarcoV for the ATC corpus, but predicted decreases in VarcoV for the BUR corpus (interaction significance: χ^2 (1) = 5.73, $p < .05$).

The acoustic predictor of final lengthening also affected vowel metrics differently for each of the corpora. Increases in final lengthening ratio significantly predicted decreases in %V for the ATC data, but predicted increases in %V for the BUR data. Corpus-by-final lengthening ratio interaction ($\chi^2(1)$ = 17.18, $p < .001$) and final lengthening ratio ($\chi^2(1)$= 5.78, $p < .05$) were both significant predictors of %V. Increases in final lengthening ratio significantly predicted decreases in VarcoV for the ATC corpus, but predicted increases in VarcoV for the BUR corpus (interaction significance: $\chi^2(1)$= 14.79, $p < .001$).

2.3.3.2. Vowel Metrics: Articulation Rate Range-Matched Data

Results indicate that AE vowel interval durations are more similar to one another than those of SAE at matched articulation rates. Because articulation rate affected vowel duration metrics in the current study, matched subsets were examined to control for rate effects. Within the selected overlapping range of articulation rates (5.7-6.5 syllables per second, see Figure 2.1), relative overall vowel duration was greater and vowel interval variability was lower for the ATC data than for the BUR data (see Table 2.6). Metric comparisons from these subsets reveal that %V was significantly higher for ATC than for BUR (*t* (294) = 4.04, *p* < .001), VnPVI was significantly lower for ATC than for BUR (*t* (293) = -2.952, *p* = .012) and VarcoV was also significantly lower for ATC than for BUR (*t* (275) = -2.11, *p* < .05) in the range-matched data. For differences between corpus metric values in the range-matched subset compared to values in the full range data sets, see Table 2.6.

Table 2.6.
Rhythm metric differences in articulation rate-matched subsets and full data set

Rhythm Metric	Rate- matched Subsets		Full Data Set		Full Data Set Metric Increase Factor
	ATC	BUR	ATC	BUR	
			Vowel Mean (SD)		
%V	42.99* (5.01)	40.61* (5.12)	43.84* (4.87)	41.86* (6.44)	Final Lengthening
VnPVI (rn)	56.55* (11.06)	59.92* (11.93)	54.58* (11.87)	63.21* (13.02)	Rate
VarcoV (rn)	58.54* (10.99)	61.70* (14.53)	57.99* (11.91)	61.89* (14.00)	Final Lengthening (BUR) Utterance Duration (ATC)
			Consonant Mean (SD)		
ΔC	5.33 (1.05)	5.28 (1.26)	4.76* (1.37)	6.05* (1.61)	Rate
CrPVI	58.20 (12.57)	58.89 (15.03)	51.61* (15.29)	67.83* (19.75)	Rate, Final Lengthening
VarcoC (rn)	56.50* (10.33)	50.42* (10.55)	56.36* (10.49)	51.79* (10.48)	Corpus, Rate

*Notes: (rn) = Rate-normalized Metric. * - Group values are significantly different, at p < .05*

2.3.3.3. Consonant Metrics: Regression Results

Regression results for consonant variability metrics indicate that the ATC corpus had higher consonant variability, which decreased less in faster speech for the ATC corpus than the BUR corpus. Additionally, the dissimilar effect of final lengthening on each of the corpora indicates that the ATC corpus and the BUR corpus exhibit different temporal patterning of consonant durations.

Examining regression model predictors of consonant rhythm metrics individually, membership in the ATC corpus significantly predicted higher VarcoC than membership in the BUR corpus ($\chi^2(1) = 10.31$, $p = .001$).

Similar to vowel regression results, increases in articulation rate reduced consonant variability for both corpora, although the degree of the effect differed by corpus. More specifically, increases in articulation rate significantly predicted decreases in VarcoC for both of the corpora ($\chi^2(1) = 3.87$, $p < .05$) and decreases in ΔC for all of the data ($\chi^2(1) = 68.67$, $p < .001$), although this effect was less for the ATC data ($\chi^2(1) = 4.88$, $p < .05$). Consistent with these findings, increases in articulation rate significantly predicted decreases in CrPVI for both corpora ($\chi^2(1) = 64.92$, $p < .001$), and this effect was less for the ATC data than for the BUR data ($\chi^2(1) = 10.78$, $p = .001$). When the effect of the corpus-by-articulation rate interaction on CrPVI is isolated, values for the faster ATC corpus are generally higher than for the BUR corpus in the same rate range (see Figure 2.4).

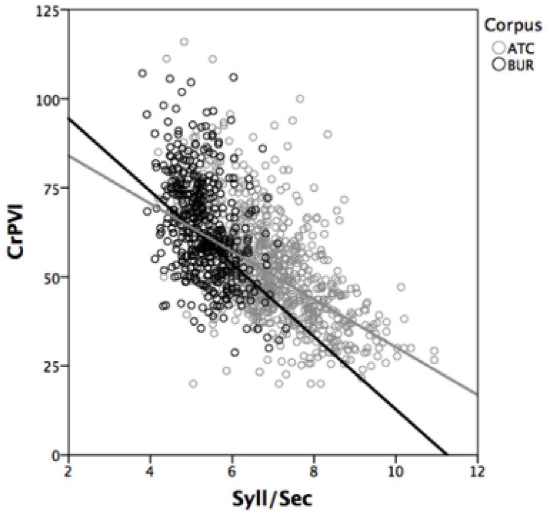

Figure 2.5. CrPVI as a function of rate only for each corpus. Individual data points and regression lines are depicted for ATC (gray) and BUR (black) corpora.

These findings suggest that there may be a lower limit to consonant reduction in both the SAE and the AE data. As articulation rates increase, vowels and consonants are reduced proportionally. However, when rate is increased beyond a certain point, consonants are no longer proportionally reduced and therefore variability in consonant durations will not decrease further. These findings indicate that AE has greater relative consonant duration variability than SAE with increasing articulation rate.

The acoustic predictor of utterance duration had the same effect on ΔC for all of the data. Increases in utterance duration significantly predicted increases in ΔC ($\chi^2(1)$ = 7.26, *p* = .01).

The effect of final lengthening on consonant metrics, similar to its effect on vowel metrics, was different for the two corpora. Increases in final lengthening ratio significantly predicted decreases in CrPVI for the BUR corpus but not for the ATC corpus (interaction significance: $\chi^2(1)$ = 15.56, *p* < .001).

2.3.3.4. Consonant Metrics: Articulation Rate Range-Matched Data

Analysis results indicate that AE consonant interval variability is relatively similar to SAE at similar rates of articulation; however, when normalized for articulation rate within this narrow range (5.7 to 6.5 syllables per second), AE has more variability in consonant interval durations than SAE. To control for rate effects on consonant rhythm metrics, articulation range-matched corpus subsets were examined. At similar articulation rates, raw variability was not significantly different between the corpora: ΔC was not significantly different between the ATC data and the BUR data (t (286) = 0.37, p = 0.71), nor was CrPVI (t (286) = -0.43, p = 0.67). However, rate-normalized overall consonant variability was higher for the ATC data: VarcoC was significantly higher for ATC than for BUR (t (294) = 5.01, $p < .001$) in this range.

2.3.3.5. Rhythm Metrics Summary

In summary, rhythm metrics indicate that AE and SAE exhibit different temporal patterning of vowel and consonant durations. AE vowel durations are more similar to one another and consonant durations are more variable than in SAE, as indicated by corpus effect on rhythm metrics over the full range of articulation rates, in addition to rhythm metric differences in articulation range-matched corpus subsets. Corpus interactions with utterance duration and final lengthening indicate further differences in vowel and consonant patterning between these registers of English.

2.3.4. Articulation Rate

Much of the reported difference in vowel and consonant interval durations was driven by articulation rate differences between the corpora. As described in Section 3.2.1, articulation rate was significantly faster for the ATC corpus than the BUR corpus. These values are similar to speaking rates reported for previous studies of AE (Farris & Barshi, 2013 and Prinzo &Lieberman, 1998, respectively) and Standard English (Dellwo & Wagner, 2003; Morgan & Fosler-Lussier, 1998; Williams & Stevens, 1972). Additionally, in the current study, the range of articulation rates for AE (4.19 - 10.94 syllables per second) is larger than the range of articulation rates for SAE (3.80 - 7.32 syllables per second) (see Figure 2.1). Since articulation rate has an effect on the majority of the rhythm metrics, the fact that articulation rate varies more in AE than SAE could play a meaningful role in their individual rhythm profiles. For example, the aforementioned English phonotactic constraints may affect rhythm metrics more at some articulation rates than at others, resulting in the differences outlined in the regression analyses above.

2.4. Discussion

This study describes a prosodic profile of Aviation English (AE), comparing it to Standard American English (SAE). Previous studies have shown that AE intelligibility is crucial to international flight safety, but none have described the sound system of this mandatory form of aviation communication. This study demonstrates that AE is prosodically distinct from SAE. AE has a more restricted pitch range, faster articulation rate, and, even when controlling for articulation rate, less variable vowel interval durations and more variable consonant interval durations than SAE.

Pitch range was examined as a measure of intonation. AE pitch range was smaller than that of SAE, supporting impressionistic claims from previous studies that AE lacks intonation (McMillan, 1998; Philps, 1991; Prinzo et al., 2011). One possible explanation for this restricted pitch range, as pointed out by Prinzo et al. (1998), is that professional air traffic controllers are trained to "eliminate emotion" from their interactions with pilots. This reference to emotion may be correlated with the observation that pitch is altered by the physiological changes that co-occur with emotional states (Williams & Stevens, 1972). The ICAO regulations directly indicate the goal of reduced intonation by suggesting that, in order to facilitate cross-cultural AE communication, NES pilots can "focus on keeping their intonation *neutral* (emphasis mine) and calm" (ICAO, 2010, 5.3.3.2.).

AE prosody also differs from that of SAE in articulation rate. Examination of the corpora indicated that AE is spoken faster than SAE. Rapid articulation rate is likely a reflection of the AE environment, which requires communicating a large amount of information to multiple interlocutors in time critical circumstances. The restricted, repetitive, predictable nature of AE enables users to speak much faster than speakers commonly do in SAE. The current study of naturally occurring speech examined rhythm metric trajectories over both a matched sample and a broad range of speaking rates, to better illustrate the interaction of articulation rate with rhythm in these data.

AE rhythm is distinct from SAE rhythm. Mean metrics of each corpus examined in this study, when plotted with data derived in other studies, indicates that AE (represented by the ATC corpus) and SAE (represented by the BUR corpus) are as different from one another as distinct languages that are described as rhythmically different, such as Dutch and Italian or British English and French (see Figure 2.X).

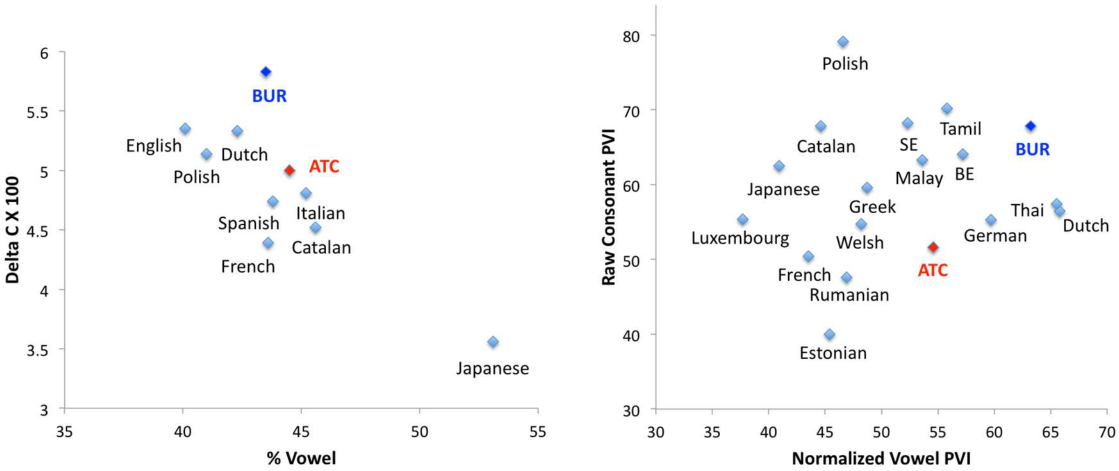

Figure 2.4. ATC and BUR metric means as compared to other studies (using data from Ramus, et al., 2000 (left) and Grabe & Low, 2002)

Examining the entire data set, regression models demonstrate that variability of vowel and consonant segments decreased with an increase in articulation rate for all the data. However, consonant variability decreased less in faster speech for AE than for SAE, a finding consistent with Dellwo (2009), who demonstrated that rate-normalized VarcoC is often correlated with speaking rate in SAE. Since the range of AE articulation rates extends far beyond that of the SAE data, the effect of articulation rate must take into account faster tokens for AE. Therefore, a smaller decrease in AE consonant variability may be the result of limited consonant interval reduction in rapid speech. Bertinetto and Bertini (2008) posited that consonant reduction in fast speech was restricted by articulatory constraints, since consonants typically involve more complex physical coordination of articulators. This restriction on consonant reduction in turn forces a decrease in vowel intervals that are not already maximally reduced, making vowel interval durations less variable (Bertinetto & Bertini, 2008; Dellwo & Wagner, 2003; Gay, 1978). The effect of this disproportionate reduction is that vowel durations become more similar and consonant durations more variable in AE.

Regression models for the full data set also suggest that cross-register factors other than articulation rate contributed to rhythm metric differences in the corpora. In particular, utterance duration predicted vowel variability for AE and final lengthening predicted vowel variability for SAE. These differences suggest that stress patterning is different in these two registers of English. The meaning of these differences must be determined through further study. However, at this juncture, it can be stated that there are corpus differences beyond rate that affect the prosodic profiles of AE and SAE.

Corpus emerged as a significant predictor for most of the rhythm metrics examined in this study, although the biggest rhythmic difference in the corpora and the most prevalent main effect was that of articulation rate. Therefore, much of the effect of corpus on speech rhythm was masked by the impact of articulation rate. For this reason, it was critical to examine an analogous portion of the corpora in which articulation rate was similar. The rhythmic differences in articulation range-matched corpus subsets suggest that AE has more similar vowel durations and more variable consonant durations than SAE. Indeed, it is apparent from the regression analyses on the entire data set that this pattern extends throughout the data. Vowel duration differences in the articulation range-matched corpus subsets could be explained by the fact that AE has fewer function words, and therefore, less vowel reduction than SAE. Less vowel reduction most likely leads to the lower vowel variability and higher percentage of vowel that were found for the AE data.

The prosodic profile of AE described above illustrates a rapid speaking style with restricted pitch range and temporally regular vowel intervals. These findings are consistent with heuristic descriptions of AE as “machine-gun”-like reported in previous studies (McMillan, 1998; Philps, 1991; Prinzo et al., 2011). Since many of the rhythmic differences between AE and SAE are due to articulation rate differences, it may be the case that AE is essentially fast Standard English. However, as demonstrated in the above findings, other acoustic factors affect the two corpora differently, creating a different rhythmic signature for AE that requires a separate description from SAE.

Cumulative differences in intonation, articulation rate and rhythm between AE and SAE indicate that all learners of AE, regardless of their native language background, will benefit from dedicated exposure to the prosody of this register of English. No doubt, most professional pilots and controllers will acquire fluency in this specialized jargon after years of practice in the field; however, the learning process could be greatly abbreviated by specialized instruction in standard phraseology, reducing the risk of lack of language proficiency contributing to an aviation accident. As AE learners, first and second language English speakers may have difficulty adjusting to the rhythmic differences in the rapid speech of AE, without the same durational stress cues for information parsing that they are accustomed to in Standard English. As speakers, novice AE users may not produce the proper prosodic patterns expected in this otherwise highly predictable language, causing time-delaying clarification requests and reiterations. This study's findings support claims from previous studies that AE prosody differs from SAE. Comparisons with natural languages indicate that these registers are as different as rhythmically distinct languages. Coupled with the flat intonation and fast speech rate of AE, these findings have implications for AE acquisition and performance. Instruction in conversational English may not suffice to prepare AE users to the standard of intelligibility necessary in radio communications.

Current international AE proficiency requirements create a challenge for many countries and airlines as they work to qualify personnel for continued work in aviation. However, this transition may be an opportunity to train pilots and controllers in the use of AE standard phraseology by exposing them to actual AE productions and acclimating them to its prosody. Past studies found that NNES (Kim & Billington, 2016) and NES (Prinzo et al., 2011) pilot observations suggest that regulators cannot assume NES English fluency entails AE fluency. These findings, together with the current results, indicate that NES aviation professionals may also benefit from training specific to AE prosody.

2.5. Conclusion

The current study demonstrates that the prosody of AE is quantifiably different than the prosody of SAE. AE has a more restricted pitch range, faster articulation rate, more uniform vowel intervals and more variable consonant intervals than SAE. Vowel and consonant variability in AE appear to be driven by an interaction between articulation rate and English phonotactic constraints as well as vowel reduction differences between the corpora. The combination of these rhythmic factors and differences in intonation and articulation rate make the prosody of AE unlike that of SAE, which has implications for segmentation and comprehension. These results predict a difference in AE performance. To test this prediction, in the next chapter, I present a study examining the ability of NESs to listen to and repeat AE.

CHAPTER III

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

This chapter uses portions of an article co-authored by Eric Pederson and Melissa Baese-Berk. Data collection was performed by Julia Trippe or under her guidance by undergraduate linguistics students Lathrop Hughes and Drew McLaughlin. Data analysis was performed by Julia Trippe. This chapter was written entirely by Julia Trippe, with Eric Pederson and Melissa Baese-Berk providing editorial assistance.

3.1. Introduction

Aviation English (AE) is the mandatory language for pilots and air traffic controllers (ATCOs) at international airports, if they do not share a first language. Proficiency in AE and conversational English (CE) are required by the International Civil Aviation Organization (ICAO), yet little is known about how AE and CE interact in language learning and usage. Requiring proficiency in both these varieties of English may be a burden on practitioners and increase miscommunication. ICAO requirements as well as AE training conventions generally assume that CE proficiency aids in AE proficiency. The current study examined AE intelligibility differences between native English speaking (NES) pilots and NES non-pilots. I use the term intelligibility here not in the narrow sense relevant to the literature, but in its most basic sense, that is, as a measure of speech recognition. This gets to the root of my research question as to whether CE users can understand AE. 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 flight safety.

3.1.1. Aviation English Description

AE is a variety of radiotelephony developed to convey critical information between pilots and ATCOs. Although AE includes both standard phraseology and “plain English”, in this study the term Aviation English (AE) is used to denote only standard phraseology and “plain English” is referred to as such. The following is a brief summary of my previous description of AE in Section 1.2.1. In order to maximize clarity and brevity in the AE environment, with multiple speakers, signal static and reduced frequency range:

- Word and phrase lexicon is severely restricted
- Function words are omitted unless absolutely necessary
- Intelligibility is enhanced by altering pronunciation (e.g. *five = fife, nine = niner*)
- Ambiguity is resolved by restricting every lexical term to a single meaning
- Aviation topics have different lexical topic identifiers (e.g. *heading*) and specific number expressions (e.g. *heading two fife zero*)

AE speakers transmitting on multiple-user radio frequencies risk being interrupted. Therefore, AE messages are produced rapidly, conveying several aviation topics in one transmission. However, the majority of pilot/ATCO interactions are predictable commands and responses that can be anticipated depending on aircraft position and phase of flight, enabling experienced AE users to select from the constrained AE phraseology to interpret transmissions. The vast majority of AE communications, though information-dense and rapidly produced, are understood and acted upon appropriately (ICAO, 2004). However, when communications go awry confusion can ensue, potentially affecting other aircraft in the vicinity and precipitating dangerous situations.

3.1.2. Regulating Aviation English

High loss-of-life accidents caused in part by communication problems (Cookson, 2011; Cushing, 1994; Jones, 2003) compelled the 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. For a detailed description of AE regulations, see Section 1.2.3. In brief, international requirements and recommendations are vague. Emphasis is placed on communicating in non-routine situations via “plain English”, which is not defined (ICAO, 2004). As a result, NESs are able to communicate relatively freely, alternating between AE and CE as they perceive the need to do so. However, CE is more ambiguous and nuanced than AE, causing miscommunication between NESs (Cushing, 1994) and NNESs (Kim & Elder, 2009). NNESs, in particular, are taxed by the additional cognitive load of translation in what is usually a stressful situation (Estival & Molesworth, 2016; Farris, et al., 2008). Studies of AE radiotelephony show that deviations from standard phraseology often result in confusion and delay (Day, 2004; Howard, 2008; Moder, 2012).

3.1.3. Testing and Training of Aviation English

3.1.3.1. Testing

ICAO regulations stipulate a mandatory level of competency in AE and CE but give vague guidelines as to how to test for that proficiency (see Section 1.2.4). In the US, the Federal Aviation Administration (FAA), which licenses and regulates pilots and ATCOs, simply requires a statement by the licensee that they “read, speak and understand English”. Prior to working as pilots or ATCOs internationally, NNESs that come to the US to train must undergo AE testing. NESs licensed in the US do not have to meet this requirement. Since the US is a member of the United Nations, they are expected to comply with ICAO regulations, but the FAA deems NES status sufficient to prove AE proficiency (Farris, 2016). My study is intended to challenge the assumption that NESs are implicitly proficient in AE.

3.1.3.2. Training

The standard for AE training has long been that radiotelephony is learned simultaneously with flight training. Although student pilots may be exposed to the published form of AE standard phraseology, little or no ground instruction is dedicated to AE. The first time pilots hear AE is often in flight while they are mastering aircraft control. If AE is not intelligible to non-pilot NESs, dedicated language training is necessary prior to flight training.

Many flight-training programs for NNESs include AE training courses designed by English-language teaching experts in consultation with aviation professionals. However, these courses mirror conversational English (CE) classes by focusing on face-to-face CE communication with aviation terminology taught by non-AE speakers, while in actual flight conditions, pilots must interpret rapid AE transmissions through static and reduced frequency range, without seeing the speaker.

3.1.4. Aviation English Intelligibility

A common determiner of whether or not languages are distinct from one another is if they are mutually intelligible. In the case of AE and CE, this determination is particularly important, since there is a tacit assumption in program design and regulations that CE users will be able to perceive and produce AE. Although AE standard phraseology vocabulary and grammar are based on English, they are sufficiently modified to make learning this jargon necessary, regardless of native language (Hansen-Schirra & Maksymski, 2013).

> The fact that the Aeronautical English is based on natural English does not mean that native speakers of English are released from 'learning' it. (Borowska, 2017a, p.139)

To further understand the relationship between AE and CE, it is necessary to determine if native CE users, not versed in AE, can understand AE and if fluent AE users, not native in CE, can understand CE. The current AE listening and repeating study addresses the first of these questions, examining the differences between NESs, with and without AE experience, repeating actual ATCO transmissions. If AE is intelligible to CE users, then relying on CE proficiency for AE use would be adequate. If AE is not intelligible to CE users, both native and NNES AE users would benefit from specific training in AE.

3.2. Method

3.2.1. Participants

Two groups of NES participants volunteered 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), age range 19-31 years. These undergraduates received class credit for their participation. Pilots were drawn from Lane Aviation Academy and Hillsboro Aero Academy in Oregon. Total pilot population consisted of 23 licensed pilots (4 female) with a mean age of 28.30 years (*SD* = 7.77), ranging from 19 to 55 (*median* = 26), and flight hours (overall piloting experience) ranging from 67 to 7000 (*median* = 350), including 4 to 2500 hours under Instrument Flight Rules (higher level of flight experience) (*median* = 56). Pilots held ratings ranging from Private Pilot to Air Transport Pilot (ATP), (indicating level of responsibility and proficiency) (see Table 1). All study participants self-reported having no hearing deficits.

Table 3.1.
Participant pilot ratings

	Visual Flight Only		Instrument Flight Qualified		
	Single-engine	Multi-engine	Single-engine	Multi-engine	ATP
Private	7	-	3	-	-
Commercial	-	-	1	10	2

3.2.2. Study Design

Intelligibility has been described as the ability to understand which words are produced, as opposed to comprehensibility, which is the ability to access word or utterance meaning (Munro & Derwing, 1995). The current study was concerned with AE intelligibility only, since non-pilots would be at an obvious disadvantage in measurements of AE comprehensibility. Therefore, the study was designed to measure correct word identification, not meaningful interpretation of the phraseology.

Most intelligibility literature describes listeners writing down what they hear (Munro & Derwing, 1995). I deemed this method inappropriate for the current study, because I wished to use these data to compare individuals from different language backgrounds, for whom transcribing would add a level of complexity. Accordingly, intelligibility was assessed via verbal, rather than written, responses to facilitate ease of response regardless of first language, enabling future comparisons with NNES populations. Verbal repetition also mirrors real world pilot/controller communication and therefore facilitates evaluation of actual usage. With this in mind, participants were asked to repeat recordings of actual ATCO transmissions of AE standard phraseology. In order to determine raw AE intelligibility, participants were not given feedback or training. This direct evaluation of AE intelligibility established a baseline for future comparisons with a broader pool of listeners from various language backgrounds. Analysis of NESs' results can also help determine what AE words and phrases are harder to understand and therefore need emphasis in AE language training. It is understood that this methodology requires an element of working memory, perhaps more than written responses do. I acknowledge that and have included a measure of working memory in the experiment design (see next paragraph). Additionally, I added a working memory task to the study to enable me to determine if there was a correlation between these measures.

Past studies (Barshi & Healy, 2002; Farris & Barshi, 2013) indicate that subjects show a sharp decrease in navigational performance when AE-style transmissions include more than three topics. Using this finding as an approximation of working memory limitations, and taking into account AE novelty for non-pilots, participants were given no more than two aviation topics in a single transmission (see Table 3.2). Additionally, the AE repetition task was designed to enable comparison between participants' performance responding to one- or two-topic transmissions. Half (42) of the selected ATCO transmissions consisted of one aviation topic and half contained two topics. Equal numbers of one- and two-topic transmissions were chosen from each of 22 (3 female) native American English speaking ATCOs. After removing transmissions that did not meet the selection criteria described below, each ATCO speaker contributed three or four transmissions, which ranged in length from two to 19 words.

Table 3.2.
Sample one- and two-topic transmissions

Type	Transmission
One-topic	*[descend and maintain three thousand]*
Two-topic	*[traffic no factor] [turn right heading two zero zero]*

Speech rate might also contribute to intelligibility. Studies have shown that speakers adjust to speech rate differences quite rapidly. Experiments in multiple languages show that listeners learn to understand compressed conversational speech after a brief period of exposure (5-10 sentences), (Dupoux & Green, 1997; Mehler, et al., 1993). Although no such studies have been performed on AE, AE has been described as being faster than CE (Farris & Barshi, 2013; Prinzo, Lieberman & Pickett, 1998). In order to assess learning over the course of the AE task, which may indicate participants adjusting to speech rate, transmissions were pseudo-randomized so that every dozen transmissions included an equal number of one- and two-topic transmissions, allowing comparison over seven sets of twelve transmissions. Eight pseudo-randomizations were distributed equally to participants.

Since the AE verbal repetition task could potentially be influenced by a participant's working memory capacity or their basic English language skills, each participant was also given a working memory (WM) task and a CE task. These tasks were designed to elicit verbal responses, similar to the AE task, so that similar skills could be measured. The need for comparable data between multiple populations dictated that the tasks be given in the same order, rather than randomized. In this way, any learning effect from task order would equally benefit all participants. The study always started with a WM task followed by a CE task, to allow participants to exercise listening and repeating skills prior to doing the AE task.

3.2.3. Procedure

All participants, both pilots and non-pilots, performed two intelligibility tasks and a WM task. They all started with a 15-minute verbal WM task to establish baseline memory 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 understood AE transmissions. Tasks were administered by computer using Psychopy software (Peirce, 2007) and were self-paced.

Participants completed language background questionnaires reporting other language and/or professional radio experience, since ability to adjust to AE radiotelephony might be aided by exposure to different languages and/or other forms of radiotelephony. Participants with college-level foreign language experience were coded as having language experience. Participants reporting any professional radio experience (other than aviation experience for pilots) were coded as having radio experience. However, neither of these measures proved to be a significant factor in any of the later analyses.

3.2.3.1. Working Memory Task

Working memory (WM) was evaluated using the Word Auditory Recognition and Recall Measure (WARRM) (Smith, Pichora-Fuller & Alexander, 2016) task, which required participants to repeat Standard English audio stimuli from one female native North American-English speaker. Five groups of five sets of monosyllabic words were presented, starting with sets of two words and increasing incrementally to six words in the final group of five sets (see Appendix B for sample word set). Subjects heard, “You will cite *target word”*, and were expected to repeat the target word and then say “first” or “second”, depending upon the half of the alphabet the initial letter fell in. After repeating this pattern for all words in the set, participants were prompted by an audible beep to repeat the target words (see Appendix B for full instructions). Trained lab technicians tallied participants’ responses for later analysis. Three balanced sets of the same 100 words were evenly distributed amongst participants.

3.2.3.2. Conversational English Intelligibility Task

The second task was a CE intelligibility task in which participants repeated ten sentences verbatim. Their responses were audio-recorded for later analysis. Sentences were selected from Harvard Sentence recordings in the online Open Speech Repository (Open Speech Repository, 2016). These sentences are approximately fifth grade reading level, phonetically balanced for SAE, from seven to ten words long. Two sentences from five NESs (2 female) were presented. A sample of static (averaging 52.27 dB) taken from an ATC recording was added to CE recordings, in order to replicate the noise level in the AE task (*M* = 53.59 dB, *SD* = 4.95). CE sound files were then normalized for intensity (70 dB). The ten sentences were randomly produced within each reproduction of the task.

Speech rate for normal CE in previous studies has been calculated to be on average 4.3 to 5.8 syllables per second (Dellwo & Wagner, 2003; Morgan & Fosler-Lussier, 1998; Williams & Stevens, 1972). The more carefully produced short CE extracts in the current study averaged 3.41 syllables per second (SD 0.67, range 2.7 - 4.5).

3.2.3.3. Aviation English Intelligibility Task

The third verbal repetition task was an AE intelligibility task in which participants repeated actual ATCO standard phraseology transmissions from the Air Traffic Control Complete corpus (Godfrey, 1994). Eighty-four pre-recorded transmissions were selected based on speaker, number of topics and terminology.[3] Participants were instructed to repeat verbatim what they heard. Their responses were audio-recorded for later analysis.

To maintain ATCO production integrity, each transmission was presented in its entirety. However, since ATCO transmissions are initiated by stating the call sign of the aircraft (e.g. *United ten ninety-six*) prior to issuing commands, study participants were asked to repeat only the commands. Text of the call sign, followed by an ellipsis, was briefly depicted on the computer screen during ATCO transmission presentation to help participants adjust to new speakers and remind them not to repeat that portion of speech.

Speech rate for the AE productions averaged 6.61 syllables per second (*SD* = 1.18), ranging from 4.77 to 9.77. These values are similar to averages of 6 and 6.6 syllables per second rates reported for previous studies of AE (Farris & Barshi, 2013 and Prinzo et al., 1998, respectively).

---

[3] Any transmissions including words with aviation meanings judged to be potentially confusing to the naive listener (e.g. *localizer*) were excluded from the study. Otherwise, the resulting transmissions were selected to represent as wide an array of terminology as possible.

3.2.3.4. Verbal Repetition Task Scoring

WM task scoring. WM task was scored as described in the WARRM methodology (Smith et al., 2016) on a scale ranging from 2 to 6 points, reflecting the number of words the participant was able to remember consistently[4]. This score was then multiplied by 16.67 to make the highest possible score 100.02, which is comparable with percentage scores for the other tasks. For example, using this method, a base score of 4.67 was transformed into a WM task score of 77.83.

Intelligibility task scoring. Scores for the CE and AE audio intelligibility tasks were the percentage of words correctly reproduced of the possible words in the entire stimuli set. The CE task consisted of ten sentences with a total of 83 words. The AE task consisted of 84 ATCO transmissions with a total of 732 words.

Words for both language variety intelligibility tasks were scored correct if they were the exact word in the same order relative to other words in the stimuli (see Table 3.3). Two trained lab technicians and the first author transcribed participants' responses and the first author scored all of the participants. Inter-coder reliability tests resulted in 99.0% transcription agreement and 98.3% scoring agreement.

---

[4] The task consisted of different levels each with five sets of words starting with two words up to six. A level was considered successfully completed if the participant repeated at least three out of five sets. They scored two points for completing the first level of two words and one point for each following completed level. If only one or two sets were repeated, the participant scored .33 or .67 for that level, respectively. Thus, if a participant correctly repeated at least three sets of two words, three sets of three words, and three sets of four words and then repeated two sets of five words, they were given a score of 4.67.

Table 3.3.
Sample points for participant response

Stimuli	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
Score									66.67%

During the AE task participants occasionally (~ 0.3% of responses) advanced through the program too quickly and accidentally skipped over ATCO transmissions. These occurrences were not included in score averages. Overall, there were 12 of these incidents: five within the pilot group and seven from non-pilots. Of these errors, eight were one-topic and four were two-topic transmissions.

3.3. Results

3.3.1. Verbal Repetition Task Scores by Group

As would be expected, WM and CE task results indicate that these populations were matched in language and memory abilities. Additionally, pilots exhibited better AE intelligibility scores than non-pilots (see Figure 3.1).

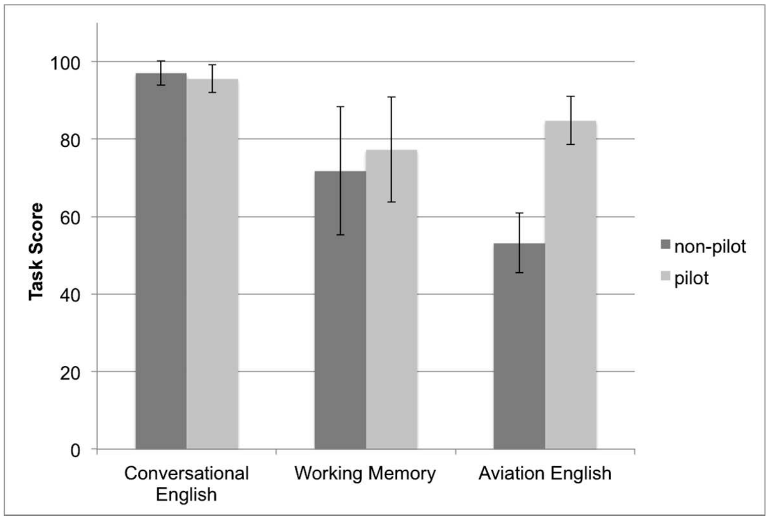

Figure 3.1. Average task scores by group. Error bars are +/- 1 standard deviation

A two-way repeated measure analysis of variance (ANOVA) was conducted on the effects of two independent variables (group and language variety) on intelligibility score. Group was a between-group factor including two levels (non-pilot, pilot) and language variety was a within-group factor with two levels (CE, AE) (see Table 3.4). Pilots' intelligibility scores were higher than non-pilots' scores (*F* (1, 94) = 129.02, *p* < .001), but this effect was due entirely to pilots having higher scores on AE than non-pilots. In general, CE score was greater than AE score (*F* (1, 94) = 475.34, *p* < .001), for non-pilots (*t* (30.13) = 21.365, *p* < .001) and pilots (*t* (35.06) = 7.2374, *p* < .001). However, there is an interaction between language variety and group (*F* (1, 94) = 156.43, *p* < .001) such that intelligibility of CE does not differ between pilots and non-pilots (*t* (44.12) = 1.52, *p* = 0.14), but pilots scored higher on the intelligibility of AE (*t* (46.69) = -15.81, *p* < .001). A two-sample t-test showed that WM scores were not significantly different for non-pilots (*M* = 71.93, *SD* = 16.74) and pilots (*M* = 77.79, *SD* = 14.02) (*t* (46.77) = -1.27, *p* > .05).

Table 3.4.
Conversational English and Aviation English intelligibility scores by group

Group	*N*	CE (*SD*)	AE (*SD*)	Mean
Non-pilots	26	97.01% (3.11)	53.20% (7.69)	75.76% (22.60)
Pilots	23	95.55% (3.55)	84.78% (6.19)	90.17% (7.38)
Mean		96.32% (3.37)	68.72% (17.30)	82.52% (18.60)

In order to determine if learning occurred over the duration of the task, AE intelligibility task score was calculated for every successive set of twelve responses for each participant. Using the Bonferroni correction in a pairwise comparison of the seven successive groups of transmissions, non-pilots showed a learning effect from the first to the third set of AE transmissions only. However, the second through seventh non-pilot

AE set scores did not differ significantly (see Table 3.5). Pilot AE task scores did not differ significantly over the entire range of AE sets.

Table 3.5.
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, $p < .05$, using the Bonferroni correction

3.3.2. Factors Predicting Aviation English Task Performance

A linear mixed effects regression was performed using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R (R Core Team, 2014) to create the best model fit of AE scores for all responses in the data. The regression accounted for random effects of specific transmission, order presented and individual participant. The full model included fixed effects of CE task score, WM task score, radio experience, language experience, age, sex, number of words per transmission, number of topics per transmission and interactions with group for each fixed effect. An examination of correlations among main effects predictably indicated collinearity between number of topics and number of words ($r^2 = 0.54$). These factors are intrinsically related, given that topics are made up of words. However, it is possible that participant groups could parse transmissions differently, either as simple word sequences or words grouped into topics. One way to examine this potential difference in parsing strategy is to examine the effects of number of words and number of topics. Accordingly, both factors were retained in the model (see Table 3.6). Collinearity among factors was examined using variance inflation factors (VIF). No main effect had a VIF exceeding 2.2, suggesting that there was not substantial collinearity among factors included in the model[5].

Table 3.6.
Linear mixed effects model summary of AE intelligibility scores by group

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
Pilot Group*Words	-0.84	0.23	13.02	< .001
Pilot Group*Topics	13.34	1.50	78.61	< .001

Note. Non-pilot group is default. Random intercepts for Subject, Transmission and Order were included in the model.

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. Other significant predictors of AE scores were number of words, and interactions of group by number of words and group by number of topics. Non-pilot scores decreased with number of words and were lower overall for two topic sentences compared with one topic. Pilot scores, on the other hand, decreased with number of words, but were unaffected by the number of topics (see Figure 3.2).

---

[5] There is no accepted rule for VIF indicating multicollinearity. My personal assessment is that a VIF over 2.5 indicates an unacceptable level of multicollinearity. In the case of these data, since VIF = $1/(1 - r^2)$, a VIF of 2.2 implies $r^2 = .45$, or ~45% of the variance in the IV is also explained by the other IVs in the model (Craney & Surles, 2002).

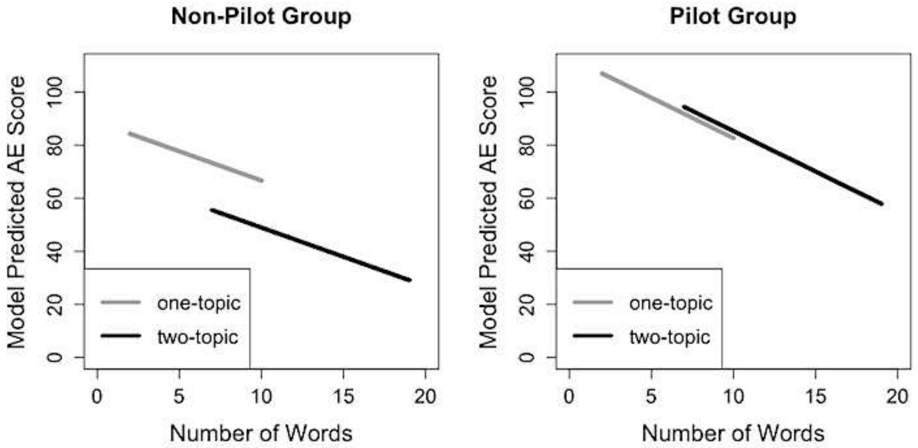

Figure 3.2. AE Intelligibility Score by group, number of topics, and number of words

When participant scores for both CE and WM tasks were included in the model, neither proved to be significant in predicting AE scores. However, when one or the other factor alone was included, they significantly contributed to the model (WM $\chi^2(1) = 4.94$, *p* = .026; CE $\chi^2(1) = 4.25$, *p* = 39). Additionally, their joint addition to the model significantly increased the fixed effects model fit from $R^2 = 0.45$ to 0.46 ($\chi^2(2) = 7.71$, p = 0.021), so both were retained in the model. VIF values for CE score and WM score in the model were 1.14 and 1.12 respectively, indicating lack of substantial collinearity between these factors.

3.3.2.1. Flight Experience Effect on Aviation English Scores

A separate regression was done on pilot group AE scores, to determine flight experience effect on AE score (see Table 3.7). Flight time was a more granular predictor of experience than pilot ratings (see Table 3.1), therefore the full model included factors from the above analysis in addition to total flight hours (TT) and Instrument Flight Rules hours (IFR). However, R^2 values for Age (0.84), TT (0.87) and IFR (0.76) indicated multicollinearity in these predictors. Therefore, IFR was chosen to best represent pilot experience with AE. Age and TT were removed from the model. IFR values were not normally distributed (*M* = 302.3, *SD* = 608.11, *median* = 56, *Range* = 4-2500), therefore IFR was natural log transformed for inclusion in the regression. Log transformation of IFR greatly improved distribution of this variable (*M* = 4.13, *SD* = 1.73, *median* = 4.03, *Range* = 1.39-7.82).

Table 3.7.
Linear mixed effects model summary of pilot AE intelligibility scores

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

*Note. Significance codes: .001 '***', .01 '**', .05 '*'. Random effects of Subject and Transmission were included in the model.*

Model fit determination using piecewiseSEM package in R (Lefcheck, 2015), gave a marginal (fixed effects) R^2 value of 0.27 and conditional (including random effects) R^2 value of 0.50. This model indicates pilots' AE scores were significantly predicted by number of words in the transmission and by flight experience. Pilot AE scores were not significantly predicted by number of topics ($\chi^2(1) = 0.83$, $p = 0.36$), CE task score ($\chi^2(1) = 0.12$, $p = 0.73$) or WM task score ($\chi^2(1) = 2.59$, $p = 0.11$).

3.4. Discussion and Conclusion

Results of this study indicate that, without training, AE is not intelligible to non-pilot NESs beyond a low threshold (53%) and acoustic learning of AE reaches its peak early at a level far below ceiling (~ 55%). Ceiling effects of conversational English (CE) scores, as well as similar WM scores amongst these two groups of NESs, make it difficult to analyze these factor effects on AE scores. However, given that these two tasks examine English speech perception and production, the scores would be expected to correlate and are indeed very nearly equivalent between the participant groups.

Examining the results, we can conclude that CE ability does not imply any appreciable AE proficiency. These findings are consistent with past studies. Moder and Halleck (2009) compared NNES ATCOs' abilities in different registers of English and found that their CE and AE performance were not correlated. Kim and Elder (2009) found that, regardless of CE fluency, NES pilots were often not comprehensible to NNES expert AE users when they abandoned AE phraseology in favor of "plain" (i.e. colloquial or technical) English.

Regression results indicate that, whereas number of words in a transmission is the primary factor in determining AE difficulty for both 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 (such as pilots) chunk information to efficiently interpret what they hear.

Pilot AE scores did not show an effect of learning over the course of the AE task; however, the regression model indicates that pilot AE proficiency is predicted by flight experience. Pilots' total hours under Instrument Flight Rules (IFR) had a logarithmic relationship with AE scores. This finding suggests the AE learning curve is steep for low time pilots and shallows out with experience. In these data, AE scores for pilots with less than 100 hours IFR averaged 83.26% correct (*SD* = 5.61) and those with more than 100 hours of IFR averaged 90.12% (*SD* = 6.07). However, this difference did not reach significance, presumably because the sample had an uneven distribution of IFR experience.

Current training protocols have pilots learn AE in flight, during which a small percentage of time is spent directly communicating with ATCOs and more time is spent listening to ATCOs interaction with other pilots. A training program combining active response and passive listening to recorded ATCO transmissions would expose pilots to both flight language experiences. Not only that, research shows that alternating passive listening with interactive listening enhances language learning (Baese-Berk, 2010; Wright, Baese-Berk, Marrone, & Bradlow, 2015) (see Section 6.2.3). This type of training protocol would enable pilots to dedicate their attention to learning AE in a focused language-learning environment, rather than having to allocate limited cognitive resources during flight training. After having acquired standard phraseology through language training and evaluation, pilots could enter the flight deck with AE skills in place.

In the current study, non-pilots did not improve AE intelligibility as quickly as previous studies of rapid conversational speech indicated they might, suggesting that AE is not comparable to fast CE and is particularly difficult for CE speakers to learn out of context. However, the demonstrated increase in low-time pilot AE scores with flight experience indicates that AE is learnable in aviation contexts. AE language is formulaic, employing a constrained lexicon and restricted phrase inventory, which makes AE easy to teach. If the focus in training were on topic identifiers -- words and phrases that identify the aviation topic of ensuing numbers -- novices may quickly learn how to recognize these rapidly produced chunks of language.

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 proficiency does not correlate with CE proficiency. The current study shows that AE is markedly less intelligible to CE speakers. Therefore, the assumption that CE proficiency automatically leads to AE proficiency is in doubt. Given the exhibited difficulty of understanding AE, it seems imperative to consider how to improve AE learning. 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.

In the current flight-training environment, NES pilots are not given any specific AE training and NNES pilot AE training focuses on CE. As a result, pilots may not be getting enough AE training before relying on it in flight. Instead of expecting NESs to learn as they fly and spending time and energy on CE training for NNESs, blocks of AE training tailored to specific phases of flight training could be integrated into the training protocol. 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 ATCO transmissions could enable pilots to effectively and confidently communicate in AE as soon as they get off the ground.

In the next chapter, I compare listening and repeating study results from NES pilots and NNES pilots. Differences in these population's performance will further develop the relationship between AE and CE performance.

CHAPTER IV

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

4.1. Introduction and Study Design

As stated in the previous chapter, 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 in Chapter 3 that native English speakers (NESs) not versed in AE are scarcely able to understand AE, it remains to be determined if fluent AE users non-native English speakers (NNESs) can understand CE. The current chapter addresses this proposition by examining NNES AE users' ability to understand CE. 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, as is the current practice, on CE. 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.

In order to establish NNES AE proficiency as compared to CE proficiency, a group of NNES pilots was given the identical oral tasks used in the previous study. Their results were compared to the NES pilot group from that study. As in the prior experiment all participants received the tasks in the same order and no feedback was given for any of the tasks.

4.1.1. Participants

All participants were pilots training or working in Oregon (with self-reported normal hearing) who volunteered for the study. The NNES pilot group was made up of Chinese flight students. The Chinese pilot (CP) group was made up of 29 (1 female) flight students training at Hillsboro Aero Academy, ranging in age from 22 to 26, with a mean age of 23.38 (*SD* = 1.08) (see Table 4.1). At the time of the study, all the Chinese flight students had been in Oregon for ten months flight training and held at least a US Federal Aviation Administration (FAA) issued private pilot license (see Table 4.2). The majority of the Chinese pilots (22) had Mandarin as their first language, all others were fluent in Mandarin, but had Southern Min (2), or another Chinese dialect (5) as their first language.

Table 4.1.
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 Pilots; EP = native English-speaking Pilots

Table 4.2.
Number of participants holding pilot ratings

	Private		Commercial		
Group	SE	SE Instrument	SE Instrument	ME Instrument	ATP
CP	18	11	-	-	-
EP	7	3	1	10	2

Note: CP = Chinese Pilots; EP = native English-speaking Pilots; SE = Single-engine; ME = Multi-engine

The NES pilot (EP) group was the same population used in the prior study (see Section 3.2.1). This group was made up of 23 (4 female) North American flight students and instructors from Lane Aviation Academy and Hillsboro Aero Academy in Oregon, ranging in age from 19 to 55 years, with a mean age of 28.30 (*SD* = 7.77) (see Table 4.1). The EPs also held at least an FAA private pilot license (see Table 4.2). Table 4.2 summarizes population descriptives: age, total flight time (TT), and instrument flight rules flight time (IFR).

4.1.2. Procedure

The same stimuli were used as in the prior experiment described in Chapter 3 (see Section 3.2.3). The following is a brief summary. Participants underwent three verbal repetition tasks, starting with a 15-minute verbal 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. Participants were allowed to take breaks, although none chose to do so.

Participants were also asked to complete a language background questionnaire reporting language experience. CPs reported number of years of English study in their home country as well as scores on the International English Language Testing System (IELTS) exam. CPs averaged 11.41 years (*SD* = 2.18) of English language training and scored on average 5.61 (*SD* = 0.32) on the IELTS exam. Students were required to have a score of 5 (out of a possible 9) to enter this program. Neither of these factors significantly predicted CP AE task score in linear regressions.

Working memory Task. Participants underwent a verbal working memory (WM) task to determine any correlation with 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.

CE intelligibility task. The second task was a CE intelligibility task in which participants were asked to repeat ten sentences verbatim (see Appendix C for instructions). The sentences are approximately fifth grade reading level, phonetically balanced for SAE, ranging in length from seven to ten words (see Appendix C for list of CE sentences).

AE intelligibility task. The third verbal repetition task was an AE intelligibility task in which participants were asked to repeat 84 ATCO utterances verbatim (see Appendix D for instructions). Half (42) of the selected ATCO transmissions consisted of one aviation topic and half contained two topics (see Appendix D for list of ATCO transmissions). Transmissions ranged in length from two to 19 words (see Table 4.3). 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. No feedback was provided during the task.

To maintain the integrity of ATCO productions, each transmission was presented in its entirety. However, since ATCO transmissions are initiated by stating the call sign of the interlocutor (e.g. “United ten ninety-six”), study participants were asked to repeat only the commands.

Table 4.3.
Number of words and items in AE transmissions

Number of Items	Mean (SD)	# Words/Item	Range
1	6.31 (1.66)	6.31	2 - 10
2	11.11 (2.63)	5.56	7 - 19

4.1.2.1. Verbal Repetition Task Scoring

Scoring proceeded as with the prior comparison, outlined in Section 3.2.3.4. What follows is a summary of that process.

WM task scoring. WM task was scored as described in the WARRM methodology (Smith et al., 2016). Possible raw scores ranged from 1.0 to 6.0, depending on number of words consistently remembered after performance of unrelated cognitive tasks. 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.

CE task scoring. Score for the CE task was the percentage of words correctly reproduced of the 83 possible words in the combined ten CE sentences.

AE task scoring. One trained lab technician and the first author transcribed participants’ responses. Inter-coder reliability tests of 5% of the data (206 out of 4116 transmissions) resulted in 99.0% transcription agreement and 98.3% scoring agreement. Response words were scored correct if they were the exact word form from the transmission in the same order relative to other words in the transmission. In the name of fairness, three ambiguous terms were scored at 1.5 points, with the correctly produced root given a point and the correct suffix 0.5 point. These terms were *cleared*, *traffic’s* and *departure's*. Percentage of words correct was determined for each response, each set of twelve responses, and all 84 responses combined for each participant. Departure's (these are contractions) cleared (yes), established (no).

Occasionally (~ 0.3% of responses), a participant advanced through the program too quickly and an ATCO transmission was accidentally skipped over. These occurrences were not included in score averages. Overall, there were 14 of these omitted trials: five from the EPs (2 of 84 trials from one subject, 1 of 84 for three others) and nine from CPs (3 of 84 trials from one subject, 2 of 84 for two others and 1 of 84 for four others). Of these, six were on one-topic (3 by CPs) and eight were on two-topic (6 by CPs) transmissions.

4.2. Results

4.2.1. Verbal Repetition Task Scores by Group

Group averages differed for all of the tasks. EPs performed significantly better on all of the tasks than CPs (see Figure 4.1). As would be expected, NES pilots scored higher on the CE task and on the WM task, since they were both English verbal repetition tasks. 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 4.1). These two task scores confirm expected differences in English proficiency.

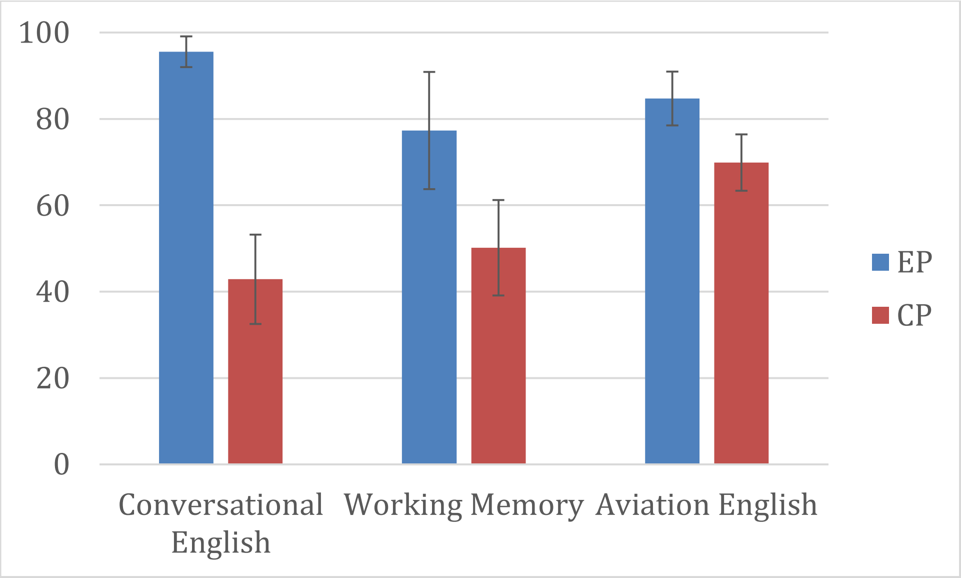

Figure 4.1. Average Conversational English, Working Memory and Aviation English task scores by Group: Chinese pilots (CP) and native English speaker pilots (EP). Error bars reflect Standard Deviations.

EPs also scored higher on the AE task than CPs. CPs' AE task scores averaged 69.46 (*SD* = 6.51), whereas EPs' AE task scores averaged 84.78 (*SD* = 6.19). As seen in Figure 4.1, CPs had less flight experience than EPs, as measured by IFR (Instrument Flight Rules) hours. Accordingly, in order to compare AE proficiency in a matched range of experience, group populations were balanced by mean IFR time (see Figure 4.2). The 16 EPs with the lowest IFR time (each with less than 76 hours) had a group average of 37.21 hours (*SD* = 25.54), which was not significantly different from CPs average of 37.19 (*SD* = 14.79) (*t* (20.69) = -0.003, *p* = 1.00). Figure 4.2 makes the comparison between groups more accessible by showing only task scores of CPs and low-time EPs. Although low-time EPs task score averages were slightly lower than high-time EP averages (see Table 4.4), none of these differences reached significance at *p* = .05 level.

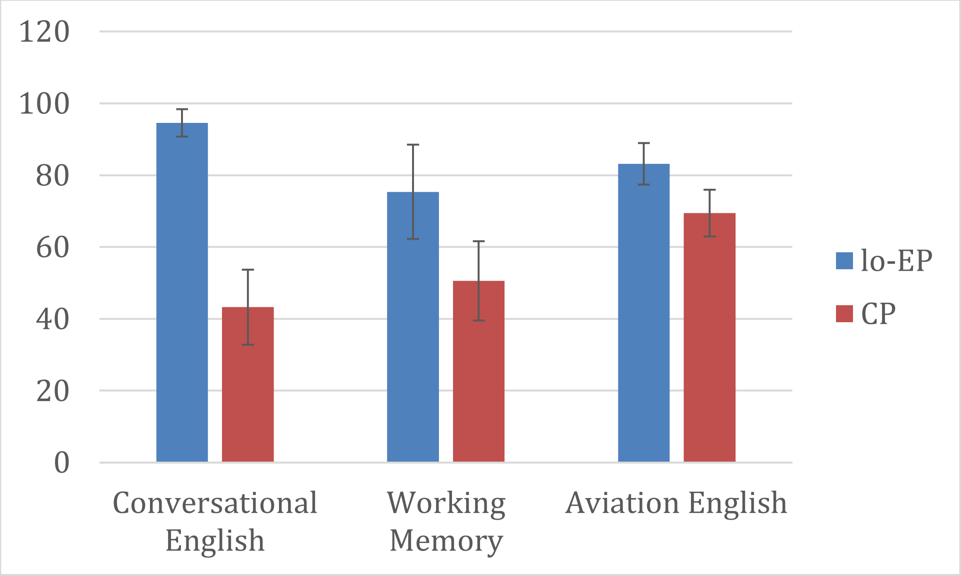

Figure 4.2. Average Conversational English, Working Memory and Aviation English task scores for similar experience levels of native English speaking pilot (lo-EP) and Chinese pilot (CP) Groups. Error bars reflect Standard Deviations.

Table 4.4.
Low-time native English speaking pilots (EP-Lo) and high-time native English speaking pilots (EP)

	Pilot Group					
	EP-Lo (N = 16)		EP-Hi (N = 7)		*significance test*	
Task	*Mean*	*SD*	*Mean*	*SD*	*t-value*	*p-value*
WM	75.36	13.14	81.75	14.60	*t* (10.48) = 1.00	*p* = 0.342
CE	94.58	3.84	97.76	1.08	*t* (19.36) = 3.05	*p* = 0.064
AE	83.10	5.87	88.46	5.82	*t* (11.46) = 2.01	*p* = 0.069

Notes: WM=Working Memory, CE=Conversational English; AE=Aviation English

Since task scores did not differ significantly between EP groups, and in order to keep the group sizes similar and have sufficient data to run a regression, I chose to keep all of the pilots in both groups in the linear mixed regression model (see Section 4.2.2). Additionally, both groups in their entirety 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 (see Table 4.5).

Table 4.5.
Mean Aviation English percentage correct over testing period by group

Group	AE1	AE2	AE3	AE4	AE5	AE6	AE7
CP	67.31^{a}	67.72^{a}	70.54^{a}	67.63^{a}	70.94^{a}	70.07^{a}	73.95^{a}
EP	82.42^{b}	83.46^{b}	84.13^{b}	84.00^{b}	85.86^{b}	86.13^{b}	87.99^{b}

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

4.2.2. Factors Predicting Aviation English Performance

As with the prior study described in Chapter 3 (see Section 3.3.2), 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 accounted for 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 significant factors that were not correlated (see Table 4.6).

Table 4.6.
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.

Examination of correlation of fixed effects using R indicated that TT and IFR ($r^2 = 0.67$), CE and Group ($r^2 = 0.79$), and number of words and number of topics ($r^2 = 0.55$) were correlated. Accordingly, I examined each of these factors for its contribution to the model. I selected IFR rather than TT, as the measure of AE exposure[6], because that is the environment in which pilots must be in constant contact with ATCOs in the IFR environment. As in the previous comparison (see Section 3.3.2), IFR values were log transformed for inclusion in the regression. Also as in the previous experiment, I retained both number of words and number of topics as factors, so that possible group differences in language parsing abilities could be discovered.

CE task score was highly correlated with Group, however it was not possible to detect an effect of the interaction of these variables on AE score, probably because the EP group is close to ceiling for CE score. Therefore, the between-group regression was done with Group, and not CE score, as a factor (Table 4.6) and a separate within-group regression was run on the CP data with CE score as a factor (see Section 4.2.3).

I examined collinearity among regression factors using variance inflation factors (VIF). No main effect had a VIF exceeding 2.19, suggesting that there was not substantial collinearity among factors included in the model[7]. 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. Regression results

---

[6] These hours of flight are in constant contact with ATCOs. All navigational decisions are made in cooperation with ATCOs. The transmissions used in the current study come directly from interactions in the IFR environment

[7] There is no accepted rule for VIF indicating multicollinearity. My personal assessment is that a VIF over 2.5 indicates an unacceptable level of multicollinearity. In the case of these data, since VIF = $1/(1 - r^2)$, a VIF of 2.19 implies $r^2 = .45$, or ~45% of the variance in the DV is also explained by the other IVs in the model (Craney & Surles, 2002).

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 as it affected each groups' 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. This combination of factors shows that, for transmissions with the minimum number of words (2), CP and EP scores were almost the same, but for transmissions with the maximum number (19), EPs scored much higher than CPs (see Table 4.7).

Table 4.7.
Main effects on AE task score between EP and CP groups

	Number of Words		ln(IFR)	WM score
Group	2-words	19-words	mean IFR	mean WM
CP	-8.24	-78.28	+3.62	+7.08
EP	-8.44	-51.68	+5.71	+10.82

In addition to the effect of number of words, pilot AE experience as measured by the natural log of their IFR time was a significant factor for predicting AE task scores. 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 Tables 4.1 and 4.7). Score on the WM task was also a significant predictor of AE score, since this score also reflects Standard English proficiency. Every percentage points 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' (see Figures 4.1, 4.2, and Table 4.7).

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

To determine the possible effect of CE proficiency on AE scores, I 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 (see Table 4.8).

Table 4.8.
Linear mixed effects model of Chinese Pilot AE performance scores

Predictor	*Estimate*	*Std. Error*	*t value*	$\chi^2(1)$	*p-value*
Intercept	94.02	5.78	16.28		
CE Score	0.37	0.08	4.65	21.58	< .001
Number of Words	-4.11	0.49	-8.42	70.82	< .001

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

The resultant mixed-effects regression model indicates that AE scores for Chinese pilots are significantly predicted by number of words in the transmission as well as 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 (compared to 4.12 in Table 4.6) in AE scores for Chinese pilots. CE task score also significantly predicted Chinese pilots' AE scores. Every percentage point in Chinese pilots' CE task score predicted a 0.37-point increase in Chinese AE scores (see Table 4.8 and 4.9). Chinese pilot AE scores were not predicted by number of topics or WM score.

Table 4.9.
Main effects on AE task score within CP group

Number of Words		CE score	
2-words	19-words	lowest (21.69)	highest (61.45)
-8.22	-78.09	+8.02	+22.74

4.3. 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 indicate that CE ability does not necessarily imply AE ability. The fact that CE scores generally predict AE scores in this study is not surprising; after all, they share some basic vocabulary and grammar. However, it appears that learning CE may be an inefficient route to AE proficiency, since other factors have more of an effect on AE proficiency. Although study results indicate that CE proficiency is correlated with AE proficiency in this population, counterexamples abound. Fully 34.5% of participants' AE and CE task scores were negatively correlated. Given a CP average CE task score of 43.23 (*SD* = 10.45) and average AE task score of 69.46 (*SD* = 6.50), five CPs scored above average on the CE task (*M* = 47.47, *SD* = 3.51) and below average on the AE task (*M* = 65.79, *SD* = 3.02), while five CPs scored below average on the CE task (*M* = 35.42, *SD* = 6.84) and above average on the AE task (*M* = 73.32, *SD* = 1.76). 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 (4.4 minutes of ATCO speech plus intervening responses, see Table 4.5), as there was for NES non-pilots using the same stimuli (see Section 3.3.2, Table 3.5). However, the regression model for all the pilot participants indicates flight experience predicts AE proficiency (see Table 4.6). This effect appears to be driven by the EP group (see Section 3.3.2.1, Table 3.7), since it was not a significant factor in the within-group analysis of Chinese pilots. Results from my previous study on the same EP group (Section 3.3.2.1) suggest that the NES pilot AE learning curve is initially steep and shallows out for more experienced pilots, 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). A correlation between flight experience and AE proficiency would probably also be found for a Chinese pilot population that included high time pilots with more exposure to AE.

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. AE proficiency was not affected by number of topics for either group, indicating that NNES pilots, as opposed to NES non-pilots (see Section 3.3.2), were as proficient at parsing AE by topic as were NES pilots. This finding is consistent with the observation that experts are able to chunk information, effectively reducing short-term memory limitations and increasing their capacity to interpret situations (Kalyuga, Ayer, Chandler, & Sweller, 2003; Sweller, 1994).

This study seeks to improve international pilot language training by enhancing the industry's understanding of NNES acquisition of AE standard phraseology. Consistent with my 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.

It must be acknowledged that CPs in this study had some familiarity with English grammar, vocabulary and pronunciation. As a reflection of English ability, CPs had a mean score of 5.61 out of a possible nine on their IELTS exam. Of the 29 Chinese pilots in my study, 19 (66%) had a score of 5.5. To clarify the level of English proficiency that this indicates: IELTS scores range from 1 to 9 (in increments of 0.5), with a score of 5 defined as a "modest user":

> The test taker has a partial command of the language and copes with overall meaning in most situations, although they are likely to make many mistakes. They should be *able to handle basic communication in their own field*. (IELTS, 2017) (Emphasis mine)

And an IELTS score of 6 defined as a "competent user":

> The test taker has an effective command of the language despite some inaccuracies, inappropriate usage and misunderstandings. They can use and understand fairly complex language, particularly in familiar situations. (IELTS, 2017)

Although it is undoubtedly helpful for NNES AE speakers to have a background in conversational English, language education for work as pilots should be in AE standard phraseology, since the rhythm and exact usage of AE are different from CE. Therefore, for nations like China, in which students have many years of written English study prior to entering flight training, practice in speaking and listening should focus on AE standard phraseology. This will enable students to develop proper comprehension and production skills for use in the flight environment.

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 entering 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 when it is recognized that the cognitive load of speaking in a second language adds to the stress that may accompany such a situation.

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 ATCO transmissions could enable new pilots to effectively communicate in AE during their first flight experiences. Dedicated AE language training will save money and time and create more fluent AE users.

CHAPTER V

AVIATION ENGLISH TASK ERROR ANALYSIS

In this chapter, I compare the intelligibility of Aviation English (AE) for native English speaking (NES) non-pilots, NES pilots and non-native English speaking (NNES) pilots. The frequencies and types of AE errors produced in AE repetitions by each of the three groups are compared.

5.1. Introduction and Study Design

NES non-pilots (NPs), NES pilots (EPs) and NNES Chinese pilots (CPs) underwent the same three audio/oral English repetition tasks in the same order. Within each task, the stimuli were randomized for each participant. The first task was a working memory (WM) task, in which participants heard several sets of single-syllable words with intervening cognitive tasks and were asked to repeat all that they could remember, in order. The second task was a conversational English (CE) sentence repetition task in which participants were asked to repeat ten fifth-grade level English sentences. The third and final task required participants to listen to and repeat actual Air Traffic Controllers (ATCOs) speaking AE.

The basic AE intelligibility study design is described in Section 3.2.2 of this dissertation. Of primary importance to the current comparison is the fact that I designed the study to be equally valid for NESs and NNESs, pilots and non-pilots. To this end, tasks were all listening and repeating exercises, presented in the same order, so as to enable a warm-up to the AE repetition task. Subjects advanced through each task at their own pace, using simple computer inputs to advance to the next trial, without feedback.

The assumption was that English speakers could understand English words, even if they were unfamiliar with the order and grammar of AE and likewise for NNESs in the CE task. I asked participants to repeat exactly what they heard, so that I could judge the ease of understanding each set of stimuli for each population.

In this chapter, I furthered the comparison of these populations by examining their AE responses in more detail. I looked at the specific errors produced by the different groups, to see if these errors were particular to one group or another. I also examined the transmissions that elicited the most and least correct responses for all of the groups.

One source of participant errors may be problems with individual ATCO transmissions. In prior comparisons (see Section 3.3.2, Table 3.6) and Section 4.2.2, Table 4.5)) using linear regressions, Transmission is treated as a random factor, in order to account for the effects of single transmissions on the relationship between group scores. Due to the possibility of a consistent influence of transmission length, number of words and number of topics were selected as main factors in the regressions, although these also could be defined as transmission characteristics. Transmission as a random factor includes effects of speaker such as: speech rate, accent and phrasing. This factor also includes effects of content, such as: familiarity of words, phrases, or constructions. These latter characteristics are expected to manifest differently for each participant group. Familiarity of words, phrases and constructions will depend upon group membership. NES non-pilots may draw upon frequency effects in CE; NNES pilots may be more dependent upon frequency effects in AE, with limited influence of CE conventions. NES pilots, given the aviation context of the task, may primarily draw upon AE, but may access CE frequency effects as well.

As seen in the regression analyses (see Tables 3.6 and 4.5), all participants had more difficulty with longer AE transmissions, and non-pilots found it more difficult to repeat two phrases than one, but what other transmission characteristics might play a part in difficulty level for all or any of the groups? It may be expected that frequency of phrases and words in the stimuli will affect the ability of all participants to repeat transmissions, since it is reported that performance in immediate recall of high frequency words exceeds that of low-frequency words (Hulme, Roodenrys, Schweickert, Brown, Martin, & Stuart, 1997; Poirier & Saint-Aubin, 1996). In addition, a frequency effect for common in-flight communications may be expected for both pilot groups. Rate of speech or ATC speaker may also have an effect on intelligibility. The particular controllers' accents or speaking styles may be harder for participants with different language backgrounds to understand (Pinet, Iverson, & Evans, 2011). To determine if any of these features are observable in the data, I performed a qualitative analysis of the highest and lowest error-inducing transmissions. Error analysis results may explain general difficulties for AE interlocutors.

It is expected that pilots are more familiar than their non-pilot counterparts with the phrasing and speech rate of the ATCO productions, given that it has been demonstrated that ATCO speech is quite rapid and repetitive. In addition, it may be expected that NESs would have more facility with understanding and producing English than NNESs. How these differences affect error production is a matter for further investigation. Particular elements in the AE speech stream may be more identifiable than others for a myriad of reasons. In Section 2.3.3 of this dissertation (see Table 2.6), I describe rhythmic differences between AE and CE such that AE exhibits less pairwise variability in consonant and vowel durations, but more overall consonant variability and similar percentage vowel duration to CE. This prosodic profile could represent a series of (temporally) equally segmented chunks of slower and faster speech (e.g. *fast-fast-fast, slow-slow-slow*) rather than a more CE-like stream of alternating stress (i.e. *fast-slow, fast-slow, fast-slow*). This pattern could arise from the production of given and new information; given information manifesting in faster speech and new information in slower. In the case of AE, these chunks are predictable. Each aviation topic identifier phrase (i.e. *climb and maintain*) precedes a series of numbers defining the appropriate target (in this case, for altitude). This model proposes that identifier phrases are old information and the following numbers are new. Therefore, I would expect identifier phrases to be produced more rapidly than numbers in AE. If this assumption is accurate, numbers will be more easily discernable than identifier phrases. Another reason numbers may be more easily reproduced by listeners is that they are a limited set of elements, whereas (non-number) words belong to a much larger set. However, this observation does not hold true for all of the populations that examined here. For instance, pilots would know that the set of possible words used in AE is much more constrained than the set of all English words. Un-coached non-pilots (as in this study) would not be privy to this knowledge. Therefore, number and word elements appear to be an important categorical distinction for error production amongst these three groups.

Further expectations regarding error production are that participants undergoing high cognitive load may be less able to repeat all of the words (properly or not) in a transmission, and that longer transmissions may be more difficult to retain in memory. Indeed, my prior study results (Tables 3.5 and 4.5) indicated that length of transmission is the main factor in decreasing participants' AE task scores. In addition, cognitive load may contribute to inaccurately reproducing or transposing elements. There also may be individual differences between participants who are willing to risk a mistake by repeating an element they are unsure of, and those who would rather omit it entirely. Indeed, this choice may be based on the perceived context. For instance, pilots have been trained to comprehend the danger in repeating back (and possibly following) the wrong command, whereas non-pilots may be less concerned about saying the wrong words. Therefore, it might be expected that non-pilots make more mistakes and pilots omit more elements due to motivational influences.

Of course, omitting or mistaking verbal elements may have to do with many external and internal factors. Transmissions for this study are from actual recordings, and therefore are embedded in static, which may reduce intelligibility. Additionally, since productions from 22 different AE speakers were used in the stimuli, some of them may have been easier or more difficult for participants to understand. Also, since the participant heard (although did not repeat) the call sign of the aircraft being addressed by the ATCO, there may have been interference from those phrases (which contain words and numbers). Particularly in longer transmissions, there may have been a working memory limitation, especially for the non-pilots who would not have a meaningful way of chunking the material beforehand (although many of them devised chunks that they repeated as the study progressed). Other sources of cognitive load such as distraction, stress, or fatigue may also contribute to a participants' ability to repeat a transmission.

In their study of Australian general aviation pilots' radio communication, Estival and Molesworth (2016) examined errors produced in native and NNES pilot communications during simulated flight scenarios with and without cognitive load conditions. Their study focused on differences in communication performance due to particular cognitive load conditions. As described in Section 2.1.1 of this dissertation, pilots' communication is primarily made up of *reading back* what controllers have said to them. The process of *readback* consists of repeating all the salient elements of a transmission, in order to indicate that commands have been heard properly. Accordingly, errors in the Estival and Molesworth study were coded only for missing or incorrect phrases in the readback. As a result, their number of incorrect items is a measure of communication accuracy, reflecting comprehension and correct readback by the pilots (Estival & Molesworth, 2016, p. 151), as opposed to my study results, which simply indicate intelligibility.

An examination of participants' responses in the current study discovered several types of errors occurring at different frequencies for each individual and population. Study participants left out words (omissions) and substituted different words for those that they heard (mistakes). These different types errors may have different sources. For instance, missing words can signify the participants not having heard, not recalling, or deeming information not worth repeating. Incorrect words, on the other hand, could be the product of mis-hearing a transmission (due to static, accent, language background or distraction, among other possibilities) or forgetting the actual words and replacing them with familiar or meaningful words. Any of these errors could also be a result of idiosyncratic choices. As described above, pilots may be less apt to make something up if they have not heard it properly, while non-pilots may just try to fill in the blank. Errors could also be divided by pertinent unit of speech, that is, either numbers or words.

Crossing these two pairs of meaningful categories, four distinct categories are thus defined: word omissions, number omissions, word mistakes and number mistakes.

This is similar to, yet different from, Estival and Molesworth's (2016) categorization of AE errors in their study of actual pilot transmissions. They chose to code errors as instances of either omissions or mistakes and either words or numbers. Similarities in categorization attest to their appropriateness, and perhaps may lead to an understanding of focus areas for further study and AE pedagogical design. However, the two studies entail different tasks, assumptions and research goals. Accordingly, error coding and results reflect study design differences.

Although both studies evaluate responses to actual ATCO transmissions, Estival & Molesworth's study (2016) analyzed pilot communication in simulated flight conditions, responding to recorded ATCO requests, in the form of readbacks, and the current study was done in the lab using recorded ATCO commands and requiring participants to repeat every word they heard. The purpose of the current study design was to enable accurate comparisons between native English and NNESs as well as fluent and novice AE speakers. Because of this methodological difference, errors could not be coded the same. For the current study, even if participants' responses included all of the salient semantic elements conveyed by ATCO, they still lost points for not repeating every word, since there was no way of determining why they left out particular elements. Accordingly, I coded every missing or incorrect word in the AE task response data as an error (Section 4.2) and subsequently categorized them as word OR number and omission OR mistake. On the other hand, Estival & Molesworth (2016) used a correctness rubric based on the accuracy of the information necessary in readback protocol. In their study, an omission or mistake could entail a single word or a phrase representing information mandatory in readback. In my current study, every single word counted as a potential error and was categorized accordingly.

5.1.1. Stimuli

Numbers make up a large portion of AE standard phraseology. In the stimuli for the current study, 43% of all of the word elements are numbers. AE numbers can signify many things, including direction, altitude, and radio frequency. AE word elements other than numbers are often predictable, frequently produced identifier phrases, such as *climb and maintain* or *turn right heading,* which provide the topic for the forthcoming numbers.

Since errors were coded for each individual element, every slot in a number sequence counted as a number (i.e. *one zero thousand* counts as three numbers). Within AE standard phraseology numbers represent new information (except for call signs). Within a topic phrase, the identifier phrase is given information in the sense that it is an oft-repeated chunk of language that serves to label the upcoming number phrase. The 57% of stimuli coded as (non-number) words are primarily derived from a limited inventory of identifier phrases for issuing commands. Many of these phrases are heard multiple times in the course of a single flight and most are so frequent that they are produced in a rapid, co-articulated, chunk-like manner.

When a pilot hears *descend and maintain,* they recognize it as a command to begin a descent and they anticipate hearing the altitude at which they are required to arrest the descent. In this sense, numbers are less predictable than most other words in AE. However, given the specific topic being addressed, even unknown number sequences are relatively predictable. For the data used in this study, only *zero* through *twelve*, *thirty,* *hundred* and *thousand* were possible selections for number slots. Accordingly, in each of the number slots, there are potentially 16 choices. However, given the topic addressed, number sequences are more predictable than this implies. Each aviation topic limits the number expressions available for selection. Given the specific number possibilities for each topic, predictability increases dramatically (Table 5.1).

Table 5.1.
Percentage frequency given possible number expressions for given topics

# Expression	Radio	Flight Level	Altitude	Heading	Traffic
0	3.61	36.36	0.84	44.44	.
1	25.13	6.06	13.92	12.96	5.56
2	17.28	18.18	6.75	12.96	5.56
3	11.38	18.18	8.86	10.19	5.56
4	4.04	3.03	3.38	3.70	5.56
5	18.57	3.03	9.28	3.70	5.56
6	4.04	3.03	2.95	3.70	5.56
7	7.83	3.03	2.95	2.78	5.56
8	4.06	4.55	2.53	2.78	5.56
9	4.06	4.55	2.11	2.78	5.56
10	.	.	.	.	5.56
11	.	.	.	.	5.56
12	.	.	.	.	5.56
30	.	.	.	.	33.33
100	.	.	19.41	.	.
1000	.	.	27.00	.	.

When a pilot hears an identifier phrase, they are able to narrow down the possible numbers in the upcoming transmission. For example, headings are always given in three digits by tens, from *0-1-0* to *3-6-0.* The final digit is always *zero,* and often (25% of the time) the first digit is, as well. Overall, *zero* occurs in 44.4% of the number slots in a heading sequence (see Table 5.1). Another common transmission is the assignment of a radio frequency. Aviation Radio frequencies range from 118.000 - 136.975, in increments on .025. These numbers are always given in the form of three digits, then *point*, followed by one to three more digits. In all possible radio frequency number sequences, *one* is the given number 25% of the time. Altitudes are similarly constrained by context. Above 18,000 feet, altitudes are given in flight levels and identified with the phrase *flight level*, followed by a three-digit number ranging from 1-8-0 to 3-9-0. Below 18,000 feet, altitudes are given in hundreds and thousands of feet. *Thousand* has the highest percentage chance of any of the possible altitude digits (27%), followed by *hundred* (19%) and *one* (14%). Direction to look for other aircraft in the vicinity (i.e. *traffic*) is given in terms of clock hands. Accordingly, *ten, eleven, twelve,* and *thirty* are possible number slot fillers in this context. Therefore, the breadth of number selection is restricted by AE topic and context.

In actual AE usage, verbiage is even more predictable than in the current study, since flight context dictates what will be said. The overall number of AE topics is limited, and within any situation, they are further reduced. Since study participants were on the ground, listening to random ATCO commands, it would have been impossible, even for pilots, to predict what they were going to hear.

Within the AE task context of repeating phrases, word or number omissions could be the result of a memory limitation or misunderstanding. In the realm of AE, omissions may also indicate that the respondent is using proper AE standard phraseology, which omits many elements of Standard English. Additionally, actual, in-flight pilot responses to ATCOs are not required to include all of the elements that the ATCO command entailed, as long as they include all critical elements of the transmission. For example, when a pilot hears, “turn right heading two zero zero”, it is appropriate to respond with any combinations of, “(turn)(right)(heading)(to) two zero zero”. In my AE task, although participants were instructed to repeat all of the words they heard, in the order that they heard them, pilots occasionally omit non-critical terminology in their responses, presumably reflecting their in-fight communication patterns. In the current study I coded these omissions as errors, effectively inflating the pilot error rate and creating a more conservative comparison with non-pilots. In the Estival and Molesworth study (2016), they were not coded as errors.

In the selection of ATC transmissions for use in this study, any transmission including a potentially confusing element was omitted from the stimuli. As explained in Section 3.2.3.3 of this dissertation, these elements included ambiguous words and abstract waypoint names.

5.2. Results

5.2.1. Overview of Analytical Methods

I used three different methods to analyze different aspects of the study data. First, for comparing the three populations' AE task scores, I performed a linear mixed regression on all of the responses by all of the participants in the study. I used one regression to estimate the effect of WM and CE task scores, native speaker status, and pilot status on AE Intelligibility scores. I then performed another linear mixed regression on the data for all three participant groups using the same factors and methodology as I used in prior comparisons of pairs of groups (see Tables 3.6 and 4.5).

Next, I analyzed AE task errors by type and frequency for each population, employing ANOVAs and post hoc Tukey tests to establish significant differences in types of error production by group. Error categorization for this analysis divided AE repetition errors by nature: omission and mistake, and category: word or number. Within the category of word omission, I further determined if the omission was in accordance with AE standard phraseology to determine if pilots consistently leave out words that are normally omitted in pilot/ATCO communication.

The third analysis is a qualitative examination of error production in AE. For this comparison, I examined patterns in error production within and between groups by isolating the highest and lowest error-inducing transmissions. These transmissions were analyzed for possible effects of frequency, length, accent/speech style and speech rate, among other factors. Lacking any statistics about types of messages in actual transmissions, the stimuli were chosen to be a roughly representative sample of ATCO transmissions. However, problematic transmissions, which included acronyms or randomly named waypoints, were removed (see footnote in Section 3.2.3.3.). I designed the AE intelligibility study to approximate the frequency of actual transmissions in the Instrument Flight Rules (IFR) environment, since all air carrier flights occur under these conditions.

5.2.2. Verbal Repetition Task Scores by Group

Group average scores differed for all of the tasks as detailed in previous chapters. In the current analysis, to examine possible language background effects, I focused on a comparison between the two populations that have the least in common and who have not yet been compared: NES non-pilots and NNES pilots. Figure 5.1 illustrates results of all three task scores for all three populations. To sum up prior results:

- Chapter 3 described NES pilots’ v. non-pilots’ task scores. Consistent with expectations, these populations scored similarly on WM and CE tasks and pilots scored higher on the AE task.

- Chapter 4 described NES v. NNES pilots' task scores. In this comparison, also consistent with expectations, NESs had higher scores on all of the verbal repetition tasks. Also of note is that, of the three tasks NNES pilots performed, they fared best on the AE task.

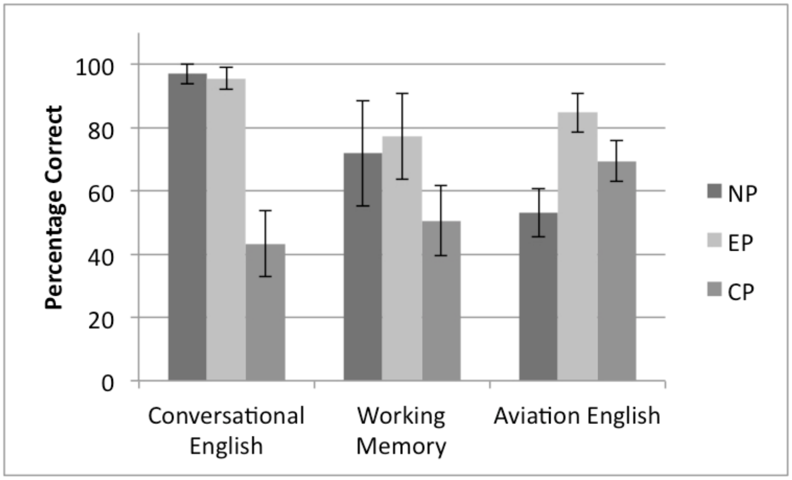

Figure 5.1. Task score results for all subjects by group. Error bars depict standard deviations. CP = Chinese Pilot, EP = native English speaking Pilot, NP = native English speaking Non-Pilot.

A comparison between NPs and CPs showed, as expected, that NPs scored higher than CPs on the CE task and the WM task (see Table 5.2), probably since they are both Standard American English (SAE) verbal repetition tasks. An analysis of variance (ANOVA) on these scores yielded significant variation among groups, $F(2, 75) = 538.3$, $p < .001$. A post hoc Tukey test showed that the NNES pilot group differed significantly from the NES pilot group at $p < .001$ and also from the NES non-pilot group at $p < .001$; the NES groups were not significantly different from one another.

Table 5.2.
Task score means by group

Group	N	CE Mean (SD)	WM Mean (SD)	AE Mean (SD)
NP	26	97.01% (3.11)	71.93% (16.74)	53.20% (7.69)
EP	23	95.55% (3.55)	77.31% (13.60)	84.78% (6.19)
CP	29	43.23% (10.45)	50.56% (11.06)	69.46% (6.51)

Notes: CE=Conversational English, WM=Working Memory, AE=Aviation English, CP=Chinese Pilot, EP=native English speaking Pilot, NP=native English speaking Non-Pilot

WM verbal repetition task scores for NESs were also higher than for NNESs (see Table 5.2). An analysis of variance (ANOVA) on these scores yielded significant variation among groups, $F\,(2, 75) = 28.0, p < .001$. A post hoc Tukey test showed that the non-native pilot group differed significantly from the NES non-pilot and the NES pilot groups at $p < .001$; the NES (pilot and non-pilot) groups were not significantly different from one another. CE and WM task scores confirm expected differences in English proficiency between NES and NNES populations.

AE task scores were not divided by native speaker status. Although NES pilots scored highest on this task, NNES pilots scored next highest and NES non-pilots had the lowest AE scores on average (see Table 5.2). An analysis of variance (ANOVA) on these scores yielded significant variation among groups, $F\,(2, 75) = 130.5, p < .001$. A post hoc Tukey test showed that all three groups differed significantly from one another at $p < .001$.

Further analysis of task score averages using a linear mixed effects regression model shows that Group and Task and their interaction significantly predict scores for all of the study participants (see Table 5.3). Fixed effects of Group ($\chi^2(2) = 265.95$, *p-value* $< .001$), Task ($\chi^2(2) = 74.712$, *p-value* $< .001$) and Group by Task ($\chi^2(4) = 506.85$, *p-*

value < .001) are significant for all combinations of group and task (see Table 5.3). Model fit using piecewiseSEM in R (Lefcheck, 2015), gave a marginal (fixed effects) R^2 value of 0.78 and conditional (including random effects) R^2 value of 0.84. No fixed effect had a variance inflation factor (VIF) exceeding 1.33. These results suggest that there was not substantial collinearity among factors included in the model[8].

Table 5.3.
Linear mixed effects model summary of AE and CE task scores by group

Predictor	*Coefficient*	*Std. Error*	*t-value*
Intercept	69.46	1.81	38.33
NP Group	-16.23	2.64	-6.16
EP Group	15.32	2.73	5.62
CE Task	-26.23	2.17	-12.10
WM Task	-18.90	2.17	-8.72
NP Group*CE Task	70.01	3.15	22.21
EP Group*CE Task	36.99	3.26	11.35
NP Group*WM Task	37.49	3.15	11.89
EP Group*WM Task	11.42	3.26	3.51

Note. CP group and AE task are defaults. Random intercept for Subject was included in the model. All t-values had p-value < .001, using pnorm function in R.

NES groups' WM and CE task scores were similar and better than the NNES group. NNES pilots scored better than NES non-pilots on the AE task, but worse than NES pilots. For a simplification of task and group effects on task scores, see Table 5.4.

Table 5.4.
Main effects of Task and Group on participant scores

Group	WM	CE	AE
CP	-18.90	-26.23	0
EP	+7.84	+26.08	+15.32
NP	+2.36	+27.54	-16.23

Note: CP=Chinese Pilot, EP=NES Pilot, NP=NES Non-Pilot

---

[8] There is no accepted rule for VIF indicating multicollinearity. My personal assessment is that a VIF over 2.5 indicates an unacceptable level of multicollinearity. In the case of these data, since VIF = 1/ $(1 - r^2)$, a VIF of 1.33 implies $r^2 = .248$, or ~25% of the variance in the DV is explained by the other IVs in the model (Craney & Surles, 2002).

5.2.3. Factors Predicting AE Intelligibility

In order to explore possible relationships between participants' task scores and other factors with their ability to understand AE, I performed a linear mixed effects regression using nlme package (Pinheiro, Bates, DebRoy, & Sarkar, 2014) in R (R Core Team, 2014). The regression accounted for random effects of specific transmission, order presented in the task, and individual participants (subject). In this model, however, since NES group scores were highly correlated with one another ($r^2 = 0.82$) and with CE score ($r^2 = 0.83$), new categorical factors of Native Speaker (Y/N) and Pilot (Y/N) were created to capture differences in the three groups. Accordingly, NES pilots were categorized as pilot level Y and native speaker level Y, NES non-pilots were pilot level N and native speaker level Y and Chinese pilots were pilot level Y and native speaker level N. The full regression model included fixed effects of native speaker level, pilot level, WM task score, age, sex, number of words per transmission, number of topics per transmission, and interaction with group for each of the fixed effects. The final model includes only significant factors that are not correlated (see Table 5.5). Examination of correlation of fixed effects using R revealed that number of words and number of topics were correlated ($r^2 = 0.50$). However, as in the previous comparisons (see Tables 3.6 and 4.5), I retained both number of words and number of topics as factors, so that possible group differences in language parsing abilities could be discovered.

Collinearity among regression factors was examined using variance inflation factors (VIF). No main effect had a VIF exceeding 2.28, suggesting that there was not substantial collinearity among factors included in the model (see VIF explanation in Footnote 8).

Table 5.5.
Linear mixed effects model summary of AE intelligibility scores

Predictor	*Coefficient*	*Std. Error*	*t value*	$\chi^2(1)$	*p-value*
Intercept	84.48	5.23	16.15		
WM Score	0.14	0.05	2.92	8.55	0.003
Pilot	17.84	2.31	7.71	306.54	< .001
Number of Topics	-10.83	3.69	-2.94	0.62	0.431
Number of Words	-3.58	0.59	-6.08	32.94	< .001
NES	-2.29	2.50	-0.92	22.61	< .001
Pilot*Topics	12.06	1.31	9.21	84.84	< .001
Pilot*Words	-0.69	0.22	-3.11	9.65	0.002
Native Speaker*Words	1.40	0.16	8.88	78.77	< .001

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

Model fit determination using piecewiseSEM package in R (Lefcheck, 2015), gave a marginal (fixed effects) R^2 value of 0.40 and conditional (including random effects) R^2 value of 0.61. Regression results indicate that a combination of significant factors affect AE scores for the entire population of the study. Members of the pilot group, whether NESs or not, did better on the whole than their non-pilot counterparts. The greatest effect on the entire participant pool was number of words in a transmission and its interaction with native speaker and pilot factors. Number of topics per transmission was also a significant predictor of AE scores, especially for non-pilots. The result of specific fixed effects is summarized in Table 5.6.

WM task score was also a significant factor for predicting AE scores for the entire pool of participants. As with the result of the previous comparison of NES and NNES pilots (see Table 4.5), it may be assumed that WM score incorporates a participant's basic ability to hold spoken language in memory, as well as reflecting their SAE proficiency. Every percentage points correct on the WM task, corresponded to a .14 percentage point increase in participants' AE scores (see Table 5.5), exactly as it did in my prior analyses.

Table 5.6.
Main effects on AE task score between EP and CP groups

	Native Speaker	Pilot	Number of Words		Number of Topics		WM score
Group			2-words	19-words	1	2	Group mean
CP	0	+17.84	-8.54	-81.13	+1.23	+2.46	+7.08
EP	-2.29	+17.84	-5.74	-54.53	+1.23	+2.46	+10.82
NP	-2.29	0	-4.36	-41.42	-10.83	-21.66	+10.05

Notes: CP=Chinese Pilot, EP=native English speaking Pilot, NP=native English speaking Non-Pilot

Although abilities captured in the task scores and language background affect AE ability, perhaps there are other ways of looking at these data that may illuminate learning and usage of AE. One such method is an analysis of errors produced by the different groups. In the prior comparisons, I have simply looked at AE scores, weighting all errors committed in AE repetitions equally. But what types of errors is each group making? Are they the same or different from one another?

5.2.4. Comparison of Aviation English Error Types and Frequency Between Groups

In comparing errors (omissions and mistakes) produced by the three groups, I looked at errors produced by ten random participants in each group, who had responded to all 84 transmissions. Errors were measured as the percentage of all possible transmission elements in that category (316 numbers, 416 words) (see Figure 5.2 and Table 5.7). As seen in the results above, more AE errors were produced by non-pilots ($M = 44.40$, $SD = 6.99$) than by pilots and more by CPs ($M = 30.89$, $SD = 7.73$) than by EPs ($M = 14.73$, $SD = 4.90$). An analysis of variance (ANOVA) on error percentages yielded significant variation among groups, $F(2, 27) = 49.92$, $p < .001$. A post hoc Tukey test showed that all the groups differed significantly from one another at *p* < .001. This relationship held true for number and word error categories as well. However, within each of the groups, more word errors were produced than number errors (see Figure 5.2 and Table 5.7).

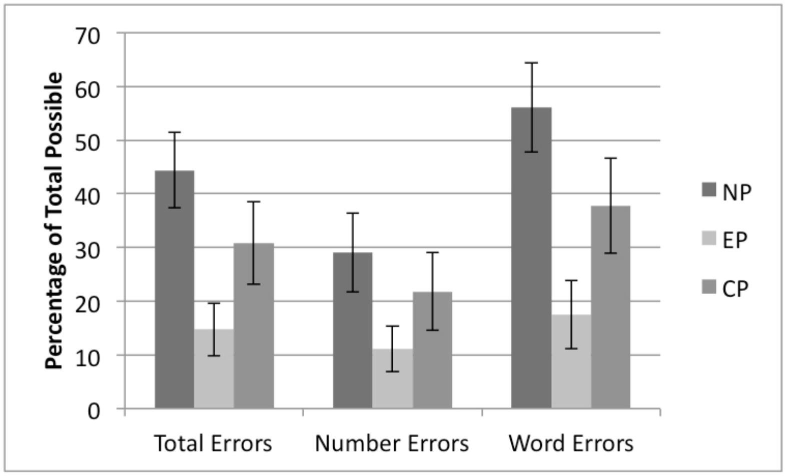

Figure 5.2. Total, number and word errors by group

Percentage of number errors by group was greater for NPs (*M* = 29.08, *SD* = 7.41) than CPs (*M* = 21.77, *SD* = 7.28) and greater for CPs than for EPs (*M* = 11.11, *SD* = 4.22) (see Figure 5.2 and Table 5.7). An analysis of variance (ANOVA) on number error percentages yielded significant variation among groups, *F* (2, 27) = 19.51, p < .001. A post hoc Tukey test showed that the all the groups differed significantly from one another at *p* < .05. Percentage of word errors by group was greater for NPs (*M* = 56.05, *SD* = 8.30) than CPs (*M* = 37.83, *SD* = 8.88) and greater for CPs than for EPs (*M* = 17.39, *SD* = 6.33) (see Figure 5.2 and Table 5.7). An analysis of variance (ANOVA) on word error percentages yielded significant variation among groups, *F* (2, 27) = 59.42, *p* < .001. A post hoc Tukey test showed that the all the groups differed significantly from one another at *p* < .001.

Table 5.7.
AE errors as percentage of possible by category, nature and group

Group	Word Errors	Number Errors	Omissions			Mistakes			Mistakes as %Errors	Total Errors
			Word	Number	Total	Word	Number	Total		
NP	56.05* (8.30)	29.08* (7.41)	36.63 (7.53)	20.86* (5.81)	29.82 (6.14)	19.41* (5.99)	8.23 (4.40)	14.58* (5.13)	32.85 (9.40)	44.40* (6.99)
EP	17.39* (6.33)	11.11* (4.22)	12.97* (5.16)	5.00* (2.71)	9.53* (3.82)	4.52 (2.58)	6.11 (2.60)	5.20 (3.82)	35.15* (11.44)	14.73* (4.90)
CP	37.83* (8.88)	21.77* (7.28)	31.29 (9.06)	14.78* (6.84)	24.16 (7.72)	6.54 (2.08)	6.99 (2.12)	6.73 (1.84)	22.67* (7.50)	30.89* (7.73)

*Note: Standard deviation in parentheses. * - Significant difference between groups at p < .05*

5.2.4.1. Number Mistakes versus Number Omissions

Percentage of number mistakes was similar for all of the groups: NPs (*M* = 8.23, *SD* = 4.40), CPs (*M* = 6.99, *SD* = 2.12), EPs (*M* = 6.11, *SD* = 2.60) (see Figure 5.3 and Table 5.7). An analysis of variance (ANOVA) on these percentages yielded no significant variation among groups, *F* (2, 27) = 1.11, *p* = .345. However, number omissions were highest for NPs (*M* = 20.86, *SD* = 5.81), then CPs (*M* = 14.78, *SD* = 6.84) and lowest for EPs (*M* = 5.00, *SD* = 2.71) (see Figure 5.3 and Table 5.7). An analysis of variance (ANOVA) on number omission percentages yielded significant variation among groups, *F* (2, 27) = 538.3, *p* < .001. A post hoc Tukey test showed that all groups differed significantly from one another at *p* < .05.

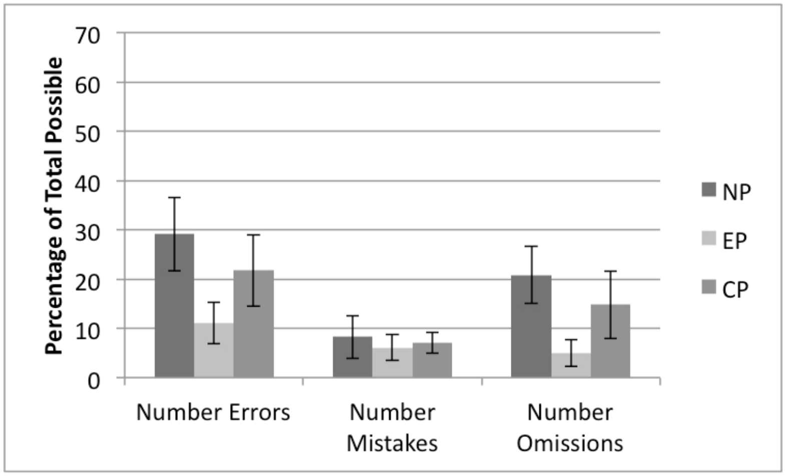

Figure 5.3. Type of number errors by group

5.2.4.2. Word Mistakes versus Word Omissions

The percentage of word mistakes for non-pilots was higher than for both pilot groups, whose percentages were similar to one another (see Figure 5.4 and Table 5.7). Word mistakes were higher for NPs (*M* = 19.41, *SD* = 5.99) then for CPs (*M* = 6.54, *SD* = 2.08) and EPs (*M* = 4.52, *SD* = 2.58) (see Figure 5.3). An analysis of variance (ANOVA) on word mistake percentages yielded significant variation among groups, *F* (2, 27) = 41.7, *p* < .001. A post hoc Tukey test showed that the NP group differed significantly from both pilot groups at *p* < .001, however the CP and EP groups were not significantly different from one another.

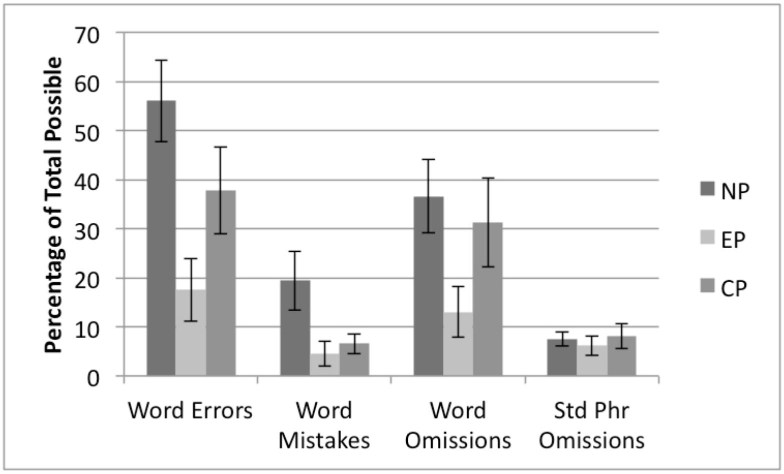

Figure 5.4. Type of word errors by group

Percentage of word omissions was highest for NPs (*M* = 36.63, *SD* = 7.53), then CPs (*M* = 31.29, *SD* = 9.06), and lowest for EPs (*M* = 12.97, *SD* = 5.16) (see Figure 5.4 and Table 5.7). An analysis of variance (ANOVA) on word omission percentages yielded significant variation among groups, *F* (2, 27) = 27.93, *p* < .001. A post hoc Tukey test showed that the EP group differed significantly from both other groups at *p* < .001, however the NP and CP groups were not significantly different from one another (at *p* < .05). Omissions that would be natural in the production of AE standard phraseology, as a subset of word omissions, were similar for all of the groups: NPs (*M* = 7.57, *SD* = 1.46), CPs (*M* = 8.14, *SD* = 2.61), and EPs (*M* = 6.16, *SD* = 1.95) (see Figure 5.4 and Table 5.7). An analysis of variance (ANOVA) on standard phraseology omission percentages yielded no significant variation among groups, *F* (2, 27) = 2.46, *p* = .105.

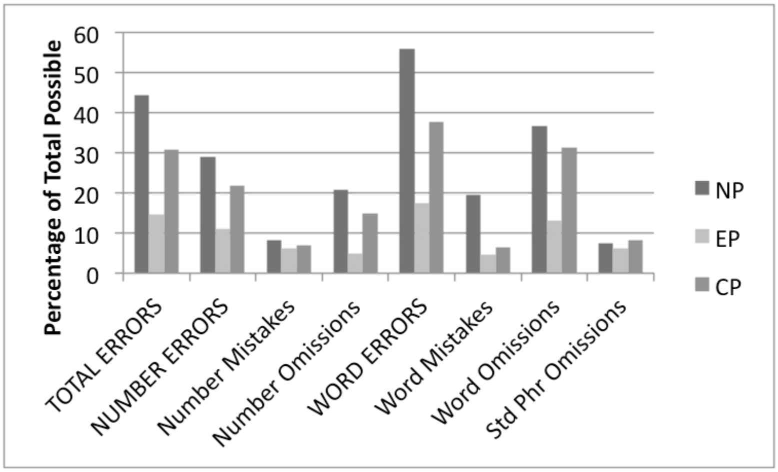

Figure 5.5. Overview of percentage of possible errors produced by group

5.2.4.3. Error Analysis Results Comparison with Estival and Molesworth 2016 Study

The following section compares the current study's results with those of Estival and Molesworth (2016), in which they looked at percentage of word / number and omission / mistake errors within subjects. For current study results, I re-analyzed the data to determine what percentage of each group's errors fall into these categories. Figure 5.6 gives the percentage of errors by possible correct for each group. As described above, non-pilots produced a higher percentage of AE errors (*M* = 44.40%, *SD* = 6.99) than did Chinese pilots (*M* = 30.89%, *SD* = 7.73), who produced a higher percentage than did native English speaking pilots (*M* = 14.73%, *SD* = 4.90).

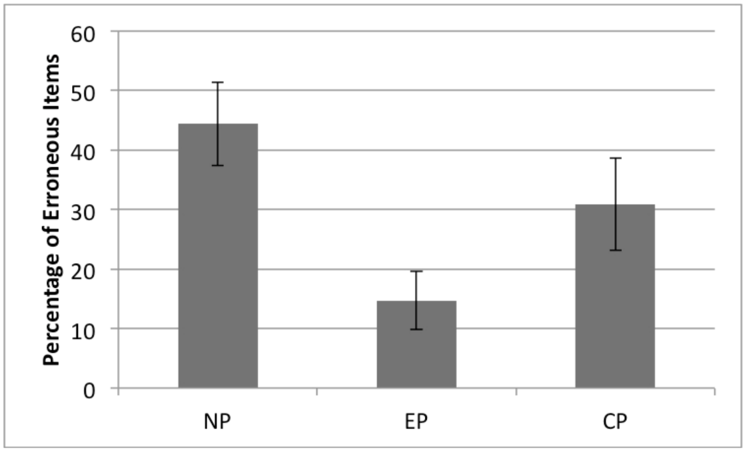

Figure 5.6. Percentage of errors (of possible correct) by group. Error bars = +/- 1 SD

Dividing individual AE errors into mistakes and omissions revealed that omissions made up the vast majority of errors within participants. NPs led in mistake production (*M* = 32.85%, *SD* = 9.40), but were close to EPs (*M* = 35.15%, *SD* = 11.44), whereas CPs produced the lowest percentage of mistakes (*M* = 22.67%, *SD* = 7.50) (see Figure 5.7). An analysis of variance (ANOVA) on mistake percentages yielded significant variation among groups, F (2, 27) = 4.80, $p = .016$. A post hoc Tukey test showed that CPs differed significantly from EPs, at $p =.018$. However, NPs were not significantly different from CPs or EPs at $p < 0.05$.

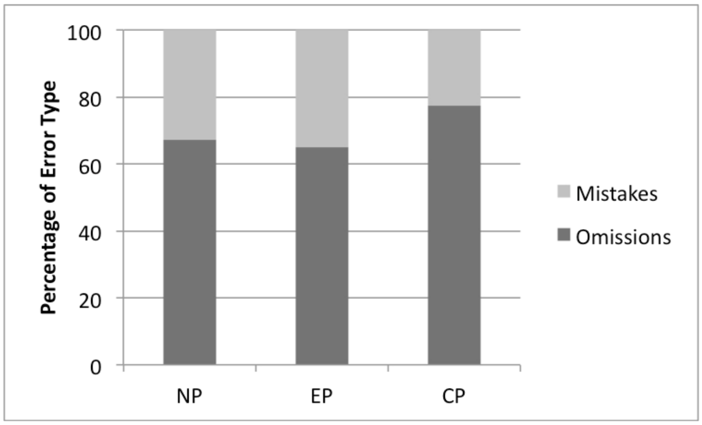

Figure 5.7. Percentage of omission and mistake errors by group

Further subdividing the data, errors were cross-referenced by nature (mistake or omission) and category (word or number). All of the groups produced similar percentages of word and number errors of omission (see Figure 5.8). Errors of omission were primarily made up of word elements: NPs (*M* = 70.05%, *SD* = 5.26), CPs (*M* = 73.18%, *SD* = 5.37), EPs (*M* = 78.20%, *SD* = 8.23). An analysis of variance (ANOVA) on omission percentages yielded significant variation among groups, F (2, 27) = 4.01, $p =$ .03. A post hoc Tukey test showed that NPs differed significantly from EPs, at $p < .001$. However, neither the NP nor EP group was significantly different from CPs ($p = 0.34$ and 0.36 respectively).

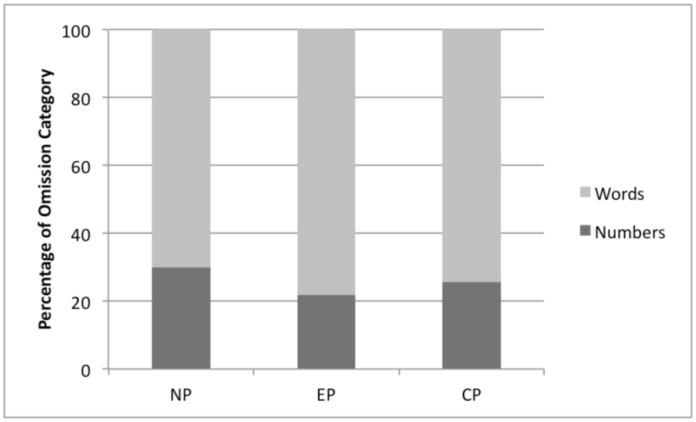

Figure 5.8. Percentage of word or number omission errors by group

Mistakes by NPs reflected almost the same pattern as omissions in that NP word mistakes made up the majority (*M* = 76.57%, *SD* = 5.18) of their total mistakes (see Figure 5.9). However, those of CPs (*M* = 55.14%, *SD* = 7.30) and EPs (*M* = 46.14%, *SD* = 14.54) were more evenly split. An analysis of variance (ANOVA) on these percentages yielded significant variation among groups, *F* (2, 27) = 25.13, *p* < .001. A post hoc Tukey test showed that non-pilots differed significantly from both EPs and CPs, at *p* < .001. However, the CP and EP groups were not significantly different from one another (*p* = 0.122).

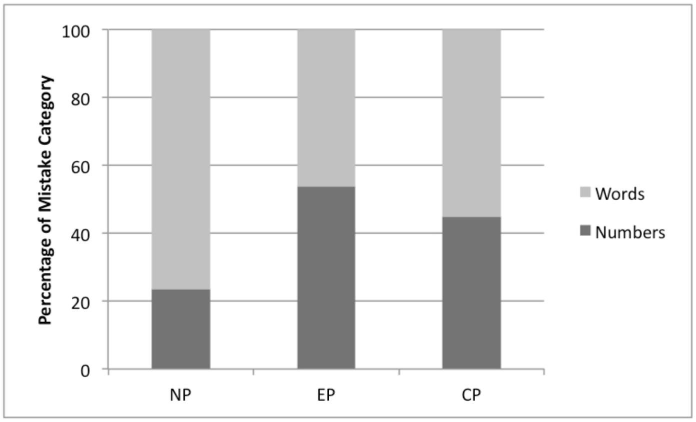

Figure 5.9. Percentage of word or number mistake errors by group

In order to contextualize these results, I compared them with those of Estival and Molesworth's study (2016). Although the studies differ in some meaningful ways, I looked at overall trends in error production. Both studies involved NESs and NNESs responding to ATCO commands. However, since my study evaluated non-pilots as well as pilots, I evaluated strict repetition whereas they used pilots only and evaluated AE readback accuracy (as described in Section 5.1.). Another difference between the studies involves the breadth of flight experience in the pilot populations. The current study purposefully examined pilots with similar levels of experience, in order to focus on language background differences. As a result, I did not have a wide enough range of pilot ratings to yield significant results for certificate level, as shown by Estival and Molesworth, but the increase in ability with IFR experience in my study indicates this pattern of experience increasing AE proficiency holds (see Table 4.5).

Despite some differences, the two studies agree on some meaningful results. Both studies' findings indicated that NES pilots produced fewer errors overall than NNESs, pilots with more advanced training produced fewer mistake errors than other pilots, all participants produced proportionally more omissions than mistakes, and more word errors than number errors.

In the Estival and Molesworth (2016) study, NES pilots produced 40% incorrect readback items v. 50% for NNES pilots. In the current study the findings were 15% error rate for NES pilots v. 44% for NNESs. Since scores are broken down differently for the two studies, these percentages would not be expected to line up exactly. I coded transmissions in single word elements, increasing the range of possible scores for each transmission. They scored in chunks of word phrases or number sequences. For instance, in the previous study, *contact tower 1-1-8 point 2* would be one word element (*contact tower*) and one number element (*118.2*), a maximum of two errors. Whereas, in the current study, this transmission consists of three word elements (*contact-tower* and *point*) and four number elements (*1-1-8 and 2*) for a total of seven potential errors.

Similar to the current study, the total of all four of Estival and Molesworth's scenarios in both conditions (with and without additional cognitive load), indicated pilots produced a higher percentage of total possible omission errors ($M = 22.77\%$, $SD = 9.43$) than mistake errors ($M = 2.72\%$, $SD = 2.34$) (Estival & Molesworth, 2016). Also in keeping with the current study's results, their participants exhibited a much higher percentage of possible word omissions ($M = 35.36\%$, $SD = 14.31$), than number omissions ($M = 10.57\%$, $SD = 6.85$). Overall, Estival and Molesworth's participants produced fewer word mistake errors ($M = 0.595$, $SD = 1.07$) than number mistake errors ($M = 5.115$, $SD = 4.31$).

The relationship between number and word mistakes for pilots in the current study were more similar to the previous study findings than were those of NPs (see Figure 5.9). The current study found pilot group values were similar to the previous study for omission and mistake error production as well, with more word omissions ($M = 22.13\%$, $SD = 11.83$) than number omissions ($M = 9.89\%$, $SD = 7.12$) and more number mistakes ($M = 6.55\%$, $SD = 2.36$) than word mistakes ($M = 5.53\%$, $SD = 2.51$), although the mistake figures are not significantly different ($t\ (37.855) = 1.33$, $p = 0.19$).

In Estival and Molesworth's (2016) study, regardless of language background, pilots with higher qualifications are less likely to make mistake errors in general (Private Pilot License~21%; Commercial Pilot License~12%). My study showed that pilots with more IFR time produced fewer errors overall than did lower IFR time pilots. These two measures of experience are comparable; pilots must earn an instrument rating to fly under Instrument Flight Rules. Both measures indicate that pilots have spent time in advanced training environments.

5.2.4.4. Qualitative Analysis of Errors by Transmission

The question arises, when examining the errors produced in participants' repetitions of ATCO transmissions, what other factors may have contributed to these errors? Were there factors that affected all groups similarly? Were there factors that only affected one group and not the others? In order to address these questions, I analyzed representative transmissions that had resulted in the highest and lowest percentage of errors in participants' repetitions, for patterns and possible sources of intelligibility challenges and aids. Since number of subjects differed by group (see Table 5.2), I determined overall performance on individual transmissions by taking the average of group AE score means for each transmission. I then determined the transmissions with the worst performance to be the sequential set (by overall average) that included the three lowest scores for each group. This resulted in fourteen lowest score transmissions (see Table 5.8). I also examined the top scoring thirteen transmissions, since this group included the top three scores for each group (see Table 5.9). I also tallied AE percentage correct scores by group for the three highest and three lowest scores (see Table 5.10).

From the set of 14 and 13 transmissions with the lowest and highest respective overall AE scores, several interesting observations can be made. These can be broken down into effects of length, familiarity of words and phrases, overall word frequency, speaker and rate of speech on the different groups. A cursory look at the data reveals that EPs' and CPs' errors were more similar to one another than to NPs'.

- *Transmission length as a factor in overall error production*

As shown in the above linear regression on these data and in previous studies (Estival & Molesworth, 2016; Farris, 2007, Farris & Barshi, 2013; Prinzo, 2008), transmission length has an effect on AE intelligibility (see Table 5.5) for all groups. Stimuli ATCO transmissions ranged from 2 to 19 words (M = 10.34, SD = 2.90). Longer transmissions' negative effect is apparent in the 14 lowest AE score set. The three longest transmissions (by word count) in the data were among the top five error-producers, and five of the ten longest transmissions were among the top ten error-producers (see Table 5.8). Eleven of the top fourteen error-producers consisted of two aviation topics (as opposed to one). Effect of transmission length can also be seen in the highest AE score set, in which all of the transmissions were 7 words or less and one topic only (see Table 5.9). The shortest transmission (2 words) had the highest overall score. All 14 highest AE scores were among the 34 shortest transmissions in the data.

Table 5.8.
Ranking of transmissions precipitating the lowest AE performance scores overall and by group

All Ss Errors	Transmission and speaker (by location identifier)	NP Rank	EP Rank	CP Rank	Average % Corr	# of Words	# of Items	Speech Rate
1	Climb and maintain one seven thousand, contact Fort Worth center one three two point eight five (DFW1)	5	3	3	38.78	16	2	5.36
2	Good rate of descent to five thousand (DFW2)	12	1	4	39.38	7	1	7.49
3	Traffic holding in position, cross runway two two right (BOS1)	2	8	6	39.74	9	2	7.37
4	Turn left heading two one zero, established on that heading the airport will be twelve o'clock and five miles (BOS2)	4	2	7	40.42	19	2	6.46
5	Traffic nine thirty to ten o'clock and three miles, southbound climbing out of two thousand two hundred (BOS3)	15	4	1	42.69	17	2	5.58
6	Turn left heading three five zero and join the final for three six left (DFW3)	9	9	2	43.96	14	2	6.67
7	Traffic twelve o'clock a mile, eastbound just about below you (DFW4)	11	5	12	45.38	10	2	7.00
8	Cleared visual approach three five right, contact tower one two six five five (DFW3)	8	6	9	46.25	13.5	2	7.02
9	Traffic twelve o'clock five miles, northbound descending out of nine thousand (DFW5)	10	7	13	47.20	11	2	6.89
10	Traffic arriving runway two two left, will hold short of your runway (BOS4)	3	12	14	47.44	12	2	6.82
11	Cleared visual approach three six left (DFW3)	14	18	5	50.48	6.5	1	8.47
12	Descend and maintain three thousand, two hundred and ten knots on your speed now (DFW6)	20	10	8	51.65	14	2	6.63
13	Proceed inbound runway two two left, hold short of runway two seven (BOS4)	13	21	7	57.05	12	2	6.55
14	Caution wake turbulence (BOS5)	1	77	23	57.27	3	1	8.47

Note: CP = Chinese Pilot, EP = native English speaking Pilot, NP = native English speaking Non-Pilot.

Table 5.9.
Ranking of transmissions precipitating the highest AE performance scores overall and by group

AE Score Rank	Transmission and speaker (by location identifier)	NP Rank	EP Rank	CP Rank	Average % Corr	# of Words	# of Items	Speech Rate
1	Contact departure (BOS6)	2	1-12	1	99.33	2	1	6.59
2	One one eight point one (DFW7)	4	1-12	7	96.75	5	1	7.14
3	Reduce speed to two zero zero (BOS5)	1	28	21	96.09	6	1	6.59
4	Reduce speed to one seven zero (DFW8)	3	26	29	94.30	6	1	6.88
5	Contact approach one two six point five (BOS2)	6	14	11	93.86	7	1	5.50
6	Contact approach one two zero point five (DFW9)	5	13	19	93.20	7	1	4.68
7	Contact approach one two six point five (BOS7)	10	1-12	4	92.88	7	1	6.43
8	Squawk three four two seven (BOS3)	9	1-12	13	91.95	5	1	4.81
9	Descend and maintain five thousand (DFW5)	22	1-12	2	90.54	5	1	6.98
10	Fly heading three three zero (BOS8)	21	1-12	8	90.11	5	1	6.76
11	Squawk three four two three (BOS9)	12	17	17	89.89	5	1	4.86
12	Turn right heading two zero zero (BOS10)	11	13	15	89.56	6	1	8.90
13	Descend and maintain three thousand (BOS4)	26	1-12	3	89.26	5	1	6.56

Note: CP = Chinese Pilot, EP = native English speaking Pilot, NP = native English speaking Non-Pilo

Table 5.10.
Characteristics of ATCO transmissions with the lowest and highest AE scores by group

		Lowest AE scores				Highest AE scores			
		First	Second	Third	Mean	First	Second	Third	Mean
NP	# Words	3	9	12	8.0	6	2	6	4.7
	Numbers	0%	22%	17%	13%	33%	0%	50%	33%
	ATCO	BOS5	BOS1	BOS4	BOS	BOS5	BOS6	DFW8	BOS
	Sp. Rate	8.47 s/s	7.37 s/s	6.82 s/s	7.55 s/s	6.59 s/s	6.59 s/s	6.88 s/s	6.69 s/s
	Correct (sd)	11.54% (16.17)	14.96% (18.31)	24.68% (12.13)	17.06%	99.36% (3.27)	98.00% (10.00)	96.15% (13.59)	97.84%
EP	# Words	7	19	16	14.0	top 12 average: 4.75 (1.22)			4.8
	Numbers	29%	26%	50%	35%	top 12 total: 49%			49%
	ATCO	DFW2	BOS2	DFW1	DFW	4DFW(2 same)/8BOS(2 same)			BOS
	Sp. Rate	7.49 s/s	6.46 s/s	5.36 s/s	6.44 s/s	top 12 average: 6.50 s/s (0.96)			6.50 s/s
	Correct (sd)	48.45% (19.62)	55.69% (26.64)	56.25% (21.98)	53.46%	top 12 average: 100.00% (0.00)			100%
CP	# Words	17	14	16	15.7	2	5	5	4.0
	Numbers	47%	36%	50%	44%	27%	40%	40%	36%
	ATCO	BOS3	DFW3	DFW1	DFW	BOS6	DFW5	BOS4	BOS
	Sp. Rate	5.58 s/s	6.67 s/s	5.36 s/s	5.87 s/s	6.59 s/s	6.98 s/s	6.56 s/s	6.71 s/s
	Correct (sd)	34.48% (8.85)	35.47% (19.14)	35.56% (12.17)	35.17%	100.00% (0.00)	99.31% (3.71)	99.31 % (3.71)	99.54%

Note: CP = Chinese Pilot, EP = native English speaking Pilot, NP = native English speaking Non-Pilot, ATCO = air traffic controller (labeled by location identifier).

- *Number elements as a factor in overall error production*

As described in the introduction and supported by the above findings, numbers in the transmissions may be more recognizable to non-pilots and more predictable for pilot participants than other words. Results of this qualitative study indicate that this is the case. Over the entire data set 316 of 732 word elements (43.17%) were numbers.

However, in the 14 lowest AE scores set of transmissions, only 59 out of 164 word elements (35.98%) were numbers, while in the highest 13 AE score set 48 out of 82 word elements (58.54%) were numbers.

- *ATCO region and speaker as a factor in overall error production*

It appears that ATCO speaker also had an effect on intelligibility. Twenty-two air traffic controllers (half and half from the Boston (BOS) and Dallas/Ft. Worth (DFW) airports) produced 3 or 4 of the 84 transmissions used in the study. All three of one DFW controller's productions were in the 11 lowest AE scores (see Table 5.8, items ranked 6, 8 and 11). Two out of four transmissions from a BOS controller (see Table 5.8, items ranked 10 and 13) were in this set, as well (however, this speaker's other two productions were in the 16 highest AE scores (see table 9 numbers 13 and 16)). Alternatively, two of four transmissions for one DFW controller and one BOS controller were in the top 14 highest AE scores (see Table 5.9, items ranked 2 and 9 and 7 and 12, respectively).

These results from individual ATCOs could imply that participants had an equally difficult time with accents from both ATCO locations. However, when location is treated as a factor, it is clear that regional accent or speech style may have played a part in intelligibility for participants. DFW controllers delivered eight of the 14 (57%) lowest AE scoring transmissions and only four of the 13 (31%) highest AE scoring transmissions (see Tables 5.8 and 5.9). This comparison indicates that participants had a harder time understanding DFW ATCOs. An acoustic analysis of the data is necessary to claim any specific reason for this apparent effect. My subjective impression based on native speaker linguist intuition is that the ATCOs with the strongest regional accents, regardless of location, were difficult for subjects to understand.

- *Speech Rate as a factor in overall error production*

Although the average speech rate (in syllables per second) for the most error-inducing transmissions (*M* = 6.91, *SD* = 0.88) was higher than that for the least error-inducing transmissions (*M* = 6.38, *SD* = 1.12), the differences between these transmissions were not significant. An analysis of variance (ANOVA) on these rates yielded no significant variation between groups (F (2, 81) = 0.996, $p < .374$). In fact, given that the range of speech rates in the selected ATCO data was 4.77 to 9.77 syllables per second, one of the fastest transmissions (8.9 s/s) was in the highest AE score set and some slower transmissions were in the lowest AE score set (see Tables 5.8 and 5.9). Overall, variation in speech rate did not seem to account for AE score performance in this study. This is a similar finding to those in Sections 3.3.2 and 4.2.1 (see Tables 3.5 and 4.4) in which there was very little learning effect for non-pilots over the course of the study.

- *Frequency and other factors in AE error production by group*

In addition to the factors described in the preceding analyses, such as transmission length (see Tables 5.8 5.9, and 5.10) and the prevalence of numbers or words in a transmission (see Table 5.10), word frequency may influence AE intelligibility. In the stimuli, words and phrases that occurred more frequently could have been more accessible for participants as the task progressed. However, this effect would lead to learning over the duration of the task and prior analysis of these data indicate that this did not occur, except for a slight increase in NP scores after the first dozen of 84 trials (Sections 3.3.2, Tables 3.5 and Section 4.2.1, Table 4.5). This trend in the NP group was thought to be the result of adjusting to prosody and other global factors of AE and not to word or phrase frequency in the stimuli, given that participants did not receive any feedback regarding the accuracy of their responses. Since the groups have varying levels of exposure to AE and CE, they are expected to have different lexicons of familiar words and phrases. Therefore, it is best to approach an analysis of frequency effect by looking at each group separately. In order to determine the most and least problematic factors for AE intelligibility by group, the ATCO transmissions that elicited the three lowest and three highest scores for each group are examined below for frequency effect as well as differing relationships to length, percentage of numbers, regional accent, and speech rate (see Table 5.10).

o *Frequency and other factors in AE error production for NPs*

Evaluating average AE scores by transmission for NES non-pilots (NPs): longer transmissions, lower proportion of number elements, and higher speech rate contributed to lower task scores. Regional accent didn't appear to be the source of difficulty for NPs, as the majority of both lowest and highest NP scores came from responses to BOS controllers. Although the three lowest AE scoring transmissions for all groups averaged more words than those of their three highest AE scores, the length difference had less of an effect on NP scores than the pilot groups' (see Table 5.10). Conversely, percentage of number elements and speech rate differed more between NPs' lowest and highest AE intelligibility scores than for pilots'.

Table 5.11.
Frequency of select words and phrases in Aviation and Conversational English stimuli

Word/Phrase	AE per million	CE* per million
one	54819	2845.94
zero	48194	27.83
two	47014	1278.90
three	37938	668.63
heading	21056	29.01
maintain	17245	51.15
visual	1724	46.14
approach	18243	134.33
visual approach	1549	.032
caution	545	14.31
caution wake	664	0.00
wake	545	48.55
wake turbulence	664	0.04
turbulence	545	3.64
traffic	4810	62.99
traffic holding	442	0.005
traffic arriving	332	0.005
hold	2451	181.95
hold short	1549	0.03
hold on	0	18.89
hold the	0	16.78
hold up	0	5.74
holding	726	106.05

*Note: * - CE values derived from the Corpus of Contemporary American English (COCA). AE values derived from ATC corpus. Two-word phrase frequency is given per million words, also.*

However, the difference in speech rate was not significant for the highest and lowest scoring transmissions. The highest- and lowest-scoring five and ten AE transmissions for NPs were compared and neither pairing resulted in a significant difference in speech rate ($p < .05$). NPs three lowest AE intelligibility scores were on transmissions averaging 8.0 words long, 13% numbers, and 7.55 syllables per second, whereas their three highest intelligibility scores averaged 4.7 words long, 33% numbers, and 6.69 syllables per second (see Table 5.10).

Next, I examined frequency as an effect on intelligibility of AE and CE words. For NPs, the transmission with the lowest average AE intelligibility score was *Caution wake turbulence* (see Tables 5.8 and 5.10). Although each of the words in this phrase are common English words, they are not very frequent in CE (see Table 5.11), and in fact, this is the only instance of any of these words in the stimuli. However, it was only the fourteenth lowest average score for all groups combined (77th for EPs and 23rd for CPs), because pilots recognize it as a common phrase in AE (see Table 5.11).

The second and third lowest AE scores for NPs came in response to the transmissions issuing traffic warnings: *Traffic holding in position, cross runway two two right* and *Traffic arriving runway two two left, will hold short of your runway* (respectively) (see Tables 5.8 and 5.10). In AE, *traffic* refers to other aircraft in the vicinity. It is possible that this alternate meaning of the common English noun *traffic*, may increase the difficulty for NPs in interpreting these phrases. The common English verb *hold* also appears in both of these transmissions. It's meaning in AE is to remain in one spot. This meaning is relatively infrequent, although not unattested, in CE (see Table 5.11).

Frequency and other factors in AE error production for EPs

EPs all achieved 100% intelligibility on twelve AE transmissions, therefore I averaged high scoring transmission characteristics over these twelve. For EPs, longer transmissions and lower proportion of numbers appeared to contribute to lower task scores. Counter to expectation, speech rate was actually slightly slower in the lowest scoring examples for EPs. Although both ATCO locations produced transmissions that elicited amongst the three lowest and three highest AE scores for EPs, the majority of lowest scores came from DFW transmissions and the highest from BOS. EPs three lowest AE intelligibility scores were on transmissions averaging 14.0 words long, 35% numbers, and 6.44 syllables per second, whereas their highest intelligibility scores averaged 4.8 words long, 49% numbers, and 6.50 syllables per second.

The lowest score for EPs (and second lowest AE score for all groups) was for the transmission: *Good rate of descent to five thousand.* Many factors suggest that this transmission should not be difficult for pilots: it is one of the shorter transmissions in the stimuli; the words are all common words in CE; and it is a standard AE phrase. However, in the population of pilot participants in this study, this is an infrequent phrase (it elicited the 4th lowest score for CPs). Most of the pilots in the study are low flight time, either working on their instrument rating or having recently acquired it. This means that they would have spent most of their instrument time in the training environment at relatively low altitudes, not in descent from cruise, which is where you would more commonly get the request to descend rapidly.

The second lowest AE score by EPs was in response to the transmission: *Turn left heading two one zero, established on that heading the airport will be twelve o'clock and five miles*. The most obvious challenge with this transmission is that it is the longest transmission in the stimuli (19 words). However, it is only the 4th and 7th lowest for NPs and CPs (respectively). EPs averaged 55.69% correct for this transmission: better than the average of all groups together, but about half of the elements in it posed problems for EPs. Examining the makeup of the transmission, both topics are phrased in AE standard phraseology (with the exception of *and*). Further, the first topic (*Turn left heading two one zero*) is very frequently encountered in flight. However, unlike the first topic (and the majority of ATCO transmissions), the second topic requires passive listening only and not salient element repetition (readback). The correct AE radiotelephony response to this transmission in actual flight conditions would be something to the effect of *left two one zero, will report airport in sight.* Pilots will not have practiced saying all the words in the second topic. And, although pilots undergoing the AE intelligibility task were instructed to repeat all the words in the transmission, their lack of daily practice producing this type of phrase presumably requires extra effort to be expended in cognitive and articulatory processes, resulting in lower proficiency.

The third lowest average AE intelligibility score for EPs was elicited by the transmission: *Climb and maintain one seven thousand, contact Fort Worth center one three two point eight five.* Given that this transmission is in standard phraseology and both topics are very frequently heard in actual conditions, it appears that length is the primary factor in level of difficulty for EPs. At 16 words, this is the third longest transmission in the stimuli.

o *Frequency and other factors in AE error production for CPs*

Examining average AE scores by transmission for Chinese NNES pilots (CPs), longer transmissions appeared to contribute to lower task scores. However, contrary to expectation and to NPs performance, speech rate was actually lower and percentage numbers higher in the lowest AE score averages for CPs. Although both ATCO locations produced transmissions that numbered among the three lowest and three highest AE scores for CPs, the majority of lowest scores came from DFW transmissions and the majority of the highest came from BOS (see Table 5.10). Their three lowest AE intelligibility scores were on transmissions averaging 15.7 words long, 44% numbers, and 5.87 syllables per second, whereas their three highest intelligibility scores averaged 4.0 words long, 36% numbers, and 6.71 syllables per second.

The lowest average AE intelligibility score for CPs was elicited by the transmission: *Traffic nine thirty to ten o'clock and three miles, southbound climbing out of two thousand two hundred.* This is another transmission that is common in AE, but not required to be repeated by AE readback protocol. The standard response to this transmission is simply *have traffic* (if they see the other aircraft) or *negative traffic* (if they do not see the other aircraft). This is also the second longest transmission in the stimuli (17 words). The combination of length and lack of practice saying these phrases caused this to be the hardest transmission to repeat for CPs (see Table 5.8).

The second most difficult transmission to repeat for CPs was: *Turn left heading three five zero and join the final for three six left.* At 14 words, length was a factor in the level of difficulty for CPs. On the other hand, this transmission uses standard phraseology in a frequently encountered set of topics. However, it was produced by the DFW controller for whom three of four transmissions were in the lowest 11 overall AE scores. Therefore, regional accent or personal speech style may have reduced intelligibility for CPs on this transmission.

The third lowest score for CPs came in response to the same transmission that elicited the third lowest scores for EPs: *Climb and maintain one seven thousand, contact Fort Worth center one three two point eight five.* As with EPs, given that this transmission is in standard phraseology and both topics are very frequently heard in actual conditions, it appears that length is the primary factor in level of difficulty for EPs. At 16 words, this is the third longest transmission in the stimuli.

5.3. Discussion

While attempting to repeat verbatim ATCO instructions in AE, all of the current study participants committed far more omission errors than mistake errors (3 to 1) (see Table 5.6) and more word errors than number errors (2 to1) (see Table 5.2). These results are consistent with prior studies (Estival & Molesworth, 2016). That being said, these proportions differed by group, and groups were more or less similar to one another in particular error categories. As would be expected, native speaking pilots produced the least number of AE errors, followed by NNES pilots and then non-pilots (see Figure 5.2), similar to Estival and Molesworth (2016) findings. Non-pilots (NPs) often exhibited different error production patterns than pilots, although, in some cases NPs results were similar to Chinese NNES pilots (CPs), and in other cases to NES pilots (EPs). Parallels in these populations' error patterns may be used to identify potential areas of difficulty for NES and NNES AE trainees.

NPs were most obviously divergent in their production of word mistakes. They produced nearly three times the number of word mistakes than did the pilot groups, who performed the same as one another in this category (see Figure 5.4). On the other hand, all participants produced a similar quantity of number mistakes (see Figure 5.4). These results are likely caused by the fact that number elements are most easily recognized by all the participants, whereas specific AE terminology is more familiar to pilots than to non-pilots. Even though all of the words in the transmission were common English words, NPs often could not recognize them without understanding the context. And, although CPs produced more word mistakes than EPs (see Figure 5.4), this difference was not significant, showing that CPs have learnt AE and recognize AE terminology as readily as native EPs.

Omission errors were more likely than mistake errors for all of the groups (see Figure 5.6), probably due to memory limitations or motivational learning-experience. This effect was stronger for NPs and CPs, who produced the same number of word omission errors as one another and almost three times the amount produced by EPs (see Figure 5.4). The benefit of the combination of English proficiency and pilot experience was most evident in this error category. In contrast, results show that CPs produced the same number of word mistakes as EPs, which was only about a third the amount that NPs produced (see Figure 5.4). It appears that CPs were more willing to leave out a word or phrase than to say the wrong one. This choice may be the result of pilots understanding the inherent danger of following an inaccurate ATCO instruction.

The high number of omissions for both CPs and NPs can probably be attributed to a memory limitation. Regression results on AE error production (Table 5.5) indicate that non-pilots suffered a detriment for multiple items in a transmission, and pilots did not. The putative cause for this behavior is that AE users have developed a chunking strategy that allows them to recognize groups of words (e.g. aviation topics identifiers) instead of having to parse each word individually. However, in longer transmissions, NNES cognitive load penalty may cause CPs to attend to one aviation topic at a time and omit an entire topic, producing only those they can correctly, but not attempting to recreate any they were unsure of. Whereas NPs, without the imperative of flight safety in mind, may rapidly interpret and repeat any verbiage they construe as reasonable English.

Another interesting result of the current study is that word omissions that could be categorized as the result of standard phraseology conventions were also omitted from non-pilot speech at the same rate (see Figure 5.4). Words not repeated by all of the groups may be less salient for at least a couple of different reasons, which are often related. They could be acoustically reduced (shorter, less voluble) or lack semantic weight (i.e. function words like prepositions). Non-pilots would likely elide these words due to lack of saliency in the speech stream. This may also be the case for pilots, or they could be following conventions. Either way, this finding supports the design choice in AE to remove function words when possible.

Transmissions that proved more or less difficult for participants also gave clues as to what AE learners may need to focus on. Although length of transmission was shown to be a significant factor predicting AE scores (see Table 5.5), results of the qualitative study show that length is not the only cause of difficulty (see Section 5.2.4.4). Certainly, participants suffer from memory limitations, especially those less familiar with AE (NPs) or CE (CPs). However, longer transmissions that included more numbers and / or more frequent phrases were less error-prone than those with more words and / or less familiar phrases (see Table 5.10 and 5.11). The selected ATCO stimuli were intended to be representative of actual transmissions. Accordingly, phrases less frequent in actual transmissions were less frequent in the stimuli set. As an example, a phrase that confounded pilots and non-pilots alike was *visual approach* (8th and 11th lowest overall AE score). This phrase is not frequent in the data or in actual flight conditions, nor is it widely represented in CE (see Table 5.11). Although aware of its existence, low-time pilots (the majority in this study) may never have heard *visual approach* in actual communications with an ATCO, since it is not issued to pilots training on instrument approaches.

Another type of familiarity arises from physically producing specific words and phrases. For example, the five *traffic* phrases (amongst the lowest scoring transmissions, see Table 5.8) would not require repetition in actual AE communication, merely an acknowledgment. Regional accent or speech style may have made particular ATCO transmissions more difficult for some participants to understand, as well. This finding suggests that training in regional AE productions would also benefit AE users, especially NNESs, who presumably have had less exposure to a variety of English accents and speech styles.

5.4. Conclusion

The results of these analyses support findings in previous chapters that Aviation English (AE) is not equally intelligible to all NESs. Without the benefit of aviation training, NESs can barely understand AE and most of what they do understand is in the form of numbers, which make up a restricted and relatively predictable set of elements in this lexicon. Once again, these results indicate that conversational English (CE) ability is not sufficient or appropriate as a predictor of AE ability.

The findings of the current study in combination with the findings of Estival and Molesworth (2016) indicate that pilots with more advanced training make fewer errors in AE speech. The importance of this finding is that individuals can gain skill in AE through training. This is a critical point, given the high levels of errors produced by pilots in both of these studies. Additionally, these circumstances in combination with this study's finding that non-pilots incurred far more word errors than pilots did in their attempts to repeat AE verbiage imply that training specifically focused on AE terminology is needed. Not only is it needed, but these findings also imply that it could be relatively simple: focusing on AE standard phraseology. The key to recognizing chunks of rapidly produced air traffic controller (ATCO) jargon is familiarity – with the context, meaning and form, so that phrases are predictable in specific situations and easily discernable from one another. Within this study we also see what could be a frequency effect for all groups. Therefore, AE training should focus on less frequent standard phraseology.

The examination of transmissions that produced the most errors can further guide training. Consistent with previous research, all participants had more trouble reproducing longer transmissions. To some extent, reducing the length of ATCO transmissions can mitigate this problem. However, the AE environment will not allow much of this. There is too much information to convey in a limited amount of time and radio frequency space. Therefore, it is incumbent upon AE language educators to expose learners to actual ATCO transmissions so that they may become accustomed to the pace, rhythm and length of standard phraseology productions.

In the next and final Chapter, implications of these findings for aviation human factors research and pilot/ATCO communication training will be discussed.

CHAPTER VI

IMPLICATIONS AND FUTURE RESEARCH

Scholars, educators and regulators around the world are currently scrutinizing the international language of aviation: Aviation English (AE). AE speakers come from every language background represented in the United Nations. Although it is based in wartime American English radiotelephony (Jones, 2003), AE cannot be said to be wholly derived from general American English, as the lexicon, syntax and grammar are unique. As non-native English speaking countries come to grips with training and testing protocols, native English speaking countries must face their own challenges in implementing international language guidelines and regulations. In the current climate of international concern about in Aviation English, it is critical for the US to take responsibility for the pilots and air traffic controllers (ATCOs) trained and employed there. In order to do this, we need to examine how US practices contribute to the problem of miscommunication in AE. Scholars have shown that the adoption of AE globally has created an environment of “language dominance” in which native English speakers (NESs) are given more flexibility and credibility in communications than non-native English speakers (NNESs) (Borowska, 2017a; Hansen-Schirra, 2013). The only way AE can be effective globally is if it is strictly adhered to, in a form reproducible by all of its users. AE is not native to any population. Everyone, even NESs, must learn it (Bieswanger, 2013; Borowska, 2017a; Estival, 2016).

Current ICAO and FAA regulations have created a legitimatized pathway to colloquial English usage in aviation environment under the auspices of “plain language” for communicating in non-routine situations. This avenue is taken by NESs more frequently than necessary and with a lack of language awareness as to what will be uninterpretable to NNES AE users. Miscommunications that result from NES conversational English (CE) use have been assumed to be the result of NNESs lack of CE fluency (Bieswanger, 2013; Prinzo et al., 2011). However, there is no reason to assume that NNESs fluent in AE standard phraseology should be able to understand CE. It has been documented that AE and CE are different varieties of English, existing in different realms of practice, with different interlocutor relationships, lexical referents and grammar (Beiswanger, 2016; Borowska, 2017c). It has also been remarked upon, but not demonstrated, that AE and CE differ prosodically (McMillan, 1998) and that untrained NESs cannot understand AE standard phraseology (SP) (Estival, 2016). In this dissertation, I demonstrated that AE is prosodically different from CE and unintelligible to fluent CE speakers, thus supporting the growing perception that requiring NNESs to have CE proficiency and allowing NESs to use CE in international airspace creates an imbalanced and unsafe environment.

6.1. Findings (What I Have Shown)

This dissertation demonstrated that AE is prosodically different from CE in ways that may make it unintelligible to conversational English speakers (see Chapter 2). It was then demonstrated the lack of AE intelligibility to NESs without aviation experience (see Chapter 3), while showing that AE is intelligible to non-native English speaker (NNES) pilots with relatively low conversational English ability (see Chapter 4). Furthermore, analysis of errors produced in repetitions of AE suggested frequency effects indicating that pilots learn AE vocabulary and phraseology throughout their flight experience (see

Chapter 5). This finding is particularly important because it implies that AE is indeed learnable through exposure.

6.1.1. Aviation English Is Prosodically Different than Conversational English

This study of AE prosody arose from an informal observation (by the author, who is a linguist and commercial pilot) that NES air traffic controllers (ATCOs) asked for clarification less frequently from NNES pilots with more native-like prosody than those with accurate segmental production and non-native-like prosody. Previous studies have described differences in vocabulary, grammar and environmental factors between AE and CE. They have also suggested that AE is faster and sounds different (i.e. more "machinegun-like") than CE. This dissertation demonstrated that consistent differences do in fact exist in the acoustic signal of these two language varieties. To demonstrate that rhythm, rate, and intonation are measurably different between AE and CE, I divided up and measured corpora of ATCOs' speech and radio broadcasters' speech using the same techniques. My analysis demonstrated that AE is produced more rapidly, with flatter intonation and has more similar sequential vowel durations and less similar consonant durations than CE.

These prosodic differences mean that AE users don't have the usual cues for disambiguating English language when listening to AE. Given that AE is produced in an environment that includes radio static, multiple speakers, and no face-to-face interactions, these missing cues can make AE difficult to parse for uninitiated first or second language English speakers. Additionally, differences in AE and CE could mean that CE experts may not necessarily have AE fluency and vice versa, which could affect assumptions underlying international regulations requiring AE and CE proficiency for pilots and ATCOs.

6.1.2. Conversational English Ability Is not a Reasonable Indicator of Aviation English Ability

This dissertation's findings support those of past studies, which indicate that AE and CE abilities are not substantially correlated (Moder & Halleck, 2009, Kim & Elder, 2015). Chapters 3 and 4 describe comparisons of three populations, representing different configurations of CE and AE language background, attempting to repeat ATCO speech.

6.1.2.1. Native English speaker non-pilots don't understand AE standard phraseology

First, I compared AE standard phraseology intelligibility to pilots and non-pilots from NES backgrounds (see Chapter 3). As expected, these populations had essentially identical CE proficiency (96% and 97% words correct, respectively) and pilots were more proficient in AE than non-pilots. However, the degree of difference was greater than might have been expected, given their commensurate CE abilities. Expert English speakers without aviation experience had very low intelligibility scores in AE (53% words correct), compared to NES pilots (85% words correct). Pilot scores also averaged lower than expected, prompting examination of within group factors and the discovery that, although all pilots were licensed to fly in the instrument environment, greater experience was still a factor in their AE proficiency level.

6.1.2.2. Non-Native English Speaker Pilots Understand AE Standard Phraseology

Next, I compared NES pilots to NNES pilots (see Chapter 4). Not surprisingly, NES pilots proved to be more proficient at both CE and AE than NNES pilots. NES pilots produced higher CE scores (96% words correct) than NNES pilots (43% words correct) and higher AE scores (85% words correct) than NNES pilots (70% words correct). These findings also demonstrated NNES pilots' superior AE proficiency compared to their CE proficiency.

Expanding language comparisons to include all three populations (see Chapter 5) demonstrated that, although NNES pilots were less proficient in CE (43% words correct) than NES non-pilots (97% words correct), NNES pilots were more proficient in AE (70% words correct) than NES non-pilots (53% words correct), which supports the hypothesis that CE proficiency is not sufficient to predict AE proficiency.

6.1.2.3. Error Patterns Across Groups

It is expected that NES groups and pilot groups would share some similar error patterns in their attempts to repeat ATCO AE productions. The error analysis supports this assumption. In addition, the two seemingly unrelated groups – NES non-pilots and NNES pilots – had some similar error patterns, possibly because they incurred higher cognitive load during the repetition task and shared some strategies that were apparent in their error production (see Chapter 5).

As demonstrated in prior studies (Barshi, 1997, Cardosi, Falzarano, & Han, 1998; Prinzo, 2008) length of AE transmission was the most consistent error predictor for all populations in this study. Another effect that appeared as an across-group error predictor was ATCO US regional accent and/or speech style. Also, as reported in prior studies, transmission speech rate was not a factor in error production for any of the populations (Farris & Barshi, 2013).

Word frequency in AE (for pilots, regardless of language background) and in CE (for non-pilots) also appeared to be a factor in repetition errors. Another interesting finding was that all pilots produced more errors when attempting to repeat transmissions that were not normally repeated in AE standard phraseology, suggesting a production frequency effect (rather than one of perception frequency).

Within group, all populations produced more errors in repeating non-number words than number words and more errors of omission than mistakes. However, pilots had equal percentages of mistakes that were numbers and words, while non-pilots produced many more word mistakes than number mistakes. These findings indicate that AE phraseology was as familiar to pilots as number terms and that number terms were more familiar to non-pilots than AE terminology.

One surprising finding in the study was pilots' single word omission errors that were classified as non-standard AE phraseology (and are therefore not included in proper readback protocol) were mirrored by non-pilots. This finding demonstrates that AE standard phraseology is well designed and that these elements are not salient in ATCO transmissions. It also suggests that the error rate I reported for pilots was conservatively high.

Regional accent or speech style was apparently more of a challenge for NNES pilots and less for NES non-pilots, with NES pilots scoring somewhere in between. These findings suggest that NESs may have more familiarity with regional accents and speech styles than NNESs. The difference between the NES pilot and non-pilot groups may be explained by the fact that these are measures of relative difficulty within each group, and

NES pilots were closer to ceiling in percentage of correct words repeated, so that regional accent had proportionally more effect on their proficiency.

For those populations with less familiarity in one or the other language varieties (AE or CE), the cognitive load of translating unfamiliar verbiage seemed to result in similar repetition strategies. NES non-pilots and NNES pilots produced a similar number of word omissions. However, the NNES pilots seemed more reticent to make word substitution mistakes than were the NES non-pilots, preferring to omit words altogether rather than say the wrong one. NES non-pilots made the most word mistakes of any population.

Overall error patterns between the groups indicate that pilots have learned AE vocabulary and phraseology from experience in the flight environment. Although NNES pilots were more error-prone than NES pilots, they were less likely to produce word mistakes than non-pilots, suggesting their level of AE understanding precluded them from stating an inaccurate intention.

6.2. Implications (Why This Study Matters)

As long as communication continues to be a factor in civil aviation accidents and aviation continues to include more multicultural interactions, there must be attempts to clarify and implement best AE practices. It is important that all AE users, regardless of language background, are confident and capable of communicating in the language of international aviation. Even the addition of data-link textual communication does not preclude the need for verbal interactions, though it may mean less crowded radio frequencies with more flexibility to speak slowly and ask for clarification. Although potential danger exists in imposing a global aviation language, it is necessary to create an environment where flight crews are aware of each other's intentions. This element of situational awareness arising from a shared radio frequency is sometimes called the "party-line effect" and can only occur if everyone is using the same language.

In an attempt to mitigate the safety threat created by imposing a global aviation language, international regulations have mandated proficiency in AE and CE for all international pilots and ATCOs. Implementation of these requirements creates different problems for NES and NNES AE users, which are not all addressed by the current regulations. First, NESs tend to deviate from AE standard phraseology under the guise of "plain English", causing interpretation problems for other NESs and for NNESs. On the other hand, the requirement for NNESs to have proficiency in conversational English is an unnecessarily heavy burden, given the subtlety and nuance of CE and the demonstrated differences between CE and AE standard phraseology.

6.2.1. Problems with Aviation English Communication

Past research shows that NESs and NNESs alike are the sources of miscommunications in the Aviation environment (Cookson, 2011; Cushing, 1994; Kim & Elder, 2009; Orasanu, Fischer, & Davison, 1997; Tajima, 2004; 1997). Two of the primary causes of these problems are deviations from standard phraseology and obscure or simply misunderstood "plain English".

6.2.1.1. Native English Speaker AE Miscommunication

Use of non-standard phraseology resulted in 101 deaths when Eastern airlines (EAL) flight 401 crashed in the Florida Everglades in 1972. As the aircraft continued descending below their assigned altitude, the ATCO queried the crew by saying, "How are *things* comin' [sic] along out there?" To which the crew, who had just resolved a landing gear issue that had distracted them from altitude awareness stated, "*Okay*" (emphasis mine). The plane crashed 30 seconds later (National Transportation Safety Board, 1973). Neither of these transmissions was in proper standard AE phraseology, which should have proceeded, ATCO: "request current altitude", pilot: "(current altitude) niner hundred". This type of interaction would have alerted both the crew and ATC that the aircraft was not at its assigned altitude of 2000 feet. As a result, inaccurate interpretation of the ambiguous referent "things" on the part of the crew and the vague response "okay" by the ATCO culminated in neither party being aware of the imminent danger.

The above scenario illustrates that NNESs are not the only AE users to suffer from ambiguous use of "plain language" in non-routine situations. Because of the "plain English" caveat, NESs often fall into colloquial or technical speech during interactions with ATC, especially when there is an added element of stress. This habit becomes more of a problem when NES pilots interact with NNES ATCOs. The fact is, most pilots and many ATCOs do not have an accurate concept of what would be helpful for a NNES, or what would be considered "plain English" by an interlocutor. Many ATCOs are quite skilled in clarifying transmissions to frightened student pilots or non-local pilots. Some even understand the type of clarity that is necessary to convey a message to a NNES. However, this is not a part of their training. Indeed, it is described as an "ethical obligation" in the ICAO language proficiency implementation guidelines (ICAO, 2004, 3.2.1.), but is not trained or tested for in any NES licensing process. The fact is this is not a simple thing to do. All of these interlocutors have different needs. Even a linguist would be hard pressed to spontaneously produce clear communication with each of these interlocutors. This is just one more reason that AE standard phraseology needs to be expanded and built on, so that, rather than be responsible for the entire repertoire of CE, AE users could command a limited, but more comprehensive, AE register.

6.2.1.2. Non-Native English Speaker AE Miscommunication

The highest loss of life aviation accident occurred on the ground in Tenerife, 1977, as a result of miscommunication between a Dutch first language (L1) pilot and a Spanish L1 ATCO. When KLM flight 4805 ran into a PanAm wide body jet on takeoff in low visibility conditions, 583 lives were lost. Miscommunication in this situation was the upshot of a series of unusual events leading to this accident. While the Pan Am 747 was still taxiing back on the runway, the KLM 747 captain conveyed to the tower ATCO, “We are *at* takeoff” (emphasis mine). To which the ATCO responded, “Okay”. Had the pilot used standard phraseology and stated, “taking the active runway”, and the ATCO queried the meaning of this statement OR paraphrased it, rather than replying, this accident might never have happened. Stress from low-visibility conditions and abnormal procedures added to the confusion for both interlocutors in this situation and may have caused the pilot to revert to the Dutch use of a preposition plus verbal infinitive meaning to be *in the act of Verb-ing*. The Spanish ATCO, perhaps assumed that the pilot had elided the word *position* (Cookson, 2011), in which case, the pilot would have meant that they were stopped on the runway, awaiting takeoff clearance from the tower.

Whatever the final analysis, had the interlocutors in the above situations used AE standard phraseology and readback protocol, these fatal accidents (and others like them) would not have taken place. The idea that interlocutors are allowed to use non-standard “plain English” in unusual situations opens the door to miscommunication (Day, 2004; Howard, 2008; Moder, 2012).

6.2.1.3. Deviations from Aviation English Standard Phraseology

NNESs are not the only ones who suffer from use of non-standard phraseology. Although it seems that the allowance for “plain English” would ease the burden of communicating in the highly formulaic, and sometimes awkward, realm of AE standard phraseology, the flexibility it affords causes problems between NES pilots and ATCOs. CE is highly nuanced and context-dependent. Multiple meanings are typical in CE, in which they are differentiated by intonation, gesture or facial expression. These types of clues are not available in the aviation environment, where there is virtually no intonation nor time for pauses, nor face-to-face contact.

The EAL 401 accident is a good example of how ambiguous CE can be. Both parties produced perfectly acceptable responses that would have been judged appropriate by any native speaker. However, in the time constrained, dynamic environment of low altitude / low airspeed flight, these transmissions were absolutely inappropriate and, it turned out, deadly.

As it stands, there is no definition of plain English. To this day, even with all of the discussion of communication problems and the implementation of international regulations for AE proficiency, there are no formal restrictions on “plain English”. Instead, to accommodate this speech style, NNESs are required to command CE proficiency. Not only does this place an unfair burden on NNES pilots and ATCOs, there is no reason to believe that it would alleviate communication problems given the idiosyncratic interpretation of “plain English”.

On the other hand, there exists the extremely functional tool of AE standard phraseology, which is rigidly defined: form and function mapped exactly. Certainly, this is the foundation of a shared language for emergent and unusual situations, as opposed to wildly variable and ambiguous conversational English. It is important, however to address the issue of AE plain English abuse. Why do NESs revert to colloquial or technical English when they are under pressure? What can we expect NNESs to do in similar circumstance? It is documented that it takes more mental energy to produce one's second language in a period of high cognitive load than one's first language (Ganushchak & Schiller, 2009). This is true whether the second language is AE or CE. In other words, NESs revert to their L1-- CE -- and NNESs try to revert to their L1 as well, as is seen in communication records of accidents involving crews sharing a non-English L1. In the case of the Korean crew of Asiana flight 214 that crashed in San Francisco a few years ago (2013), transcripts of the last few minutes of flight deck communications prior to impact reveal that the crew was speaking to one another in Korean and to the tower controller in AE (NTSB, 2014). The cognitive load of translation during incidents like these increases work load for NNES crews, even more so if the interaction does not conform to AE standard phraseology.

NNES crews are not given the option of communicating in their L1, unless they are interacting with an ATCO who shares that language. Even in these cases, they are discouraged from doing so, since that deprives other nearby crews of information critical to their situational awareness (see Chapter 1). If, however, they are interacting with a NES who is also under duress, they are more than likely going to be confronted with interpreting CE, in all its confusing variability. This is not simply unfair but is a recipe for disaster. We need to supply these professionals with a form that suits all of their needs, a form that can become relatively automatic to all of them, so that, when their cognitive resources are needed for other important activities, they have language at their command to convey their needs and intentions.

6.2.1.4. Cognitive Load Increases Miscommunication

As described above, when situations require trouble-shooting or are stressful for other reasons (Tenerife was so foggy that the tower controller could not see the aircraft on the runway), communication is more difficult (Ganushchak & Schiller, 2009), adding to the already high cognitive demands of the moment. Studies show that this effect is particularly onerous for second language (L2) speakers (Estival & Molesworth, 2016; Farris & Barshi, 2013; Farris, et al., 2008). There will be moments in aviation that are high workload, and high intensity. These will arise for all crews in airspace all over the world. We can only do so much to mitigate the communication demands during these circumstances. Making AE as predictable and consistent as possible is one way to decrease the cognitive load of production and perception. Accordingly, AE standard phraseology must be the same all over the world. Efforts to this end have been ongoing and are often spurred on by accidents such as Tenerife, from which language reform arose to limit possible interactions in the highly dynamic takeoff environment (Prinzo, Campbell, Hendrix, & Hendrix, 2010a). In order to mandate terminology, it needs to be streamlined and automatic, as much as possible. For that to happen, it has to be simplified and predictable. These are the tenets of AE standard phraseology.

6.2.2. Problems with Current AE Requirements for Non-Native English Speakers

Current regulations have sidestepped the issue of regulating "plain language" by demanding that NNES AE users exhibit proficiency in CE. Although this may alleviate a part of the problem, it creates others. NES AE users are usually not required to prove "plain language" proficiency and they are typically absolved of the need for recurrent language testing, since they score a "native" level 6 on their AE exams. This score does not indicate in any way their ability to negotiate meaning or to clarify communication in a multilingual environment (Bieswanger, 2013; Borowska, 2017b; Farris, 2016). On the other hand, regulations require NNESs to prove they can communicate in CE about a wide variety of aviation-related issues (Kim & Elder, 2015).

6.2.2.1. Problems with Assumptions Underlying ICAO Regulation

Recent studies have pointed out the lack of reasonableness in the assumptions underlying International AE regulations. Kim and Elder (2015) undertook a study of professional Korean pilots and ATCOs to examine NNES AE users' opinions about the validity of the regulations. Their findings indicate that Korean aviation professionals believe that CE is not pertinent to their work. The participants also point out that one of the biggest problems with AE is NES pilots' use of non-standard phraseology. They suggest that what is needed is expanded standard phraseology (Kim & Elder, 2015).

In a separate study, Moder and Halleck (2009) examined professional Chinese ATCOs and found that there is no correlation between CE and AE ability. These findings are consistent with those of this dissertation. Differences in AE terminology and prosody (see Chapter 2) would indicate that NNESs of CE might find less similarity than difference in these varieties. Indeed, my findings of higher AE proficiency than CE proficiency in NNES pilots would indicate that understanding and speaking CE could be an added burden, rather than a benefit for these populations, since they could be expecting CE-like productions and misunderstand AE.

6.2.2.2. Issues Reported with Training and Testing NNES AE Users

Scholars in ICAO member-states throughout the world are reporting problems with new training and testing tools. Many studies of AE learning tools report that texts do not show an awareness of the aviation environment or use AE terminology, sometimes having no reference to standard phraseology (Alshabeb, Alsubaie, & Albasheer, 2017; Melnichenko & Melnichenko, 2009; Zolfagharian & Khalilpour, 2015). Most AE textbook reviews also cite a lack of listening and speaking exercises, regardless of the ICAO emphasis on these tasks.

The findings of this dissertation indicate that, given the prosodic differences between CE and AE and demonstrated experiential AE acquisition, AE learners should be exposed to the unique prosody of AE in an environment similar to that in which they would be using it.

6.2.2.3. Aviation English Miscommunication Between NESs and NNESs

As illustrated above, NES and NNES users of AE have different communicative needs and challenges. One of the greatest problems for NES AE users is understanding AE produced with different accents or speech styles from their own, whether of a NNES or a NES from different regions (see Section 5.2.4.4). Studies show that NES pilots and NES ATCOs perceive NNES accent is one of the greatest contributors to AE miscommunication (Estival & Molesworth, 2016; Prinzo Campbell, Hendrix, & Hendrix, 2011; Tiewtrakul & Fletcher, 2010). This opinion is more typical of pilots than ATCOs, perhaps because ATCOs interact daily with AE users from diverse language backgrounds, whereas pilots may not. Although it was not the focus of the current study to examine the effect of accented AE exposure, data indicate that NESs had less difficulty with US regional accent and speech style than NNESs (see Section 5.2.4.4). Since NESs have presumably had more exposure to different American regional accents, this finding supports the theory that exposure enhances intelligibility and has implications for AE training (see Section 6.2.3).

6.2.3. Implications of Dissertation Research for Aviation English Training

Listening and repeating are key elements of the pilot's role in AE communication. Results from the second experiment in this dissertation indicate that NESs are not automatically able to listen to and repeat AE (see Section 3.3). Alternatively, NNES pilots are able to perform this task with much greater proficiency (see Section 4.2). As indicated by their low CE task scores (see Section 4.2.1) the NNES pilots have far less proficiency in CE than NES non-pilots in this study, therefore it is probable that their experience in aviation and exposure to AE enabled them to perform better on the AE task. These findings support Second Language Acquisition (SLA) theory that language is learned through exposure to spoken language (Pierrehumbert, 2003; Trofimovich & Baker, 2006). As further support of this theory, pilots with more flight experience fared better on the AE task than those with less experience, indicating that this learning is not done sufficiently before entering into the flight environment. Clearly more AE exposure prior to flight would be of benefit.

Second language acquisition literature can help guide development of AE training protocol. Findings from studies on accented English perception are particularly useful in the AE environment. AE scholars advocate the need for training NESs in accented AE (Borowska, 2017c; Farris, 2016), especially because of the NES tendency to lack linguistic awareness or accommodate their NNES interlocutors (Borowska, 2017a; Hansen-Schirra, 2013) (see Section 1.2.6). It has also been noted by AE researchers that NES pilots who have sufficient experience with different regional AE accents can interact proficiently in those environments. Prinzo, et al., stipulate that:

> Currency in flight time in the theater of operation is critical to understanding accented English and will influence how easily controllers from different countries are understood (2010a).

In support of this assertion, perception research shows that NES subjects exposed to NNE speech are able to make accent-specific generalizations across different speakers (Bradlow and Bent, 2008). However, pilots may not have to spend time in each particular area they intend to fly in prior to gaining proficiency in that specific accent. Other perception research indicates that subjects exposed to a variety of NNE speech from different language backgrounds are able to generalize learning to accents they have not been exposed to (Baese-Berk, Bradlow & Wright, 2013). These findings suggest that NES AE learners exposed to a variety of accented AEs could adapt readily to different speakers. Additionally, this exposure could be experienced on the ground, prior to flight training or prior to entering international air space.

AE training protocol development can also be informed by the area of speech perception research focusing on passive versus interactive language exposure. This research shows that language learners benefit from periods of passive exposure alternating with periods of active response or repetition. Baese-Berk showed that participants asked to repeat a novel phonetic contrast did not learn the category as well as those merely listening to it. These results suggest that learners may suffer from trying to produce language prior to sufficient exposure (2010). In a further study, Wright, et al. discovered that learners benefited from periods of repetition combined with periods of passive exposure, indicating that “practice-plus-exposure combinations may tap a general learning mechanism that facilitates language acquisition and speech processing” (2015, p. 928). Since only a small percentage of flight time is in direct communication with ATC and slightly higher percentage is in passive listening to ATC interacting with other pilots, these hours of communication could be estimated and used as the basis of a training program in which pilots are exposed to passive listening and active response to recorded actual ATCO transmissions. 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 to learning AE while learning to fly in the airplane. With this type of training, pilots could enter the flight deck with AE skills in place.

The studies undertaken in this dissertation examined populations of AE users in the US. Analysis of AE prosody was done on American English speaking ATCOs speech. I compared intelligibility between American English speaking pilots and non-pilots and NNESs training in the US responding to American English speaking ATCOs. Since many of the world’s pilots and ATCOs that make up the international community of AE users are trained in the US, it is critical that findings from this research are used to improve their training and strengthen regulations to make AE a reliable and safe form of communication.

6.2.3.1. Necessary to Focus on Aviation English Standard Phraseology in Training

Currently, in the US, AE standard phraseology is not specifically taught; it is learned in practice while learning to fly the aircraft. Often there are AE classes for NNESs training in the US, but it is taught in the form of “plain English”, manifest as CE with aviation terminology. However, this dissertation has demonstrated that AE is learnable without CE fluency. It has also demonstrated that AE proficiency continues to develop through the early years of a pilots’ career. Additionally, as illustrated in the above accident scenarios, deviation from AE standard phraseology causes miscommunication for NESs and NNESs alike. Therefore, AE standard phraseology should be taught prior to as well as simultaneously with flight training and its tenets adhered to by all AE users. A strong early foundation in AE will allow users to rely on it even under duress.

This dissertation demonstrates that AE and CE are prosodically different. Prosodic differences may be great enough that these two varieties of English are not mutually intelligible (see Chapter 2). Indeed, the study of NES and NNES pilot and non-pilot populations described in this dissertation suggests that this is the case (see Chapters 3, 4 and 5). NNES pilots were able to repeat AE transmissions much more proficiently than NES non-pilots, although their CE proficiency was demonstrated to be much lower. This finding indicates that AE standard phraseology can be learned without CE proficiency.

This dissertation also demonstrated that pilots do not exhibit adequate AE skills prior to licensure (see Chapter 4). Although the focus of this study was on pilots who were still acquiring their ratings, they were all licensed to fly and most of them were rated as instrument pilots. All of them had spent time flying in the Instrument Flight Rules (IFR) environment, which is primarily the domain of commercial aviation. The small portion of pilots with higher flight time who participated did much better at repeating AE transmissions, indicating that learning continues to take place as pilots gain flight experience. Since AE proficiency is gained through exposure and pilots show a lack of proficiency early in their careers, it should be possible to supplement language exposure during flight with dedicated language training on the ground, if the exposure is sufficiently realistic.

Flight training is time consuming and expensive. Learning AE while learning to fly is a waste of resources. Learning CE in the classroom while learning to fly is also a waste of resources, both fiscal and mental. Instead, time in the classroom could be dedicated to AE. Pilot trainees already are tasked with learning multiple aircraft systems, meteorology, aerodynamics and AE. The idea that they could have enough remaining cognitive resources to learn conversational English simultaneously is impractical.

With better training in AE, NES pilots would be less prone to lapse into “plain English”. Differences in vocabulary, grammar and prosody between AE and CE imply that switching between these language varieties takes effort for both interlocutors, adding to whatever confusion or challenge initiated the interaction (cf. Section 1.2.3.1.).

6.2.4. Expansion of Aviation English Standard Phraseology

Instead of mandating CE proficiency, the AE standard phraseology should be extended into non-routine and emergency situations that are frequent enough to be codified. And when “plain English” is called for, strict compliance to a defined grammar and lexicon should be mandated by all participating ICAO member-states.

6.2.4.1. Aviation English is Designed to be Clear and Concise

AE standard phraseology is the ideal basis for developing a language to describe non-routine situations in aviation. It is clear and concise, as opposed to non-standard “plain English” which can be ambiguous and generally requires more time and energy to understand, especially for NNESs. AE is already shared by the international aviation community and simply needs to be expanded into situations currently requiring exposition. The same tenets that served to make AE standard phraseology simple and clear could be consistently used in new constructions. The same lexicon and grammar, created to eliminate ambiguity, could be extended to support novel utterances.

6.2.4.2. High Cognitive Load Decreases Language Proficiency

Mandating CE usage in conditions including confusion, delay, and potential danger, when workload and cognitive load increases, is not productive, since it is exactly these conditions in which proficiency in EL2 decreases (Farris et al., 2008). Farris et al. simulated aviation navigation study indicates that fluency decreases most for low proficiency EL2 subjects in cognitive load conditions. The same study also found that fluency was worse for NESs in the high cognitive load condition.

6.2.4.3. International Scope of Aviation Community

Not only is conversational English potentially ambiguous, but it differs depending on language background. Patterns and pronunciation from one’s first language are carried over into the grammar and production of a second language (Flege, 1987). Findings in this dissertation indicate that US regional accents and speech styles affect intelligibility for NES pilots. This effect is manifold in international airspace where crews and ATCOs from all over the world interact. There are currently more NNES AE users than NESs. In order to avoid ambiguity, emergency communications, in particular, need to be codified. There is no room for confusion in the time-critical, high-stakes realm of international aviation.

6.2.5. Relevance of This Work to More General Issues in Human Factors

As indicated in this study of aviation language, communication is not always given its due weight in the examination of human factors in a technical task environment. The goal of pilots and controllers is to move aircraft from point A to point B in a timely and safe manner, which involves multiple layers of human / machine interaction. In such a dynamic, complex environment, verbal communication may appear to be the least taxing and most natural of the required tasks. However, this assumption of naturalness remains untested. Language is learned, over years of exposure and practice. On the surface, there is little or no conscious effort in adults' production and perception of language. However, all communication comes at a cost. Even fluent, successful communication takes energy. Variables that increase the cost of communication in aviation include code switching / translation, lexical access, visual access, and audibility. With the addition of each of these factors, the cognitive load of communication is increased, effectively reducing AE users' cognitive resources that may be needed for the performance of other critical flight tasks.

A frequently cited caveat in flight training is "aviate, navigate, communicate". This mnemonic is meant to remind pilots that their first job is to keep the plane in the air, then to get to their destination and, finally, to let someone know how they are planning on accomplishing these goals. However, aviating and navigating require verbal input from controllers (organizing the flow of air traffic) and pilots of other aircraft (sharing the airspace), in addition to fellow crewmembers. Communicating is a necessary part of every aspect of aviation. The cognitive load of this communication must be assessed, analyzed and, when possible, reduced, so that AE users can focus more of their attention and energy on actions that will keep the plane in the air and on its way.

As with any other activity, the increased cognitive load of communication is more apparent for NNES AE users. This population obviously has to engage in translation and face possible increased pronunciation and perception challenges. Cognitive load of communication is not so obvious for NES AE users, who are seemingly engaging in a language that they have spoken all of their lives. However, as can be seen in the phonological analysis of AE v. CE in this dissertation, NESs are actually engaging in a different variety of that language, with specific grammar and terminology distinct from what they are indoctrinated in. Every time a NES AE user transitions from speaking to their fellow NES crew-members in CE to speaking to ATC in AE, they are code-switching which requires cognitive resources.

For pilots and controllers of all language backgrounds, every time they have to strain to hear or interpret an interlocutor, they are using cognitive resources. Every time they have to create a novel utterance or recall an infrequent or poorly learned construction in AE, they are using cognitive resources. And, as has been illustrated in Chapter 1 (Section 1.2.3.1.), whenever there is confusion in the conveyance of a concept in AE, cognitive resources are taxed to negotiate meaning between interlocutors. Increased cognitive load of communication creates confusion and delays, affecting the smooth flow of activity on the flight deck and in the control room and potentially of air traffic in general.

Much of the difficulty involved in AE communication could be mitigated through consistent, predictable AE use. Through proper training, AE users can be acclimated to production and perception of AE phraseology and accented AE. Non-routine circumstances could be addressed through memorizable chunks of precise language. AE use should be efficient and stress-free, so that flight crews and controllers can do their jobs.

In the broader field of human factors in human-machine interaction and safety critical environments, communication should be a primary field of scientific inquiry, on a par with ergonomics and attention span. Currently, communication is often overlooked in these assessments. Even in forums specifically dedicated to aviation human factors, communication analysis is severely underrepresented. In a recent conference of aviation human factors experts, only 4 out of 123 presentations addressed communication issues (International Symposium on Aviation Psychology, 2017).

As in medicine or business and many other fields in today's global economy, aviation requires individuals from diverse language and cultural backgrounds to work together to attain specific goals. Aviation's safety-critical, time-sensitive environment is perhaps more demanding, but no less important to participants in other fields where professional duties require clear, concise communication. For this to be feasible, there should be no doubt as to the meaning of an AE utterance. Just as the movements of control levers on the flight deck must lead to a single mechanical response, so must the words uttered by a controller or pilot have one conceivable result. And, just as pilots must learn the proper physical action in each flight situation, so must they learn the proper verbal action in any scenario. In order for pilot actions and words to be predictable and appropriate, their training protocol must address human factors that may prohibit or deter them from that performance.

What I have shown here, in elaborating on the phonological and intelligibility differences between AE standard phraseology and plain English is that the assumption that individuals can communicate via radio using colloquial forms of English in addition to standard phraseology does not take into account the real differences in these language forms. Recognition of these differences indicates the need for specific training in AE.

6.2.5.1. Revising Training Protocol

This dissertation illustrates one relevant method for examining AE training efficiency. The studies in this dissertation indicate that training in actual ATC productions, including regional accents will enable AE users to acquire fluency interacting with diverse AE interlocutors. If left to learning on the job, AE users may indeed attain a level of automaticity, which could fail them in critical situations.

> It is a commented fact that cosmetic behaviors crumble under stress and a reversion to native behaviors takes place. The safety concern about dealing with cross-cultural interfaces through adaptation is obvious; in stress situations adaptation may become ineffective. (Merritt & Maurino, 2004, p. 156)

Therefore, targeted AE language training needs to be given so that AE users have a solid base of experience to draw upon.

6.2.5.2. Towards Designing New Training Protocols

Differences in prosody between AE and CE and the suboptimal environment of pilot / ATCO communication, indicate that AE training should be based on hearing and speaking AE in realistic conditions including head phones, radio static, realistic speech rate, and different ATCO / pilot accents and speech styles. Findings demonstrated in this dissertation suggest that frequency also effects AE proficiency (see Section 5.2.4.4). Accordingly, emphasis in training should include a focus on infrequent AE standard phraseology.

AE language training on the ground in a stress-free, dedicated environment would be less expensive and more efficient than passively acquiring AE while flying. Levels of language training appropriate to a student's current flight training goals could be integrated into the flight training protocol, just as other training modules on aircraft systems and meteorology are staged throughout current training protocols. This type of language training would focus on AE perception and production through headphones in realistic conditions, using actual ATCO transmissions with static, multiple speakers, and different language backgrounds.

6.3. Future Directions and Suggestions

Several areas of study are indicated to enhance the findings of this dissertation. Prosodic information should be further analyzed for rhythm patterns of different information categories, such as given v. new, numbers v. words, and frequent v. infrequent, standard phraseology v. plain English. Intelligibility studies should be followed up using targeted populations of different L1 AE users and different flight experience levels.

6.3.1. Balanced, Usage-Based Needs Analysis of Aviation English

This dissertation has outlined possible sources and solutions to particular aspects of the AE communication problem. Mainly it has demonstrated the inequity in the current regulations and the need for developing expanded AE standard radiotelephony that is accessible to AE users from all language backgrounds. It has also suggested possible improvements to AE training protocols that are inexpensive and effective. However, it has not addressed exactly how AE be expanded. In order to address this problem intelligently, a needs analysis must be undertaken focusing on how to address AE concepts currently addressed via "plain English".

Of primary concern in regard to pilots and ATCOs trained in the US is compliance with ICAO phraseology. Since compliance is still a matter for each individual state to decide, there are a number of ICAO terms that have yet to be incorporated into the US regulatory FAA AE standard phraseology. Work on streamlining and creating a consistent lexicon has been ongoing and continues to refine international AE standard phraseology. However, more could be done at the grassroots level to ensure that new AE users learn the proper form. One problem unique to the US is that we may have international students training at regional US airports, that are not required to speak in ICAO condoned AE. ATCOs at these airports may not be cognizant of ICAO phraseology and are not required to use it (Prinzo & Campbell, 2008). Even ATCOs at international airports regularly stray from AE standard phraseology. Sahliger and Renn state that their "… research has shown that ATCOs start to change words and phrases after 20-30 minutes" because "… it is extremely difficult to repeat the exact same phrases over and over again in a situation where all other tasks demand high awareness and professional handling of difficult situations" (2013, p. 140).

6.3.1.1. Developing AE Standard Phraseology for Non-Routine Events

First, we must establish where and when "plain English" is used. Many studies mention non-routine situations in which it may be necessary to address concepts that are not codified in the standard phraseology. For instance, Zolfagharian and Khalilpour list

"… sick passenger, runway incursion, excursion, shortage of fuel, fire, dangerous goods, pressurization, landing problems, strikes, communication failure and etc." among areas to be covered in plain AE (2015, p. 49). However, (especially NES) AE users resort to plain English for a myriad of other reasons, as well. Commercial pilot and human factors specialist Cesar Holzem enumerates reasons for deviations from standard phraseology including: distraction, confusion, over-simplification, justification, deception, and over-compliance (2013).

Non-compliance with AE protocols is an issue that would best be addressed through language awareness education (Beiswanger, 2013; Borowska, 2017a; Farris, 2016; ICAO, 2010; Jones, 2003). Since this is normally a problem introduced by NESs, it should be a mandatory part of training curricula for NES AE users. It is also important to address language awareness across the L1 spectrum. It has been reported that when NNES ATCOs defer to what they believe to be the superior English language skills of NESs, they may not be able to interpret CE comments by NESs. A 1995 American Airlines accident occurred in Columbia because the NNES ATCO did not understand the transmissions from the NES flight crew. Rather than insisting on clarification, the ATCO deferred to the flight crew who proceeded to fly into a mountain, killing 151 people (McMillan, 1998, Prinzo & Campbell, 2008). NNES AE users need to be empowered to question and clarify non-standard AE usage.

There are non-routine situations yet to be addressed in AE standard phraseology that could be identified and addressed using the AE standard phraseology paradigm. Some of these situations happen frequently enough that new standard phraseology could be developed. For novel productions, and clarification requests, tools that ATCOs and pilots currently use could inform AE users about proper “plain language” usage. Borowska (2017b) cites “paraphrasing, repetition, grammar and vocabulary simplification, comprehension check, clarification request, confirmation check” as methods used by effective NESs interacting with NNESs in AE (Borowska, 2017b, p. 144).

Another issue to be addressed in the expansion of AE standard phraseology is that not all details of a situation need to be addressed. Part of language awareness training could be educating AE users to prioritize their needs and communicate them clearly. A fine example of NNES use of strict AE standard phraseology in an emergency situation was the successful avoidance of a midair collision by two Japanese airliners in German airspace (Koble & Roh, 2013), whereas other flights analyzed in this same study described NESs unsuccessfully using colloquial English for clarification with their NNES interlocutors, adding more confusion in emergency situations.

6.3.2. Wrap-Up

Linguistic study of Aviation English, such as in this dissertation, is increasingly important to ensure global flight safety (Borowska, 2017c). Given the specialized environment, training, regulations and vocabulary of this language, it cannot be seen merely as another English for Specific Purposes (ESP) and must be studied independently. Forces that combine to make AE unique include cultural and native language differences across myriad backgrounds. Indeed, international AE communication more often than not involves NNESs from different language backgrounds interacting with one another in AE. Here is certainly an opportunity to simplify critical messages and address imperative communication, outside of the social and cultural jockeying that occurs in conversational communication. We have the opportunity to find a pure reason language, based in mutual need for clarity and concision.

As an international community, we have a long way to go to change AE practices. But, we must continue to strive to make AE work equally well for all pilots and ATCOs. Leveling the field for all AE users means that NESs will have to become aware of how their language affects others and be willing to use simplified jargon to obtain clarification. But each member-state must buy in to the ICAO standard and the standard must be one of reliable, reproducible communications that everyone can have confidence in. There will still be individual differences in AE production and level of proficiency, but with proper training in accents and speech styles as well as terminology and syntax, AE users will be comfortable interacting with professionals from all language backgrounds. In fact, focusing on a more restricted AE plain language, will free up energy that has historically been less effectively dedicated to developing broad conversational English skills and valuable resources can be dedicated to AE training to enhance perception and production skills.

APPENDICES

Appendix A. Formulae for Rhythm Metrics

Eq. (A.1) $$Percent\,V = (dur_V / dur_{C+V}) \cdot 100$$

Eq. (A.2) $$deltaC = 100 \cdot \sqrt{\frac{n\sum_{i=1}^{n} C_i^2 - \left(n\sum_{i=1}^{n} C_i\right)^2}{n(n-1)}}$$

Eq. (A.3) $$VnPVI = 100 \cdot \left[\frac{\sum_{i=1}^{n-1} |x_v - x_{v+1}|}{n-1}\right]$$

Eq. (A.4) $$CrPVI = \left[\frac{\sum_{i=1}^{n-1} |x_c - x_{c+1}|}{n-1}\right]$$

Eq. (A.5) VarcoV = $(\Delta V / mean_V) \cdot 100$

Eq. (A.6) VarcoC = $(\Delta C / mean_C) \cdot 100$

Legend

dur = duration

V = vowel

C = consonant,

Δ = standard deviation

n = number of intervals

i = current interval

Appendix B. Working Memory Task

1. Instructions

After you hear each word, repeat the word you heard.

If you are unsure of the word then take your best guess.

You will then judge whether the word starts with a letter from the first or second half of the alphabet.

Say "first" if the word starts with the letters A through M. Say "second" if the word starts with the letters N through Z.

Press any key to continue.

Keep repeating and judging words until you hear a beep.

After the beep, repeat all of the words that you can in the set, in order, if possible.

You do not have to recall the judgments, just the words.

Press any key to continue.

We will start with 5 sets of two words, then increase the number of words until we reach 6 words in a set.

Please let the experimenter know if you have any questions.

Press any key to begin.

2. Sample Score Sheet

Randomization One Alphabet

Example for Alphabet: 2 block

Track #	Word	Recognition Response	Alphabet - First Half/2nd Half	Recall Response	
77	WEST		Second		
	FLAG		First		
78	ACHE		First		
	TUB		Second		

PROCEED IF SUBJECT IS ABLE TO DEMONSTRATE UNDERSTANDING OF THE INSTRUCTIONS

Randomization One Alphabet: 2 block

Track #	Word	Recognition Response	Alphabet - First Half/2nd Half	Recall Response	
2	GRACE		First		
	WHAT		Second		
3	FIST		First		
	SHOVE		Second		
4	RICE		Second		
	CALF		First		
5	MOON		First		
	THAT		Second		
6	BAR		First		
	RISK		Second		

Randomization One Alphabet: 3 block

Track #	Word	Recognition Response	Alphabet - First Half/2nd Half	Recall Response	
7	FLIP		First		
	BOOST		First		
	SAY		Second		
8	THREAD		Second		
	FALL		First		

	SIGN		Second		
9	**HAND**		First		
	WILD		Second		
	UP		Second		
10	**GOAT**		First		
	EAR		First		
	NUDGE		Second		
11	**PHONE**		Second		
	YET		Second		
	EARTH		First		

Randomization One Alphabet: 4 block					
Track #	**Word**	**Recognition Response**	**Alphabet - First Half/2nd Half**	**Recall Response**	
12	**WITH**		Second		
	BONE		First		
	PEARL		Second		
	MERGE		First		
13	**FUDGE**		First		
	POND		Second		
	IF		First		
	ROPE		Second		
14	**SAID**		Second		
	YEAST		Second		
	GERM		First		
	LOSS		First		
15	**OF**		Second		
	DEAD		First		
	KEEP		First		
	TREE		Second		
16	**CHIN**		First		
	PINK		Second		
	TAPE		Second		
	GIVE		First		

Randomization One Alphabet: 5 block					
Track #	Word	Recognition Response	Alphabet - First Half/2nd Half	Recall Response	
17	BATHE		First		
	USE		Second		
	IS		First		
	CHEEK		First		
	WALK		Second		
18	OUT		Second		
	BOOK		First		
	RAG		Second		
	STAR		Second		
	CLOWN		First		
19	FARM		First		
	PURSE		Second		
	THUMB		Second		
	FEEL		First		
	OAK		Second		
20	OWN		Second		
	FIT		First		
	VINE		Second		
	AT		First		
	LID		First		
21	LEAVE		First		
	AND		First		
	WIG		Second		
	SPIN		Second		
	GLOVE		First		

Randomization One Alphabet: 6 block					
Track #	Word	Recognition Response	Alphabet - First Half/2nd Half	Recall Response	
22	TOY		Second		
	HANG		First		
	REACH		Second		
	CLUB		First		
	HOME		First		
	NOOSE		Second		

23	**COW**		First		
	STOVE		Second		
	NEST		Second		
	ARM		First		
	ROD		Second		
	LOSE		First		
24	**URGE**		Second		
	HAT		First		
	YEARN		Second		
	DULL		First		
	NET		Second		
	CRUTCH		First		
25	**GAVE**		First		
	ELSE		First		
	ODD		Second		
	THIS		Second		
	BOAT		First		
	WHERE		Second		
26	**SHOULD**		Second		
	FEW		First		
	HOOF		First		
	TELL		Second		
	BURN		First		
	POOL		Second		

Appendix C. Standard English Performance Task

1. Instructions

You will hear 10 sentences. Please listen to the sentence and, after the prompt, repeat what you heard word for word.

When you are done speaking, press the space bar to hear the next sentence. The next sentence starts immediately. If you need more time, wait to press the space bar.

To move on from this screen, press the space bar.

2. Sentences

The small pup gnawed a hole in the sock.
The ship was torn apart on the sharp reef.
The fish twisted and turned on the bent hook.
Wipe the grease off his dirty face.
Slide the cATCOh back and open the desk.
The blind man counted his old coins.
She danced like a swan, tall and graceful.
A castle built from sand fails to endure.
Carry the pail to the wall and spill it there.
Cut the cord that binds the box tightly.

Appendix D. Aviation English Performance Task

1. Instructions

You will hear 84 Air Traffic Control transmissions. Please listen to the controller and, after the prompt, repeat what you heard word for word.

You will see the first few words of each transmission printed on the screen. This is the call sign of the aircraft being addressed. Please do not repeat the call sign, only what follows.

When you are done speaking, press the space bar to move on to the next transmission. Remember, the next transmission starts immediately. If you need more time, wait to press the space bar.

To move on from this screen, press the space bar.

2. ATCO utterances (call signs in parentheses)

(Delta nine forty six) descend and maintain three thousand, turn right heading two zero zero
(Delta four fifty five) descend and maintain three thousand till established now
(American eight fifty eight) when you reach four thousand reduce speed to one eight zero
(Delta seven fifty eight) descend and maintain five thousand, traffic's eleven o'clock three miles
(American twelve eighty six) reduce speed to one eight zero maintain four thousand
(Citation nine four golf) turn left heading three five zero and join the final for three six left
(Delta six ninety one) cleared visual approach three five right, contact tower one two six five five
(ASEA nine twenty two) descend and maintain three thousand report the airport in sight
(American two fifty six heavy) descend and maintain three thousand two hundred and ten knots on your speed now
(Southwest thirty four) climb and maintain eight thousand fly runway heading
(American twelve eighty six) climb and maintain one seven thousand contact fort worth center one three two point eight five
(Southwest two twenty eight) contact approach on one two five point two maintain four thousand
(Delta three nineteen) traffic twelve o'clock a mile, eastbound just about below you
(American six zero three) descend and maintain five thousand contact approach one two five point two

(American eight fifty eight) descend and maintain six thousand maintain speed two three zero
(American twelve seventy four) reduce speed two one zero traffic twelve o'clock four miles
(American fourteen oh six) descend and maintain five thousand fly heading zero eight zero
(American fourteen zero one) descend and maintain five thousand expedite your descent
(American eleven eighty three) traffic twelve o'clock five miles, northbound descending out of nine thousand
(Citation eight tango tango) heading one three zero descend and maintain seven thousand
(Swift six one) descend and maintain five thousand contact approach one one eight point one
(Cessna zero one charlie) advise traffic in sight, last transmission was broken
(Delta one ninety nine heavy) descend and maintain three thousand good rate of descent
(United eight ninety eight heavy) reduce speed to one nine zero turn right heading one eight zero
(Five nine uniform) proceed inbound runway two two left hold short of runway two seven
(U S Air six fifteen) traffic arriving runway two two left will hold short of your runway
(Ryan six seventy two) descend to four thousand turn right heading one five zero
(Ryan six seventy two) reduce speed to one seven zero turn right heading one five zero
(JetLink thirty eight forty five) traffic nine thirty to ten o'clock and three miles, southbound climbing out of two thousand two hundred
(AirNova eight fifty seven) fly heading three one zero contact approach one one eight point two five
(Continental seven seventeen) departure's radar contact turn left heading one zero zero
(American five thirty one) turn right heading one five zero contact approach one two zero point six
(American two ninety one) departure radar contact turn left heading one zero zero
(American two ninety one) turn right heading one six zero contact approach one two zero point six
(Pem thirty four ninety eight) turn right heading two seven zero runway two two right cleared for takeoff
(Cessna three eight foxtrot) wind two five zero at one four runway two two right cleared for takeoff
(Bizex four fifty six) traffic holding in position, cross runway two two right
(American two ninety one) turn left heading one four zero and contact departure
(Centurion three zero november) cancellation received squawk one two zero zero
(Centurion three zero november) turn left heading two one zero established on that heading the airport will be twelve o'clock and five miles
(Delta four thirty three) traffic no factor, turn right heading two zero zero
(November five alpha sierra) radar service terminated, squawk 1-2-0-0
(Delta three ninety four) contact tower one two six point five five

(Delta ten forty five) descend and maintain three thousand
(Continental five thirty one) reduce speed to two one zero
(American eight fifty eight) turn right heading two five zero
(American six eighty four) continue right turn heading one two zero
(Chaparral seven sixty four) reduce speed to one seven zero
(Delta ten sixty seven) cleared visual approach three six left
(ASEA nine twenty two) contact the tower on one two four point one five
(American eight ten) start reducing to one hundred and eighty knots now
(Continental ten thirty four) contact fort worth center on one three three point three
(Southwest three thirty seven) climb and maintain one seven thousand
(Delta ten nineteen) descend and maintain five thousand
(T W A three nineteen) fly heading one eight zero
(Continental five thirty one) good rate of descent to five thousand
(Delta seven fifty eight) when able reduce speed to one seven zero
(American six oh two) at five thousand reduce speed to two one zero
(Falcon nine nine two) turn left heading two seven zero
(American eleven eighty three) descend and maintain five thousand
(Delta eleven fourteen) one one eight point one
(Delta ten sixty seven) descend and maintain five thousand
(Arrow five six quebec) contact approach one two zero point five
(Bizex three twenty nine) turn right heading one three zero
(November two two zero one charlie) follow that traffic to runway two two left
(Shamrock ninety one seventeen heavy) reduce speed to two zero zero
(United eight ninety eight heavy) caution wake turbulence
(Navajo five nine uniform) descend and maintain three thousand
(Zantop nine twenty) descend and maintain three thousand
(Bizex four sixteen) reduce speed to one five zero
(Delta two oh eight) turn left heading three zero zero
(Shamrock ninety one seventeen heavy) climb maintain one zero thousand
(U S Air three seventy three) squawk three four two seven
(Northwest triple three heavy) contact boston approach one one eight point two five
(Continental triple three) squawk three four two three
(Delta eighteen thirty three) turn right heading two seven zero
(Nera thirty seven twelve) fly heading three three zero
(MedEx two ten) runway two two right cleared for takeoff
(Clipper five forty seven) contact departure
(Nera thirty seven sixty one) runway two two right cleared for takeoff
(JetLink thirty eight fifty) contact approach one two six point five
(Centurion three zero november) descend and maintain one thousand five hundred
(U S Air seven eighty one) turn right heading two zero zero
(United ten ninety six) contact approach one two six point five

REFERENCES CITED

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

Alshabeb, A. M., Alsubaie, F. H., & Albasheer, A. Z. (2017). English for Specific Purposes: A Study Investigating the Mismatch between the "Cutting Edge" Course Book and the Needs of Prince Sultan Air Base Students.

Aoyama, K., & Guion, S. G. (2007). Prosody in second language acquisition. *Language experience in second language speech learning: In honor of James Emil Flege, 17*, 281.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica, 66*(1-2), 46-63.

Baese-Berk, M. M. (2010). Learning novel phonetic categories in perception and production. *The Journal of the Acoustical Society of America, 128*, 2489.

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America, 133*(3), EL174-EL180.

Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S., & Kostadinova, T. (2003). Do rhythm measures tell us anything about language type. *Proceedings of the15th International Congress of Phonetic Sciences*, Barcelona, Spain, 2693-2696, ISBN 1-876346-48-5.

Barshi, I. (1997). *Effects of linguistic properties and message length on misunderstandings in aviation communication.* Unpublished doctoral dissertation, University of Colorado.

Barshi, I., & Healy, A. F. (1998). Misunderstandings in voice communication: Effects of fluency in a second language. *Foreign language learning: Psycholinguistic studies on training and retention*, 161-192.

Barshi, I., & Healy, A. F. (2002). The effects of mental representation on performance in a navigation task. *Memory & Cognition, 30*(8), 1189-1203.

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

Bertinetto, P. M., & Bertini, C. (2008). On modeling the rhythm of natural languages. *Quaderni del Laboratorio di Lingistica, 7*(1-23).

Bieswanger, M. (2013). Applied linguistics and Air Traffic Control: Focus on language awareness and intercultural communication. In S. Hansen-Schirra, & K. Maksymski (Eds.), *Aviation communication: Between theory and practice*: (pp. 15-32). Peter Lang GmbH, Internationaler Verlag der Wissenschaften.

Bieswanger, M. (2016). Aviation English: Two distinct specialised registers? *Variational Text Linguistics: Revisiting Register in English, 90*, 67.

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer. [Computer Program]. Version 5.4. 06. Retrieved February 21, 2015.

Borowska, A. (2017a). Is there any dominant culture in global aeronautical settings? *ББК 74.484. 4я431 С 83*, 137.

Borowska, A. (2017b). Aeronautical English: an analysis of selected communication strategies used by native English speakers in interaction with operational level 4 personnel. *Proceedings of National Aviation University, 71*(2).

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

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America, 121*, 2339.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707-729.

Cardosi, K., Falzarano, P., & Han, S. (1998). *Pilot-Controller Communication Errors: An Analysis of Aviation Safety Reporting System (ASRS) Reports*. Retrieved from

Cookson, S. (2011). Zagreb and Tenerife: Airline accidents involving linguistic factors. *Australian Review of Applied Linguistics, 32*(3), 22.1-22.14.

Craney, T. A., & Surles, J. G. (2002). Model-dependent variance inflation factor cutoff values. *Quality Engineering*, *14*(3), 391-403.

Cushing, S. (1994). *Fatal words: Communication clashes and aircraft crashes*. Chicago, IL: University of Chicago Press.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language, 2*(3), 133-142.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive psychology, 24*(3), 381-410.

Day, B. (2004). Heightened awareness of communication pitfalls can benefit safety. *ICAO Journal, 59,* 20-22.

Dellwo, V. (2006). Rhythm and speech rate: a variation coefficient for ΔC. *Language and Language-processing*, 231-241.

Dellwo, V. (2008). *The role of speech rate in perceiving speech rhythm.* Paper presented at the Speech Prosody Conference, Campinas, Brazil.

Dellwo, V. (2009). *Choosing the right rate normalization methods for measurements of speech rhythm*. Presented at the Associazione Italiana di Scienze della Voce (AISV), Venice.

Dellwo, V., & Wagner, P. (2003, August). *Relationships between rhythm and speech rate*. Paper presented at the 15th ICPhS, Barcelona. Retrieved from: https://pub.uni-bielefeld.de/publication/1785384

Dennis, W., (2015, November 26). Chinese ATC to adopt English-only policy. *AIN online.* Retrieved from: https://www.ainonline.com/aviation-news/air-transport/2015-11-26/chinese-atc-adopt-english-only-policy

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of phonetics*.

Duez, D. (1985). Perception of silent pauses in continuous speech. *Language and speech, 28*(4), 377-389.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: effects of talker and rate changes. *Journal of Experimental Psychology: Human perception and performance, 23*(3), 914-927.

Estival, D. (2016) Aviation English: A linguistic description, in D. Estival, C. Farris and B. Molesworth *Aviation English: A lingua franca for pilots and air traffic controllers* (Routledge Research in English for Specific Purposes). Taylor and Francis.

Estival, D., & Molesworth, B. (2011). A study of EL2 pilots'radio communication in the general aviation environment. *Australian Review of Applied Linguistics, 32*(3).

Estival, D., & Molesworth, B. (2016). Native English speakers and EL2 pilots: An experimental study. In D. Estival, C. Farris, & B. Molesworth (Eds.), *Aviation English: A lingua franca for pilots and air traffic controllers* (pp. 140-181). London: 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).

Farris, C. (2016). Aviation language testing. In D. Estival, C. Farris, & B. Molesworth (Eds.), *Aviation English: A lingua franca for pilots and air traffic controllers*. London: Routledge.

Farris, M. C., & Barshi, I. (2013). *Misunderstandings in ATCO communication: Language, cognition, and experimental methodology*: Ashgate Publishing: Surrey, England.

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.

Federal Aviation Administration (FAA) (2017). Aeronautical Information Manual, US Department of Transportation. Retrieved from http://www.faa.gov/atpubs.

Flege, J. E. (1987). The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*(1), 47-65.

Ganushchak, L. Y., & Schiller, N. O. (2009). Speaking one's second language under time pressure: An ERP study on verbal self-monitoring in German–Dutch bilinguals. *Psychophysiology, 46*(2), 410-419.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *The Journal of the Acoustical Society of America, 63*(1), 223-230.

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

Gorman, K., Howell, J., & Wagner, M. (2011). Prosodylab-aligner: a tool for forced alignment of laboratory speech. *Canadian Acoustics*, *39*(3), 192-193.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology*, *7*(515-546).

Guion, S. G. (2005). Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Studies Second Language Acquisition*, *27*(04), 503-533.

Guion, S. G., Flege, J. E., Liu, S. H., & Yeni-Komshian, G. H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics, 21*(2), 205-228.

Hansen-Schirra, S., & Maksymski, K. (2013). Linguistic Dominance in Air Traffic Control. In S. Hansen-Schirra, & K. Maksymski (Eds.), *Aviation communication: Between theory and practice*: (pp. 83-94). Peter Lang GmbH, Internationaler Verlag der Wissenschaften.

Hazrati, A. (2015). Intercultural Communication and Discourse Analysis: The Case of Aviation English. *Procedia-Social and Behavioral Sciences, 192*, 244-251.

Henton, C. (1995). Pitch dynamism in female and male speech. *Language & Communication*, *15*(1), 43-61.

Hieke, A. E., Kowal, S., & O'Connell, D. C. (1983). The trouble with " articulatory" pauses. *Language and Speech, 26*(3), 203-214.

Hinrich, S. W. (2008). *The Use of Questions in International Pilot and Air-traffic Controller Communication*: ProQuest. Retrieved from https://shareok.org/bitstream/handle/11244/7069/English%20Department_20.pdf?sequence=1.

Holzem, C. (2013). So much to say, so few words - Why pilots deviate from Standard Phraseology. In S. Hansen-Schirra, & K. Maksymski (Eds.), *Aviation communication: Between theory and practice*: (pp.55-76). Peter Lang GmbH, Internationaler Verlag der Wissenschaften.

Howard, J. W. (2008). “Tower, am I cleared to land?”: Problematic communication in aviation discourse. *Human Communication Research, 34*(3), 370-391.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1217.

International Civil Aviation Organization (ICAO) (2001). *ICAO Annex 10: Communications Procedures*. Montreal: International Civil Aviation Organization.

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

International Civil Aviation Organisation (ICAO) (2007). *Manual of radiotelephony*, *4th ed*. ICAO Document 9432-AN/925. International Civil Aviation Organization: Montreal, Quebec, Canada.

International Civil Aviation Organization (ICAO) (2010). *Manual on the Implementation of ICAO Language Proficiency Requirements, 2nd ed*. International Civil Aviation Organization: Montreal, Quebec, Canada.

International Civil Aviation Organization (ICAO) Annual Reports last retrieved January 30, 2017, from http://www.icao.int/about-icao/Pages/annual-reports.aspx.

International English Language Testing System (IELTS) 2017. Retrieved from https://www.ielts.org/about-the-test/how-ielts-is-scored on July 20, 2017.

International Symposium on Aviation Psychology (2017). Retrieved from https://isap.wright.edu/sites/default/files/ISAP2017-Final.pdf

Johnson, K. (2004). *Massive reduction in conversational American English.* Paper presented at the Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium.

Jones, R. K. (2003). Miscommunication between pilots and air traffic control. *Language problems & language planning, 27*(3), 233-248.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational psychologist, 38*(1), 23-31.

Kendall, T. (2013). *Speech rate, pause and sociolinguistic variation: studies in corpus sociophonetics*: Palgrave Macmillan.

Kerns, K. (1991). Data-link communication between controllers and pilots: A review and synthesis of the simulation literature. *The International Journal of Aviation Psychology, 1*(3), 181-204.

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*. https://doi.org/10.1093/applin/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.

Kim, H., & Elder, C. (2015). Interrogating the construct of aviation English: Feedback from test takers in Korea. *Language Testing, 32*(2), 129-149.

Koble, S., & Roh. P. (2013) Linguistic Characteristics of Aviation English and their Practical Use -An Analysis. In S. Hansen-Schirra, & K. Maksymski (Eds.), *Aviation communication: Between theory and practice*: (pp.43-54). Peter Lang GmbH, Internationaler Verlag der Wissenschaften.

Lang, H. (1975). A computer based analysis of the effects of rhythm modification on the intelligibility of the speech of hearing and deaf subjects.

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

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language and Speech*, *43*(4), 377-401.

Lozito, S., Verma, S., Martin, L., Dunbar, M., & McGann, A. (2003). The impact of voice, data link, and mixed air traffic control environments on flight deck procedures. *Air Traffic Control Quarterly, 11*(4), 293-310.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465-494.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General, 134*(4), 477.

McMillan, D. (1998). *Miscommunications in Air Traffic Control.* Masters Thesis, Queensland University of Technology. Retrieved 2-7-2018 from http://www.aviationenglishclasses.com/libary_contents/Say%20again%20Communication%20Issues.pdf

Mehler, J., Sebastian, N., Altmann, G., Dupoux, E., Christophe, A., & Pallier, C. (1993). Understanding compressed sentences: The role of rhythm and meaning. *Annals of the New York Academy of Sciences*, *682*(1), 272-282.

Melnichenko, N., & Melnichenko, S. (2008). *What is this thing called radiotelephony?* Paper presented at the ICAO Language Proficiency Requirements Implementation Workshop, Almaty, Kazakhstan.

Merritt, A., & Maurino, D. (2004). Cross-cultural factors in aviation safety *Cultural Ergonomics* (pp. 147-181): Emerald Group Publishing Limited.

Moder, C. L. (2012). Aviation English. In B. Paltridge, & S. Starfield (Eds.), *The handbook of English for specific purposes* (Vol. 120, pp.227-242). Hoboken, NJ: John Wiley & Sons.

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

Morgan, N., & Fosler-Lussier, E. (1998). Combining multiple estimators of speaking rate. *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, 2*, 729-732. doi: 10.1109/ICASSP.1998.675368

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning*, *45*(1), 73-97.

National Transportation Safety Board (NTSB) (1973). Report number NTSB-AAR-73-14. Washington, DC. Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR7314.pdf

National Transportation Safety Board (NTSB) (2014). NTSB/AAR-14/01, PB2014-105984, Notation 8518, Adopted June 24, 2014. Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1401.pdf

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and performance, 24*(3), 756.

Orasanu, J., Fischer, U., & Davison, J. (1997). Cross-cultural barriers to effective communication in aviation. *Cross-cultural work groups*, 134-160.

Ostendorf, M., Price, P. J., & Shattuck-Hufnagel, S. (1995). *The Boston University radio speech corpus*, (LDC96S36). CD-ROM. Linguistic Data Consortium, Philadelphia, USA.

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

Pallier, C., Sebastian-Gallés, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory & cognition, 26*(4), 844-851.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, *162*(1), 8-13.

Philps, D. (1991). Linguistic security in the syntactic structures of air traffic control English. *English World-Wide, 12*(1), 103-124.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Doctoral dissertation, Massachusetts Institute of Technology).

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech, 46*(2-3), 115-154.

Pinet, M., Iverson, P., & Evans, B. G. (2011). Perceptual Adaptation for L1 and L2 Accents in Noise by Monolingual British English Listeners.

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.

Poirier, M., & Saint-Aubin, J. (1996). Immediate serial recall, word frequency, item identity and item position. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 50*(4), 408.

Prinzel III, L. J., Shelton, K. J., Jones, D. R., Allamandola, A. S., Arthur III, J. J., & Bailey, R. E. (2010). Evaluation of Mixed-Mode Data-Link Communications for NextGen 4DT and Equivalent Visual Surface Operations. *Air Traffic Control Quarterly, 18*(2), 177.

Prinzo, O. V. (2008). The computation and effects of air traffic control message complexity and message length on pilot readback performance. *Measuring Behavior 2008*, 188.

Prinzo, O. V., & Britton, T. W. (1993). *ATC/pilot voice communications-a survey of the literature* (No. DOT/FAA/AM-93/20). Washington, DC: Federal Aviation Administration, Office Of Aviation Medicine.

Prinzo, O. V., & Campbell, A. (2008). *US Airline Transport Pilot International Flight Language Experiences, Report 1: Background Information and General/Pre-Flight Preparation*. Retrieved from

Prinzo, O. V., Campbell, A., Hendrix, A. M., & Hendrix, R. (2010a). *United states airline transport pilot international flight language experiences, report 2: Word Meaning and Pronunciation*. Retrieved from

Prinzo, O. V., Campbell, A., Hendrix, A. M., & Hendrix, R. (2010b). *US airline transport pilot international flight language experiences, report 3: Language experiences in non-native English-speaking airspace/airports* (DOT/FAA/AM-10/9). Washington, DC: Federal Aviation Administration, Office Of Aerospace Medicine.

Prinzo, O. V., Campbell, A., Hendrix, A. M., & Hendrix, R. (2010c). *US airline transport pilot international flight language experiences, report 4: Non-native English-speaking controllers communicating with native English-speaking pilots* (DOT/FAA/AM- 10/12). Washington, DC: Federal Aviation Administration, Office Of Aerospace Medicine.

Prinzo, O. V., Campbell, A., Hendrix, A. M., & Hendrix, R. (2010d). *US airline transport pilot international flight language experiences, report 5: Language experiences in native English-speaking airspace/airports* (DOT/FAA/AM-10/18).Washington, DC: Federal Aviation Administration, Office Of Aerospace Medicine.

Prinzo, O. V., Campbell, A., Hendrix, A. M., & Hendrix, R. (2011). *US Airline Transport Pilot International Flight Language Experiences, Report 6: Native English-Speaking Controllers Communicating With Non-Native English-Speaking Pilots* (DOT/FAA/AM-11/04). Washington, DC: Federal Aviation Administration, Office Of Aerospace Medicine.

Prinzo, O. V., Hendrix, A. M., & Hendrix, R. (2008). *Pilot English language proficiency and the prevalence of communication problems at five US air route traffic control centers* (DOT/FAA/AM-08/19). Washington, DC: Federal Aviation Administration, Office Of Aerospace Medicine.

Prinzo, O. V., Lieberman, P., & Pickett, E. (1998). *An acoustic analysis of ATCO communication* (DOT/FAA/AM-98/20). Washington, DC: Federal Aviation Administration, Office Of Aerospace Medicine. Retrieved from: http://www.dtic.mil/docs/citations/ADA353962.

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/.

Rakas, J., & Yang, S. (2007). *Analysis of multiple open message transactions and controller-pilot miscommunications.* Paper presented at the The Seventh USA/Europe Air Traffic Management Research & Development Seminar.

Ramus, F. (2002). Language discrimination by newborns: teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition*, *2*(1), 85-115.

Ramus, F., Nespor, M., & Mehler, J. (2000). Correlates of linguistic rhythm in the speech signal. *Cognition*, *75*(1), AD3-AD30.

Robertson, F. A., & Johnson, E. (1988). *Airspeak*: Prentice hall international.

Roy, A. (2004). Secure aircraft communications addressing and reporting system (ACARS): Google Patents.

Sahliger, M., & Renn, O. (2013) Communication needs in a high risk environment. In S. Hansen-Schirra, & K. Maksymski (Eds.), *Aviation communication: Between theory and practice*: (pp. 131-143) Peter Lang GmbH, Internationaler Verlag der Wissenschaften.

Smith, M. R., Cutler, A., Butterfield, S., & Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech, Language, and Hearing Research, 32*(4), 912-920.

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.

Suksiripakonchai, W. (2012). ICAO Language Proficiency Requirements and Training Guidance for Thailand. *SDU Research Journal Humanities and Social Sciences, 8*(1), 175-184.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction, 4*(4), 295-312.

Tajima, A. (2004). Fatal miscommunication: English in aviation safety. *World Englishes, 23*(3), 451-470.

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics, 25*(1), 1-24.

Tallyntyre, S., (2014, June 16). Spanish pilots ordered to speak only English. *The Local.es***.** Retrieved from: https://www.thelocal.es/20140616/spanish-pilots-ordered-to-speak-english

Tiewtrakul, T., & Fletcher, S. (2010). The challenge of regional accents for aviation English language proficiency standards: A study of difficulties in understanding in air traffic control–pilot communications. *Ergonomics, 53*(2), 229-239.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28*(1), 1-30.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: first language and second language studies. *Journal of Phonetics*, *35*(4), 501-522.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America, 52*(4B), 1238-1250.

Working Auditory Recognition and Recall Measure (WARRM). (Smith, Pichora, Fuller, Wilson & Alexander, under revision)

Wright, B. A., Baese-Berk, M. M., Marrone, N., & Bradlow, A. R. (2015). Enhancing speech learning by combining task practice with periods of stimulus exposure without practice. *The Journal of the Acoustical Society of America, 138*(2), 928-937.

Zolfagharian, A., & Khalilpour, J. (2015). Evaluation and content analysis of “English for aviation for pilots and air traffic controllers” textbook as an ESP book. *International Journal of Language Learning and Applied Linguistics World, 8*(2), 36-50.