Examining Computer-Generated Aeronautical English Accent Testing and Training

by

Hyun Su Seong

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Systems Design Engineering

Waterloo, Ontario, Canada, 2025

Examining Committee Membership

Supervisors	Dr. Shi Cao Department of Systems Design Engineering Committee Chair of the Collaborative Aeronautics Program, WISA Univ. of Waterloo
	Dr. Suzanne Kearns Aviation, Faculty of Environment Founding Director of WISA
Internal Member	Dr. Siby Samuel Department of Systems Design Engineering Director of the User Performance Lab Univ. of Waterloo
Internal Member	Dr. Lisa Aultman-Hall Department of Systems Design Engineering Professor and Chair of Systems Design Engineering
Internal-External Member	Dr. Katherine White Department of Psychology Developmental Research Area Head Univ. of Waterloo
External Member	Dr. Philippe Doyon-Poulin Department of Mathematical and Industrial Engineering Polytechnique Montréal

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Objective. This thesis focused on the persisting problem of language-related issues, in pilot-air traffic controller (ATC) communication, particularly in relation to foreign accents interfering with pilots' understanding. It examined the effect of foreign accents embedded in human and computer voice (HV, CV), as well as demographic background on the level of understanding of the participants.

Background. Studies focusing on the impacts of foreign accents in Aviation English (AE) are scant. Accents have been identified as one of the main contributors to miscommunication between pilot-ATC radiotelephony communication in the air, thereby endangering flight safety. It is necessary to examine how to train ab initio and returning pilots on extracting accurate meanings from an accented instruction coming from ATCs. This thesis introduces a Text-to-Speech (TTS) supported by artificial intelligence for such training.

Method. Multiple studies (a total of six) were conducted: 2 literature reviews, 4 empirical studies. For the empirical studies, 50 participants from the University of Waterloo who had experiences with flight or had experience in listening to pilot-ATC communications were recruited. They were put into two Voice Groups (HV and CV) one of which played only human voices and the other TTS. They completed two rounds (Round 1 and 2) of listening tests that contained both Aviation Script (AS; scripts read in foreign accents that were related to aviation context) and Neutral Scripts (NS; non-aviation scripts read in foreign accents with no contextual background). The foreign accents used in the listening tests along with native-accented English were three of the ICAO's main languages: Arabic, Spanish, and French. Scores were analyzed according to the Script Types (NS, AS), Accents (Arabic, Spanish, French), Rounds (1 and 2), and Demographic Profiles (Age, Gender, Years of Speaking English, Flight Hours, Flight Ratings, Language Background, Familiarity with Arabic, Spanish, French, and Aviation English).

Results. For the empirical studies, in the HV group, participants improved their scores from round 1 to 2 in the AS portion of the tests. In the CV group, participants improved their scores in NS. Examination of demographic information showed that non-native English speakers (NNES) tended to perform more poorly on average than native English speakers (NES). Being familiar with Aviation English was beneficial for completing listening tests. Also, having a higher flight rating was beneficial. Having more years of speaking English was only partially advantageous. Post survey results were analyzed, and it was found that participants in the CV group found the speech mostly

unnatural. Those in the HV group also expressed difficulty in understanding due to accents but mentioned that the speech was clear, and scripts were representative of real-life pilot-ATC communication. Participants expressed foreign accents interfered with their process of logical deduction when choosing answers on the tests. Participants – regardless of whether they belonged to the HV or CV group – found NS difficult and challenging due to lacking contexts when answering questions on the tests. For AS, participants were able to piece together information using contextual knowledge related to aviation.

Condusion. Accents do interfere with pilots' understanding in radiotelephony communication by making extracting content challenging, which in turn makes interpreting messages or instruction difficult. This is an important finding as it will affect situational awareness to a certain extent when making decisions on the fly. Pilots have to multi-task whenever possible to keep the passengers safe and to find the best route to get to a destination that maximizes fuel efficiency but minimizes passenger wait times. Communication plays a large role in deciding the fate of an aircraft's journey. In this logic, accents can be said to be at the core of this overarching issue with language in the context of aviation. Therefore, training with a new technology such as TTS, along with other educational resources, could confer a valuable experience and exposure to pilots who are either beginning or re-starting their language training.

Acknowledgement

I want to express my gratitude to my supervisors Dr. Cao and Dr. Kearns for the guidance in accomplishing my PhD at the University of Waterloo. I have learned many important aspects of conducting research throughout my journey as a PhD student. I want to also extend my gratitude to the other examiners Dr. White, Dr. Aultman-Hall, Dr. Samuel.

I want to thank voice-over volunteers who provided their recordings of the scripts.

I want to thank my family from the bottom of my heart for their patience and unending support throughout the journey fraught with various challenges. Lastly, I want to thank my cat Kong, who has been and still is with me throughout the most turbulent times and stayed up late while I was writing my thesis.

Table of Contents

List of Figures

In the Appendix

List of Tables

In the Appendix A

In the Appendix B

List of Abbreviations

AE	Aviation English
AELTS	Aviation English Language Test Services (ICAO)
AeE	Aeronautical English
A.I	Artificial Intelligence
AICc	Akaike Information Criterion (corrected for small sample size)
ΔAICc	Delta AICc value
aGVIF	Adjusted Generalized Variance Inflation Factor
AR, ARB	Arabic (Accent)
AQP	Advanced Qualification Program
AS	Aviation Script
ASRS	Aviation Safety Reporting System (NASA)
ATC	Air Traffic Control
ATCO	Air Traffic Control Officer
ATIS	Automatic Terminal Information Service
CARs	Canadian Aviation Regulations
CBE	Competency-based Education
CBET	Competency-based Education & Training
CBA	Competency-based Assessment
CE	Colloquial English
CFIT	Controlled Flight into Terrain
CPL	Commercial Pilot's License
CRM	Crew Resource Management
CV	Computer Voice
CVAS	Computer Voiced Aviation Scripts
CVG	Computer Voice Group
CVNS	Computer Voiced Neutral Scripts
dB	Decibel
DV	Dependent Variable (Response Variable)
EBT	Evidence-based Training
EFL	English as Foreign Language
ELPAC	English Language Proficiency for Aeronautical Communication Test (EU)
EN, ENG	English
EPTA	English Proficiency Test for Aviation (Korea)
ER	Evidence Ratio
ES	Spanish (Accent)
ESL	English as Second Language
ENGSP	English for Specific Purposes
FCLTP	Flight Crew Licensing and Training Panel
FH	Flight Hours
FLR	Flight Ratings
FR, FRE	French (Accent)
F.AR, ES, FR, AE	Familiarity with Arabic, Spanish, French, Aviation English
GEN	Gender
HV	Human Voice
HVAS	Human Voiced Aviation Scripts
HVG	Human Voice Group
HVNS	Human Voiced Neutral Scripts
ICAEA	International Civil Aviation English Association

ICAO	International Civil Aviation Organization
IPA	International Phonetic Alphabet
ITU	International Telecommunications Union
IV	Independent Variable (Predictor Variable)
KSA	Knowledge, Skills, Attitude (CBE)
KSS	Karolinska Sleepiness Scale
LBG	Language Background
Log.Lik	Log Likelihood
LOSA	Line Operations Safety Audit
LPR	Language Proficiency Requirement
M/C	Multiple-Choice Questions
M0	The best model (most parsimonious) (AIC)
M1,2,3…	The next best models after M0 (AIC)
MPL	Multi-Crew Pilot License
NASA	National Aeronautics and Space Administration
NES	Native English Speaker
NNES	Non-Native English Speaker
NOTECHS	Non-technical Skills
NS	Neutral Script
OF	Old Female
OM	Old Male
PAE	Plain Aeronautical English
PANS-TRG	Procedures for Air Navigation Services - Training
PPL	Private Pilot's License
PE	Plain English
PL	Plain Language
RTC	Radiotelephony Communication
SAE	Simple Aviation English
SARPs	Standards and Recommended Practices (ICAO)
SID	Standard Instrument Departure
SOP	Standard Operating Procedure
SP	Student Pilot's License
SPH	Standard Phraseology
SPSCS	Samn-Perelli Crew Status Check
TEM	Threat and Error Management
TS	Total Score
TTS	Text-To-Speech
W_i	Akaike Weight
YM	Young Male
YF	Young Female
YSE	Years of Speaking English

Chapter 1. Introduction

When the Covid-19 pandemic hit in 2020, the demand and supply of air travel took a direct hit. A recent report by ICAO (International Civil Aviation Organization), published on January 27th, 2023, provided the world passenger traffic evolution statistics. Although the decline in total passengers around the world can be seen since 2019, the overall sharp decline which the pandemic caused is steeper than the financial crisis in 2008. In 2020, there was an overall reduction of 50% of seats and 2,703 million passengers were lost (ICAO Air Transport Bureau, 2023). The preliminary estimates they project for 2021 show a slight decrease in both statistics, showing that the economic impacts felt by the industry cannot be said to have recovered completely.

Airlines are a critical part of today's economy. The industry is traditionally considered an oligopoly market where a few competitors hold a large share of the pie, hence the barrier to entry for newer airlines and low-cost carriers is high. The competition within this oligopoly market can be aptly described as a "red ocean" environment, meaning there is only little to salvage for newer airlines in terms of profits. However, the situation itself can serve as a motivator for improvement as each airline is committed to bring not only the best service passengers look for, but also to lead the growth of job market in its periphery. It is through profits they can add useful services for passengers and other clients who look towards them for transporting goods and services across borders. It is through profits they can help boost a region's local economy through job creation. Job creation could involve the improvement of training methods for flight crew or ground-side employees. Considering the overall role of the airlines in today's economy, including job creation, investing in new training methods and adding new services, their impacts towards the community they directly or indirectly serve cannot be underestimated. In other words, the economic well-being of communities is intertwined with the fate of airlines.

It is important to understand what passengers think about aviation in general. After 9/11, passengers held fear of their plane being hijacked. This fear can crystallize into, figuratively speaking, a latent threat that can upset the demand and supply of air travel. A decreased load factor shown on an airline's balance sheet as a result of an external factor such as 9/11 may make both future customers and investors frown, which would in turn bring negative perception towards the airline in question. This in turn implies that airlines may need to produce a marketing and pricing strategy that targets safety-conscious customers, which may be time-consuming and resource-demanding depending on that airline's financial situation. After the pandemic, this belief may have gotten bolder which means that safety-conscious customers would avoid travelling by airplane for the near future. Furthermore, the shortage of pilots makes it difficult for airlines to recruit competent pilots. Looking at the current situation, the industry is riddled with problems of short supply of qualified pilots and less demand for air travel pushed by the pandemic stress. The slogan that ICAO promotes – "Uniting Aviation" – seems desperately fitting in today's harsh climate.

Current training standards are focused on "getting more flight hours", that is, the assumption is that the more hours a pilot has flown, the better his or her performance will be. But this traditional approach does not show other aspects of training, which could also be a valuable indicator of how a trainee will fare well in other aspects of flight. Hours are just numbers; these do not tell whether a trainee is competent enough in communication or in other non-technical skills (NOTECHS). Furthermore, just because a trainee has fulfilled hourly requirements of flight training does not indicate that he or she will perform well against unique standards set out by the airline they are part of. Rather than judging a trainee's flight skill by hours, weight should be placed on a set of core competencies, some of which could be denominated in hours. These competencies may align with the industry standards that continuously evolve which in turn may be based on the rate of technological change. This concept is termed Competency-based Education (CBE), sometimes referred to as Evidence-based Training (EBT). It has been adopted in Australia and is beginning to be adopted slowly in Canada.

It is important to realize that the inherent issues present in aviation are multi-faceted. For instance, there is no concrete definition of 'competencies' and what they purport to represent. A frequent question that is raised is whether competencies that are abstract such as communication can be quantified and directly measured. Not only that, airlines might have different sets of competencies that they specifically emphasize or look for based on their unique organizational culture. Whether flight schools across Canada know the detailed descriptions of these competencies is another concern. The standardization of competencies has not been widely communicated nor published elsewhere, although IATA (International Air Transport Association) and ICAO have provided their own versions on a set of competencies. This gives rise to problems that affect the quality of training, thereby training standards.

However, with the rise of CBE, language in aviation has now finally been brought into a spotlight. There is a sublanguage that is used by pilots and air traffic control officers (ATCOs). *Aviation English* (or *Aeronautical English*, abbreviated as AE or AeE hereafter) is an English for

Specific Purpose (ENGSP). This is linguistically different than the colloquial English people speak daily in their lives. It is further divided into a) Standard Phraseology (SPH) – a group of coded phraseologies that provide clear and succinct meanings when spoken, which should be used throughout a flight operation; and b) Plain Language (PL) – the opposite of SPH which means non-coded and colloquial use of English. Every trainee who is aspiring to become a pilot in the future or a returning pilot must go through ICAO's Language Proficiency Requirements (LPR), which is a test that assesses a pilot or ATCO's English verbal and comprehension skill across six categories: Pronunciation, Structure, Vocabulary, Fluency, Comprehension, Interaction. A rating scale is used to rank a candidate's skill level: 1 (pre-elementary) to 6 (expert). A trainee is expected to have level 4 (operational) across all categories, otherwise he or she is not fit to fly. Experts (level 6) are not re-tested, but those who achieved level 5 ("Extended Level") or 4 ("Operational Level") should be re-tested at least once every 6 and 3 years respectively (Friginal et al., 2020, p. 55).

What is concerning about pilot training regarding AE is that its training curricula are vastly different. ICAO does not oversee the creation and administration of AE tests; the making of a test is the responsibility of Civil Aviation Authority (CAA) of ICAO members (Friginal et al., 2020, p. 57). This leads to a multitude of training materials, either by a third-party company or an instructor who provides a private lesson, the quality of which will be also wildly different from a country to country.

It is also not well known how AE tests are created. That is, test constructs (types of questions and how these are created, for example) are somewhat difficult to find. Lack of resources on AE teaching and testing makes it even more challenging to develop more precise test constructs. According to Alderson (2010), language tests in aviation have implications that extend beyond just improving a pilot's language skills, which simply means that using AE correctly and effectively could prevent a serious incident or accident from happening. Borowska (2017) argued in favour of producing more meaningful research in AE, which is inadequate compared to other fields in linguistics (p. 45). Furthermore, Borowska (2017) argued that "scientifically, nothing has been done in this field apart from some lexicographic work or representations of aviation phraseology" (p. 49). On top of that, both Borowska (2017) and Friginal et al. (2020) noted that language-related issues in aviation have not been frequently documented or analyzed in detail, leading to a gap in knowledge in several aspects and ultimately an underestimation of negative impacts of language. These present a new challenge for the human factors experts who are well-versed in linguistics.

Research that thoroughly examines a specific aspect of pilot-ATC communication or dialogue can yield new insights, which in turn can be used to develop robust AE learning material.

For pilots, AE may come across as difficult not only because of standard phraseology (being precise when communicating), but also the presence of foreign accents embedded therein. ICAO Doc 9835 (2010) defines accent in a way that all speakers have an accent that is tied to a geographical place or inherited from a phonology of the mother tongue. In AE, accent has not been documented well but there are findings that differences in accent cause communication breakdown. Baugh and Stolzer (2018) found that both non-native and native accents (ex. southern accent in the U.S) can have a negative effect in the level of understanding in pilot-ATC conversation. This is in line with the finding in Tiewtrakul and Fletcher (2010) that accents influence the level of understanding, although in various degrees based on the interlocutors' linguistic background.

Academic fields such as psycholinguistics, acoustics, or communications seem to have more articles on accents, but in aviation literature, it does not seem so. Although such studies in the other fields can be helpful, they are not directly related to the concerns of aviation industry professionals. With CBE rising in aviation training, flight instructors would be interested in how to effectively teach ab initio pilots with a new technology to correctly extract accurate meanings from accented instructions coming from ATCOs. With the advancement of artificial intelligence (A.I) and text-to-speech (TTS) systems, incorporating new technological invention into a learning or training curriculum can be a fruitful endeavour that benefits both instructors and new pilot trainees in the long-run. Therefore, this thesis focuses on computer-generated accent and whether it can be used complementarily to real human accents when it comes to training listening skills for pilots.

A more specific question is whether the computer-generated accent can be said to be "recommendable" to trainees who engage in AE training. The word "recommendable" encompasses many different meanings, which could be affected by listeners' perception of A.I speech in general, vocabulary knowledge, or their linguistic background or actual flight experiences. Do listeners understand as well under "computer accent" compared to "human accent"? If not, what can be said about the role of a computer accent then? What backgrounds of listeners mostly affect their performance under the computer accent compared to human accent? Answering these two central questions could help the author answer whether introducing the computer accents to today's pilot AE training is a worthy endeavour.

1.1 Objectives & Overall Research Questions

Aviation English is relatively a new subject of study (Borowska, 2017, p. 45), therefore it is necessary to look into its past (how it became what it is today) and the present (what is mainly being studied in literature). To understand its depth of history, it is important to understand first what CBE is and how that has been impacting today's training standards for pilots. With the introduction of CBE, which accompanied changes in how a training method is created and delivered, language training is likely to be improved as well. This discussion pertaining to the change or improvement of training methods (or standards) is ultimately related to flight safety. Documenting the past and history of AE along with CBE would provide a reader with holistic understanding of why this research carries great importance in today's most innovative industry (aviation). In an occupation such as an aviation policy makers, not being aware of the past endeavours (i.e. knowledge of past training policies) to improve flight safety will result in the creation of a near-sighted policy that could affect the lives of passengers.

This thesis contains multiple studies (a total of six studies): Two literature reviews and four studies with data. The literature reviews contain a detailed overview and summary of CBE and how it might influence today's AE training curriculum. It explored the history and characteristics of CBE, advantages and disadvantages of its implementation. The four studies involve data obtained from listening tests consisting of two types of scripts (referred to as 'Script Types'): 1) Neutral Script (referred to as NS in which no background information is given and sentences are random by nature); 2) Aviation Script (referred to as AS in which there is a three-part conversation piece between a pilot and ATCO. Therefore, it has background context). There were two rounds which participants completed (Round 1, 2). The Script Types and Rounds were further categorized under the two types of voices: a) Computer Voice; 2) Human Voice, which were then referred to as the 'Computer Voice Group' (only used computer voices for playing both NS and AS) or the 'Human Voice Group' (only human voices).

Below outlines each study done in the thesis in the order.

Study 1. What is the relationship between CBE and Aviation English? (Chapter 2: Section 2.1)

- How is traditional pilot training different from competency-based training (CBE)?
- How is AE taught and evaluated?
- What could be said about AE in relation to CBE?

Study 2. What is the relevant literature to AE and Accent? (Chapter 2: Section 2.2)

- Is there sufficient literature on the impacts of foreign accent on pilot-ATC communication?
- What lessons can be learned from other disciplines that have investigated foreign accents?

Study 3. Are the scores in each Voice Group (CV, HV) clearly different? (Chapter 5)

- Is there an improvement in the scores of both Voice Groups?
- Do the Script Type and Rounds influence the scores in each Voice Group?

Study 4. Does demographic information of participants have any influence on the total scores in NS or AS, and in each Voice Group? (Chapter 6)

- Which of the following demographic profiles or information about the participants (Age, Gender, Years of Speaking English, Flight Hours, Flight Ratings, Language Background, Familiarity with Foreign Languages (Spanish, Arabic, French), Familiarity with Aviation English) can be said to be influential to the total scores (representing their comprehension) in NS and AS? That is, do some of the predictors in this list belong to a best model equation? (Chapter 6)

Study 5. In relation to Study 4 above, does having familiarity with the specific accents (ESP, ARB, FRE) influence the scores? (Chapter 7)

Study 6. On the Post Surveys, how did the participants in each Voice Group rate the naturalness of accents and the conversation they heard overall?

1.2 Potential Contributions of the Work

The work in this thesis is both novel and important, and its potential contributions can be described as follows:

1) It will help in raising awareness on the effects of foreign accents in pilot-ATC communication. For too long the language issues in aviation, especially regarding foreign

accents, have not been investigated properly. Perhaps because of its perceived difficulty in disentangling the double meanings contained in the words used in a paragraph or because there are too many variables to consider to properly set up experimental procedures with language issues (ex. a participant not willing to disclose their accent because he or she is shy). Reasons could vary. However, there is evidence that foreign accents cause miscommunication. This should not be overlooked as it is heavily intertwined with flight safety. Thus, it would be a worthwhile effort to investigate foreign accent-related issues in pilot-ATC communication, whether it be employing a new testing method, analyzing a corpus of AE, or improving educational technology in this field. Knowledge accumulated this way will help future researchers derive insights which, in turn, improve recommended standards and practices related to AE.

2) It shows how the test constructs (question types) were created to assess participants' accent listening skills. Although there are testing materials available for AE in general, such as reading comprehension and verbal activities, it is rare to come across unique testing materials that concern multiple foreign accents. The adjective 'unique' in this case refers to the originality of the scripts and questions, that is, they were created from scratch. This thesis contains Appendix E, F, and G, which contain the scripts, questions, and answers used on the actual listening tests. Chapter 3 is dedicated to filling this gap as well. These will be helpful for future researchers who might want to conduct a relevant study in this field.

3) It provides an overview of the past and present of CBE and AE. Based on the author's observations, the connection between these two distinct fields has not been clearly demarcated in the surveyed research papers. A discussion on how CBE could be implemented by incorporating AE is included to build a bridge between the two. It hopefully generates some considerations on how CBE could be best utilized with respect to AE.

4) The experimental procedure employed in this thesis may provide some guidance on how to set up a more sophisticated procedure. Although this thesis does not investigate the effect of training transfer from computer voice to human voice, it is nonetheless one of the first attempts to include foreign accents as a main subject of research in the context of aviation. Readers may find that there could have been other ways to create an experimental procedure that could have been more ideal choices than the one employed in this thesis.

However, there is a value in such realization that it may help future researchers in making more objective assessment in this field of study.

5) The thesis, when viewed wholistically, carries an important message regarding the training of ab initio pilots (or possibly for pilots who are returning to the industry) in listening skills. As repeated throughout the thesis, foreign accents will always be present in Aviation English because speakers from various geographical regions will have their own unique accents whenever they converse. The author thinks that there could be other studies that may take inspiration from this thesis, and make more adjustments in their own experimental designs, and find new results that could in turn provide new ideas for developers in educational technology (perhaps including strength of foreign accents in their new technology).

Chapter 2. Literature Review Studies

2.1 Study 1: CBE and AE

2.1.1 The Rise of Competency-based Education

2.1.1.1 What is Competency-based Education?

Competency-based Education (CBE)[1] is a training curriculum with an aim to teach a *set* of competencies required for demonstrating professional skills instead of focusing solely on one aspect of training. A competency is defined as a combined set of knowledge, skills, and attitude (KSA) that enables a trainee to participate in an organizational context[2] (Kearns et al., 2017). It is derived from job requirements (Mansikka et al., 2017, p. 78). A competency is treated as a single component of an overall professional skill. A competency can be diverse across organizations (ex. health care, nuclear, environmental, aviation) for a similar duty. For instance, a simple competency that reads "ask for clarification" may have different meanings in both health care and aviation. In the former, a doctor would be asking a series of questions to a patient about his or her previous medical conditions to identify health complications that may affect the treatment. In the latter, a pilot would be requesting to repeat the instruction because it was difficult to understand due to ATCO's thick foreign accent.

Competencies, which are then a group or set of competencies, are standardized (that is, each competency included therein has been negotiated and reviewed by experts) and written as texts with appropriate language for training (Kearns et al., 2017). In simple words, a set of competencies are discretized "knowhows" that are written on a list which shows what an expert would do if he or she were given a task. If a competency is a single keyword of a job duty, then competencies would be a job description. If everything that has been discussed above is put into a boundary, then logically this whole "package" is called CBE and it becomes a training manual. From an instructor's standpoint, CBE contains a set of competencies, wherein a competency outlines an easy-to-understand unit of learning outcomes for a trainee to absorb, adjust, and apply. See **Figure *1*** for a visualization of CBE and its components.

---

[1] Sometimes in literature, it is also known as CBET (Competency-based Training and Education) or CBA (Competency-based Assessment). It is also referred to as EBT (Evidence-based Training).
[2] Specifically pertains to aviation industry.

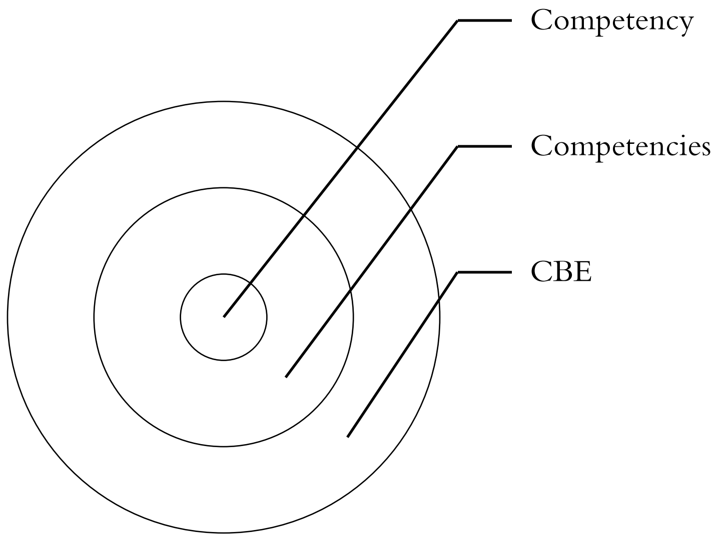

Figure 1. A diagram showing the components of CBE.

CBE was one of the subjects discussed at the House of Commons in the Government of Canada (Sgro, 2019). One of the most pressing concerns in aviation is that the demand for pilots is increasing, but the supply is stagnating. The demand acts as a pressure for the industry (Valenta, 2018). Demand is simply the passengers' willingness to purchase flight tickets and travel, which is the main source of revenue for airlines (i.e., Revenue Passenger Miles is a metric that indicates traffic for an airline, and it is calculated by multiplying passengers onboard with travelled distance). When the supply of qualified pilots is low, the price of the tickets will increase. Although this can be a good short-term trend for producers (airlines), the price-conscious consumers (passengers, either leisure or business travellers) will not be too happy about the rising price of a ticket. On top of that, passengers are also conscious about flight safety. So, airlines need to continuously invest in training a next generation of pilots to not only keep up with demand, but also to uphold the highest flight safety standards.

Today's pilot training philosophy is largely derived from that of World War II (Todd & Thomas, 2013). The outdated training philosophy and the pressure of demand call for a major overhaul in how an ab initio pilot is trained. Not only that, today's aircraft is structurally different from those used in WWII (ex. an aircraft with computerized cockpit or an electric aircraft surfacing as the new generation of aircraft). Pilots who will eventually fly an electric aircraft must be knowledgeable about its characteristics. Thus, when training pilots, it is necessary to account for the type of aircraft that is used for training. Not only that, since the ratification of the Chicago Convention in 1947 and the deregulation of airline industry in the mid 1970s, more airlines emerged in the market. The growth of the market accompanied the publication of many flight manuals and standard operating procedures (SOP) related to fatigue management or Crew Resource Management (CRM) which were eventually adopted and promulgated by ICAO.

But it is not just the type of aircraft that is important for training. Language also matters. It is not an exaggeration that flight safety is primarily a communication issue in the industry. For instance, it is not hard to find language-related flight accidents that cost many lives (ex. Tenerife Airport Disaster, Avianca 052). Other aspects of flight operation (such as CRM mentioned above) may have gotten enough public attention, but for language-related matters, not so much. Friginal et al. (2020) have stressed the need to document the language-related issues that persist in pilot and ATC communication. This is very important because it could reveal unexpected insights that can pave the way for innovation in language training. It is important to note that language has many aspects that are worthy of further investigation such as where the primary and secondary stresses are in a word, and other aspects of prosody (ex. loudness, rhythm, and finally the accent). The accent has been found to cause miscommunication in aviation (Tiewtrakul & Fletcher, 2010), but only a very small number of articles investigated this issue (Dissanayaka et al., 2023). All things considered, CBE questions the traditional way to train pilots and helps in revealing any other aspects of flight operation that are worthy of attention (such as language and accent).

Traditional pilot training mostly emphasized hours of practice and therefore is based on a quantitative input (Defalque, 2017, p. 3). Kearns et al. (2017) argued that hours alone cannot assess the overall quality of a candidate because it cannot explain specifically what a candidate has learned during the hours of practice. Numbers, compared to a record of written texts, only quantify "how much time a trainee spent learning a skill" and cannot fully explain or describe "how a trainee has learned and applied a skill"[3]. Therefore, a focus must be given to a set of behavioural objectives (a competency) that can be observed, measured, and improved. This does not imply that hours should

[3] This is an important point. Crandall et al. (2006) noted that aviation, along with nuclear and health services industry, are characterized by unusual occurrences that bear important consequences for safety in general. These industries are directly related to humans; small or large incidents in those industries can lead to large casualties if left untreated or undiscovered. Procedures based on strict protocols are necessary to protect the lives of those involved (in aviation, it would be passengers). Thus, it is crucial to document and record competencies from the onset of pilot training and store these competencies in a database that can be referred to for comparison of performances across candidates, used to upgrade existing competencies, or combined to give birth to a new set of competencies.

not be included. In fact, hours are an important metric that can provide useful and objective data on whether a trainee's certain skill has improved over time. Thus, time (hour) is an integral part of CBE that will serve the role of a denominator to each KSA (**Figure 2**).

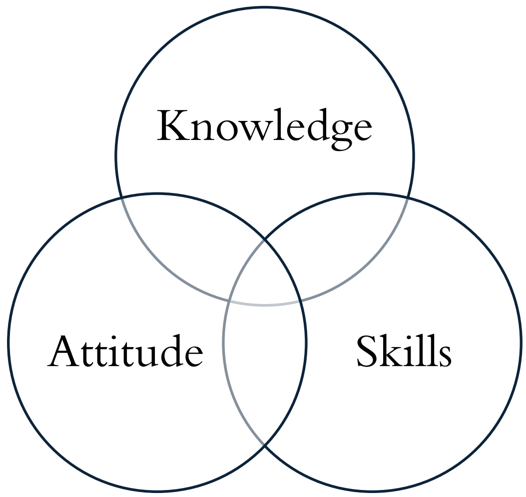

Figure 2. Illustration of the CBE using a Venn Diagram. The Universe (U) is CBE, which contains K (knowledge), S (skills), A (attitude). The centre is defined by $K \cap S \cap A$, where a trainee's competency lies. A complex interaction among KSA allows a competency to be created and absorbed by a trainee.

2.1.1.2 History of CBE

Historically, CBE was influenced by many previous policies such as FCLTP (Flight Crew Licensing and Training Panel), CRM (Crew Resource Management), NOTECHS (Non-technical Skills), AQP (Advanced Qualification Program), TEM (Threat and Error Management) (Kearns et al., 2017). It is important to review these policies since these helped in forming foundational philosophies of CBE (**Figure 3**).

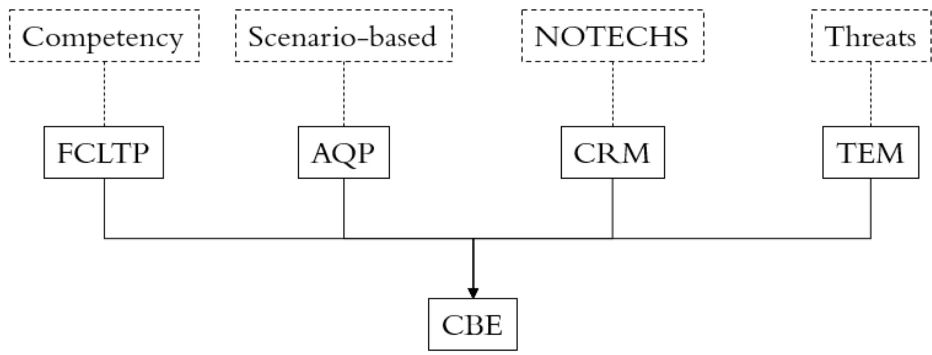

Figure 3. Various policies that have shaped CBE.

FCLTP was responsible for creating the *Procedures for Air Navigation Services – Training* (PANS-TRG), which outlined designs for CBE as well as making amendments to *Annex 1 – Flight Licensing* (ICAO RASG-APAC/1, 2011, p. 1). FCLTP was first held in Montreal in 2003 and its objective to establish a clear outline of competencies for pilots was influenced by the necessity to reduce aircraft accidents (ICAO Doc. 9868, 2016). Its outcome was a list of competencies and observable performance criteria for multi-crew pilot license (MPL) (ICAO RASG-APAC/1, 2011). Other notable achievements under FCLTP are the inclusion of threat and error management (TEM) and flight training simulation (ICAO, 2010, p. 10). According to Kearns (2017), the concern of FCLTP was on the hours as a unit of (flight training) measurement. The obstacles and challenges faced by individual trainees are not similar, thus hours cannot be said to express a trainee's competency in flying (Kearns et al., 2017). This fact shows that experts were already grappling with hours as a problem from the beginning of the development of CBE.

In simple words, it puts more emphasis on "what a trainee has encountered and learned during *x* hours of training". It is rather inappropriate to conclude that "this trainee *probably* has mastered all skills in *x* hours, and we do not need to look further". What seems to be overlooked is the fact that a challenge or a problem has many facets that can upset a trainee's understanding (see **Figure 4** below), which needs to be analyzed by an instructor if a proper intervention is necessary. If a trainee is having trouble with Problem A ('understanding accented speech from an air traffic control officer (ATCO)') and especially Facet 1 ('weather advisory'), the instructor in question must impart knowledge to the trainee on how weather reporting is structured in aeronautical communication and teach them how to interpret cues even though a foreign accent is embedded (ex. “First, keep yourself composed and collected even if there’s accent. If you hear a ‘S’, ‘H’, ‘R’, ‘A’, then it means ‘shower and rain’. The ATCO might not pronounce each alphabet letter individually and go with ‘SHIRA’ or ‘SRA’, so be aware of it.”). A very specific strategy is needed for each obstacle or challenge faced by trainees. The work by FCLTP was not in vain; it provided a way for educationalists to further decompose a problem into smallest competency units so that trainees could understand, learn, and apply.

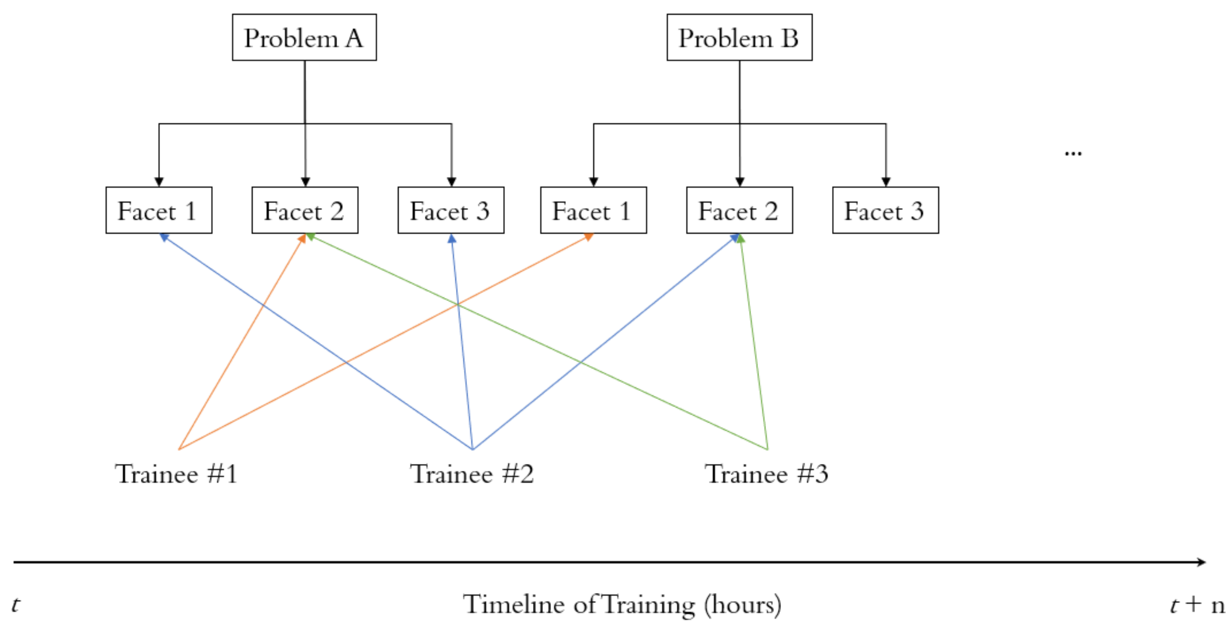

Figure 4. The individual facets of a problem encountered by trainees over training hours.

AQP was another policy that had a lasting effect on CBE. Although it is limited in its scope of application such as not being applicable to ab initio (Kearns et al., 2017), it nonetheless added elements that would later resurface as “flexibility” – a core characteristic of CBE. According to Kearns (2017), it is comparable to CBE in that it allows an instructor to explore ways to modify learning schedules determined by a task analysis. Both Transport Canada and FAA state that participating in AQP is voluntary (FAA, 2006; Transport Canada, n.d.b). This is because AQP puts more emphasis on cooperative skills with other crew members. It also emphasizes the use of scenario-based training, wherein trainees learn that accidents are a chain of unforeseen events that can transition into either latent or fatal threats to flight safety[4]. Finally, another benefit of AQP is it focuses on CRM which promotes the importance of communication – a critical aspect of teamwork. Flying an aircraft while maintaining effective communication with team members is an arduous task, but in the long run it will not certainly tarnish the reputation of an airline that markets itself as the safest airline in the world.

The last two policies are CRM and TEM. As can be seen clearly, these are derivatives of FCLTP and AQP. However, it is nonetheless important to state what these are. CRM emphasizes five competencies that are of non-technical nature: Communication, Leadership and Teamwork, Problem Solving and Decision-making, Situation Awareness, Workload Management (ICAO Doc. 9995, 2013). Although these are difficult to quantify (meaning it is challenging to measure them objectively) (Kearns et al., 2017), these are required for piloting a complex aircraft while managing flight crew simultaneously. TEM fits very well with CRM, as it teaches a pilot or a crew member to identify latent (hidden) threats before they become a larger problem during flight. When finding those threats, operators (usually an expert in human factors or a pilot with many years of experience) conduct line operations safety audits (LOSA) by closely observing the behaviour of flight crew in the aircraft. They would document such behaviours on a checklist. Knowledge and evidence gathered in this way can become a valuable resource for future pilots who may not realize that there are various types of hidden threats in piloting an aircraft or in activities that require cooperation.

2.1.1.3 On the Low Supply of Pilots

The historical development of CBE is, however, much more complex than just stating the influences of the policies above. In fact, it is ongoing due to the impact of demand-supply of pilots, and this has an important implication to pilot training in general. Canada has appealed the stagnating supply of competent pilots in 2019 to the ICAO Assembly. In the working paper of the ICAO 40th Assembly (2019), serving worldwide demand for air travel requires an additional 500,000 pilots, along with 858,000 cabin crew and 622,000 maintenance personnel (p. 2). For Canada, the commercial aviation industry has conferred economic benefits worth $35 million in gross domestic product and 144,000 in employment (ICAO 40th Session of the Assembly, 2019, p. 4).

[4] This connects to the Swiss Cheese Theory. In this theory, an impact of a small issue can be grown into a larger issue that directly pose threat to safety (i.e., a hole that lets a threat pass through easily). Language is a perfect example, as it is considered an underrepresented issue in aviation (Baugh & Stolzer, 2018).

A labour market report published by Canadian Council for Aviation and Aerospace shows how many pilots and aircraft maintenance technicians the industry needs (see **Table 1**)[5]. Unfortunately, only 70% of new pilots remain in the aviation industry and approximately 1,200 licenses are issued yearly (CCAA, 2018, p. 18). But this does not illustrate the whole picture. With other industries recruiting students from various disciplines (including those who majored in aviation or aeronautics) and less domestic students pursuing pilot training (or those who stop in the middle), the number of new pilots staying in the industry could be smaller.

Table 1

Number of Pilots and Aircraft Technicians needed by 2025

Category	Statistics
Pilots[a]	7,300
Aircraft maintenance technicians	5,300

Note. The information from this table was obtained from Canadian Council for Aviation and Aerospace's labour market information report, which was published in 2018 (the latest report at the time of access).

[a] Needed by 2025.

While the passenger load factor in March 2023 has increased (**Figure 5**) and reached beyond that of pre-pandemic March 2019 (Statistics Canada, 2023, p. 2), Canada still has not seen an improvement in increasing the supply of pilots. The point is that each airline will not survive if it cannot keep and retain talented pilots for a long time, and major airlines are always looking for qualified pilots for themselves as well. What is an already oligopoly market is now facing more severe competition in expanding hiring pools.

[5] It must be noted that this report was published in 2018. And the statistics for additional pilot supply is from 2019. It does not account for the impacts of COVID-19 and other inherent challenges faced by the industry. Therefore, the statistics presented in Table 1 may not be accurate as of 2023. But they are listed here for informational purposes.

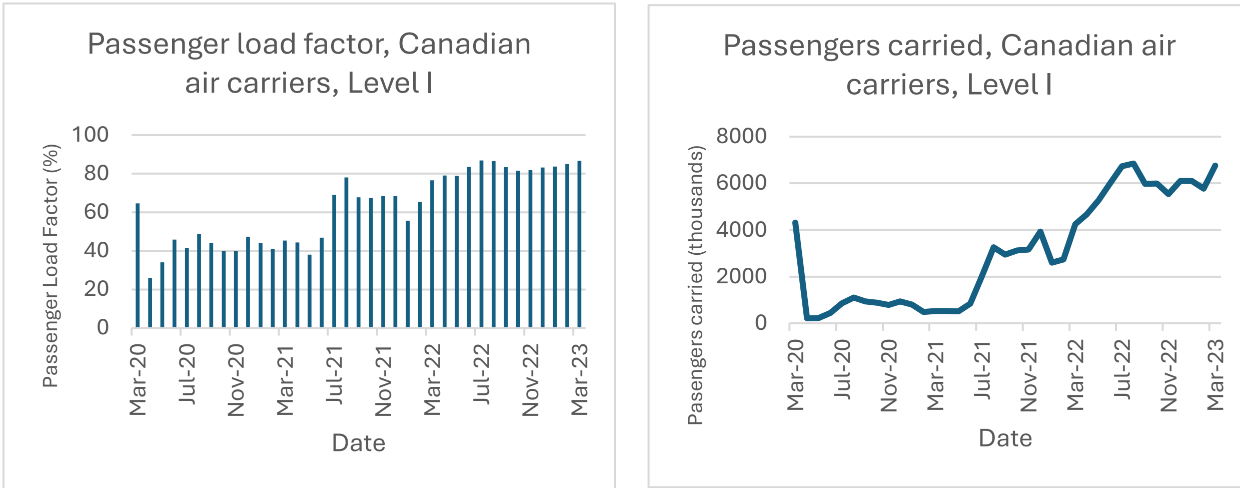

Figure 5. Passenger load factor and passenger carried in 2023 by Canadian airlines, Level 1. Source: Statistics Canada.

Perhaps what drives the low supply is the fact that pilot training is expensive. This fact alone may present the occupation as undesirable, even if the wage is likely to rise in the future (i.e., the barrier to entry is high for students and trainees alike). The industry is in a dilemma that is hard to get out of; on one side there is a shortage of quality pilots and the other side there is a cost for training and retaining. From a trainee (or student) perspective, the dilemma is also similar: too expensive on one side and too risky (for loss of jobs due to unforeseen natural or health disaster) on the other side. In brief, the economic facts and the fear of uncertainty justify the development of CBE, that is, to train and retain quality pilots to prepare for potential impacts of demand-supply relationship.

Taking the economical consideration above, it is necessary to think about how to best train pilots with "cost-effective" in mind. In terms of language training, it is unfortunately expensive as well (ICAO Circular 323 & ICAEA, 2009). Therefore, a careful analysis of learning needs of each pilot trainee is important. With a traditional pilot training (which places more emphasis on hours flown), it is not possible to strengthen each trainee's KSA. CBE, on the other hand, is predicated upon the idea of flexibility and customization of learning outcomes, which can help a flight instructor build the most effective training method.

2.1.1.4 Advantages and Disadvantages of CBE

An advantage of implementing CBE is that there is no time limit. That is, the time it takes to master a competency is not more important than being able to demonstrate a competency in reality (Joyner, 1995). In other words, the importance lies in the actual attainment of a skill that can ultimately improve flight safety, rather than vaguely recalling how many competencies a trainee knows.

A second advantage is flexibility (or self-paced), which enables customization of learning outcomes based on a specific task. To make this work, instructors need to be aware of the specifics of the learning needs of their students. Joyner (1995) mentioned that an instructor in CBE is not "the traditional authoritarian holder of knowledge" anymore, but rather someone who can provide learning resources that are compatible to learning styles and strengths of each student. This is an important point because it changes the direction of the communication. In traditional sense, a teacher who holds "authoritarian mindset" would impart knowledge to the student (top-down). That knowledge, regardless of whether it is correct or not, would end up crystallizing in the student's mind which can later manifest as a precursor to a serious incident or accident. However, by adopting the mindset of a teacher who is focused on communication, the alignment of learning resources and students' needs can be configured effectively. It also means (by continuing to adopt this mindset) the instructor has implicitly confirmed with the trainee that the foremost priority is to help the trainee learn necessary and required competencies to become a pilot which today's industry wants. This helps eliminate an unnecessary bottleneck where a training is delayed due to an instructor figuring out alone what a trainee should need.

However, CBE has garnered criticism as well. It is criticized for oversimplification of training (Franks et al., 2014), which does not address other aspects of training that could potentially influence the training outcome. An example would be the relationship between the instructor and the trainee. If not careful, it may hinder trainee's learning progress if they are hostile to one another. Another example would be delivery of training sessions which can be either classroom teaching, online learning, or blended learning. If it is online learning, does the training session provide asynchronous tool for trainees who do not have online access? CBE must take into consideration "peripherals" that might affect the learning progress of trainees, not just emphasizing the value it provides. A well laid-out educational plan can be amended later, but the opposite may result in poor training that can be absorbed by the trainee which will manifest as a bad habit (ex. skipping a step in the pre-flight checklist because the pilot believes that it never caused any problem).

Second, myriad definitions of CBE exist (Todd and Thomas, 2013; Mansikka et al., 2017). There is also a criticism that elements that make up a 'competency' have not been well-established (Thomas & Richards, 2015). This confuses scholars and those who are interested in including it into their education (Todd and Thomas, 2013). Understanding 'competency' depends on an organization's culture and values they uphold. If an organizational culture is more oriented towards individualism where a competency in being self-reliant and independent is seen as more valuable, then "seeking help for a problem" might be shown as a sign of weakness. Then how should this be incorporated into a curriculum for pilots who study and train in different countries? A pilot's role, apart from flying safely, is to facilitate communication and teamwork which requires being collaborative and answering time-limited questions on the spot. In relation to the delivery of training sessions, what tools (or learning strategies) should be used for students who are from countries that value individualism over collectivism? Although these questions may seem trivial, a long-lasting effect of training grounded in a misfit educational strategy can manifest unbeknownst to the trainee during work. This can potentially manifest in the form of an occupational hazard in the future.

Third, there is a lack of a method to properly assess competencies (Franks et al., 2014). This creates logical fallacy of CBE's definition (if there is a universally agreed upon definition, that is). A competency, as aforementioned, is the observable performance. How should the performance be compared when each trainee goes through unique phases of learning and has experiences that are unlike the experiences of other trainees? This trail of question ultimately leads to this problem: Is there a way to quantify or measure performances objectively? There have been studies related to fatigue that looked at KSS (Karolinska Sleepiness Scale) and SPSCS (Samn-Perelli Crew Status Check) (Gander et al., 2015), but it remains elusive as to how a massive performance data that is recorded as either a checklist or in written texts can be quantified further. It remains debatable whether all aspects of a competency should be quantified or left to a subjective evaluation from an instructor, whatever that method may be.

2.1.1.5 CBE Development and Framework

A development of a curriculum based on CBE starts with an outline of competencies. These are usually represented in a tree diagram or a table that specify three elements: 1) Competency Units; 2) Competency Elements; 3) Performance Criteria (ICAO Doc. 9868, 2016). Competency units are the main objectives: What the instructor wants to teach. Competency elements are sections (such as unit 1,2, and 3 in a textbook, for example) that contain specific activities. These activities, known as performance criteria, are texts which describe the detail of steps to be taken by the trainee. **Table 2** below shows the general competency units outlined in ICAO Doc 9868.

Table 2

Competency Framework for Instructors

Competency Units	Competency Elements	Performance Criteria
1. Manage Safety	1.1. Provide safe training environment. 1.2. Provide safety intervention.	a. Communicate hazards. b. Provide support for mental stress.
2. Prepare the training environment	2.1. Abide by the syllabus that's previously agreed upon. 2.2. Manage equipment and the facility itself.	a. Make certain that the scenarios employed are realistic. b. Maintain the equipment to be operational.
3. Manage the trainee	3.1. Recognize trainees' needs. 3.2. Support trainees through coaching.	a. Do know what the trainee's learning needs are based on their background. b. Become flexible based on the trainee's performance and learning needs.
4. Conduct training	4.1. Show credentials to earn trust. 4.2. Stimulate realism in training	a. Clearly state training objectives. b. Provide answers to questions the trainee may have.
5. Perform trainee assessment	5.1. Build an evaluation method. 5.2. Provide effective feedback.	a. Be transparent about how evaluation works. b. Ensure positive reinforcement and encouragement.

6. Perform course evaluation	6.1. Administer self-evaluation. 6.2. Conduct course evaluation.	a. Look back on the strengths and weaknesses of the instructor him/herself. b. Analyze the course feedback.
7. Continuously improve performance	7.1. Seek ways to improve the course itself. 7.2. Allow for personal development	a. Be open to feedback. b. Keep broadening knowledge base.

Note. Notice that most of the competency units start with an action verb. The descriptions of performance criteria have specific objectives (mental stress, trainee background etc), whereas those of the competency elements are generic.

The table above is a general outline of a training, and it may look different depending on the type of occupation. For example, if this table is used to create a training module for flight crews who will always be confronted with a challenge of multitasking, TEM may be incorporated in the competency unit 6. For crew members (flight attendants), NOTECHs may be included. Because this is a general outline, it can be modified to fit a specific training module.

It is noteworthy that it does not specifically include Aviation English anywhere in the table above. The author of the thesis is of the position that language communication be included as part of a competency element 4.1. This is because language is the intertwined with flight safety. For example, a new competency element for new pilots could be "Attain AeE proficiency exam result that is at least above level 4". A performance criterion related to this competency element could be "how to maintain uninterrupted, clear, and concise communication". Because AE is described to be elliptic (short sentences which contain hidden assumptions and meanings), how to communicate without being too verbose is a priority for trainees.

Or if a flight instructor wants to create a new competency framework that only focuses on AeE training, then customization to the above table can be done. Table 3 below is an example of a competency framework that focuses on Aviation English-based communication.

Table 3

An Example of a Competency Framework for Pilots (Language)

Competencies [a]	Reference Material	Observation & Evaluation
1. ENSURE LEXICAL COMPETENCE		
1.1. Identify operational words and codes	ICAO Doc 9835, sec. 3.4.2	
a. Able to identify abbreviations and jargons.		Satisfactory / Not Satisfactory
b. Able to identify topographical features.		Satisfactory / Not Satisfactory
c. Able to identify METAR codes.		Satisfactory / Not Satisfactory
…	…	…
2. FAMILIARIZE COMMUNICATIVE FUNCTIONS		
2.1. Be ready to manage a communication with ATC.	Borowska, 2017;(ICAO Circular 323 & ICAEA, 2009), sec. 4.4.1	
a. Able to give orders.		Satisfactory / Not Satisfactory
b. Able to read back.		Satisfactory / Not Satisfactory
c. Able to ask for clarification (requests to resolve ambiguities, paraphrasing, repetition etc).		Satisfactory / Not Satisfactory
…	…	…
3. MAINTAIN PROPER GRAMMAR STRUCTURE		
3.1. Control the consistent use of proper grammar and syntax when speaking.	ICAO Doc 9835, sec. 4.6.3	
a. Able to detect grammatical errors.		Satisfactory / Not Satisfactory
b. Able to determine when to change tenses.		Satisfactory / Not Satisfactory
c. Able to use linking verbs in the correct situation.		Satisfactory / Not Satisfactory
…	…	…

[a] Includes competency units (bolded), elements (index number), and performance criteria (letter bullet points).

The benefit of creating a competency framework is twofold: 1) It ensures that the criteria are standardized (referring to the overall format of the table above where it specifically states the performance criteria under competency units and elements), leading to less errors when trainee data is analyzed (ex. no missing information or unruly format structure when handling data); 2) New competency units, elements, and performance criteria can be added anytime when new ICAO

Standards and Recommended Practices (SARPs) are issued, ensuring flexibility in planning a lesson. It is also helpful for flight instructors when revisiting a pilot candidate's profile, if that candidate in question returns to complete the Aviation English re-examination. In conclusion, having the framework streamlines the decision for instructors on what to teach and what to update based on relevant KSA possessed by the trainee in question.

2.1.1.6 Conclusion

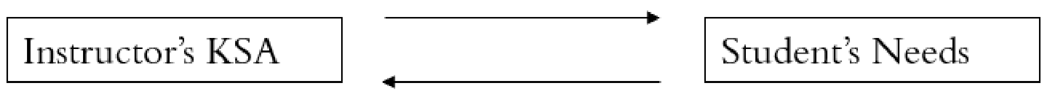

Figure 6. Overlap of the instructor's KSA and a trainee's needs.

Overall, CBE is a consultative and collaborative training curriculum where the exchange of information is in both directions, rather than in one direction (from a teacher to student and vice versa; see **Figure 6** above). A traditional pilot training curriculum emphasizes hours over any competency, which leads to a misconception that "hours tell everything about what a trainee went through". Hours cannot be said to express or describe the specific KSA a trainee has obtained throughout their training. CBE, in contrast, allows self-paced learning with no time limit that ensures students practice a skill until they feel comfortable. This is great for teaching language-related skills to pilot trainees, as language learning is a time-consuming and costly task that involves, for example, interaction with foreign accents and understanding double meanings. Language issues have been persistent in aviation since commercial aviation was born. Thanks to the rising importance of CBE, language issues inherent in aviation can be more effectively dealt with.

CBE is not without its disadvantages which are mainly centred on its vague definitions and assessment techniques. However, if the industry is to gain benefit from employing qualified pilots and become financially sustainable in the long term, they must look at the range of KSA each pilot holds and complement these elements with organizational support. For instructors, one must focus on an individual competency held by a trainee and determine whether they need more practice or need specific lessons that can improve their prior KSA.

2.1.2. Aviation English in the Present

2.1.2.1 The Brief History of Aviation English (AE)

To understand AE, it is important to know about its history. The origin of AE is a part of the history of electric communication and can be traced back to morse code. The British then invented Q-Code, which involved three letters that represented a specific meaning when transmitted. This was originally used in maritime communication (Borowska, 2017, p. 26). The Q means 'question'. Examples of Q-Codes are QDR (magnetic bearing), QFU (runway magnetic orientation), and QTE (true bearing). The characteristic of AE is that every transmission must have a clear, purposeful meaning with no room for other interpretation. It can be said this characteristic is inherited from Q-Codes.

Then comes the International Telecommunications Union (ITU), which is a special United Nation agency (established in 1865) responsible for maintaining standards of technological applications that send and retrieve information. It adopted the Phonetic Alphabet. These are codes with specific pronunciations which help two parties communicate alphabets used in English. Not only is this widely used in aviation, but also in amateur radio and North Atlantic Treaty Organization (NATO). See **Table 4** below.

Table 4

The Current Version of the Phonetic Alphabet for use in Aviation Communication

A	Alpha	N	November
B	Bravo	O	Oscar
C	Charlie	P	Papa
D	Delta	Q	Quebec
E	Echo	R	Romeo
F	Foxtrot	S	Sierra
G	Golf	T	Tango
H	Hotel	U	Uniform
I	India	V	Victor

J	Juliet	W	Whiskey
K	Kilo	X	X-Ray
L	Lima	Y	Yankee
M	Mike	Z	Zulu

When the World War II (1939-1945) occurred, the American and British military began to develop their own versions of the alphabets (Borowska, 2017, pp. 27–28). At the time, there was no unifying phonetic alphabet. But these two versions became similar as time went by. During the war, aircraft became more sophisticated, which later gave birth to commercial aviation. With the rise of the airline industry, it necessitated the need for a unifying language that could be spoken in air-to-ground and air-to-air communication.

The year of 1944 saw the gathering of fifty-five nations that decided to make an international framework for civil aviation (ICAO, n.d.). The result was the Chicago Convention, outlining many standards that member states should abide by. This eventually led to the creation of a governing body known as International Civil Aviation Organization (ICAO) in 1947. However, in the original 1944 convention document, there was no concrete agreement on what language should be spoken in the air. The language of the air fell into obscurity.

Table 5

A list of aviation accidents that were related to language

Accident	Date	Description	Source
Avianca 052	1989	Probable cause includes "the lack of standardized understandable terminology for pilots and controllers for minimum and emergency fuel states".	National Transportation Safety Board, 1991, p. 76.
Flying Tiger Line 66	1989	The pilots, upon approaching Kuala Lumpur Airport, misinterpreted "descend two four zero zero" as "descend to four zero zero". Flying below prescribed height, it crashed, killing all four passengers. Kuala Lumpur ATC used a non-standard phraseology, which resulted in the above accident.	Aviation Safety Network, n.d.
China Northern Airlines Flight 6901	1993	The pilots did not understand the meaning of "pull up" from GPWS.	Tajima, 2004.

American Airlines Flight 965	1995	The air traffic controller was not fluent in plain English (non-aeronautical) but knew that American pilots' requests to fly directly to destination (when in fact they were off course) did not make any sense.	Tajima, 2000.
Chakri-Dadri Mid-Air Collision	1996	Kazakhstan Airlines 1907 and Saudi Arabian Airlines 763 collided mid-air over Chakri-Dadri in India. The contributory cause is the Kazakh pilot's "inadequate knowledge of English language" which led to "wrong interpretations of ATC instructions".	Ministry of Civil Aviation, n.d., p. 7.
Tenerife Airport Disaster	1997	Identified "Inadequate language" as one of the contributing factors. Stipulated that concise and clear use of aeronautical language must be adhered to.	Netherlands Aviation Safety Board, n.d., p. 55.
Helios Airways 522	2005	Rather than a miscommunication between a pilot-ATC, one of the contributors to the accident could be the captain's German accent.	Hellenic Republic Ministry of Transport & Communications, 2006.
Helsinki-Vantaa Aerodrome Runway Collision Incident	2011	Two airliners (one ready to depart, one landing) on the same runway risked collision. English language proficiency on the flight crew of one of the airliners was found to be insufficient.	Safety Investigation Authority Finland 2011, p. 17.

Note. It should be noted that the use of non-standard phraseology is one of many factors that contributed the accidents above. For instance, Avianca 052 had to fly in circles until Kennedy Airport had to resolve high traffic issues in poor weather. In the case of American Airlines 965, the pilots did not fully report their position with the nearest VOR, which was a requirement. Singling out one factor is not possible in any aviation accidents.

It was not until in 2003 that ICAO began to address language issues after a series of unfortunate accidents. Among the accidents listed in the **Table 5** above, there were three major accidents that were influential to the creation of standards related to persisting language issues in the industry: Tenerife Airport Disaster and Charkhi-Dadri mid-air collision. The Charkhi-Dadri mid-air collision gave rise to a resolution called A32-16 (Friginal et al., 2020, p. 43).

A32-16 is known as "Proficiency in the English language for radiotelephony communications". It contains two preambles outlining that English proficiency and comprehension are contributing factors to accidents and that there should be a way to ensure that all flight crew maintain (English) proficiency when communicating via radio. These are followed by the recommendation from the assembly. The recommendation reads as follows (ICAO 32nd Session of the Assembly, n.d.):

> (The Assembly) Urges the Council to direct the Air Navigation Commission to consider this matter with a high level of priority, and complete the task of strengthening the relevant provisions of Annex 1 and Annex 10 with a view to obligating Contracting States to take steps to ensure that air traffic control personnel and flight crews involved in flight operations in airspace where the use of the English language is required, are proficient in conducting and comprehending radiotelephony communications in the English language.

However, there were concerns with the resolution: 1) It leaves out those whose native language is not English; 2) It is mainly focused on the oral proficiency (Friginal et al., 2020, p. 44). Friginal et al. (2020)'s argument is that it should also emphasize reading and listening, which is essential for training. Reading is important for pilots and ATCOs because they always need to follow with regulations that are printed on a manual. Listening is especially important for both as they must rely on two-way communication without a common visual aid that bridges them together.

Currently, there are a total of six languages used in ICAO: Arabic, Chinese, English, French, Russian, and Spanish. However, considering that the Chicago Convention was signed and ratified in English (and that the U.S and Britain at the time were considered hegemonic powers in Europe and Americas respectively), English would remain as the *de facto* language of communication for important matters.

2.1.2.2 On the semantics of the term "Aviation English"

AE is a specialized language that serves a specific purpose. It belongs to English for Specific Purposes (ENGSP). It simply means its syntax and grammar are distinct from those of Colloquial English (CE). AE contains a codified lexicon and sentences which incorporate specific jargons, and it is characterized by ellipsis. That is, every word or sentence has specific meanings that can only be understood in operational settings. This is to avoid repeating the same instruction or to reduce verbosity. This is not a characteristic found in CE which people speak daily. When talking about AE in general, it is wholistically referred to as standard radiotelephony phraseology, and it is bound by a set of strict rules that govern grammar (morphology and syntax) and pronunciation. The purpose is to be concise, precise, and directly to the point, and it precludes idiomatic phrases (Emery, 2014). In this sense, AE cannot be said to fully belong to CE because the former contains specialized linguistic features not commonly shared or found in the lexical and grammatical pool of everyday English speakers.

PE can be used in a flight operation. This is because pilots and ATCs cannot use standard phraseology to accurately describe the nature of the immediate threat to safety. According to Emery (2014), ICAO acknowledges that using only the standard phraseology is limited, thus language proficiency requirement (LPR) reinforces the usage of PE to fill in the gap in meanings not fully explained by the elliptical nature of the former. It also serves another purpose, which is to fix misunderstanding (Coertze et al., 2014). Whether a pilot is speaking AE or PE, the careful selection of words to form a sentence to communicate threat, whether it is apparent or latent, is crucial so that ATC can provide the most appropriate instruction at that moment.

There is no detailed discussion on establishing a standard definition of AE, leading to a confusion as to whether AE is a correct term to use. Even though the term "Aviation English" itself is used widely (ICAO Circular 323 & ICAEA, 2009; ICAO Doc. 9835, 2010), it is somewhat vague because the word "Aviation" as a prefix is too broad in scope[6]. For example, Borowska (2017) advanced an argument that AE should be different from Aeronautical English (AeE) and thus should avoid using these terms interchangeably (p. 64-69). A main reason is that AE is a much broader term that contains other specialized sublanguages such as aviation legal sublanguage, cabin crew sublanguage and AeE itself (Borowska, 2017). Not only that, the use of AE for this sublanguage is misleading because it presumes that other national languages can be named in the same fashion (Borowska, 2017, p. 64). For instance, this would give rise to Aviation Korean, Aviation French, and Aviation Russian. When saying "Aviation Korean", does it mean it specifically refers to the special language used by a Korean pilot and ATCO? Or does it refer to a special language used by a Korean aircraft maintenance engineer and supervisor? The question then becomes "what distinguishes these aviation languages from the pilot-ATC communication?" A specific term that refers to the pilot-ATC communication seems necessary without referring to the special languages used by other professions in the domain of aviation.

[6] Aviation could encompass the following occupations: Aircraft Maintenance Engineers, Aviation Management Service Providers, Air Traffic Controller, Vehicle Operator, Flight Crew, Ground Crew, Emergency Services, Duty-free Shops etc. Members of each occupation listed above may have their own sublanguages that are used for communication.

Aeronautical English (AeE), on the other hand, is more fine-tuned as it refers to the operation of an aircraft. Borowska (2017) argued that AeE contains both standard phraseology (SPH) and Plain Aeronautical Language (PAE), which are used in communication between pilot and ATC (p. 67). The definition of "Aeronautical" is stated as follows (Cambridge Dictionary, n.d.):

Aeronautical /ˌeərəˈnɔːtɪkl/

Relating to the science of designing, building, and operating aircraft.

Why is aeronautical a better word? Because the definition above relates to the 'airworthiness' of an aircraft. A pilot who has a license to fly, must apply for, and have in possession an airworthiness certificate of his or her own aircraft. He or she at least should know about aircraft's airworthiness, that is, how the aircraft in question was built and designed to specification. By knowing the specifications of an aircraft in detail, a pilot learns whether some parts of the aircraft must be upgraded, replaced, or even decide whether to fly this aircraft at all. If problems arise due to equipment malfunction in a cockpit during flight, a pilot who is unfamiliar with a layout of his or her aircraft's specifications may not be able to pinpoint a problem or accurately describe it to ATCO. If this were to happen during a commercial flight, then knowing the 'totality' of the aircraft is crucial. It could be said that having a knowledge of the aircraft's airworthiness is a key to maintaining a meaningful conversation with an ATCO. And the definition above specifically states "operation", a descriptor that naturally fits pilots.

All things considered, "Aeronautical" is a more suitable descriptor than "Aviation". For the remainder of the thesis, both AeE and AE are used interchangeably for convenience.

2.1.2.3 Features of AeE

A distinctive feature of AeE is that there are communicative functions unique to pilots and ATCs. According to ICAO Doc 9835 (2010), there are four defined communicative functions of aeronautical radiotelephony communications[7]. These are (ICAO Doc. 9835, 2010, Section 3.4.6):

---

[7] Notice that ICAO Doc 9835 uses "aeronautical" instead of "aviation".

1) Triggering actions
2) Sharing information
3) Managing the pilot-ATC relationship
4) Managing the dialogue

The **Table 6** below lists topics under each communicative functions, which differ based on whether they are spoken by a pilot or ATCO.

Table 6

Communicative Functions of AeE

<table>
<tr><th>Communicative Functions</th><th>Topic Markers</th><th>Pilot</th><th>ATCO</th></tr>
<tr><td rowspan="6">Triggering actions</td><td>a. Orders</td><td>▪ State (non-)compliance with an order.</td><td>▪ Provide (amended, negative, alternative) order.
▪ Cancel order.</td></tr>
<tr><td>b. Requests and offers to act</td><td colspan="2">▪ Agree / refuse to act.
▪ Announce reluctance / unwillingness to act.
▪ Accept / refuse an offer to act.
▪ Request action.</td></tr>
<tr><td rowspan="2">c. Advice</td><td>▪ Request advice.</td><td>▪ Provide advice.</td></tr>
<tr><td colspan="2">▪ Suggest a solution, other ways to correct a problem.</td></tr>
<tr><td>d. Permission / Approval</td><td>▪ Seek approval / permission</td><td>▪ Provide, withdraw, disallow approval / permission</td></tr>
<tr><td>e. Undertakings</td><td colspan="2">▪ Provide service, assist, contact, report, relay.
▪ Agree to undertaking or decision.
▪ State a spontaneous decision to act.</td></tr>
<tr><td rowspan="4">Exchanging information via sharing</td><td rowspan="2">a. Information on the present facts</td><td>▪ Ask about an instruction.</td><td>▪ Provide instructions.</td></tr>
<tr><td colspan="2">▪ Identifying preferences and needs.
▪ Describing a state, action, procedure, process, problem.
▪ Declare a problem.</td></tr>
<tr><td>b. Information for the future</td><td colspan="2">▪ Describe an unexpected event, action.
▪ State an intention to act.
▪ Ask about an intention.
▪ Request a prediction.
▪ Warning.</td></tr>
<tr><td>c. Information for the recent past</td><td colspan="2">▪ Announce a change in action (an action can be also nearly complete) or event.</td></tr>
</table>

	d. Information on the past	▪ Report a problem or accident. ▪ Provide deductive reasoning or explanation of past actions.
	e. Necessity	▪ Declare a compulsory action (only the controller). ▪ Declare an inevitable event or action. ▪ Ask or state necessity.
	f. Feasibility	▪ Ask or announce (un)feasibility or (in)capacity.
Pilot-ATCO Relation	▪ Thanking, complaining, reprimanding, apologizing, addressing concerns, encouraging. ▪ Respond to a complaint, reprimand, greeting, leave-taking, satisfaction, or dissatisfaction.	
Dialogue Management	▪ Self-correction, checking understanding, clarify misunderstanding, Read back, Acknowledge, Providing repetition and confirmation.	

Note. Adapted from Appendix B of ICAO Doc 9835 (2010).

Table 7 below lists common AeE vocabularies that are used by both pilots and ATCOs. From the point of view of a person without aviation background, the words may sound and 'feel' the same. But their meanings are defined in the confines of the operational context. For example, 'Negative' may mean 'No' or 'Incorrect'. But it also can mean 'Permission Denied'.

Table 7

Common AeE vocabularies

AeE Vocabularies	Meaning
"Roger."	I am aware of your transmission [a].
"Acknowledge."	Notify whether you understood what I told you just now.
"Affirmative."	Yes.
"Confirm."	Check the instruction, information, clearance.
"Correction."	I correct what I just told you because there was an error.
"I say again."	I repeat so that you can understand clearly.
"Immediately."	Your action should take place right now because it has safety implications. Act now.
"Read back."	Repeat what I just told you.
"Words twice."	Repeat words twice because it is difficult to communicate.
"Negative."	No, Not correct, Not capable, Permission denied.

"Recleared."	A modification on the part of or all of the last transmission has been made. Follow this new transmission instead.
"Expedite."	Act in accordance with a given instruction immediately.
"Over."	Transmission ends but a response from the receiver is required.
"Contact."	Communicate with ~
"Disregard."	Take no notice of ~
"SQUAWK Ident."	Press "ident" on the transponder [b].

Note. Adapted from NAV Canada IFR Phraseology (2022).

[a] "Roger" is mainly spoken by the ATC.
[b] "SQUAWK" is a transponder phraseology. Requires inputting frequencies into a transponder that helps ATCs identify precise location of the aircraft in question.

However, Borowska (2017) also acknowledged that "the linguistic classification of AE specialized subsets is not so straightforward" (p. 70). This could be because more research has yet to determine what is the most ideal definition for describing a sublanguage that is spoken in aviation community. Lingua Franca, a term referring to a widely adopted and spoken language in a society where natives have their own mother-tongue, is often used as a descriptor when introducing AE (ex. lingua franca of the sky). However, there are two subsets within AE: Standardized Phraseology (SP or SPH) and Simplified Aviation English (SAE) (Borowska, 2017, p.70). Other than the generic descriptor "lingua franca" to just refer to Aviation English (or Aeronautical English), there is no concrete taxonomy for fully describing (and confirming the existence of) the two subsets aforementioned. Borowska (2017) argued that SP and SAE should not be treated as "languages" because they are used for communication in a specific profession (i.e., pilot-ATC communication). Rather, Borowska (2017) referred to the two subsets as "specialized tools". Although discussing a creation of linguistic taxonomy is beyond the scope of this thesis, it would nonetheless help (if there were agreed-upon specific words or even abbreviations to refer to these subsets) researchers to clearly explain concepts without repeating too much the linguistic descriptions of AeE.

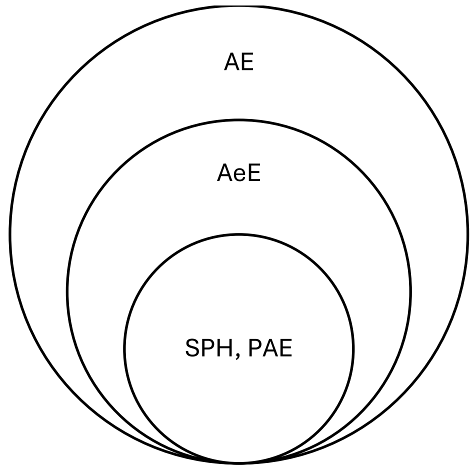

Figure 7. The relation between AE (Aviation English), AeE (Aeronautical English), SPH (Standard Phraseology), and PAE (Plain Aeronautical English). Adapted from Borowska (2017).

2.1.2.4 Teaching and Learning AeE

AeE was developed in response to many accidents that have occurred due to ineffective communication, where English skills were a major contributor (Kraśnicka, 2016). Tajima (2004), who investigated aviation accidents related to AeE, argued that such tragedy is the result of "the lack of shared situation model". A shared situation model could be about flight safety, or the "last line of defense" against an unexpected accident while flying from place A to B. What creates a shared situation model is the language, in this case, AE. If one of the parties does not speak AE fluently nor understand its embedded meaning (or hidden assumptions which require quick semantic analysis. This is only possible when language is completely understood by both parties), then a shared model of what constitutes flight safety could be absent. This, if left not clarified either by hearer or speaker, will lead to communication breakdown.

Considering the fact that there are more speakers whose mother-tongue is not English (ICAO Doc. 9835, 2010, Section 2.5.1), language problems are bound to occur in contexts which native English speakers (NES) could not have imagined. This raises a serious question regarding the creation of SARPs related to AeE: Is it appropriate to use NES as a central "model" or figure (ICAO Doc. 9835, 2010, Section 2.5.2)? It would not be appropriate to recommend every pilot to acquire accents and pronunciation that are on the same level in every linguistic aspect as that of NES. This is impossible to carry out because of the influence of the first language. If this near-sighted standard is adopted (highly unlikely), it will add more time to complete lessons, put undue financial pressure, and impose unnecessary cognitive load on non-native English speakers (NNES).

Furthermore, an instructor who teaches AeE has a lot of responsibilities that are worthy of discussion. The reason is largely because the requirements are stricter than just being an ESL[8] teacher who provides learning materials and teaches basics of English grammar (syntax), vocabularies, and phonology. Anyone, if he or she has commendable skills in English language (reading, writing, and speaking), can become or self-declare as an ESL teacher or English tutor. But this is usually not the case for AeE instructor. Below are some of the necessary qualifications below (ICAO Circular 323 & ICAEA, 2009):

1. Master's or diploma in Teaching English as a Foreign/Second Language (TEFL or TESOL).
2. Knowledge in aeronautical radiotelephony communication.
3. Knowledge in operational characteristics of aviation.
4. Knowledge in ICAO Doc 9835 and ICAO LPR.
5. Cross-cultural awareness.
6. Able to focus on operational aspect of the sublanguage that is critical to safety rather than on correcting its structure.
7. Able to pass on knowledge simultaneously.
8. Able to cooperate with industry subject matter experts.
9. Able to support professionally regarding matters related to students.

It is important to note that all of these are beyond the requirements than a traditional ESL teacher may have. Not only must the AeE instructors have experience in operational settings related to aviation communication (ex. who had been employed at an airport, airlines, or done extensive linguistic research in AeE), but also be aware of how cultural backgrounds of a student may affect his or her desire to learn and improve his or her AeE skills. The latter requires empathy and the ability to understand cultural differences in operational context – that is, the aviation industry.

[8] English as Secondary Language. ICAO Cir. 232 (2009) states that AeE is ENGSP. It belongs to a group of language which serves a specific purpose in highly regulated profession such as health care, finance, and hospitality. ESL teaching does not contain a broader and specific scope of teaching than AeE teaching (p. 23). It is more oriented towards helping NNES become more familiar with English.

Language acquisition is a complicated process, and it is fraught with various challenges. An L2 learner (a student who learns a second language) is faced with a) How to adopt nearly perfect pronunciation; b) How to write without grammatical errors; c) How to find someone and engage in conversation for practicing; d) How to read and answer questions in that language; e) How to improve listening when a native speaker speaks faster than expected; f) How much one should spend in buying language books etc. The very first question, however, that precedes all the considerations above is "why one should learn a specific language". What particular purpose does it serve to learn a language? Is it for job hunting or is it to retrace a lost family history? Learning a new language is a time-consuming process and can be an expensive endeavour depending on an availability of resources (whether there are free resources nearby or a fellow language speaker to practice with). ICAO Circular 323 & ICAEA (2009) also takes a note of this by stating that language acquisition[9] is a "long and costly business" as human behaviour comprises of "memorization, personality, age, culture, sensory perception, past experience, motivation, and social interaction".

Another notable challenge is overcoming ambiguities in syntactic structure, which can cause misunderstanding (Seiler, 2009). An example of a sentence that can cause this is the following: "A boy went to a garage with his bicycles." Does it mean that the boy was going towards a garage by riding his bicycles or did he just enter the garage on his own two feet while carrying his bicycle on his right (or left)? Also, does it mean that he went to the garage which stores his bicycles? There would be less ambiguity if the sentence above was written with more details: "A boy went to a garage while walking on his own two feet with a bicycle on his right". This sentence is clearer and explicit than the previous, as it clarifies the ambiguity that lies in what the boy does. It must be reminded that AeE sentences contain ellipsis, which means that unnecessary words are eliminated to avoid verbosity and cluttering of communication. Unfortunately, this adds a cognitive load on a hearer, who may not understand fully the deeper meaning of a sentence being spoken. Ambiguity might be present when ellipsis is embedded in a sentence. For NNES pilots, a topic of a discourse should be made explicit and must be direct (ICAO Doc. 9835, 2010, Section 5.3.3.5). Speaking in

[9] ICAO Doc 9835 differentiates the definition of language learning and language acquisition (sec. 2.6.2). The difference lies in the question of who (infant or adult) learns language. Learning a language during childhood is more advantageous than learning during adulthood since a child may have more time and opportunity to be malleable to accepting lexical varieties, developing a lexical dictionary subconsciously, or exposed to more situations that are beneficial for learning that particular language. However, in this thesis, language learning will be synonymous with language acquisition.

an unclear way (using jargons specific to a region or even worse, code-switching) will only make a discourse complicated and this will return as a latent threat to flight safety.

There are other errors that could lead to communication breakdown in AeE (see **Figure *8*** below). These errors are divided into mainly three areas: Speaker, Channel, and Hearer. However, in real operational settings, it is not so easy to pinpoint whether an error can be categorized as shown in the figure below. This makes communication a challenging subject to conduct research in because a simple sentence may have multiple meanings depending on the context or situation in which it is spoken. Coertze et al. (2014) summarizes this very well; the researchers described that unexpected events could still occur even if the pilots are aware of possible risks during flight or are attentive to safety concerns, and this precarious situation, if it were to happen, makes language "the final safety net". If the net is not woven properly, that is, if proper use of language (being accurate and precise) is not maintained or there is lack of comprehension, a significant improvement in flight safety may not be seen.

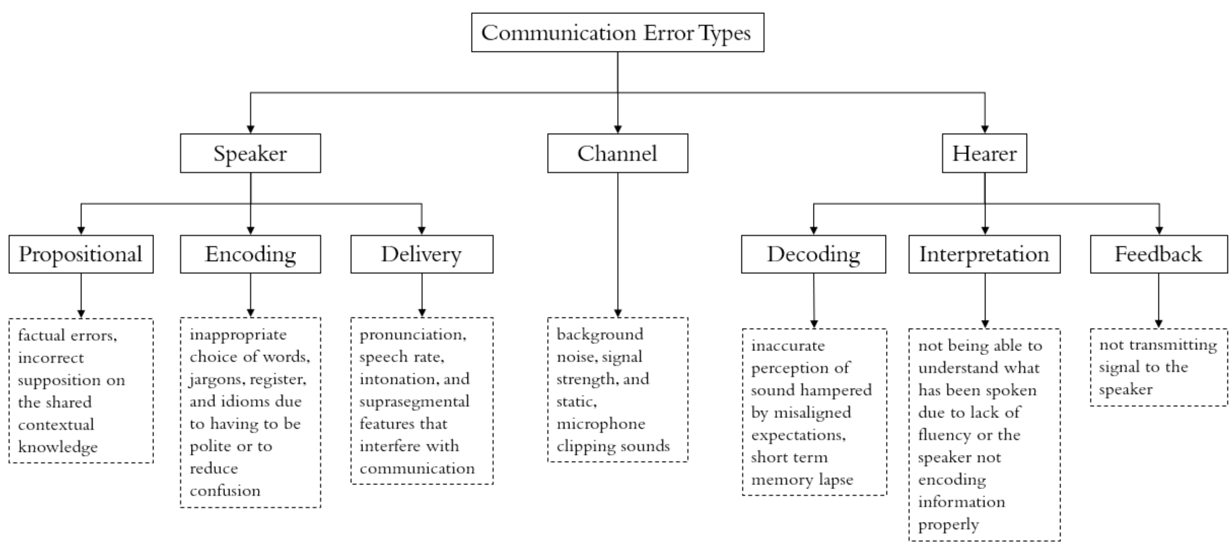

Figure 8. Communication errors (or failures). Adapted from ICAO Doc 9835 (2010).

2.1.2.5 Current Outlook on the AeE Teaching

There are several AeE tests that are being used. Table 3C below shows the tests that are currently being used.

Table 8

Summary of AE Tests

Names of AeE Exams	Where	Types of Questions
EPTA	Korea	• Tape-mediated questions • Interview
Transport Canada Formal & Informal Assessment	Canada	• ICAO SPH Questions • Plain Language Questions • (Aviation-related & Non-aviation related)
Eurocontrol ELPAC	Ireland, UK, Norway, Germany, Netherlands, Belgium, Luxembourg, Czechia, Poland, Belarus, Lithuania, Latvia, Estonia, Finland, Moldova, Romania, Austria, Switzerland, Slovenia, Hungary, North Macedonia, Albania, Greece, Cyprus, Turkiye, Croatia, Kazakhstan, Saudi Arabia, Sudan, Algeria, Morocco, Indonesia, Taiwan, Brazil, Colombia	• Paper 1 Test: Listening Comprehension • Paper 2 Test: Oral Interaction If wanting to achieve Level 6: • Paper 3 Test (or "Level 6 Test")

ICAO does not strongly advocate one AeE exam over another, nor does it provide their own English training services. Rather, it maintains a database of which exams meet the SARPs and LRPs. This database is called Aviation English Language Test Services (AELTS), and it categorizes exams into two levels: 'Conforms' and 'Partially Conforms'. As of May 21st, 2024, there is only one exam that has met the standards: ELPAC (English Language Proficiency for Aeronautical

Communication Test). This raises a question whether other exams that are administered elsewhere across the world can be said to be acknowledged by ICAO. ICAEA (International Civil Aviation English Association) has a frequently asked question section, where they explain that one needs to be careful in choosing private testing service providers that advertise "ICAO approved" or "CAA approved". This warning shows that test constructs employed by private providers could vary in terms of quantity and quality.

In terms of test constructs, ELPAC has three tests: Paper 1,2, and 3 (Eurocontrol, 2021). Its sole purpose is to assess a candidate's "ability to communicate effectively in unusual or unexpected situations and in the context of aeronautical communication using plain English" (ELPAC, 2021, p. 4). Paper 1 is a 35-minute online listening comprehension test. It offers both ATC and Pilot portion. A sample test is also provided, wherein a participant listens to either a short or long transmission of a pilot-ATC conversation and fills numbers (altitude, speed etc.) in a blank text box. There is a total of five parts in this test with varying types of questions and length of recordings, but most of them involve filling in the blank text boxes that ask them to fill in missing words, numbers, or complete sentences. There is a time limit, so a participant must answer the questions "on the go". There is no option to go back once the test is completed.

In Paper 2 test, oral communication is the focus. The instructors or raters look for whether a participant can use standard phraseology (SPH), switch from SPH to PE, disentangle miscommunication, provide opinions, and hold conversation about aviation-related topics (ELPAC, 2021, p. 8). There are three tasks in this test: Task 1a and 1b, Task 2, Task 3. Task 1a (for pilots only), a participant plays the role of a pilot and communicates with ATC (an instructor or assessor in this case). In Task 1b, the instructor plays the role of a supervisor to the pilot (the participant) and the pilot must report details on the events that occurred in Task 1a to the supervisor. Task 2 requires the participant to describe what is happening in a picture provided by the instructor. Task 3 is a final assessment where the participant and the instructor discuss a wider range of aviation-related topics.

Paper 3 test is reserved for those who seek to attain level 6 on the LPR ratings. The test focuses on "communicative proficiency", which emphasizes negotiation and dealing with difficult situations that involve ambiguities (ELPAC, 2021, p. 10). For example, idiomatic usage of the language, identifying ambiguities, asking for clarifications by negotiating meanings are tested. There are four tasks: Task 1, Task 2, Task 3, and Task 4. Task 1 involves describing two pictures and listening to a recording of a person describing one of the pictures. The participant must choose which picture the person was referring to. Task 2 involves listening to a radiotelephony communication that contains a misunderstanding and then summarizing what happened. And the instructor asks the participant to explain it to someone who doesn't know about aviation or is an NNES. So, paraphrasing (that involves controlling the use of technical jargons) is tested here. Task 3 involves listening to a radiotelephony communication and is similar to Task 2, but the instructor places more emphasis on whether the participant can use the language effectively to accommodate (explain to someone who is not familiar with aviation or NNES). Task 4 involves listening to an informal conversation (thus idioms are used) about a particular event in the cockpit or operations room. The instructor asks what those idioms mean and what is happening in the conversation.

As can be seen, the ELPAC test comprises mostly of listening comprehension and oral communication skills. There is no writing or reading component. Because in the cockpit, pilots are not able to see ATCOs via screen. They rely heavily on the conversation they hold with ATCOs to get to the destination. Although oral communication is important, more emphasis must be placed on listening comprehension, as that contains information and context that can help a pilot formulate sentences or make forward arguments to ATCOs nearby.

2.1.2.6 On the Rating Scale and its Implications

AE has a strict testing and rating requirement. ICAO provides Language Proficiency Requirements (LPRs), which are divided into six categories of assessment: Pronunciation, Structure, Vocabulary, Fluency, Comprehension, Interaction. Across these categories are six levels to which a trainee is assigned: Level 1, 2, 3, 4, 5, and 6. Level 4 is the "operational level", which is the minimum level required for communicating on the radio. Level 5 is "extended", and Level 6 is "Expert". A pilot who is on Level 4 is required to complete a re-test after three years, whereas Level 5 is six years and Level 6 is indefinite (Borowska, 2017, p. 33). Those who are rated on Level 1, 2, or 3 are below the language requirement of the ICAO (Borowska, 2017, p. 33). Table 9 shows the Transport Canada's language proficiency requirement scale as per the Canadian Aviation Regulations (CARs) 421.06(4) (Transport Canada, n.d.-a). It is noteworthy that the six categories are the same, but the descriptions might be slightly different.

Table 9

Summary of the Six Categories of Assessment

Categories	Description
Pronunciation	Controlling stress on syllable utterances, recognizing pattern and phoneme, understanding accents.
Structure	Having grammatical knowledge (tenses, modality), upholding appropriate usage of sentences to accurately express meanings, maintaining context relevancy.
Vocabulary	Using words relevant to context (speed, range, accuracy), memorizing new words and being able to use them in a context, understanding morphological applications.
Fluency	Maintaining pace throughout conversation, controlling the use of hesitations and fillers, relying less on repetition or rehearsals before speaking.
Comprehension	Increasing processing fluency (ability to understand), breaking down complex conversation into meaningful discrete units (often with accents or background noise), making inferences.
Interaction	Establishing shared understanding between a speaker and listener, maintaining and supplying new information and responses, catching misunderstandings.

Note. This table was adapted from ICAO Doc 9835 (pp. 2-8 – 2-9). The description column contains the author's summary in his own words.

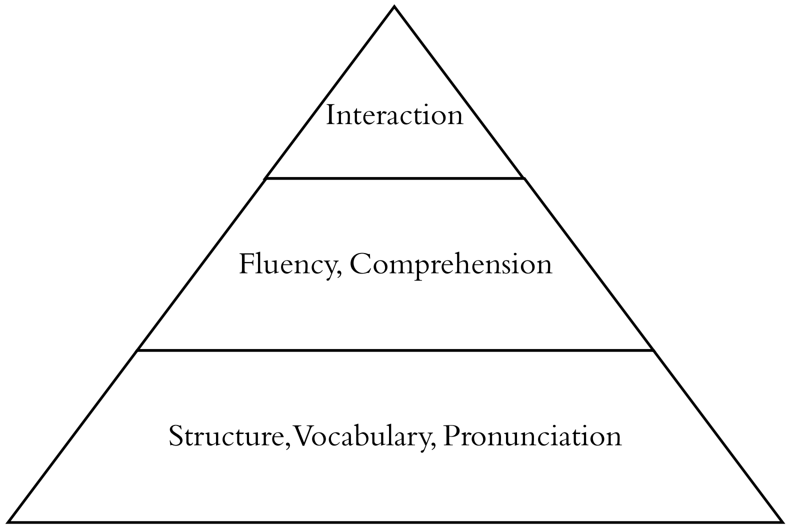

Figure 9. Pyramid of the Categories of Assessment. Adapted from ICAO Doc 9835 (2010).

It is interesting that ICAO Doc 9835 represents these categories of assessment as a pyramid, with the top being Interaction, the middle filled by Fluency and Comprehension, and the bottom with Structure, Vocabulary, and Pronunciation (See **Figure *9*** above). It is noteworthy that pronunciation forms the basis of Fluency and Comprehension (along with Structure and Vocabulary). This hints that controlling the accents when speaking is important. If it is not controlled, then miscommunication (due to not understanding what has been said) could occur. Although this pyramid summarizes well the relationship between the categories of assessment, it does not explain, for example, how one's vocabulary or pronunciation can be improved. See **Figure *10*** below.

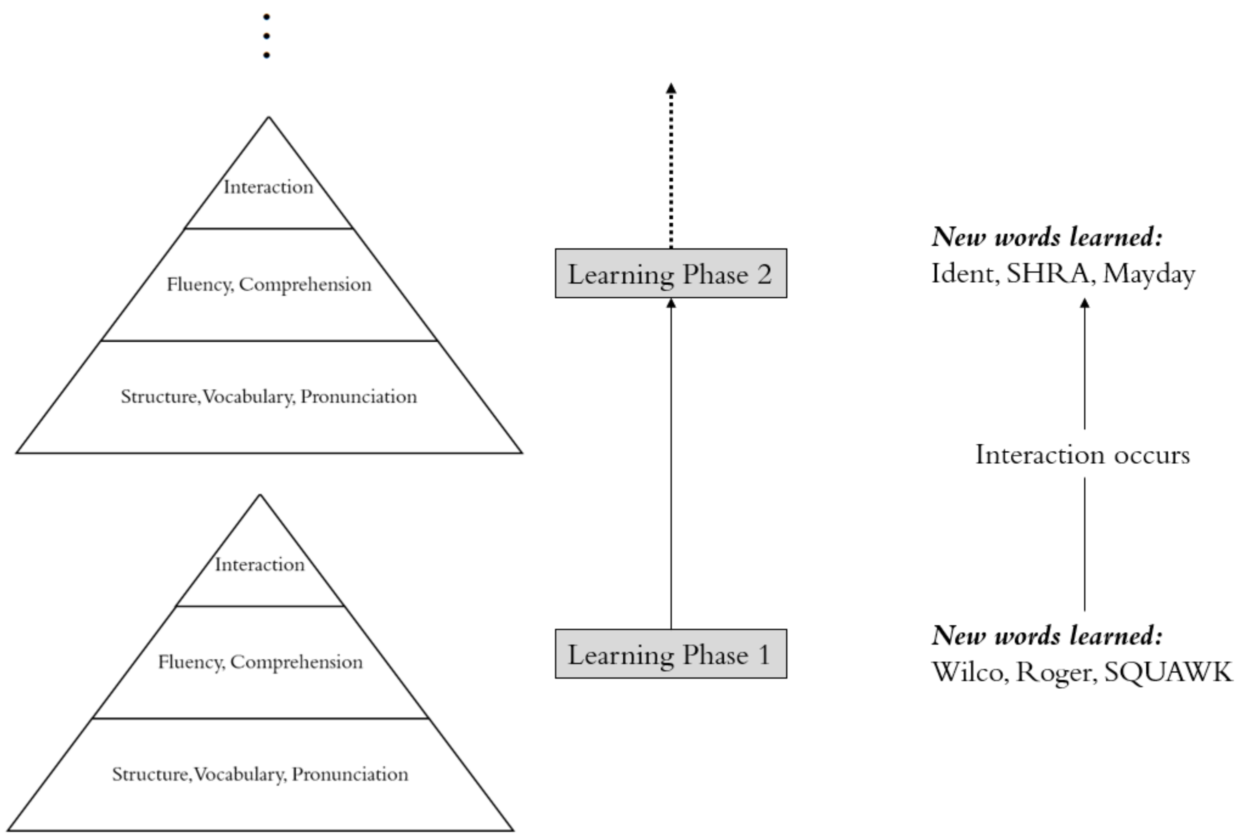

Figure 10. The extended pyramid of the categories for Vocabulary.

Figure 10 includes learning phases whereby people learn new words after an interaction. For example, an individual learns new words through interaction, perhaps from a fellow pilot or from a senior pilot. This individual may then search for the meanings of these words and see example sentences that use these words. After comprehending and developing fluency around those words, he or she may then begin using those words in a conversation (Interaction). From that subsequent interaction, the individual will learn another set of new words. Then he or she will likely repeat the process described above. This way, the individual maintains lexical competency (and thus strengthening lexical mapping), thereby contributing to language maintenance via constant updates.

The above figure can be made for Pronunciation as well. For example, a NNES having a dialogue with NES could learn that his or her accent might need to be controlled in order to have a more fluid conversation next time (Learning Phase 1). When meeting another NES, the NNES will try to adopt native-like pronunciation but may realize that it would be too difficult (Learning Phase 2). So, he or she finds the middle ground between what could be a "good enough" and "not good enough" pronunciation (Learning Phase 3). However, it must be reminded that changing the accent is very challenging. This is because (doing so) would mean the deliberate change of one's own way of pronunciation which has been a part of him or her for many years. To get to the NES level of pronunciation (that is, the elimination of unique accent that one has) is an endeavour that is time-consuming and often impossible. The point of the figure above is that language maintenance is a continuous process.

Another point of discussing the pyramids above is to question whether the current LPRs (i.e., the Six Categories of Assessment) are adequate for capturing the dynamics within a language variety. When there are so many different dialects, and specifically considering how speech sounds and patterns are made, creating a unified standard test construct to govern all these linguistic aspects is going to be challenging.

The author is of the position that introducing a new category that regulates the use of colloquial language or code-switching might be beneficial to pilots. Although code-switching is not the focus of this thesis, discouraging the use of a native language when there are other foreign airplanes flying in the vicinity could reduce stress for (foreign) pilots. Suppose that a South African ATCO starts speaking in Afrikaner when communicating with a South African pilot flying in vicinity. A British pilot overhearing this conversation via radio will be confused as to what they are saying, interfering with the understanding of the air traffic in that area. When there is already a clear evidence that code-switching even in colloquial conversation causes listening anxiety (Sumalinog, 2018), it will be more difficult for pilots to understand conversation in a language that he or she is not familiar with especially during flight. Ultimately, this impacts situation awareness. When accent is already bothering pilots for the duration of a flight, introducing code-switching might worsen stress that originates from long-haul flight.

The above discussion also leads to the way LPRs were created. Unfortunately, there is lack of sufficient information on the development of the LPRs (Knoch, 2014). The six categories of assessments are basically the foundations for making test constructs. Kim and Elder (2015) found that the English test (EPTA) completed by Korean pilots and ATCOs had irrelevant test constructs. This finding is of significance because it shows that there is a discrepancy in what is tested and what should be tested according to ICAO LPRs (i.e., six categories). Test constructs should be broadly based on the six categories. As stated in (ICAO Circular 323 & ICAEA, 2009), placement tests must assess skill areas that are related to ICAO LPR scale with more emphasis on oral exercises (p. 18). In terms of oral skills, ICAO Doc 9835 (2010) specifically states that for radiotelephony communication to occur effectively, 'Pronunciation' on the LPR must be highly prioritized (sec. 5.3.3.6 (a)). When the distance between the test constructs and ICAO LPRs is increasing, what would be the effect of that deteriorating test quality? With minimal information on the validity and reliability of tests (Alderson, 2010; Kim & Elder, 2015), it is hard to grasp whether ICAO LPRs achieve the objectives as they are set out to do. In conclusion, LPRs may need to be updated to reflect the changing conditions of the industry and linguistic evolution of a language.

2.1.2.7 On the Cross-Cultural Communication and Accents

One of the qualifications that an AeE instructor should hold is having awareness of cross-cultural communication. This means that an instructor has a responsibility in identifying whether cultural backgrounds of trainees can hinder their ability to effectively learn AeE, which requires more direct form of communication by stating intentions clearly. For instance, power distance across culture is varied. Can a subordinate person (a first pilot) immediately talk back to a person of higher rank (captain)? Can an individual from a country that values individualism fit in with an organization that promotes collectivism? These questions are important as they can relate to the outcome of a trainee's performance in training. Also, considering the cultural background of a learner is one of the core philosophies in CBE.

Hazrati (2015) provides an example of a real-life example of a discourse between a Korean pilot and ATC (Korean Air Lines Boeing 707, August 2nd, 1976). The plane crashed due to CFIT (Controlled Flight into Terrain) with a total of five casualties. ATC told the pilot to follow SID 11 (Standard Instrument Departure, which is to be immediately followed after takeoff). The pilot asked what SID stood for, to which the ATC clarified. However, the pilot's action was silent after. Hazrati (2015) argued that Korea's high-power distance might have prevented the pilot for asking clarification.

Speaking standard phraseology seems to be regarded as an arduous task by NNES speakers based on the figure above. This is not surprising since learning a new language is stressful (ICAO Doc. 9835, 2010, Section 5.3.3.1). As discussed previously, the problem of finding a native speaker to practice speaking, writing, and reading with is already challenging and costly. Not only that, trying to emulate the pronunciation of a NES puts immense pressure on a NNES. Adding to the fact that there are no visual cues during flight, it can be a daunting task that needs to be overcome by constant practice. In brief, it is a serious uphill battle for NNES as they tend not to rely on contextual information, but rely more on pronunciation to carry out their flight duties (ICAO Doc. 9835, 2010, Section 5.3.3.6).

Accent is briefly mentioned in ICAO Doc 9835 (2010) along with dialect. It stated that intelligibility of accent can be controlled by reducing large information into segments (speaking in pieces), using breaks, and decreasing speed of the speech (ICAO Doc. 9835, 2010, Section 5.3.3.7). Furthermore, it stipulated that speakers, regardless of NES or NNES, need to make their accent (or dialect) understandable to the aeronautical community (ICAO Doc. 9835, 2010, Section 5.3.4.1). It also acknowledges that there is no "single acceptable accent", therefore speakers should: a) Know that there are cultural influences to communication; b) Be open to accepting accents and; c) Come up with a method to maintain effective flow of conversation to prevent misunderstandings (ICAO Doc. 9835, 2010, Section 5.3.4.3).

ICAO is aware of the language-related issues such as accents faced by pilots and ATCOs. It would have been helpful if there was a standalone policy explaining further why accents need to be given more attention. In all likelihood, radio telephony communication (RTC), or simply pilot-ATC communication, will have foreign accent embedded because majority of the countries are not English-speaking countries. There is evidence that listening to accents is found to be detrimental to understanding (Baugh & Stolzer, 2018; Sumalinog, 2018). Thus, a listening training for accent could be beneficial to ab initio pilots who are learning about aeronautical communication for the first time. The more they are exposed to foreign accents in their earlier stages of training, the more comfortable they will be able to deal with accented instructions coming from ATCOs.

Table 10

Language Proficiency Scale of Transport Canada

Level[10]	Pronunciation	Structure	Vocabulary
Expert	Pronunciation, rhythm, intonation, stress may have regional variations, but do not negatively affect with level of understanding.	Basic, complex grammar and sentence structures are under control.	Can handle the proper use of vocabularies in its idiomatic forms or with nuance and sensitivity to convey messages in both familiar and unfamiliar situations.
Operational	Pronunciation, rhythm, intonation, stress have regional variations, and these affect the level of understanding sometimes.	Basic, complex grammar and sentence structures are controlled to the extent that they can be used creatively, but errors can be found in unexpected situations, but rarely do these affect the meaning contained therein.	Can usually handle the proper use of vocabularies in work-related situations and common topics. And can paraphrase effectively even though the range of vocabulary for describing usual or unexpected situations are lacking.
Below Operational	Pronunciation, rhythm, intonation, stress have regional variations, and these affect the level of understanding frequently.	Basic, complex grammar and sentence structures contain errors even in predictable circumstances and thus affect the meaning.	Limited in using vocabularies properly and paraphrasing due to insufficient vocabulary knowledge.

Note. Adapted from (Transport Canada, 2019). Sentences in the table above have been rephrased by the author. The table only shows Expert, Operation, and Below Operational levels. See next page for the other three categories of assessment.

[10] The LPR scale used by Transport Canada is unique to Canada's aviation. It is only slightly different from ICAO's LPR in that the former has three levels instead of six as defined in the latter. However, the six categories of assessment (Pronunciation, Structure, Vocabulary, Fluency, Comprehension, Interaction) remain the same.

Level	Fluency	Comprehension	Interaction
Expert	Can speak naturally with controlled flow. Can use emphasis, markers, and connectors.	Cultural subtleties and linguistic cues are accurate, making comprehension consistently accurate across all situations.	Appropriately responds to and easily interacts with both sensitive to verbal and non-verbal cues.
Operational	Can speak with formulaic speech and sometimes spontaneous speech. Can use emphasis, markers, and connectors, but in limited fashion. These do not negatively affect the level of understanding.	Comprehends mostly across work-related or common situations and when accent is present. But when complexities are involved in a conversation, the speaker may ask for clarification.	Properly and quickly responds in the manner that is informative and continues to engage in information sharing in uncertain situations or check for clarification when misunderstanding occurs.
Below Operational	Can speak but spotted with pauses, inappropriate phrasing. Hesitations or slow processing fluency can cause ineffective communication.	Comprehends often across work-related and common situations and when accent is present. But may not understand completely when complexities arise in a conversation.	Able to maintain exchanges of information on common situations but insufficiently deals with uncertain situations.

2.1.2.8 Conclusion

By knowing the definition and brief history of AeE, it is not hard to imagine why language is an important issue. Unfortunately, it has been overlooked and was not given proper attention. It must be known that language issues are difficult to resolve, hence why they need to be given more spotlight. This is largely because languages contain numerous meanings depending on who says it, where something is said, and how it is said. Languages can be a point of contention or a precursor to communication breakdown. In the context of aviation, a pilot or ATCO must carefully choose words to ask for requests or instructions. A shared mental model of what constitutes flight safety can be only established if the two parties to the communication speak the same language. If there is a mismatch, a communication breakdown or miscommunication could potentially endanger the lives of passengers on board.

2.2. Study 2: Themes in Aviation English and Accent Literature

2.2.1 Introduction

A look into AeE literature provides two valuable insights: (a) How AeE is being tested and evaluated by researchers; (b) How accents are investigated by studies outside aviation literature; (c) How accent studies can have societal implications. This literature review may be a good resource for a researcher who is interested in conducting human factors experiment or who knows very little about the link between linguistics-human factors. In fact, some of these studies share similarities in how a research experiment is set up. For example, obtaining interviews with ATCOs or pilots would be a good idea if a researcher wants to know directly whether current LPRs need improvement or not (see Kim & Elder, 2015). Another example would be using a method of analysis which uses pilot-ATC recordings to examine specific cases (ex. cases of misunderstanding, identifying the most used words from a single conversation, level of stress etc).

2.2.2 Procedure

Unfortunately, studies involving AeE accent are scant (Dissanayaka et al., 2023, p. 80). There are not many articles that directly examine the impacts of foreign accents on pilots' or ATCO's level of understanding (or processing fluency). And rarer is the creation of original test constructs (questions) to assess how participants do under different circumstances. It would have been helpful if such was not the case because this could have helped establish some guidelines for setting up basic experimental methods for future researchers. Based on observation, accent related studies frequently appear in fields that are unrelated to aviation studies, like psycholinguistics or acoustics. Only few were related to aviation in general (Tiewtrakul & Fletcher, 2010; Clark, 2017; Wu et al., 2019; Dissanayaka et al., 2023). The focus in this review was on finding the study topics, real-life implications, and the types of methods employed in their research.

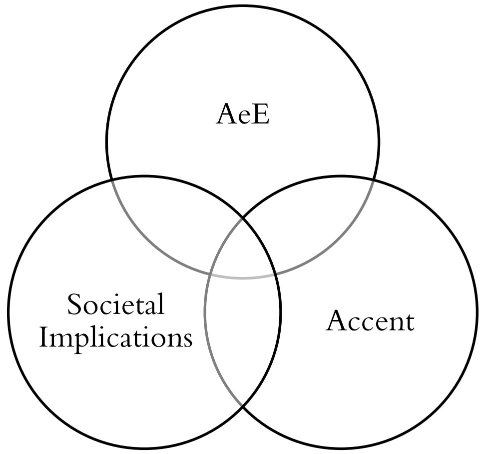

Figure 11. Venn diagram of AeE, Societal Implications, and Accent.

To fill the gap of knowledge in this specific area, the author created the following themes (see **Figure 11** above): AeE, Accent, and Societal Implications. These themes were used to make **Table 11**, **Table 12**, and **Table 13** containing summaries of some notable AeE-related articles. The selected articles had at least inspired this thesis or provided a practical implication that was worthy of discussion.

2.2.1 Theme 1: AeE

The most influential AeE articles to this thesis are shown in **Table 11** below. Although the field each article belongs to is different, they are interrelated in a sense that they focused on improving AeE skills of pilot or ATCO trainees. For instance, Moskalenko and Didenko (2018) developed a computer software that can help in training future pilots' English listening skills. The software had six main tasks that asked a participant to translate an instruction from ATCO to English, reply to an instruction, translate information provided by a pilot into English, understand information from ATIS (Automatic Terminal Information Service) or ATC clearance, and finally fill in the words to fragmented sentences (Moskalenko & Didenko, 2018, p. 192). This study would be of interest to linguists and psychologists conducting research in cognitive task analysis.

Table 11

Table of AeE-related Articles

Article	Field	Study Topics	Methods and Tasks
* Tiewtrakul and Fletcher, 2010	Human Factors	Communication errors and challenges (when foreign accent is involved).	Conversation analysis of voice recordings of regional accents in Thai airspace.
Sullivan and Girginer, 2002	Linguistics	Usage of AeE in Turkish airspace.	Analysis of voice recordings of pilots and ATCs, observation from the tower, interviews.
* Baugh and Stolzer, 2018	Aeronautics	Communication problems due to language in general aviation.	Archival approach; analysis of NASA ASRS data[a].
Moskalenko and Didenko, 2018	Information Technology	Development of a computer software for training listening skills of pilot trainees.	Recording performance of pilot trainees.
Kim and Elder, 2015	Linguistics	Examination of underlying constructs of an English test for pilots and ATCs in Korea and ICAO LPR.	Questionnaire and interviews of aviation personnel.

Note. The articles with an asterisk (*) in the first column are related to aviation and have investigated accent empiricially or identified accent as one of major issues in radiotelephony communication.
[a] ASRS (Aviation Safety Reporting System) data is a database where aviation personnel can submit their flight experiences that caused safety concerns. Their submission omits personal details and thus the database is non-punitive.

An important concept that sprung from the analysis of the AeE-related research above seemed to be its emphasis on 'information asymmetry'. That is, what pilots see, feel, think while flying cannot be experienced by ATCOs on the ground. As a result, ATCO may not fully understand the situation a pilot is facing, and this could possibly lead to providing an ineffective instruction based on incomplete information. On the other hand, pilots work with information derived from his or her surrounding environment that is not available to ATCOs, so their "sudden" actions while flying (i.e., deviating from a flight path or an altitude) could be perceived by the ATCO as a deviation from SOP (Standard Operating Procedure). Thus, information asymmetry could endanger the safety of any flight. This is an important implication for human factor experts because it implies that relying on the so-called "comprehensive training package" alone cannot improve performance. A deeper issue such as this is yet to be investigated and requires expertise from academics from interdisciplinary backgrounds. Reducing information asymmetry could be a key to reducing flight incidents or accidents.

Hazrati (2015), who investigated aviation accidents that arose from cultural differences, raised an interesting point that cultural differences of the people who speak English as mother-tongue and those who do not need to be given more attention. This is an important point because language is an asset of culture and thus inherits all cultural customs of which an interlocutor is a part. For example, Korean has honorifics which are used when speaking to an elderly person. On the contrary, English does not have honorifics. The existence of a linguistic characteristic that is not found in another language could possibly impact the flow of communication.

The study by Tiewtrakul and Fletcher (2010) is a good example that supports the arguments discussed above. The researchers investigated voice recordings of pilots and ATCOs and analyzed their speech while taking note of the influence of foreign accents. It is noteworthy that they quantitatively measured impacts of accents on the level of understanding based on three types of communication parties: Thai Pilot – Thai ATCO (local group), NES Pilot – Thai ATCO (English group), and NNES Pilot – Thai ATCO (non-native group). They found that the non-native group was having the most trouble in understanding each other, whereas the English group were better than the former, and the local group had the least trouble (Tiewtrakul & Fletcher, 2010, p. 234). This showed that accents can affect how or what a pilot understands. The researchers also noted the role a language background can play in maintaining efficient communication (Tiewtrakul & Fletcher, 2010, p. 238). It can be said that there is truthfulness in the argument that intercultural knowledge is a key to solving information asymmetry between a pilot and ATCO.

Have there been more language-related issues in aviation in the past? This is an important question to answer because it could give insights into how language issues have affected the behaviours of pilots and ATCOs. This was investigated by Baugh and Stolzer (2018). By extracting language problems in aeronautical communication from NASA's ASRS (Aviation Safety Reporting System) database, the researchers noted proficiency issue with standard phraseology (Baugh & Stolzer, 2018, p. 8). Accent was also noted to be related to the frequency of repeating requests, trouble in understanding calls (Baugh & Stolzer, 2018, pp. 8–9). This finding confirms that accents do interfere with processing fluency (i.e. comprehension), as echoed by Tiewtrakul and Fletcher (2010).

The study by Kim and Elder (2015) adds a unique insight by incorporating the opinions of the domain experts regarding the constructs of an AeE test in Korea. The researchers asked what they thought of EPTA (English Proficiency Test for Aviation; administered in Korea), whether it was helpful in improving radiotelephony communication skills. They found that there were more negative answers from the respondents who cited its test validity as the main concern (Kim & Elder, 2015, p. 137). The ICAO LPR was also perceived as negative, with most respondents citing the three reasons: (1) ICAO seemingly attributing NNES's poor English as a leading cause of accidents; (2) Unfair treatment to seniors just because they are not as fluent in English as the newer generations; (3) NES not honouring the radiotelephony conventions (Kim & Elder, 2015, p. 145). The negative responses obtained in the researchers' study were not surprising as there are doubts on whether the set of competencies outlined by ICAO and AeE tests can be considered insufficient in terms of influence and adequacy (Tajima, 2004; Kraśnicka, 2016).

Reason (2) is interesting because it counters the argument that newer generation of aviation professionals are "better" performers than the seniors who have built know-hows for many years in the industry. One of the reasons that information asymmetry could occur might be attributed to this finding. Suppose a pilot needs a specific instruction from an ATCO, but the ATCO is inexperienced and instead gives an instruction that does not align well with the pilots' information. Then the probability of an unfortunate incident or accident occurring might increase because such instruction might be construed by the pilot as either 'negative' or 'not fitting the situation he or she is in'. The finding perhaps suggests that information asymmetry can be overcome if the know-hows from the seniors are bequeathed to the next generation of aviation professionals. In turn, the newer generation can help seniors get better in AE by providing newly updated knowledge. From the seniors' perspective, they will play the role of an advisor who provides wisdom and undocumented knowledge that help an inexpert make a more effective decision. This resonates with a discussion on enabling two-way communication in a learning environment guided by CBE.

Sullivan and Girginer (2002) documented the usage of AeE in Turkish airspace. They noted an interesting case of code-switching between a Turkish pilot and ATCO, such as Turkish being used more frequently in a conversation (Sullivan & Girginer, 2002, pp. 401–402). Some instances included numbers not being read individually which was improper (ex. "triple six" instead of "six six six") (Sullivan & Girginer, 2002, p. 401). These findings point out that non-standard phraseology is being used during real operations. This causes confusion to those pilots – especially if they are NNES – who are listening on the same frequency. It could also cause information asymmetry by confusing ATCOs as well if they happen to be NNES, because they may not have known that such variation in AeE existed, leading to mistaking an aircraft with a similar sounding aircraft.

In summary, the articles above are explained in relation to why information asymmetry might occur. A discussion on the impacts of information asymmetry is also provided. There are five conclusions to consider:

- First, there is a need to develop competency in intercultural interlocution and knowledge for pilots and ATCOs, which will help them understand the differences among cultural norms embedded in languages other than English.
- Second, accents do interfere with understanding. This is obvious given the fact that accents make it harder to interpret what is being said. A deflection in the standard pronunciation of a word can cause miscommunication between NES and NNES, NES and NES, and NNES and NNES.
- Third, an uninterrupted transmission of know-hows from seniors in the industry to new professionals can help the latter make an effective decision. From efficiency standpoint, this is ideal because the new professionals can 'learn-on-the-go' without having to, for example, ask repeatedly about what to do in situations x, y, z. Otherwise, they would have to struggle with trial and error (this would be very risky considering that fatal accidents can occur even with a minor deviation from SOP).
- Fourth, an unconventional usage of radiotelephony communication must not be used. This has a specific implication on making an AeE test, especially on the usage of unfamiliar abbreviations for callsigns (ex. "triple three" instead of "three-three-three") or colloquial conversations spoken in a foreign language (ex. code-switching). It is possible that unconventional abbreviations may be understood by ATCOs working in busy airports where it is the hub and spoke for major international airlines, but this would not hold in many parts of the world that receive fewer tourists. Identifying such non-standard phraseology could be included in a new AeE test to observe whether a participant can detect it as a problem. As Park (2018) pointed out, providing authenticity in contents is necessary to enable trainees to use cognitive strategies to handle realistic communications. In brief, authentic content that reflects such usage (i.e., a real-life scenario evidence-based training) can be helpful to pilot (or ATC trainees). Zhao et al. (2017) also noted that a customized program in AeE would benefit future pilots.
- Fifth, a development of a computer software or program for AeE teaching and testing is crucial. To meet rising demands for air travel, more airlines will start looking to recruit

competent pilots (and airports will be looking for competent ATCOs). Since NNES are the majority when it comes to communication in the air (Borowska, 2018), an educational tool that can be utilized for teaching many aspects of AeE (accent, understanding ATIS, translating instructions etc.) is needed. A simple computer program that includes some form of interaction (such as listeners playing a script and answering questions) could help them learn faster than a traditional classroom setting. This can be more cost effective from the perspective of a trainer because computer programs only require a personal computer. From an instructor's point of view, a computer program can be coded (or updated) with new sets of information to enrich the contents of the already existing curriculum. An important update from the aviation industry such as a change in training policy or an addition of a new training standard may trigger this action to happen.

2.2.2 Theme 2: Accent

The articles related to accents were mostly found in Psychology (see Table 12 below). This was surprising as the author previously thought that accents were not of interest to psychologists but had more to do with linguistics. However, the study topics investigated by each article overlapped most of the time and the methods and tasks employed in the articles were similar as well. The main research experiment style in these articles is having participants do one or more auditory and visual tasks (Stroop test, recalling a specific word or phrase, vocabulary knowledge test, memorization) under few specific conditions (ex. background noises). Their scores in those tests would then be used to conduct hypothesis testing with quantitative information. But it is noteworthy that these experiments were not carried out in a similar fashion most of the time, as there were some additional materials, parameters, or hypotheses appearing in the later stages of experiments. The precision of results from the studies needs to be carefully interpreted in relation to aviation.

Table 12

Table of Accent-related Articles

Article	Field	Study Topics	Methods and Tasks
Lev-Ari, 2015	Psychology	Working memory, top-down processing, comprehension.	Verbal working memory task, Eye-tracking.

McLaughlin et al., 2018	Psychology-Linguistics	Working memory, inhibition, rhythm perception, vocabulary in adverse listening conditions.	A group of auditory and visual recognition tasks with noises (phrase recognition, Stroop test, recalls, vocabulary tests).
Cristia et al., 2012	Psychology	Accent adaptation, speech perception, aging.	Analysis of literature on accent adaption based on age cohorts.
Dragojevic and Giles, 2016	Communication	Language attitudes, accent, processing fluency, background noise.	Auditory task (listening to PAE in American English, Punjabi English accent with noisy conditions introduced later)
Banks et al., 2015	Acoustics	Vocabulary, inhibition, working memory, accented speech	Speech recognition test, Stroop test, vocabulary knowledge test, working memory tests.
* Dissanayaka et al., 2023	Psychology	Accent, politeness markers, information density, speech rate, communication errors	Correct readbacks identified from the recordings of radio transmissions at four airports.
* Wu et al., 2019	Psychology	Accent, communication errors, omissions, mistakes	Correct readbacks identified from the recordings of radio radio transmissions at Sydney Airport.
* Clark, 2017	Independent research supported by British CAA	Overview of Aviation English (including accent)	Accents impose stress, loss of situational awareness, and prolonging delay in radio transmissions.
Note. This table shows types of training materials used, parameters measured (study topics), what accents were used (if provided). The articles with an asterisk (*) in the first column are related to aviation and have investigated accent empirically along with other linguistic parameters.			

The difference in the methodological approach between the AeE articles (**Table 11**) and Accent articles (**Table 12**) seemed to be that the former covered both previous aviation accidents and conducted either a discourse, conversational, or quantitative analysis. Their scope of analysis was much more varied, but their findings were generally in unison that training standards must be improved to account for foreign accents. On the other hand, Accent articles analyzed numerical data with much more scrutiny. That is, their methods involved participants who were assigned to tasks or groups. Their scope of analysis was relatively fixed, but they were generally applicable to human behaviours under specific contexts, which may not be related to aviation.

Dragojevic and Giles (2016) conducted an experiment to see whether listeners' attitudes changed when they listened to a story read in the standard American English accent versus Punjabi English accent. One of their hypotheses was interesting, which was finding out whether the listeners viewed favourably the American accent or someone with a foreign accent. Their results showed that listeners felt more friendly and sociable towards American English accent than Punjabi English accent (Dragojevic & Giles, 2016, p. 19). They concluded that when processing fluency (i.e., the level of understanding) is interfered with, it influences attitudes towards an unfamiliar accent more than a familiar accent (Dragojevic & Giles, 2016, p. 20). This is related to the finding of Tiewtrakul and Fletcher (2010) that accents caused more cases of misunderstanding in the English group than the local group.

The study by McLaughlin et al. (2018) provides an insight into how an experiment can be set up for accent-related research. First, they used a group of auditory and visual tasks. Second, their experiment included background noise in each of the four listening conditions, two of which contained non-native accent. One of the tasks was a phrase recognition test in which the participants were asked to guess what they heard and input their guesses in a box that appeared on a computer screen. Using the scores from variety of tests, they conducted statistical analyses. One of the important points raised from this study was that some listeners may have a larger vocabulary than others, helping them decode the meaning of a speech affected by noises (McLaughlin et al., 2018, p. 1567). Although somewhat distant from the findings from this study, this point could possibly explain why some NNES pilots and ATCOs do code-switching during aeronautical communications, even if noises are embedded. For example, in one's own native language, a pilot may have more larger vocabulary knowledge than in the second language (ex. More words available to use to convey specific meanings). Their native language allows them to convey more precise meanings when communicating with ATCOs, therefore not having to worry about making mistakes. On the other hand, when communicating in English, they have to worry about decoding implicit meanings (ex. listening in English, translating it into their native language, decide what each word means).

The study by Banks et al. (2016) had a similar approach to that of McLaughlin et al. (2018) in that it included a background noise. One of their findings was that vocabulary knowledge could help in figuring out meanings in accented auditory inputs (Banks et al., 2015, p. 2022). This suggests that the more vocabulary one knows, the better he or she understands the speech with a foreign accent and a noise.

Lev-Ari (2015) investigated whether speech from a non-native speaker induces a listener to rely more on context. This article provided an insight that listeners who have high working memory are more reliant on contextual information to deduce what has been said in accented speech (Lev-

Ari, 2015, p. 7)[11]. It is also interesting that when a listener hears a voice, an expectation is created regarding the speaker's gender which can influence the interpretation of the speech (Lev-Ari, 2015, pp. 2–3). These findings show that listeners take linguistic cues from various hearing conditions, including the identity of a speaker. This means that the speaker's identity might influence their comprehension.

The findings of Lev-Ari (2015) pair well with those of Hazrati (2015). Intercultural dialogue between two speakers from different cultures will encounter confusing moments (ex. using a wrong or unfamiliar word by a non-native speaker to point to an object). A listener with a high working memory will adapt faster to the non-native speaker's speech by referencing some expectations about this speaker. For example, Konglish (a portmanteau for English and Korean) is a group of loanwords from English that are Koreanized and used by most Koreans. One example word is "handeupone" (literally "hand phone"). A NES having a conversation with a Korean speaker would not understand initially what this word means, so he or she will develop an expectation around this Korean speaker as 'someone who uses Konglish to refer to a cellphone'. This expectation will let him or her know that next time when in conversation, more Konglish words could be encountered. If this process is repeated many times, then the NES in question would become more familiar (or "naturalized") with Konglish words (see **Figure *10*** for a relevant discussion).

An example in an aviation context would be a NES captain who has many experiences in flying to, for example, ICN (Incheon International Airport, Korea) from LAX (Los Angeles International Airport, U.S). The captain would have listened to accented speech from Korean ATCOs so many times that he or she would be able to interpret meanings.

Cristia et al. (2012) examined how people from various age groups (infancy to older adults) perceive and adapt to accented speech. A few important points were raised in this article. First, accents incur processing costs (that is, impose greater difficulty) which manifest as cognitive strains and linguistic challenges (Cristia et al., 2012, p. 8). Second, although accents can have an impact on accuracy and processing speed, listeners across all ages can adapt to accents regardless of which age cohort they belong to (Cristia et al., 2012, p. 9). Third, although there remains a few differing opinions regarding the "format of adaptation", listeners might store knowledge on newly acquainted

[11] In the researcher's article, there is an example of the word 'pie'. When a non-native speaker says 'pie', then a native speaker could think one of many different pies that are perhaps not traditionally defined as a 'pie' in his or her culture. The researcher mentioned brownie (Lev-Ari, 2015, p. 3).

accents (Cristia et al., 2012, p. 9). This finding resonates with the those of Bank et al. (2016) and McLaughlin et al. (2018). This suggest that even if a listener does not understand fully when listening to accented speech, one can rely on vocabulary knowledge to find parts of meanings.

Lastly, both articles by Wu et al. (2019) and Dissanayaka et al. (2023) investigated number of correct readbacks from the live recordings of pilot-ATC radio transmissions. These two articles specifically focused on accents held by the interlocutors, hence investigating whether accents can interfere with their processing fluency. Whereas the former focused on live recordings from one airport (YSSY; Sydney Australia), the latter used live recordings from four airports (YSSY; Sydney, VHHH; Hong Kong, KLAX; Los Angeles, RJTT; Tokyo). The methodology of both articles is similar in that they used live recordings, not recruiting participants to have solve auditory or visual tests. The results, however, were somewhat contradictory. Wu et al. (2019) found that NNES pilots who had accents made more communication errors than NES pilots. On the other hand, Dissanayaka et al. (2023) found that NES pilots made more communication errors than NNES (or accented) pilots. This could be because the former only investigated one airport, whereas the latter investigated four airports, the two locations which were not a part of English hemisphere. Furthermore, Dissanayaka et al. (2023) also noted that messages that were lengthy with four or more items accompanied greater number of errors. Although NNES face more difficulties in learning AeE, NES are also not exempt from encountering difficulties when having a conversation in AeE.

In summary, this section has shown how an individual perceives and adapts to accented speech. A listener's attitude changes when speech is accented, perhaps unfavourably. A NES may form an expectation around a non-native speaker, using these expectations access lexical storage by mapping words that sound similar to the word spoken by a NNES. It seems that vocabulary knowledge matters when decoding meanings of accented speech. This has an important implication for aviation because having a larger vocabulary can make it advantageous, even in the prevalence of accents, to describe situations in detail, thus helping in preventing incidents that lead to a fatal accident. Also, as a pilot, it is necessary to be aware of foreign accents that mask a standard pronunciation of an AE word, which is as important as just learning the meaning of that AE word and when to use it. Therefore, learning aviation vocabulary while being aware of the influence of foreign accents is desirable.

2.2.3 Theme 3: Societal Implications of AeE and Accent

Table 13

List of Articles that provide societal implications of AeE and Accent

Article	Field	Methods	Implications
Lev-Ari and Keysar, 2010	Social Psychology	Comparison of 'truthfulness' of statements spoken by non-native and native speaker.	Accented speakers will be viewed less favourably than native speakers without accents.
Shum et al. 2021	Psychology-Linguistics	Use of High-Variability Perceptual Training Program to adjust to voiced/voiceless stop consonants in English.	Audio listening to native speakers can enhance NNES phonology.
Hamza and Fei, 2018	Linguistics	Analysis of corpus derived from pilot-ATC communication from Malaysian airspace.	Deviation from SOP and unclear instructions impact processing fluency, further lengthening interactions.
Parohinog and Meesri, 2015	Linguistics	Needs analysis of aviation students on AeE.	A combination of both synchronous and asynchronous tool would be helpful.
Tripp and Baese-Berk, 2019	Linguistics	Speech rate, rhythm, intonation of AeE compared to Standard American English.	Is being fluent English enough to do well as a pilot?

There are a few notable implications that are worthy of discussion when it comes to discussing societal implications of either AeE or accent. In the study by Lev-Ari and Keysar (2010), the researchers investigated whether trivial statements spoken by a speaker with mild or heavy accent were judged less favourably than a speaker with no accent. Their finding showed that listeners rated speech with accents as less truthful and the impact of the presence of a mild or heavy accent was not a significant influencer in processing the speech (Lev-Ari & Keysar, 2010, p. 1094). They also examined whether awareness of the difficulty might change such perception towards accented speakers (that they are less credible), but this only worked for mild speakers and not those with heavy accents (Lev-Ari & Keysar, 2010, p. 1095). The researchers point out that this is a worrisome finding as non-native speakers who are trying to find jobs in a foreign country or are eyewitnesses are bound to be judged less credible (Lev-Ari & Keysar, 2010, p. 1095). In simple words, NNES are challenged by the unfavourable judgments held by NES because of their accents.

Although the study was not related to aviation in particular, the above findings are worthy of discussion in the context of aviation. For instance, a stressful situation might arise where a pilot who is heavily accented will confuse even an experienced ATCO, leading the latter to believe that the pilot does not what he or she is talking about (ATCO can see a radar with flight information on a screen, so fact checking is not a problem). This is true in a reversed situation where an ATCO with a heavy accent will not help in maintaining situational awareness of a captain or pilot in a precarious situation that the ATCO does not fully yet know about. As aforementioned, pilots rely heavily on radiotelephony communication to avoid obstacles on the ground, get approach clearance, or to clarify the meaning of instructions. Although it is not known whether NNES and NES attribute "credibility" to one another when engaging in communication in aviation, considering that the industry is culturally diverse there could be unnoted cases where this "credibility issue" might occur.

Shum et al. (2021) used a gamified high-variability perceptual training with feedback to help NNES from Hong Kong (middle childhood, early adolescence, young adulthood) adopt and adjust to the voiced (/b, d, g/) and voiceless stops (/p, t, k/) in American English. Their experiment (listening to NES speaker's audio and guessing words) was divided into a pre-test and post-test (divided by time periods) with a randomized design that allowed comparison of the participants' performances in each subsequent test. The findings showed that the participants across all age groups considered in the experiment had improved their recognition of the voiced and voiceless stops (Shum et al., 2021, p. 337). Adults were more likely to detect linguistic cues (intervocalic divisions) that enabled them to notice contrasts in consonants than younger listeners (Shum et al., 2021, p. 337). That is, listeners vary in how they apply cognitive strategies to accurately learn. This finding elaborates the conclusion from Cristia et al. (2012): Although listeners adapt to accents over time, the way each age cohort employs cognitive and linguistic strategies to do so might differ.

A similar discussion on using cognitive strategies to understand accented speech can also be found in Hanulíková et al. (2012). By analyzing how listeners' brains react to the speaker's identity with accented speech, they concluded that native speakers could use the accent of the speaker as a sign of 'non-nativeness' and adjust quickly to ensure that the communication between them flows smoothly. This is alluding to another interesting point made in this article that native speakers who listen to accented speech may overlook grammatical errors committed by non-native speakers

(Hanulíková et al., 2012). In relation to AeE, overlooking detail in communication is fatal because the weight attributed to each instruction coming from an ATCO might be different[12].

It is important to discuss linguistic aspects of AeE. Trippe and Baese-Berk (2019) specifically looked at the comparison of speech rate, rhythmic properties, and intonation between Standard American English and AeE via quantification of these prosodic elements. Their assumption was that prosody of AeE makes it more challenging to learn (Trippe & Baese-Berk, 2019, p. 33). Their data were quite interesting, as they compared corpora that contained radio broadcasters' speech and that of ATCOs. They found that: (a) the rate of articulation is faster in AeE; (b) has a smaller pitch range than that of Standard American English; (c) AeE differs from Standard American English in terms of rhythmic properties (influenced by variability in consonants and vowels along with articulation rate) (Trippe & Baese-Berk, 2019, p. 39). In brief, AeE is not a language that can be learned easily. It is not surprising to find that the prosodic elements of AeE come across as something that is hard to grasp.

Radiotelephony communication in the air requires pilots and ATCOs to avoid using words that could have ambiguous meanings, so they need to carefully choose words to ensure miscommunication does not occur. This elliptical nature may impact how sentences could be formed, thereby influencing how they are put together and ultimately the intervals (durations) of consonants and vowels. But when accents are added on top of this, it makes pilot-ATC communication more confusing.

It is also noteworthy that every speaker has different consonant and vowel variabilities. Unifying the pronunciation of these variabilities is impossible because it (unifying or in other words, 'standardizing') means that consonants such as plosives, affricates, fricatives, approximants, nasals (and including aspirations) must be adapted to those of a NES speaker. But what is a NES speaker? Who defines what a proper pronunciation of an English word should sound like? As discussed elsewhere on this thesis, when there are internal varieties of English as well as regional dialects that

[12] Compare "Roger" and "Wilco" in AeE. They are not necessarily the same. The former indicates that a pilot received the last transmission of an instruction, but it serves the role of an acknowledgement. The latter references the intention to do something in the future as a follow up, which means "will comply". In this sense, "Wilco" has more weight than "Roger" because it has a connotation that a future action will be carried out. Due to similar nuances contained therein, these words could possibly be mixed up. In similar fashion, "Pan-Pan" and "Mayday" do not have the same gravity.

spring from these varieties, the focus should be on the conveying of exact meanings in aeronautical communication even if it is accented or characterized by varied prosodic elements.

Miscommunication results from other causes as well. Hamzah and Fook Fei (2018) investigated frequencies of misunderstanding in the audio recordings of pilot-ATC communication from three different radio frequencies in Malaysian airspace. The researchers used conversational analysis technique, which involved categorizations of communication breakdowns (procedural deviation, problematic instruction, other). They found that miscommunication could occur when both pilots and ATCOs are not comfortable with standard phraseology, possibly leading them to revert back to English (Hamzah & Fook Fei, 2018, p. 207). They also found that intonation, unclear instructions, lack of proper radio etiquette, use of one or more pronouns, and incorrect pronunciation of words could lead to confusion (Hamzah & Fook Fei, 2018, p. 199). Overall, these findings indicate that the general attitude of interlocutors when facing difficulties with standard phraseology also matter.

Whereas Kim and Elder (2015) explored the opinions of those who are employed in the industry, Parohinog and Meesri (2015) investigated the needs of Thai aviation students regarding AeE education. Based on their finding that students were having the most difficulty with grammar (represented as 'structure' on ICAO LPR), vocabulary, and interaction, they advocated the use of both online synchronous and asynchronous tools to help them learn better (Parohinog & Meesri, 2015, p. 266). This research is similar in nature to Moskalenko and Didenko (2018)'s, as each research considers the use of computer tools to teach AeE. A needs analysis is helpful in observing and recording the thoughts of participants (one of the methods used in CBE), as they are directly involved in the learning phase of AeE. Knowledge extracted from this can be used to fuel the development of a computer program that aids in experiential learning, which can be cost-effective in the long term.

In summary, the societal implications of AeE and accent were considered simultaneously. AeE and accent are inseparable. Therefore, the focus is to look at an individual's ability to convey clear and exact meaning in aeronautical communication. It is inappropriate to put labels such as 'hard to hear' or 'more (less) truthful' based on the level of accents of a speaker. Such attribution of status is misleading and does not help in improving overall flight safety.

2.2.4 Discussion

In this chapter, key articles that overlap between AeE and Accent were reviewed. Their fields, study topics, methods (or tasks), and societal implications were covered in detail. Each insight from these articles have been explained in relation to aviation context. It specifically explored how such finding could impact aeronautical communication and training of future pilots.

In terms of methodologies, both AeE and Accent research employ tasks that divide participants or a group of auditory data into two or more divisions. A stimulus such as background noises was considered. For quantitative studies, the analytical methods vary; experiments can be within or between-subjects or mixed depending on the dependent variable that is generated by the variety of tasks or tests chosen for an experiment (ex. test scores). For qualitative studies, conversational analysis (transcribing audio recordings of pilot-ATC communication), literature reviews, interviews and questionnaires were used. Although the studies examined in this section employ distinct methods, it nonetheless gives an idea to this thesis that audio recordings of accents could be useful for setting up an experimental study, which in turn can be used for making a listening test.

It is noteworthy that the study participants are exposed to a variety of tasks such as verbal recall tasks, fill-in-the-blanks, questionnaires, surveys, interviews, phrase recognition, and vocabulary tests. There is no dependency on only a single method. Although tasks administered in these tests differ, the data derived from these tests (test scores) can yield new information that either confirms findings of contemporary research in AeE training or not.

It is noteworthy that being fluent in English does not mean that one will have a good command of English and perform always better than a NNES. If this statement were true, then it would be equivalent to saying, 'a NES who has a full grasp on English is also a good writer' which is not always true. Also, internal varieties of English (British, American, Antipodean, Caribbean, Indian, South African) have unique prosodic elements with salient features (such as vowels) that are not recognized by listeners from other language backgrounds. Not only that, the way speech is generated when using AeE (i.e., ellipsis) can put undue pressure on the speaker, as pilots or ATCOs with limited lexical mapping (due to lack of field experience) can increase the duration of communication unintentionally.

2.2.5 Limitations and Gaps of the Studies Above

So far, the literature review conducted above has identified which parameters and methods have been used for analysis. An in-depth discussion is also provided regarding the results of the literature in relation to Aviation English. However, it is worthwhile to discuss the limitations of the studies above because this could give insights on why this thesis is a valid and novel effort.

First, there are only few accent-related articles that produced empirical results in the aviation domain (Tiewtrakul & Fletcher, 2010; Clark, 2017; Wu et al., 2019; Dissanayaka et al., 2023). The author of the thesis does not find this surprising for two reasons: a) When measuring or quantifying the impacts of accents, more scrutiny is needed to define what accent is, which accent(s) is collected for analysis, and how it is utilized for the purposes of research; b) Studying accents in aviation is very challenging. When studying accents, not only is knowledge in linguistics needed, but also in acoustics, phonetics, and cognitive science. The main point is that there is a higher barrier to entry, since studying AeE requires understanding what pilots do, how they communicate with ATCOs, and what types of training they do at flight schools.

Second, the accent-related studies surveyed here are mostly concentrated in the field of psychology. A typical method in this field seems to be testing vocabulary knowledge or taking notes on general know-hows from a participant via a series of experimental tasks and documenting the results. Various tools such as Stroop test, eye tracking, working memory test, High Variability Perceptual Training Program are used. Unfortunately, a main concern is that the research was done in a non-aviation context. Therefore, it is not clear whether all results from psychological research can be said to be applicable to aviation.

In summary, a researcher who is well-versed in aviation, human factors, and linguistics is needed to investigate the complex problem set forth by foreign accents in AeE. Here, it is necessary to highlight another novel aspect of this thesis. Aviation is a field that is at the forefront of innovation. It is a field that embraces interdisciplinary research ranging from engineering to arts and humanities. However, there is evidence that not much analytical work has been done on the impacts of accents in AeE. It does not make sense that a field that dearly values innovation does not have adequate amount of research in this regard. A researcher who is well-versed in aviation topics that knows how to effectively train ab initio pilots' AeE skills with a new technology (such as A.I-backed text-to-speech system) is more likely to bring in fresh ideas to the field.

It is not difficult to find articles that describe what is currently happening in AeE research, what has been done experimentally, and what could possibly be a new avenue or an inspiration for future research. There are also articles that provide analyses based on the collection of textual or auditory data, from which researchers project their ideas on how language testing in aviation should be amended. However, these articles do not specifically address how foreign accents might interfere with a pilot's understanding of either a short or long interlocution, especially in pilot-ATC communication. Furthermore, there are no specific articles that explain how a person's linguistic and aviation background can influence their listening skills in AeE. To be more specific, research that solely focuses on accents alone while employing an advanced technology such as TTS to assess listening abilities of participants is not heard of at the time of writing this thesis. This is the novelty of this thesis, which if published, could show how technology can be used to train future pilots especially for language training.

2.2.6 Conclusion

In conclusion, AeE and accent are intertwined. The impact of accented speech cannot go unnoticed in pilot or ATCO training, as it poses serious threats that can lead to fatal accidents which in turn will ultimately tarnish the reputation of airlines. The literature covered in this section provides several implications: (a) Being a NES is not immune to the challenges of using AeE; (b) Tolerance towards accented speech depends on the cognitive strategies employed by listeners; (c) ICAO LPRs need improvement; (d) A computerized learning tool can be used to teach AeE effectively; e) AeE is structurally different from plain English.

A main takeaway from the methodology standpoint was that it suggested ideas for setting up an experimental procedure for research. The following is the list of ideas (or tools) that were used in the thesis:

a. **Questionnaire.** This is used to extract demographic information from participants such as their language background, programs they study in, years of flight experiences, and languages they speak most frequently.
b. **Accent listening tests.** A round or two of listening tests that incorporate scripts embedded foreign accents. This can test a participant's ability to extract accurate meanings from accented speech. The dependent variable is the score, and independent variables are

demographic information, rounds, and script types based on its characteristics. The scores from these two tests would then show whether they have improved from Round 1 to 2.

c. **Post survey.** This survey is used to collect what each participant thought about the naturalness of accent used in the listening tests.

Chapter 3. Material Creation

3.1 Introduction

As identified in the previous chapter, there were three main tools which needed to be developed to carry out the research studies. First was the questionnaire. When looking to assess a participant's listening skills that involve foreign accents, demographic profiles of each participant may provide useful information on their overall experience. These could be used for statistical analyses or to gain more understanding about each participant's background pertaining to, specifically, their familiarity with foreign languages. Therefore, a Preliminary Survey (or demographic survey) was created which can be viewed in the **Appendix D**.

Second was the accent listening tests. ELPAC, as discussed in the literature review, emphasizes listening and oral proficiency. Although the oral proficiency is as equally important as listening, considering that pilots have to multi-task while on duty, listening skill must be given more weight. It is the listening skill that enables a pilot to understand information from ATCs, which in turn helps him or her communicate information, and use that information to ask for requests, clarification, or predictions. The listening skill, therefore, is directly related to the future actions or plans of the pilot. It must be also noted that one of the novelties of this thesis lies in the inclusion of foreign accents in the listening tests. The accents were recorded not just by human voices (HV), but also by computer voices (CV). Therefore, the work in thesis closely reflects real world pilot-ATC communication – that is, speech embedded in foreign accents.

The delivery of the listening tests was also configured. First, regarding the distribution of the listening tests, Qualtrics was chosen which is an online platform for making digital surveys. Second, regarding the presentation of listening tests, pre- and post-test design was chosen. Based on the observations from the contemporary literature (Tamjid & Hassanzadeh, 2012; Moskalenko & Didenko, 2018; Shum et al., 2021; Larandang et al., 2023), a pre- and post-test was a popular choice. However, a more important reason for choosing the pre- and post-test design was that the author wanted to observe whether participants' scores could be improved in either of the Voice Groups (CV, HV). A question of "Are they improving only in CV, HV, or both?" holds significance because, if a noticeable change in the score is found, it could be evidence in favour of recommending computer voices as an alternative training tool for language. As aforementioned in Chapter 2, language training is expensive and using human voices requires voice actors or actresses who need to be paid for their service. From a training organization's perspective, a cost-effective training solution that minimizes costs but maximizes training efficiency is a top priority. In essence, the thesis also sheds light on this aspect of research.

Third, a post survey was created which asked a question: "Overall, how did you find the speech in the pilot-ATC conversations (AS) and neutral/non-aviation scripts (NS) in terms of accent and the way of talking?" A Likert scale from 1 to 5 was used, with 5 being 'very natural' and 1 'very unnatural'. The results of the scores indicated by the participants on the post survey would reveal interesting insights regarding the use of foreign accents.

The material creation was a significant work, which took many months. However, this step was necessary because there were no widely known testing resources to draw inspiration from for foreign accent testing especially in pilot training. Thus, an original set of scripts and questions had to be developed completely. To do this, the author gathered relevant documents, websites, research, and listened to live audio recordings of pilot-ATC communication to generate ideas and finally create a unique set of scripts and questions. These scripts were divided into two Script Types: a) Aviation Scripts (scripts that mimic actual pilot-ATC communication; AS); b) Neutral Scripts (scripts that have no background information or context; NS). AS can be seen in the **Appendix E**. NS is in the **Appendix F**.

The rest of the chapter explains the detailed procedures of making each material used in the thesis. These would be helpful to future researchers who might be interested in conducting similar research that involves computer voices, or in general, assessing listening test skills of participants when accents are embedded in the speech.

3.1 Voice Over Volunteers (for HV Group)

Voice over volunteers who had moderate level of accent as judged by the author in Spanish, Arabic, and French were recruited. The author recruited a total of 16 voice-over volunteers from around the world. To be specific, four people in each accent (English, French, Arabic, Spanish[13]) were recruited with a condition that they be two middle-aged and two young people. Thus, for each accent, there was Old Male (OM), Old Female (OF), Young Male (YM), and Young Female (YF) (see ***Figure 12*** below). Volunteers who needed a parental consent to volunteer were excluded

[13] These are four of the main languages used in ICAO. Two other languages are Russian and Chinese.

from recruitment. The volunteers were instructed to record both NS and AS in one setting. Recordings were edited in Audacity software (version 2.4.2).

ENG	ARB	ESP	FRE
• OM	• OM	• OM	• OM
• OF	• OF	• OF	• OF
• YM	• YM	• YM	• YM
• YF	• YF	• YF	• YF

Figure 12. Gender was all equally distributed for each accent. These were used to make a gender pair (ex. OM x OF). The reason for having two age cohorts (middle-aged, young) was to make the scripts more realistic. Also, there are female pilots or ATCOs who work in the industry.

3.2 Text-To-Speech (for CV Group)

For computer voices, the author used Murf which was a company specializing in providing artificial intelligence-backed (A.I) voice overs. It was the only company that provided the age setting with the text-to-speech (TTS) system (middle-aged and young), which fit the purpose of the study. Before choosing the age, the author had to choose some speakers to make the scripts.

For Arabic and Spanish, there were not many speakers to choose from (only five in each group). For French, there were ten speakers to choose from. The author listened to all speakers and selected some with the most natural pronunciations (5 for Arabic and Spanish, and 6 for French). After choosing speakers (and their ages) in each of the following languages (Spanish, Arabic, French, English), the author generated some sample speech from each speaker (except English speakers) and asked five people from the university to rate the speakers' strength of accent from 1 to 10 (1= very weak, 10 = very strong) via online (Teams meeting). When they were asked, they were instructed not to look at their screens but only listen (the author wanted to make sure that they could only listen without looking at the screen where the sentences were visible).

The author was looking for a moderate level of accent. Each of the raters listened to a sample speech from each speaker across the three language groups and rated it from 1 to 10. The average scores for each speaker showed that their accent strength could be safely concluded moderate (min=4.4, max=9.4), which meant that all were equally likely candidates. However, after listening to more recordings for some speakers, the author found out that it was best to leave some of them out as they had sounded a little unnatural and felt "robotic". In this case, another speaker among the rated ones was chosen instead. In summary, speakers were very carefully selected after repeated listening.

Finally, both AS and NS were put into the online platform (Murf), which were then read by the selected speakers' voices. The recordings had the same speech rate, pitch, and the same volume upon downloading. The resulting recordings also followed the distribution of genders shown in ***Figure 12***. Recordings were edited in Audacity software as well.

3.3 Aviation Scripts

The aviation scripts (AS) consisted of 64 scripts (32 in Human Voice Group and 32 in Computer Voice Group). The contents of the scripts in each Voice Group were made to be different and unique. This was done to prevent the mistake of mixing script recordings when embedding them on Qualtrics. To make the scripts and multiple-choice questions, various resources were used (See **References used to write the scripts** in **Appendix E**). This step ensured that the questions were not too far from reality. Because the scripts and questions were different for each Voice Group (HV, CV), the author sought comments on improving the quality of and normalizing the difficulty level across the scripts and questions in both Voice Group from an experienced pilot (see ***Acknowledgement*** in **Appendix E**). This additional step ensured that the scripts and questions in both Voice Group were of the same quality based on similar test constructs.

Each script had three lines that were ready to be read by a gender pair (ex. OF and YM). One of the two genders played a role of a pilot and another the ATC. For example, **Table 14** shows an actual script used in the study.

Table 14

Script 9 from Aviation Script (Computer Voice Group)

ID	Gen.	Script 9
ATC	OF	Roncesvalles two-four-two-three, Iqaluit Tower. Wind three-one-zero at seventeen, visibility less than three-quarters.
Pilot	YM	Roger. Request RVR for one-six, Roncesvalles two-four-two-three. Over.
ATC	OF	Roncesvalles, Iqaluit Tower, RVR four thousand. Squall in the vicinity.

Note. This table shows a gender pair of Old Female (OF) and Young Male (YM). The ID column shows which gender is associated with which role (pilot or ATC). A gender pair was equally distributed to each line of scripts for both Human Voice and Computer Voice scripts.

Below is an actual copy of the question used for Script 9 above (**Figure 13**).

1. What is the aircraft callsign?
a. Roncesvalles 2234
b. Roncesvalles 2432
c. Roncesvalles 2342
d. Roncesvalles 2243
e. Roncesvalles 2423

2. What was the weather information provided by the ATC regarding the said runway?
a. R16/040, VCTS
b. R16/0400, VCSQ
c. R16/4400, VCSH
d. R16/0400, VCTS
e. R16/4000, VCSQ

3. Suppose the pilot confused the definition of runway visual range with that of visibility. What would be the rule of thumb for remembering the two concepts?
a. RVR is measured in feet, Visibility is in statute mile
b. RVR is measured in statute mile, Visibility in feet
c. Both RVR and Visibility are measured in feet
d. Both RVR and Visibility are measured in statute mile
e. RVR is measured in kilometres, Visibility is in metres

Figure 13. A sample multiple-choice question set for Script 9 in AS. Each question was worth 1 point.

Each AS was followed by three multiple-choice questions that were inspired by the three levels of situation awareness (Endsley, 2017). In the **Figure 13** above, question 1 tested the knowledge gained from the script (which in this case was the callsign of the aircraft). For instance, gathering information correctly is an important skill for pilots because it will help in preventing miscommunication with ATCO if the latter requests the pilot to provide a specific piece of information to confirm something.

Question 2 tested comprehension. In real life, this would be a skill that can affect whether a pilot is competent enough to carry out a long-haul flight. Having a clear mental model established on the complete understanding of the current situation is necessary for flight safety. But a situation never remains the same due to external influences, and neither does information therein. Information received via communication from an ATCO also changes its shape throughout the duration of the flight, which could be addons to the changed state of previous information. This entails updating information constantly for the duration of the flight. For instance, jargons like VCTS (thunderstorm in the vicinity) or R16/0400 (runway 16, runway visual range 400) would require the understanding of their definitions, given any circumstance in the script. They may change based on the circumstances presented in the scripts.

Question 3 was a prediction type question that builds on the previous mental model and information. It can be considered an amalgamation of all question types above into a single question that require both prior knowledge (including both general and technical knowledge in aviation), circumstantial evidence, and the flow of conversation in the script. Therefore, a participant must be knowledgeable about what is happening in the script, but also be able to guess what may be the best approach the pilot in question or ATC can do in the unfolded event or in an imaginary situation. A synthesis of knowledge of past and present is needed to solve this question.

After all questions were made, they were embedded along with AS recordings on Qualtrics. All AS questions can be viewed in **Appendix G**.

3.4 Neutral Scripts

The neutral scripts (NS) were similar in the format (three lines per script) but had separate, individual sentences that were not related to each other (**Table 15**). These were Harvard Sentences, which were phonetically balanced sentences that did not depend on a larger context. For example, one sentence might be "The hogs were fed chopped corn and garbage". Harvard Sentences, being phonetically balanced, were used for creating voice overs, which in turn was used for mobile devices and internet-related systems (IEEE, 1969). Harvard Sentences were also used in studies related to linguistics (Nigam et al., 1992; Oh, 2011; Banks et al., 2015). They were chosen because their predictability was low (Banks et al., 2015, p. 2017).

The NS had fill-in-the-blanks questions. Fill-in-the-blanks is a common type of question for participants, as can be seen in the accent-related research (McLaughlin et al., 2018). In Banks et al. (2015), the sentences were read aloud (and in this article, some vowels in the sentences were replaced from standard British accent to a regional accent). Taking inspiration from the previous research, the author decided to create three fill-in-the-blanks questions that could be played by the participants when they were doing the tests online. They would then type what they heard (words) in the missing blanks (indicated by [?]; **Figure 14**).

After all questions were made, they were embedded along with NS recordings on Qualtrics. All NS and its questions can be viewed in **Appendix F**.

Table 15

Script 1 from Neutral Script (Human Voice Group)

Gen.	Script 1
OF	The birch canoe slid on the smooth planks.
YM	The boy was there when the sun rose.
OF	The small pup gnawed a hole in the sock.

Note. The first sentence 'The birch canoe…' has no relation with the second sentence 'The boy was there…'. Each line of script is independent of the other in this script.

1. The [?] [?] slid on the [?] [?].
2. The [?] was [?] the [?] [?].
3. The small [?] [?] a [?] in the [?].

Figure 14. A sample question set for Script 1 in NS. The maximum grade that can be achieved is 1 for each sentence (a total of 3 points altogether).

3.5 Online Tests

After making all the materials above, both NS and AS and their questions were embedded on Qualtrics. In the end, four separate listening tests were successfully made (Computer Voice Group Round 1 and 2, Human Voice Group Round 1 and 2). The flow control was checked for each test before publishing them online.

Chapter 4. Experimental Method

4.1 Participants

Participants were 50 undergraduates or graduate students at the University of Waterloo who were enrolled in various programs. Most of the participants were students enrolled in AVIA 100 course at the university, which taught fundamentals of aviation. A participant recruitment form (see **Appendix H**), information letter, consent form, and preliminary survey (demographic survey) were all uploaded in this course to recruit students. However, due to the low response rate, students outside the course were also recruited via different avenues. These students were those who contacted the author via email by seeing the 'Call for Study Participants' webpage which was maintained by the university, where the same recruitment information was uploaded. Some students also contacted the author via the recruitment poster uploaded on the website of the Waterloo Institute for Sustainable Aeronautics. Students were also recruited via the university's student chapter of the Human Factors and Ergonomics Society, where the links to the two webpages were provided. Lastly, LinkedIn was used for recruitment.

In terms of recruiting, a thorough and strictly selective procedure was applied: a) They were familiar with pilot-ATC communication (or those who watched videos on live radiotelephony communication); 2) They were students at the University of Waterloo. All participants' (i.e., students) emails had to be a University of Waterloo email address to maintain homogeneity, validity, and security of the data, consistent with Tamjid and Hassanzadeh (2012) and Sumalinog (2018). Otherwise, they were sent a 'termination' email saying that another volunteer has filled the position[14].

For those students who were recruited outside the course, they were sent the participant recruitment form again to make sure that they fit the eligibility requirements. They were also given opportunities to ask the question about their eligibility as well. If they did not provide any response, the author acknowledged this as them not fitting the requirements (as a result, 21 students never replied or followed up). Some voluntarily identified themselves not to be eligible after giving them all the necessary information, in which case they were simply let go. If they either showed a strong

[14] The author could not confirm whether they were spam emails or not. The best course of action was to not invite them to the study.

interest to continue by expressing that they fit the criteria or followed up with the author, then the author acknowledged that they fit one of the requirements and continued the study procedure.

It is not difficult to find similar studies in EFL (English as Foreign Language) that have recruited less than 50 students (Tamjid & Hassanzadeh, 2012; Sumalinog, 2018; Medina et al., 2020; Larandang et al., 2023). Tamjid and Hassanzadeh (2012)'s study divided 40 students equally between two groups and carried out pre- and post-test. Inspired by this study, the author divided the participants as follows:

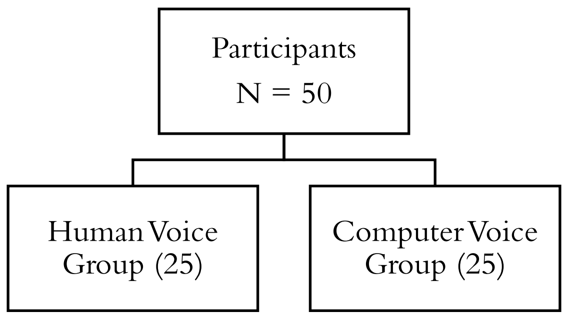

Figure 15. Division of participants into two Voice Groups.

In **Figure 15**, an equal number of participants were split into two groups: 25 in Human Voice Group, 25 in Computer Voice Group. Participants in HV Group only listened to the scripts (NS and AS) read in real human voice, whereas those in CV Group encountered the scripts read in computer voice. The scripts can be found in the **Appendix F** (NS) and **Appendix E** (AS).

In the Human Voice Group, the mean age of the participants was 20.76 (SD=3.06). There were 17 males and 8 females. There were 13 NES and 12 NNES. The average flight hours were 33.86 (SD = 63.69). Missing values in the preliminary survey were filled in by re-contacting participants via email. One of the participants did not reply to the follow-up emails from the author regarding a missing blank (textbox) for the Years of Speaking English (Question 9 in the Preliminary Survey; **Appendix D**). In that case, the missing blank was imputed with Multivariate Imputation by Chained Equation (Buuren & Groothuis-Oudshoorn, 2011). The average Years of Speaking English was 16.58 (SD=6.11) with the imputed mean.

In Computer Voice Group, the mean age of the participants was 20.72 (SD=3.57). There were 13 males and 12 females. There were 13 NES and 12 NNES. The average flight hours were 27.80 (SD=58.91). The average Years of Speaking English was 17.92 (SD=2.91). There were no missing values in this group.

4.2 Experimental Procedure

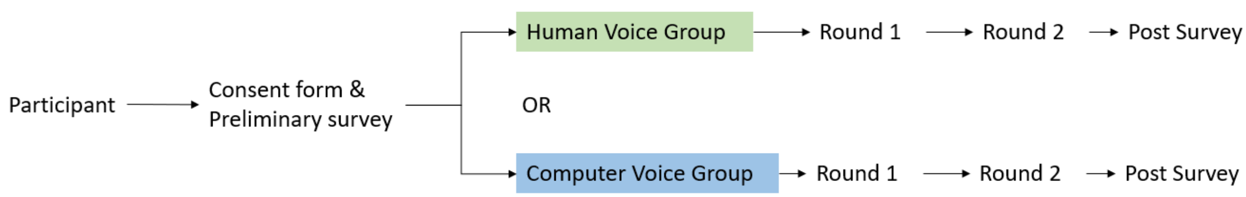

Figure 16. The overall experimental procedure.

The overall study is a between-subject design (**Figure *16***). Participants, before completing the two rounds of the listening test, were asked to fill out a consent form and preliminary survey. The participants were then randomly assigned to either CV or HV Group on a first-come-first-served basis (or whoever completes those two forms first), which indicated to them via email. This step was necessary for two reasons: a) To avoid a mistake in mixing up which participants belonged to which Voice Group when carrying out data analysis; b) To meet one of the requirements in the ethics application wherein the participants' personal ID such as names and email addresses were not allowed to be recorded directly. No single participant belonged to both Voice Groups. Once a participant belonged to one of the Voice Groups, they were instructed to do Round 1 and 2 listening tests. In this sense, the "inner" part of the experiment is within-subject design because every participant in each Voice Group completed Round 1 and 2. There was no time limit for the tests, but the strict restriction was applied to the playing of scripts which was limited to only once. Following ELPAC's example, participants were warned not to press the 'back' button to go back to the previous questions.

Each Round had 8 scripts to be played in total (i.e., 4 AS + 4 NS). Each AS had three multiple choice questions. Each NS had three fill-in-the-blanks questions. Thus, each Round had a total of 24 questions ready to be answered by participants. At the end of Round 2, a participant would have had completed a total of 48 questions. The 32 scripts (which are a mix of both AS and NS) in each Voice Group were juxtaposed and coded on Qualtrics in a way that each script appeared on a test with equal chances (see **Appendix C** for a detailed information). The order of accents appearing in Rounds 1 and 2 were different as well. All four accents were randomly arranged with no accent found twice consecutively (ex. No English NS after English AS). Participants did not know which Script Type (NS, AS) or accent (English, Arabic, Spanish, French) was going to appear first when they were doing the tests.

When a participant finished Round 1, they were shown their scores and the questions they answered correctly or incorrectly. Re-playable scripts were also coded directly on the score summary page on Qualtrics, so that they could listen to the scripts repeatedly. This was the "feedback phase" so that they could prepare themselves for Round 2. The scripts used in Round 1 were not reused in Round 2. Once a participant finished all the listening tests (Round 1 and 2), they were asked to fill out a post survey indicating whether they found the accents used in the tests natural or not. For students enrolled in AVIA 100 course, a "bonus mark form" that allowed them to earn bonus marks in the course was returned with their time of completion. All participants were given remuneration of $20 via online payment.

The scores were manually graded by hand and saved in a spreadsheet, as the fill-in-the-blanks were not calculated correctly by the online platform. This allowed the author to double-check the score calculations.

Once the scores were obtained, a complex dataset was obtained. Therefore, multiple studies were carried out. Chapters 5, 6 and 7 show the results of the statistical analyses based on different set of research questions.

Chapter 5. Study 3: Changes in the Scores based on Rounds and Script Types

5.1 Introduction

In this chapter, a comparison of the total scores from each of the Script Types (NS, AS) in Round 1 and 2 within each Voice Group (CV, HV) was carried out. The purpose was to observe whether the scores, in either Voice Group, categorized further by the Script Types, improved or not. The total scores from each round were also analyzed as well to see the overall trend.

Six sub-research questions were generated (denoted by **RQ#**):

- **RQ1**. Is there evidence of change in the scores of the Aviation Scripts between Round 1 to 2 in both Voice Groups?
- **RQ2**. Is there evidence of change in the scores of the Neutral Scripts between Round 1 to 2 in both Voice Groups?
- **RQ3**. Is there evidence of change in the Total Scores from Round 1 to 2 in both Voice Groups?
- **RQ4**. Does Factor 1 (Script Types) have an influence on the test scores?
- **RQ5**. Is there an interaction between Factor 1 (Script Types) and 2 (Rounds)?

5.2 Procedure

In the Human Voice Group, each participant's total score from the Neutral Scripts (NS) in Rounds 1 and 2 was compared (Total R1 NS vs. Total R2 NS). The same was done to the total scores from the Aviation Script (AS; Total R1 AS vs. Total R2 AS). Finally, the total scores of each participant in Rounds 1 and 2 were also tabulated and compared (Total R1 vs. Total R2). The same procedure was also applied to the scores in the Computer Voice Group. The comparison was carried out with paired-samples t-tests. The results of the analysis would answer the RQ 1,2 and 3.

For RQ 4 and 5, two-way repeated measures ANOVA was carried out for both Voice Groups (CV, HV). In the following section, results for the Computer Voice Group (CVG) are discussed first, followed by those for the Human Voice Group (HVG).

5.3 Results and Discussion: Computer Voice Group (CVG)

Table 16

Descriptive Statistics of the Total Scores by Rounds and Script Types (CVG)

Descriptive Statistics

	Valid	Mean	Std. Deviation	Minimum	Maximum
Total R1 NS	25	5.720	1.009	4.250	8.000
Total R1 AS	25	5.920	2.235	3.000	11.000
Total R2 NS	25	7.350	1.512	4.000	10.000
Total R2 AS	25	6.560	2.434	1.000	10.000
Total R1	25	11.640	2.583	7.750	17.000
Total R2	25	13.910	3.380	6.250	19.250

.

First, to see whether two factors (Script Types, Rounds) have had any influence on the test scores, two-way repeated measures ANOVA (2 x 2) was conducted. The result is shown below in **Table 17**.

Table 17

Repeated Measures ANOVA (CVG)

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Rounds	32.206	1	32.206	20.258	< .001	0.458
Residuals	38.154	24	1.590			
Script Types	2.176	1	2.176	0.704	0.410	0.029
Residuals	74.121	24	3.088			
Rounds ✻ Script Types	6.126	1	6.126	2.941	0.099	0.109
Residuals	49.984	24	2.083			

Note. Type III Sum of Squares

Based on the table above, there was a significant effect of Rounds on the scores ($F[1,24] = 20.26$, $p < 0.001$, $\eta^2_p = 0.458$). There was no significant effect of the interaction between Rounds and Script Types on the scores ($F[1,24] = 2.941$, $p = 0.099$) or Script Types ($F[1,24] = 0.70$, $p = 0.41$). Thus, **RQ4** and **RQ5** were not supported. The results suggest that 45.8% of variance in the scores were associated with the Rounds. This finding supports the general upward trend in **Figure 17**.

Table 18

Post Hoc Comparisons (CVG)

Post Hoc Comparisons - Rounds ✻ Script Types

		Mean Difference	SE	t	Cohen's d	p_{tukey}
Round 1, Aviation	Round 2, Aviation	-0.640	0.383	-1.670	-0.339	0.351
	Round 1, Neutral	0.200	0.455	0.440	0.106	0.971
	Round 2, Neutral	-1.430	0.433	-3.306	-0.758	0.010 **
Round 2, Aviation	Round 1, Neutral	0.840	0.433	1.942	0.446	0.226
	Round 2, Neutral	-0.790	0.455	-1.737	-0.419	0.317
Round 1, Neutral	Round 2, Neutral	-1.630	0.383	-4.253	-0.864	< .001 ***

* p < .05, ** p < .01, *** p < .001

Note. P-value adjusted for comparing a family of 6

Post hoc paired t-tests were then performed with Tukey's method to adjust for family-wise errors (**Table *18***). A valid comparison was the pairs that had the same Script Types (NS, AS). The row at the bottom (Round 1 Neutral vs. Round 2 Neutral) is significant ($p < 0.001$).

Table 19

Paired-Samples T-test Results for the Total Scores (CVG)

Paired Samples T-Test

							95% CI for Mean Difference			
Measure 1	Measure 2	t	df	p	Mean Difference	SE Difference	Lower	Upper	Cohen's d	SE Cohen's d
Total R1 NS	- Total R2 NS	-5.732	24	< .001	-1.630	0.284	-2.217	-1.043	-1.146	0.277
Total R1 AS	- Total R2 AS	-1.387	24	0.178	-0.640	0.461	-1.592	0.312	-0.277	0.201
Total R1	- Total R2	-4.501	24	< .001	-2.270	0.504	-3.311	-1.229	-0.900	0.192

Paired Samples T-Test

							95% CI for Mean Difference			
Measure 1	Measure 2	t	df	p	Mean Difference	SE Difference	Lower	Upper	Cohen's d	SE Cohen's d

Note. Student's t-test. R1,2 = Round 1, 2; NS = Neutral Scripts; AS = Aviation Scripts.

A paired samples t-test revealed a significant difference in the means of the scores between Round 1 NS and Round 2 NS (t(24) = -5.73, p < 0.001, d = -1.15) (**Table *19***). It was also found that the mean difference in the total scores was significant between Round 1 and Round 2 (t(24) = - 4.50, p < 0.001, d = -0.90). In summary, **RQ1** was not supported, but **RQ2** and **RQ3** were. This means that participants performed better in Round 2 than 1 in NS, and in terms of the total score they performed better overall in Round 2 than in 1, significantly improving their scores with computer-voiced accents.

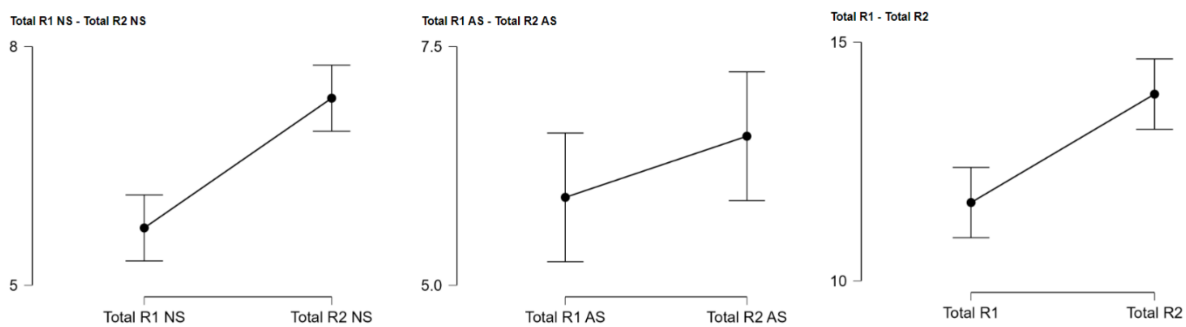

Figure 17. Descriptive plots for Computer Voice Group (CVG). Total R1 NS vs. Total R2 NS (left), Total R1 AS vs. Total R2 AS (centre), Total Score R1 vs. Total Score R2 (right). Y axis is the average score. X axis shows the total score(s) grouped by the Script Types and Rounds. For the graph on the right, the x axis is the comparison between the total scores in R1 and R2. R1, 2 = Round 1, 2. NS = Neutral Scripts. AS = Aviation Scripts.

As can be seen from the **Figure 17** above, the Total R1 NS vs. Total R2 NS (left) shows the two mean scores far apart from each other. But for Total R1 AS vs. Total R2 AS (centre) showed them somewhat closer, which indicated that there was only a small improvement in the scores from Round 1 to 2. The third graph on the right (Total Score R1 vs. Total Score R2) showed a similar trend as that of the graph on the far left.

5.4 Results and Discussion: Human Voice Group (HVG)

To answer the same questions posed for the computer voice, ANOVA and paired-samples t-test were carried out for the Human Voice Group. The descriptive statistics of the scores are shown in the **Table *20*** below. **Table *21*** shows the results of the ANOVA.

Table 20

Descriptive Statistics of the Total Scores by Rounds and Script Types (HVG)

Descriptive Statistics

	Valid	Mean	Std. Deviation	Minimum	Maximum
Total R1 NS	25	9.690	1.296	5.500	11.500
Total R1 AS	25	5.160	2.154	2.000	9.000
Total R2 NS	25	9.420	1.183	6.250	11.250
Total R2 AS	25	6.440	1.938	2.000	10.000
Total R1	25	14.850	2.595	9.000	19.500
Total R2	25	15.860	2.458	11.250	21.000

Same as the analysis laid out in the Computer Voice Group, to answer the research question of whether the two factors (Script Types, Rounds) have had any influence on the test scores, two-way repeated measures ANOVA (2 x 2) was conducted. The result table is shown below.

Table 21

Repeated Measures ANOVA (HVG)

Within Subjects Effects

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Rounds	6.376	1	6.376	3.042	0.094	0.113
Residuals	50.296	24	2.096			
Script Types	352.501	1	352.501	119.709	< .001	0.833
Residuals	70.671	24	2.945			
Rounds ∗ Script Types	15.016	1	15.016	7.010	0.014	0.226
Residuals	51.406	24	2.142			

Note. Type III Sum of Squares

Based on the **Table *21*** above, there was a significant effect of Script Types on the scores ($F[1,24] = 119.7$, $p < 0.001$, $\eta^2_p = 0.83$). Thus, **RQ4** was supported. There was also a significant effect of the interaction between Rounds and Script Types ($F[1,24] = 7.01$, $p < 0.05$, $\eta^2_p = 0.23$). **RQ5** was also supported. However, Rounds was not found to be significant ($F[1,24] = 3.04$, $p = 0.09$). The results suggest that 83% of variance in the scores were associated with the Script Types. This finding supports the disparate trend between NS and AS shown in **Figure *18***.

To examine which pairs were significant based on the interaction effect, a post hoc analysis was performed. The **Table 22** below shows the results.

Table 22

Post Hoc Comparisons (HVG)

Post Hoc Comparisons - Rounds ✻ Script Types

		Mean Difference	SE	t	Cohen's d	p_{tukey}
Round 1, Aviation	Round 2, Aviation	-1.280	0.412	-3.109	-0.756	0.016 *
	Round 1, Neutral	-4.530	0.451	-10.043	-2.675	< .001 ***
	Round 2, Neutral	-4.260	0.449	-9.487	-2.515	< .001 ***
Round 2, Aviation	Round 1, Neutral	-3.250	0.449	-7.238	-1.919	< .001 ***
	Round 2, Neutral	-2.980	0.451	-6.607	-1.759	< .001 ***
Round 1, Neutral	Round 2, Neutral	0.270	0.412	0.656	0.159	0.913

* p < .05, ** p < .01, *** p < .001

Note. P-value adjusted for comparing a family of 6

As can be seen, the most meaningful pairs (scores denominated in the same Script Types) were Round 1 NS vs. Round 2 NS and Round 1 AS vs. Round 2 AS. The findings on this table confirmed that there was no noticeable change in the scores from Round 1 to 2 for NS (bottom row; $p = 0.91$). However, a change was observed between Round 1 and 2 AS (top row; $p < 0.05$).

Table 23

Paired-Samples T-test Results for the Scores (HVG)

Paired Samples T-Test

							95% CI for Mean Difference			
Measure 1	Measure 2	t	df	p	Mean Difference	SE Difference	Lower	Upper	Cohen's d	SE Cohen's d
Total R1 NS	- Total R2 NS	1.288	24	0.210	0.270	0.210	-0.163	0.703	0.258	0.171
Total R1 AS	- Total R2 AS	-2.356	24	0.027	-1.280	0.543	-2.401	-0.159	-0.471	0.279

Paired Samples T-Test

Measure 1	Measure 2	t	df	p	Mean Difference	SE Difference	95% CI for Mean Difference		Cohen's d	SE Cohen's d
							Lower	Upper		

Note. Student's t-test. R1,2 = Round 1, 2; NS = Neutral Scripts; AS = Aviation Scripts.

In **Table 23**, A paired samples t-test revealed a significant difference in scores between Round 1 AS and Round 2 AS ($t(24) = -2.36$, $p < 0.05$, $d = -0.5$). It did not reveal a significant difference in scores between Round 1 NS and Round 2 NS ($t(24) = 1.29$, $p = 0.21$). In summary, **RQ1** was supported, but **RQ2** was not. Regarding **RQ1**, this means that participants performed better in Round 2 than 1 in AS, which is a stark contrast compared to the result obtained in the Computer Voice Group.

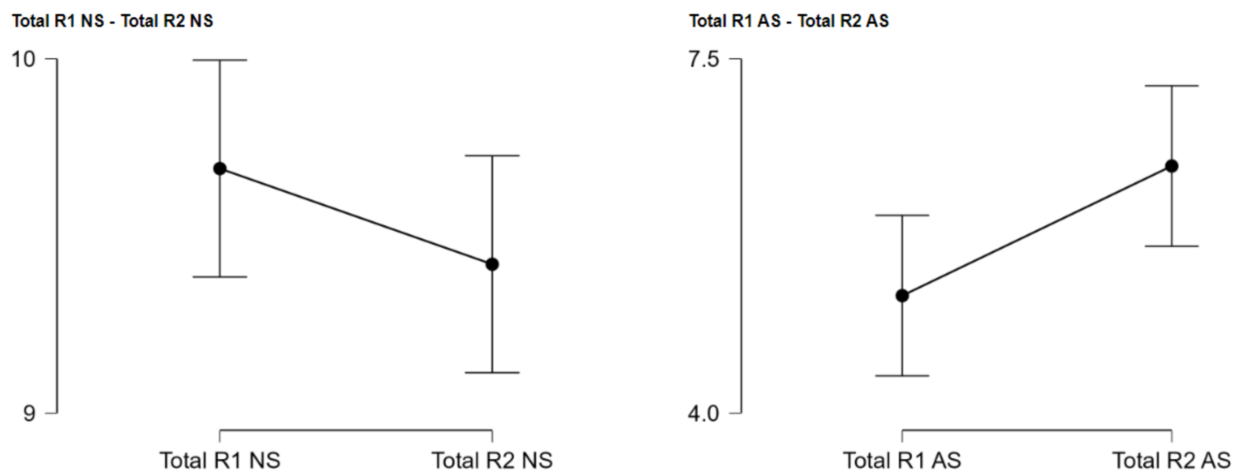

Figure 18. Descriptive plots for Human Voice Group. Total R1 NS vs. Total R2 NS (left); Total R1 AS vs. Total R2 AS (right).

As can be seen in the **Figure *18*** above, the graph on the left showed the total NS scores decreasing (slightly) from Round 1 to 2. The graph on the right showed the total AS scores increasing from Round 1 to 2.

A speculative reasoning behind the very small decrease in the total scores for neutral script (graph on the left) might be the subtle effects of accents. Human voices, compared to the computer voices which have constant accent with no fluctuations in suprasegmental features, have the opposite (i.e., there are fluctuations or variations in how they speak due to intonations). Even if the scripts were short in NS, if a participant hears a word that was spoken with certain deflections in the expected pronunciation, then there is a potential to be confused. This could explain the nearly flat slope of the graph on the left. On the other hand, there is background information to help in interpreting scripts in AS, so there might be less confusion.

The differences between Total Scores R1 and R2 violated normality assumption (Shapiro-Wilks Test, W = 0.91, $p < 0.05$). Wilcoxon signed-rank test was performed instead, the results of which is shown in the **Table *24*** below.

Table 24

Wilcoxon signed-rank Test Results for the Total Scores (Round 1,2; HVG)

Paired Samples T-Test

Measure 1	Measure 2	W	z	df	p	Rank-Biserial Correlation	SE Rank-Biserial Correlation
Total R1	- Total R2	113.500	-1.318		0.191	-0.302	0.225

Note. Wilcoxon signed-rank test.

A Wilcoxon signed-rank test revealed that there was no significant difference in the total scores between Round 1 and 2 (W = 113.5, p = 0.19). Therefore, **RQ3** was not supported. There was no evidence of a significant change in the total scores from Round 1 and 2 in the Human Voice Group. However, it is still noteworthy that the **Figure *19*** below showed an upward slope, which meant that the scores, on average, were higher in Round 2 than in Round 1.

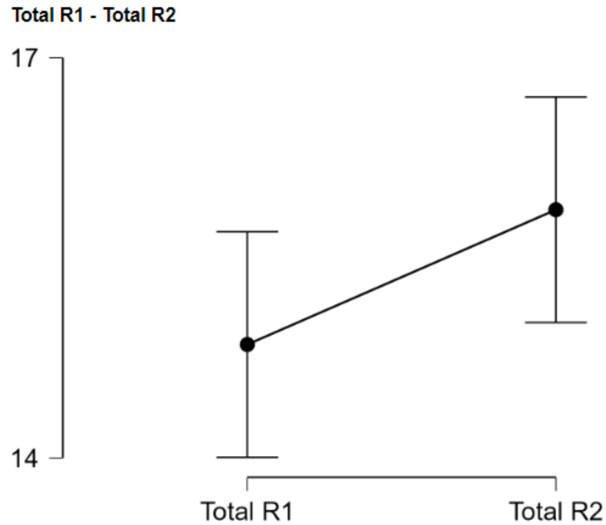

Figure 19. Descriptive plots for the Total Scores in Round 1 and 2 in the Human Voice Group.

5.5 Conclusion & Limitations

To summarize the overall findings, the scores from the NS improved in the Computer Voice Group. Also, the total scores from Rounds 1 and 2 improved in that group. But in the Human Voice Group, only the scores from the AS improved between Round 1 and 2.

There are several limitations that are worthy of discussion. First, there was a lack of control over the reviewing of activities between the two rounds of listening comprehension tests. This is primarily due to how the overall procedure was set up as the tests were administered online. For example, the duration of time that this phase lasted (that is, the time it took for a participant to finally do Round 2) was not controlled, which could have been controlled in the following fashion: "Wait for one or two days and then do the Round 2". Moreover, the author could have again explicitly specified that they should do the tests in a quiet place where there were no outside noises (ex. construction noises). The author, however, made sure that these two aspects were put in place: a) In between the Round 1 and 2, there was a feedback phase where participants were able to see their scores for Round 1 with re-playable scripts and questions (although there was also a lack of control over how many times the audio was played during this period); b) Participants should not press "back" button to go back to the previous question (which would cause a serious error in the progression of the tests).

Second, the participant demographics mainly consisted of university students who were starting to learn about the fundamental knowledge of aviation. Although they may have had a prior exposure to (or listened to pilot-ATC communication via media platforms), their knowledge of the Aviation English may vary and not have been crystallized. It needs to be reminded that AE is learned via not just the memorization of some technical jargons, but also via the operational knowledge about flight and the managerial aspect that does not immediately concern pilots (ex. Traffic management, scheduling, unruly passengers). Simply speaking, there is a possibility that variance that arises from the varied knowledge structure of participants may have been infused into the statistics obtained in this study.

Third, the author tried to reduce the background noises as much as possible for the human voice recordings (computer voices had clean sounds with no background noises when downloaded). Although real-life pilot-ATC communication contains background noises that make the communication difficult to listen to, adding noises was not the main topic of the study.

Fourth, it must be reminded that accents are not standardized but rather vary across geographical locations. African French do not necessarily share the same set of phonemes as those of standard French spoken in France. Human Voices used in this thesis were gathered from volunteers in different locations across the world. Computer Voices were downloaded from one company, whose inventory of phonology may be different than that of another company.

All things considered, although there were some statistically significant results associated with both Script Types and Rounds in each Voice Group, the results in this study should be cautiously interpreted so as not to make an erroneous conclusion that a person trained in computer voice will do better when tested with human voice or vice-versa.

Chapter 6. Study 4: Demographic Influences on the Total Scores

6.1 Introduction

In this chapter, the relationship between demographic profiles or information provided by the participants on the Preliminary Survey (see **Appendix D**) and the total scores in each Voice Group was examined separately. The purpose was to understand the relationship between the total scores in each Script Type and the predictors. Stated in another way, the study was done to identify which of the following predictors – Gender, Age, Flight Hours, Flight Rating, Years of Speaking English, Language Background, Familiarity with Spanish (and French, Arabic, Aviation English as well) – could influence the test scores based on the Script Types.

The reference levels for the categorical variables are indicated in parenthesis as follows: Gender (Female), Flight Rating (None), Language Background (NES), Familiarity with Spanish, French, Arabic, Aviation English (Excellent).

The total score (response variable) was the sum of the scores in Round 1 and 2 in each of the Script Types (NS, AS). Multiple linear regression was used. The results of this study are valuable for flight instructors because they provide insights on how such predictors could impact the overall performance of flight trainees.

Below are global equations for the total scores in the Human Voice Group, categorized by the Script Types:

$$(1)\ Total\ Score_{HVNS} = GEN + AGE + YSE + LBG + F.ESP + F.ARB + F.FRE$$

$$(2)\ Total\ Score_{HVAS} = GEN + AGE + FH + FLR + YSE + LBG + F.ESP + F.ARB + F.FRE + F.AE$$

The response variable in the equation (1) would be read as: "Total scores from the neutral scripts (NS) portion in the Human Voice". This was calculated by adding the Rounds 1 and 2 NS score. For the Computer Voice, the total score was calculated the same by adding the NS score of Round 1 and 2. Because the NS did not require or test aviation knowledge, predictors such as F.AE (Familiarity with Aviation English), FLR (Flight Ratings), and FH (Flight Hours) were not included. Equation (2) is AS (aviation scripts). The meanings of the abbreviations for the predictor variables are shown in the **List of Abbreviations**.

For the Computer Voice Group, the global model equations were also formulated in the same fashion:

(3) $Total\ Score_{CVNS} = GEN + AGE + YSE + LBG + F.ESP + F.ARB + F.FRE$

(4) $Total\ Score_{CVAS} = GEN + AGE + FH + FLR + YSE + LBG + F.ESP + F.ARB + F.FRE + F.AE$

6.2 Procedure

Because there were ten predictors, finding a model that best described data was a challenge. Fitting in many predictors could result in the inflation of the adjusted R-squared, which could lead to making judgmental errors in reporting a false-positive p-value or choosing a wrong model that is not well-representative of the data. Simply speaking, the most complex-looking model equation does not always mean it contains an accurate description of the data. Thus, it was necessary to quantify the model selection uncertainty.

In this analysis, Akaike Information Criterion (AIC) was used to find the most parsimonious model. AIC compares "multiple competing models" by taking model selection uncertainty into account (Symonds & Moussalli, 2011). It does so by producing an AIC value, which helps in ranking models based on the amount of the information loss subsumed by each model that attempts to estimate a true unknown model. The model with the smallest AIC score is the 'best approximating model' (Symonds & Moussalli, 2011). In this section, only the best models (hereafter referred to as M0) are shown and discussed. A model selection table (and other calculated statistics) was provided only if there were multiple competing best models (i.e. second-best model, third-best model etc.) that might reveal interesting insights. These tables are shown in the **Appendix B**.

Small sample size correction was applied to AIC, which is referred to as AICc. The correction was done by estimating the ratio between the sample size *n* in each Voice Group (25 in each) and the parametres *k* in the global model (i.e., the model that is the most complex and has all the predictors fitted in): n/k. If n/k is less than 40, AICc is ideal (Burnham & Anderson, 2004, p. 445). For this analysis, 25/10 = 2.5, which was less than 40. The author uses the term AIC and AICc interchangeably throughout the analysis (the actual calculations involved AICc which can be seen in Appendix B. Only the term AIC or AICc is considered synonymous in the studies).

Before performing AIC on the global models, a generalized variance influence factor (GVIF) was checked to ensure that there was no collinearity among predictors. A threshold of 2.24 for detecting collinearity among predictors was chosen based on the discussion in Truong et al. (2019). A step to removing such predictors was the same as discussed in Madlock-Brown et al. (2022), which was taking a predictor out one at a time. Alias was also checked in R, following the procedure set out in Nkwamesiga et al. (2021). Finally, regression assumptions were checked rigorously as well.

Lastly, for each analysis, outliers were identified. Identifying outliers is often overlooked and not publicly discussed when conducting research (Aguinis et al., 2013; Osborne & Overbay, 2019). Transforming the response variable was not followed because doing so could affect the final interpretation of statistical analyses by altering the distribution and thereby its original relationship with other predictors (Osborne & Overbay, 2019). By following the best practice of handling outliers shown in Aguinis et al. (2013), the type of outliers, if found (using outlier detection threshold based on studentized residuals with a threshold value set at 2), was described in detail. Therefore, the linear regression results were written for two separate cases: a) With outliers (that is, all data points included); b) Without outliers.

6.3 Results and Discussion: Human Voice Group

6.3.1 Effects of Demographics on the Total Score from NS (HV)

To determine which of the demographic characteristics could influence the total scores related to the neutral script (NS) portion of the human voice listening tests (HVNS), the sum of the total scores of NS from rounds 1 and 2 ($Total\ Score_{HVNS}$) were regressed on the following predictors: Gender, Age, Years of Speaking English, Language Background, Familiarity with Spanish, Familiarity with Arabic, Familiarity with French (see Equation (1) in pg. 90). Predictors related to flight experiences such as Flight Hours, Flight Ratings, and Familiarity with Aviation English were not included in this analysis because the questions for the neutral scripts did not require the knowledge of aviation.

Regression assumptions were checked on the global model and were all satisfied. However, the normality assumption was violated (Anderson-Darling Test, A = 0.90, p-value < 0.05), therefore an outlier diagnosis was conducted by analyzing studentized residuals with a threshold value set at 2. A visual representation of the outliers was checked to determine the nature of these outliers using olsrr package (version 0.6.0) in R. Two outliers were found. Confirming that they were not error outliers or interesting outliers by checking the data sheet once more, the author suspected they may be 'model fit outliers' which: a) could change the fit when conducting a regression analysis; b) could also introduce other meaningful predictor variables (Aguinis et al., 2013). As a result, the two outliers were removed to assess the impact of their absence. Assumptions were checked again on the global model without the outliers, and they were all satisfied.

AIC was performed on the global model. The most parsimonious model (i.e., the best model M0), was significant overall (F[1,21] = 4.819, p < 0.05, $r^2 = 0.148$). One significant predictor was included in the best model, which was Language Background (NNES) (β = -1.099, t = -2.195, p < 0.05). In simple words, the coefficient of -1.099 is the change in the score associated with a participant who identified as a non-native English speaker.

The evidence ratio analysis of the AIC generated models revealed that the best model (M0) was nearly two times better than the second-best model (M1). M1 contained Age as another predictor along with Language Background in its model equation, but it was not found to be significant.

6.3.2 Effects of Demographics on the Total Score from AS (HV)

The sum of the total scores from the aviation script (AS) portion of the human voice listening tests ($Total\ Score_{HVAS}$) were regressed on the following predictors: Gender, Age, Language Background, Flight Rating, Familiarity with Spanish, Familiarity with Arabic, Familiarity with French, Familiarity with Aviation English. Two predictors – Flight Hours and Years of Speaking English – were excluded from the above global model due to causing collinearity. The procedure to get rid of predictors was modelled after that of Madlock-Brown et al. (2022), in which the researchers repeated the removal of variables with the highest GVIF values until no predictor crossed a threshold value. The threshold value for GVIF was 2.24, based on the discussion of GVIF provided in Truong et al. (2019). After eliminating these two predictors, the assumptions were all satisfied.

AIC was performed on the resulting global model. The most parsimonious model determined by AIC contained the two predictors: Flight Rating with three levels (Student Pilot License, Private Pilot License, Commercial Pilot License), Language Background (NNES). Regression on this AIC model was significant overall (F[4,20] = 5.376, p < 0.01, $r^2 = 0.422$). This shows that the model explains 42.2% of the variance in $Total\ Score_{HVAS}$.

One level of Flight Rating was significant: Private Pilot's License (β = 4.117, t = 3.353, $p < 0.01$). Other levels of the same predictor were not significant. However, it was noted that the higher the license, the larger the coefficients became (SP < PPL < CPL). Language Background (NNES) was not significant either. The positive coefficient of 4.117 is the change in the score associated with a participant who held a Private Pilot's License.

An outlier diagnosis based on studentized residuals was conducted again. Two data points were detected, one of which was a data point that was also excluded from the neutral script analysis above. The author explained earlier that this data point was possibly a model fit outlier. To confirm whether this was true, the outliers were removed to assess how their absence affected the model fit. A new global model without the outliers was respecified and assumptions were checked again. Age predictor was dropped due to causing collinearity. After removing the said predictor, assumptions were all met. Then, AIC was performed on it.

The best model with no outliers was significant overall ($F[6,16] = 9.093$, $p < 0.001$, $r^2 = 0.688$). Although the overall significance (of the hypothesis testing) compared to that of the previous best model (that includes the outliers) did not change at all, the adjusted R-squared was improved. This best model without the outliers showed two predictors in its equation: Flight Ratings with three levels (Student Pilot's License, Private Pilot's License, Commercial Pilot's License), Familiarity with French with three levels (Moderate, Poor, No Experience).

All three levels in Flight Rating were significant: Student Pilot's License (β = 4.603 t = 3.989, $p < 0.01$), Private Pilot's License (β = 5.054, t = 5.576, $p < 0.001$), Commercial Pilot's License (β = 8.103, t = 3.405, $p < 0.01$). A stronger positive coefficient is shown in succession, which shows that participants who had a higher flight rating (thereby having more flight experience) were able to score more. For the other predictor (Familiarity with French), only the Moderate level (β = -5.372, t = -3.195, $p < 0.01$) and Poor level (β = -2.998, t = -2.939, $p < 0.01$) were significant, but their coefficients were progressively worse when having more familiarity with French. This could mean that as someone who had more fluency in French (thereby having more knowledge of French than English in general), they could have faced a harder time processing English-only conversations.

Based on the suggestions from Aguinis et al. (2013) article, it is important to think about what the outliers say. It is noteworthy that the threshold for detecting outliers was set at an absolute value of 2, not 3. There was no error in data entry as the author sent multiple emails to participants to confirm that they had read the recruitment form for the study and asking them to contact the author if they had any question. The blanks left by participants on some demographic questions on the preliminary survey were filled in as a result. Therefore, this was not an error outlier. Judging by the data point's other meta-data such as demographic information, there was nothing that stood out, so it was not an interesting outlier from that perspective. Rather, this was the case of a model fit outlier, which improved the model fit but cannot be said to be influential. This was because the significance (or hypothesis testing result) before and after the removal of outliers did not change.

6.4 Results and Discussion: Computer Voice Group

6.4.1 Effects of Demographics on the Total Score from NS (CV)

To determine which of the demographic characteristics could influence the test scores related to the neutral script (NS) portion of the computer voice listening tests, the sum of the total scores of neutral scripts (NS) from rounds 1 and 2 ($Total\ Score_{CVNS}$) were regressed on the following predictors: Gender, Age, Years of Speaking English, Language Background, Familiarity with Spanish, Familiarity with Arabic, Familiarity with French. Predictors related to flight experiences such as Flight Hours, Flight Ratings, and Familiarity with Aviation English were not included in this analysis because the questions for the neutral scripts did not require the knowledge of aviation. Regression assumptions were checked on the global model and were all satisfied.

The most parsimonious model after applying AIC contained two predictors: Familiarity with Spanish (Poor, No Experience levels) and Language Background (NNES). The best model was significant overall (F[3,21] = 11.58, p < 0.001, $r^2 = 0.57$). This shows that the model explains 57% of the variance in $Total\ Score_{CVNS}$.

Only one of the two levels within the predictor Familiarity with Spanish was significant: No Experience (β = 3.13, t = 2.983, p < 0.01). Language Background (NNES) was significant (β = -3.01, t = -5.132, p < 0.001). The positive coefficient for the No Experience level in Spanish could imply that having a greater English vocabulary, not Spanish, might have played a beneficial role in doing the tests such as helping them recall words better. Language Background (NNES), again, had a negative coefficient.

An outlier diagnosis was carried out based on studentized residuals of the global model. Two outliers were removed, and the assumption checks were completed for the respecified global model, which were satisfied. AIC was performed on this global model and it found the most parsimonious model with three predictors: Familiarity with Spanish (Poor and No Experience), Language Background (NNES), and Years of Speaking English. The best model was significant overall (F[4,18] = 14.08, p < 0.001, $r^2 = 0.704$). This shows that the model explains 70.4% of the variance in $Total\ Score_{CVNS}$ after accounting for the outliers.

For the Familiarity with Spanish predictor, the No Experience level was significant (β = 3.10, t = 3.525, p < 0.01). Language Background (NNES) was also significant (β = -2.75, t = -5.165, p < 0.001). In both cases with and without outliers, the two same predictors were significant. A participant who was a NNES would not perform well. A participant who identified as having no experience level in Spanish, ceteris paribus, tended to increase the score by 3.10 points.

6.4.2 Effects of Demographics on the Total Score from AS (CV)

The sum of the total scores from AS portion of the computer voice listening tests ($Total\ Score_{CVAS}$) was regressed on the following predictors: Gender, Age, Years of Speaking English, Language Background, Familiarity with Spanish, Familiarity with Arabic, Familiarity with French, Familiarity with Aviation English. Two predictors – Flight Hours and Flight Ratings – were excluded from the above global model due to collinearity and alias. After eliminating these two predictors, the assumptions were all satisfied. AIC was performed on the resulting global model. The most parsimonious model determined by AIC contained the predictor Familiarity with Aviation English with four levels: Good, Moderate, Poor, No Experience. The best model was significant overall (F[4,20] = 4.059, p < 0.05, $r^2 = 0.34$). This shows that the model explains 34% of the variance in $Total\ Score_{CVAS}$.

Three levels in Familiarity with Aviation English were significant: Moderate (β = -4.275, t = -2.268, p < 0.05), Poor (β = -5.800, t = -2.773, p < 0.05), No Experience (β = -8.733, t = -3.616, p < 0.01). It's interesting to note that the negative coefficients were becoming larger as the levels downgraded. This shows that being familiar with Aviation English was beneficial to participants when doing the listening tests.

Outliers were checked with studentized residuals of the global model. Two data points were removed, and the resulting global model was respecified, and assumptions were checked again. One predictor – Language Background – was dropped due to showing high GVIF values. After the assumptions were satisfied, AIC was performed on the resulting global model. The most parsimonious model determined by AIC contained two predictors with several levels: Familiarity with Aviation English (Good, Moderate, Poor, No Experience), Familiarity with Arabic (Moderate, Poor, No Experience). Regression on this AIC model was significant ($F[7,15] = 8.037$, $p < 0.001$, $r^2 = 0.69$). This shows that the model explains 69% of the variance in $Total\ Score_{CVAS}$, after accounting for outliers. Judging by the improvement in the adjusted r-squared, these two outliers were therefore model fit outliers.

Four levels in Familiarity with Aviation English were significant: Good ($\beta = -3.667$, $t = -2.266$, $p < 0.05$), Moderate ($\beta = -4.295$, $t = -3.096$, $p < 0.01$), Poor ($\beta = -6.754$, $t = -4.662$, $p < 0.001$), No Experience ($\beta = -9.256$, $t = -5.497$, $p < 0.001$). Two levels in Familiarity with Arabic were significant: Moderate ($\beta = 7.667$, $t = 2.251$, $p < 0.05$), No Experience ($\beta = 8.000$, $t = 3.377$, $p < 0.01$).

The magnitude of the negative coefficients progressively become larger from good level to no experience level in the F.AE predictor. Having Arabic proficiency at the moderate level or no experience has no stark difference, but having no experience tended to score slightly better. Judging by the coefficients for these two levels in Arabic accent, have either moderate or no experience can be said to have been advantageous when doing the listening tests. Although it is a speculation regarding why the coefficients were found to be high, from a linguistic perspective, a person whose mother tongue is English and is being educated as an ab initio pilot in what is primarily an English-speaking nation (Canada) would have a larger vocabulary knowledge than someone who is not well-versed in English but in Arabic. Reasons could be speculated, but in brief, a caution is needed when interpreting the results for Familiarity with Arabic.

6.5 Limitations

This study shares the limitations described in Study 3. In brief, although some careful measures have been taken to confine the activities in the interim, other external factors such as time or testing environment could have been at least partially controlled to some extent if a little more detail was added. The results in this study should therefore be cautiously interpreted so as not to make a strong conclusion that the predictors could be said to be influential.

The study used ten specific demographic profiles chosen from the Preliminary Survey employed in the thesis. There could be other demographic characteristics that might have been used or given more consideration if it were carried out by other researchers. Also, the study was carried out in an environment where English was mainly spoken. Nonetheless, the results of this study have practical implications with respect to establishing criteria for needs analysis – a part of CBE which is discussed in Chapter 2. Using the knowledge that language background, for example, instructors can identify the most ideal training scenarios for holders of different flight ratings.

In conclusion, a discussion on the potential impact of the following limitations (lack of control of reviewing of activities in between the tests, varied demographic information, noises, accent varieties, and the choice of demographic profiles) on the study results was carried out. Hence, the findings of this study should be taken with careful consideration of the above limitations.

6.6 Condusion

This study looked separately at what demographic characteristics influence the total scores in each Script Type within each Voice Group. In the Human Voice Group, the scores related to the NS portion of the tests were influenced primarily by a participant's language background. On the other hand, the scores related to the AS portion were influenced primarily by the flight ratings they held and their familiarity with French and Language Background. Although the actual result for French was shown to be negative in the study, perhaps on the brighter side, it can be argued that having French language skill as a pilot might be more beneficial in the real-world. Many African and Caribbean nations (which are popular destinations for tourism) speak a variation of French (ex. Creole). Being a NNES is certainly a disadvantage, as there can be challenges in communication that are not commonly encountered by a NES.

In the Computer Voice Group, the scores related to the NS portion were influenced by familiarity with Spanish and Language Background. The scores related to the AS portion were influenced by the Flight Ratings, Familiarity with Aviation English, and Arabic. Although having no experience in Spanish was positive, it must be reminded that the study was done in an environment where English was mainly spoken. Language background was also shown to be influential. On the other hand, for AS, being familiar with Aviation English was beneficial, which showed a clear pattern that the more experienced in AE, the better the scores were. Although having familiarity with Arabic was beneficial as well, the pattern is not clear. Thus, it must be cautiously interpreted that having familiarity with Arabic (whether good or bad) does not necessarily mean that one is going to accurately decipher the meaning in a heavily accented speech. In fact, in the real world, there are many dialects of Arabic. This may be the case for other languages tested in the study as well (ex. Latin-American Spanish vs. Iberian Spanish, Mainland French vs. French Creole).

The key to training pilots is to customize each training plan for different pilot cohorts according to their demographic characteristics. This can help them cover what type of training is most needed based on where their strengths and weaknesses lie. If a pilot needs to practice listening skills when there is Spanish accent, then instructors could create a training plan solely based on this requirement. The successful competition of such training depends on the overall aptitude and tenacity of the trainee.

Chapter 7. Study 5: Demographic Influences on the Total Scores from Each Accent

7.1 Introduction

In this chapter, the relationship between demographic information provided by the participants on the Preliminary Survey (see **Appendix D**) and the total scores in each of the three accent (ESP, ARB, FRE) in the Voice Groups was examined.

For example, the author wanted to know to what degree the total scores from the Neutral Scripts read in Spanish accent in the Human Voice were influenced by the following predictors: Gender, Age, Years of Speaking English, Language Background, and Familiarity with Spanish only. This procedure was also repeated for the Computer Voice Group. Similar to the study done in Chapter 6, the global model equations were set up first which are shown in the **Table 25** below.

This research question was carried out because the author wanted to see the degree of influence a specific accent had on the response variable (the total score that was from the same accent). For instance, if the response variable was the Total Score of the Neutral Script that had human Spanish accent, then the author wanted to know how much this score was influenced by Familiarity with Spanish (F.ESP), not including F.ARB, F.FRE, or F.AE. This helped in reducing the number of predictors, which could possibly cause overfitting.

An implication of the results in this study is that flight instructors (who might want to investigate a more detailed statistical analysis as a follow-up to Study 4, for example) may be interested in knowing whether his or her group of flight students need more practice in which accent. If F.ESP had a larger negative beta coefficient than those of other accents (meaning a trainee did not perform well in this accent), than it might imply that more attention is perhaps required on the training for Spanish accent. To do this, the flight instructor in question may improve the quality of the questions sets and scripts for Spanish accent.

7.2 Procedure

Table 25

Table of the global model equations for Study 3

Voice Groups	Accents	Script Types	Global Model Equations to undergo AIC

HVG	ESP	NS	$Total\ Score_{HVNS,ESP} = GEN + AGE + YSE + LBG + F.ESP$
		AS	$Total\ Score_{HVAS,ESP} = GEN + AGE + YSE + LBG + FH + FLR + F.AE + F.ESP$
	ARB	NS	$Total\ Score_{HVNS,ARB} = GEN + AGE + YSE + LBG + F.ARB$
		AS	$Total\ Score_{HVAS,ARB} = GEN + AGE + YSE + LBG + FH + FLR + F.AE + F.ARB$
	FRE	NS	$Total\ Score_{HVNS,FRE} = GEN + AGE + YSE + LBG + F.FRE$
		AS	$Total\ Score_{HVAS,FRE} = GEN + AGE + YSE + LBG + FH + FLR + F.AE + F.FRE$
CVG	ESP	NS	$Total\ Score_{CVNS,ESP} = GEN + AGE + YSE + LBG + F.ESP$
		AS	$Total\ Score_{CVAS,ESP} = GEN + AGE + YSE + LBG + FH + FLR + F.AE + F.ESP$
	ARB	NS	$Total\ Score_{CVNS,ARB} = GEN + AGE + YSE + LBG + F.ARB$
		AS	$Total\ Score_{CVAS,ARB} = GEN + AGE + YSE + LBG + FH + FLR + F.AE + F.ARB$
	FRE	NS	$Total\ Score_{CVNS,FRE} = GEN + AGE + YSE + LBG + F.FRE$
		AS	$Total\ Score_{CVAS,FRE} = GEN + AGE + YSE + LBG + FH + FLR + F.AE + F.FRE$

Note. HVG (Human Voice Group), CVG (Computer Voice Group), ESP (Spanish), ARB (Arabic), FRE (French), NS (Neutral Scripts), AS (Aviation Scripts), GEN (Gender), YSE (Years of Speaking English), LBG (Language Background), F.ESP, ARB, FRE, AE (Familiarity with Spanish, Arabic, French, Aviation English).

Each predictor was fitted into the global model equations (Table 25), categorized by the Script Type (NS, AS) and accents (ESP, ARB, FRE) in each Voice Group (HV, CV). For each global model, accents were indicated by the subscripts. The response variable was the sum of the scores in both round 1 and 2 that were from the accent being investigated. If Spanish was chosen for the accent to be investigated, others (ARB, FRE) were not included. If the Script Type was NS, the following predictors (Flight Hours, Flight Ratings, and Familiarity with Aviation English) were not included in this analysis because the questions for the neutral scripts did not require the knowledge of aviation.

Regression assumptions including collinearity were checked on the global models prior to conducting AIC. Once the assumptions were passed, AIC was performed to find the best models for each accent.

7.3 Results and Discussion: Human Voice Group

7.3.1 Effects of Demographics on the Total Score from NS with Spanish Accent (HV)

The most parsimonious model (i.e., best model) was found by using AIC, which was not significant overall (F[1,23] = 2.641, p = 0.12). The model contained only one predictor which was Age, but it was not significant (β = 0.07, p = 0.12). AIC generated some models whose delta AICc values were close to the best model (also referred to as M0). A fully averaged model was examined along with 95% confidence intervals of the coefficients contained in each of these models to see whether there were other models worthy of consideration. As Burham and Anderson (2004) noted, models that have delta AICc values below 2 are worthy of further investigation.

The second-best model (M1) was a null model with no predictors included (see **Table B6**). As Snipes and Taylor (2014) pointed out, a null model says that a response variable (or test score in this case) is a result of a random process and it cannot be predicted from the predictors at hand. Based on the evidence ratio (ER), M0 was 1.07 times better than M1 in representing the true unknown model. A full model averaging did not show any significant coefficients, nor did the 95% confidence intervals show that Age was consistently significant in other models (see **Table B7**). Then the question of why this predictor (Age) was fitted into the best model arises. The presence of Age could possibly mean that the experiences (specifically, knowhows) that come with age cannot be ignored. Wisdom, knowledge, or any other relevant experiences that are crystallized could possibly help in navigating complex instructions embedded in foreign accent.

An outlier diagnosis was conducted based on studentized residuals. Two outliers were discovered. The author posited that these could be model fit outliers because of knowing that the fully averaged model did not show any significant predictor previously. The two outliers were excluded to assess the impacts of their absence and to confirm whether this could be a valid observation. A new model without the outliers was re-specified and AIC was performed on it after checking the assumptions. The new best model was significant overall (F[3,19] = 5.391, p < 0.01, $r^2 = 0.375$), which confirmed that they were model fit outliers. Two predictors were included in this new best model: Age, Familiarity with Spanish with two levels (Poor, No Experience). Age was significant (β = 0.09, t = 3.219, p < 0.01), but the two levels for Spanish were not. This supported the author's view that Age was indeed an influential predictor in the first place.

A model selection showed that the AICc values were close between M0 (the best model) and M1 (the next best model) (see **Table B8**). M1 contained an extra predictor which was Language Background (NNES). Strictly speaking based on the evidence ratios, M0 is 1.02 times better than M1, which was not a large ratio. Among all the models considered, only Age remained consistently significant (see **Table B9**). M1 was also significant overall (F[4,18] = 5.273, p < 0.01, $r^2 = 0.437$). However, Language Background (NNES) was not significant (β = 0.31, p = 0.09). M2 was also significant overall (F[4,18] = 4.576, p < 0.05, $r^2 = 0.394$), but Gender was not significant (β = 0.23, p = 0.22). M0, M1, and M2 are theoretically the likely candidates of a true unknown model that best represents the data at hand.

7.3.2 Effects of Demographics on the Total Score from AS with Spanish Accent (HV)

In similar fashion, the sum of the scores from aviation scripts (AS) that specifically played human Spanish accent in both round 1 and 2 ($Total\ Score_{HVAS,ESP}$) was regressed on the following predictors: Gender, Age, Language Background, Years of Speaking English, Flight Ratings, Familiarity with Spanish, Familiarity with Aviation English. Flight Hours was dropped from the global model as it showed a high GVIF value (4.28). After the assumptions were met, AIC was performed to produce the most parsimonious (i.e., the best) model. The best model was not significant (F[1,23] = 3.822, p = 0.06). Language Background (NNES) was the only predictor shown in the best model, but it was not significant (β = -1.032, p = 0.06).

An outlier diagnosis was conducted. A single outlier was removed to assess its impact. The new model without the outlier was respecified and AIC was performed on it after checking the assumptions. The new best model was not significant overall (F[1,22] = 2.709, p = 0.11). Language Background (NNES) was again shown as the only predictor, but it was not significant (β = -0.83, p = 0.11). It was not a model fit outlier because it did not introduce other meaningful predictors in the model equation and change hypothesis testing (significance). It could be construed as an interesting outlier, as the participant of this data point had no flight license and indicated 'Poor' level in Spanish but scored the highest (Total HV NS ESP score = 6).

Table B10 and **Table B11** show the AIC statistics. The ΔAICc values were close between M0 and M1. M1 was the null model, which meant that the scores were a result of a random process that could not be explained by the predictors. The reason M0 was chosen to discuss in more detail than M1 was because Language Background has been appearing quite frequently in the previous analyses. The presence of this predictor is possibly hinting that its influence should not go unnoticed.

7.3.3 Effects of Demographics on the Total Score from NS with Arabic Accent (HV)

The sum of the scores from neutral scripts (NS) that specifically played human Arabic accent in both round 1 and 2 in ($Total\ Score_{HVNS,ARB}$) was regressed on the following predictors: Gender, Age, Language Background, Years of Speaking English, Familiarity with Arabic. The normality assumption was violated (Anderson-Darling Test A = 1.2329, $p < 0.05$). Thus, an outlier diagnosis was conducted to identify data points, which were identified as leverages. These two were removed to assess the impact of their absence.

After the assumptions were met, AIC was performed to produce the most parsimonious (i.e., the best) model. The best model was a null model. It is therefore necessary to consider other models generated by AIC to see which models had delta AICc values below 2. Burham and Anderson (2003) argued that models that have delta AICc values in between 1 and 2 are worthy of investigation due to having substantial support for making inferences (p. 446). There are two more models: M1 (with Gender only), M2 (with Language Background only).

Examining the fully averaged model, no predictors were found to be significant (see **Table B12**). This means that M1 and M2 were not significant. However, the presence of the following predictors Gender and Language Background in M1 and M2 seemed to provide a hint that they could potentially be influential (see **Table B13**). Simply speaking, M0, M1, and M2 are models that have substantial support, but rather than looking at M0 alone, both M1 and M2 provide a hint as to which predictor could be said to be "worthy of attention". Although no meaningful relationship was found with the given data, a researcher who might be interested in a future study could look at whether fluency in Arabic can influence listening test scores that play human Arabic accents.

7.3.4 Effects of Demographics on the Total Score from AS with Arabic Accent (HV)

The sum of the scores from aviation scripts (AS) that specifically played human Arabic accent in both round 1 and 2 in ($Total\ Score_{HVAS,ARB}$) was regressed on the following predictors: Gender, Age, Language Background, Years of Speaking English, Flight Rating, Familiarity with Arabic, Familiarity with English. Flight Hours had a high GVIF value (3.98), so it was dropped. Assumption checks showed that there was a violation of normality (Anderson-Darling Test, $A = 0.85452$, $p < 0.05$). An outlier diagnosis based on studentized residuals detected two outliers. One of the outliers was the same in the previous analysis for the Arabic accent, which the author suspected to be an interesting outlier at this point. They were removed to assess the impact of their absence.

The original global model was respecified without the two outliers and assumptions were checked once again (the predictor Years of Speaking English was dropped due to having high GVIF value of 2.52). Once the assumptions were all passed, AIC was performed to find the best model. The best model was not significant, which was a null model (M0) with no predictors. This was the same result as that of the previous analysis for the neutral script. AIC generated only one model that had delta AICc value below 2 – the second-best (M1) – which contained Language Background (NNES) as the only predictor. However, Language Background (NNES) was not significant in M1. Again, a researcher who might be interested in testing Arabic accent could look at whether it could lead to improvements in test score if it is conducted in an Arabic speaking environment.

Going back to that one participant who was also removed for further analysis in the aviation script portion as well, the participant had the highest score (Total HV AS ARB score = 5) even though the participant was a NNES and indicated "No Experience" in Arabic. But the participant was a student pilot and had indicated "Moderate" level in Aviation English. The author thinks that by learning Aviation English, one can improve listening skills in English which could help expand one's vocabulary knowledge, thereby making the decoding of accented instructions easier to some extent. An interesting outlier such as identified here provides an unexpected knowledge (Aguinis et al., 2013), which in this case, was a possible explication on how a NNES participant could do quite well in the listening test without knowledge in Arabic.

7.3.5 Effects of Demographics on the Total Score from NS with French Accent (HV)

The sum of the scores from aviation scripts (NS) that specifically played human French accent in both round 1 and 2 in ($Total\ Score_{HVNS,FRE}$) was regressed on the following predictors: Gender, Age, Language Background, Years of Speaking English, Familiarity with French. This global model passed all assumptions and AIC was performed on it to derive the best model (M0).

The best model was not significant overall (F[1,23] = 3.946, p = 0.06). The best model contained Language Background (NNES) only, which was also not significant (β = -0.50, p = 0.06).

The next best models were investigated as well. The second-best model (M1) was null model. The third-best model (M2) contained Age, Language Background. The fourth-best model (M3) contained only Years of Speaking English. Based on the evidence ratios, M0 was nearly 2 times better than M1, 2.1 times better than M2, and 2.4 times better than M3. All these AIC generated models had ΔAICc values below 2.

M2 (with Age and Language Background as predictors) was not significant overall (F[2,22] = 2.607, p = 0.10), but showed that only Language Background (NNES) was significant (β = -0.56, t = -2.17, p < 0.05). This supported the author's view that this predictor could be an influential predictor. M3 was not significant overall (F[1,23] = 2.093, p = 0.16).

An outlier diagnosis was performed. Two outliers were excluded, which were the same outliers identified in the analysis of the neutral Arabic scripts previously. The global model was respecified without the outliers and the assumptions were checked and passed. AIC was performed on it to derive the best model. The best model was significant overall (F[3,19] = 4.591, p < 0.05, $r^2 = 0.329$). The best model contained one predictor (Familiarity with French) with three levels, only one of which was significant: Poor (β = -0.98, t = -3.59, p < 0.01). A poor understanding of French language tended to change the response variable (score in the French accented section of the tests) by -0.98 points.

Looking at the AIC generated models for the case without outliers (**Table B14**), M0 (the best model) had an evidence ratio of 1.03, meaning that M0 was 1.03 times more likely to be the parsimonious (or closer to the true unknown model) than M1, but it seemed that M1 could also be a model with substantial support. M1 contained the following predictors: Familiarity with French and Gender. M2 contained Familiarity with French and Language Background, rather than Gender. M3 only had Language Background as a predictor. In the three models (M1, M2, M3), only the 'Poor' level in Familiarity with French did not cross zero in its 95% confidence intervals (see **Table B15**). From a participant's point of view, having 'poor' level in French was not advantageous because it has a negative coefficient.

Knowing that Familiarity with French ('Poor' level) was significant in M0, M1, and M2, it is noteworthy to investigate M3. M3 contained only the Language Background (NNES) (β = -

0.49, t = -2.21, p < 0.05), which was significant overall (F[1,21] = 4.886, p < 0.05, $r^2 = 0.15$). This showed that, when outliers were accounted for, being NNES can make the listening tests more challenging to solve. This conclusion was the same as that drawn with the previous best model that did not exclude outliers above.

7.3.6 Effects of Demographics on the Total Score from AS with French Accent (HV)

The sum of the scores from aviation scripts (AS) that specifically played human French accent in both round 1 and 2 in ($Total\ Score_{HVAS,FRE}$) was regressed on the following predictors: Gender, Age, Language Background, Years of Speaking English, Flight Rating, Familiarity with French, Familiarity with Aviation English. Flight Hours was dropped due to having high GVIF value (7.10). Assumptions were checked and AIC was performed on the global model above. The best model (M0) was significant overall (F[3,21] = 4.706, p < 0.05, $r^2 = 0.317$). This showed that the model explained 31.7% of variance in the response variable. The best model contained one predictor (Flight Rating) whose three levels were all significant: Student Pilot's License (β = 1.17 t = 2.17, p < 0.05), Private Pilot's License (β = 1.57, t = 2.91, p < 0.01), Commercial Pilot's License (β = 2.57, t = 2.39, p < 0.05). Having a higher flight rating or license could certainly improve test scores, which increased along with flight experiences.

An outlier diagnosis was performed. Three outliers were detected. The outliers were examined by looking at the data sheet to determine what type of outlier they were. None of them had any characteristic that stood out, although one of the data points had the lowest score (Total $Total\ Score_{HVAS,FRE}$ = 0). Removing these outliers resulted in the violation of homoscedasticity and normality assumption (Anderson-Darling Test, A = 0.84898, p-value = 0.024), therefore further analysis was not pursued. This showed that these could possibly be categorized as the outliers that were needed in the dataset (normal data).

7.4 Results and Discussion: Computer Voice Group

7.4.1 Effects of Demographics on the Total Score from NS with Spanish Accent (CV)

To examine how influential the predictors were on the test scores from the neutral scripts (NS) which *specifically* played recordings in computer Spanish accent ($Total\ Score_{CVNS,ESP}$), the sum of the Spanish scores in rounds 1 and 2 was regressed on the following predictors: Gender, Age, Language Background. Years of Speaking English, Familiarity with Spanish. This was the global model. Other languages were not included other than Spanish to see whether being familiar with Spanish accent beforehand could be advantageous when taking the listening tests. Flight Hours, Flight Ratings, and Familiarity with Aviation English predictors were not included in this analysis because the questions for the neutral scripts did not require the knowledge of aviation. Regression assumptions were checked on the global model and were all satisfied. The most parsimonious model was found by using AIC which contained two predictors: Familiarity with Spanish (Poor, No Experience) and Language Background (NNES). The model was significant overall (F[3,21] = 4.206, p < 0.05, $r^2 = 0.286$). This shows that the model explains approximately 29% of the variance in $Total\ Score_{CVNS,ESP}$.

None of the two levels in Familiarity with Spanish were significant. However, Language Background (NNES) was significant (β = -0.7756, t = -2.732, p < 0.05).

An outlier diagnosis on the global model revealed that there was one outlier (Case A), two leverages (Case B), and one influential point (Case C). A separate examination which involved taking these data points out was carried out.

In Case B, the two leverages were removed. However, it did not pass the normality check (Anderson-Darling Test, A = 0.793, p < 0.05), therefore the regression was not pursued further.

In Case C, the influential point was removed. Assumptions were checked prior to using AIC to find the most parsimonious model (i.e., the best model). Regression was performed on the best model, which was significant overall (F[3,20] = 4.13, p < 0.05, $r^2 = 0.289$). Familiarity with Spanish with two levels (Poor, No Experience) and Language Background (NNES) were included in the best model. However, only Language Background (NNES) was significant (β = -0.7046, t = -2.478, p < 0.05). It is noteworthy that there is only a slight difference in the adjusted R-squared compared to the best model that included all the cases above. The significance did not change. A model selection table for this specific case is shown in **Table B1**, which shows all other next best models that have substantial support for producing inferences.

In Case A, the outlier was removed. Assumptions were checked as well. Regression was performed on the best model, which was significant (F[4,19] = 6.25, p < 0.01, $r^2 = 0.477$).

Familiarity with Spanish with two levels (Poor, No Experience), Gender (Male), and Language Background (NNES) were included in the best model. Although the two levels in Familiarity with Spanish were not significant, Gender (β = -0.5244, t = -2.171, $p < 0.05$) and Language Background (NNES) (β = -1.0536, t = -4.211, $p < 0.001$) were. The adjusted R-square did improve but it did not change the hypothesis testing as the significance of the overall regression was still held. The outliers were therefore model fit outliers, but not influential enough.

7.4.2 Effects of Demographics on the Total Score from AS with Spanish Accent (CV)

In the similar fashion, the sum of the scores from the aviation scripts (AS) that specifically played Spanish accent in both round 1 and 2 ($Total\ Score_{CVAS,ESP}$) was regressed on the following predictors: Gender, Age, Language Background, Years of Speaking English, Flight Ratings, Familiarity with Spanish, Familiarity with Aviation English. Flight Hours was dropped from the global model as it showed a high GVIF value (4.267). After the assumptions were met, AIC was performed to produce the most parsimonious (i.e., the best) model. The best model was regressed, which was significant overall ($F[1,23] = 9.886$, $p < 0.01$, $r^2 = 0.2702$). This shows that the model explains 27% of the variance in $Total\ Score_{CVAS,ESP}$. Years of Speaking English was a significant predictor in the best model (β = 0.290, t = 3.144, $p < 0.01$). No outliers, leverages, or influential points were detected with studentized residuals.

7.4.3 Effects of Demographics on the Total Score from NS with Arabic Accent (CV)

The sum of the scores from the neutral scripts (NS) that specifically played Arabic accent in both round 1 and 2 in ($Total\ Score_{CVNS,ARB}$) was regressed on the following predictors: Gender, Age, Language Background, Years of Speaking English, Familiarity with Arabic. After the assumptions were met, AIC was performed to produce the most parsimonious (i.e., the best) model. The best model was regressed, which was significant overall ($F[1,23] = 10.19$, $p < 0.01$, $r^2 = 0.277$). This shows that the model explains approximately 28% of the variance in $Total\ Score_{CVNS,ARB}$. Language Background (NNES) was a significant predictor in the best model (β = -1.1314, t = -3.192, $p < 0.01$).

An outlier diagnosis based on studentized residuals found that there was one leverage. It was removed from the data to examine the impact of its absence. After checking the assumptions once more with the global model without the identified leverage, AIC was performed on it to find the best model. The best model was significant ($F[1,22] = 10.72$, $p < 0.01$, $r^2 = 0.297$). This shows that model explains 29.7% of the variance in $Total\ Score_{CVNS,ARB}$. Language Background (NNES) was again the only significant predictor (β = -1.1958, t = -3.275, p < 0.01). The improvement was seen in the adjusted R-squared compared to the best model that included the leverage (hence, the outlier was a model fit outlier). However, there was no change in the significance and therefore did not change hypothesis testing. Although the point was identified as a leverage, its impact was not too deleterious.

7.4.4 Effects of Demographics on the Total Score from AS with Arabic Accent (CV)

The sum of the scores from the aviation scripts (AS) that specifically played Arabic accent in both round 1 and 2 in ($Total\ Score_{CVAS,ARB}$) was regressed on the following predictors: Gender, Age, Flight Hours, Language Background, Years of Speaking English, Familiarity with Arabic, Familiarity with Aviation English. Flight Rating was removed from the global model as it was causing alias. After meeting the assumptions, AIC was performed on it to find the best model. The best model included two predictors: Flight Hours and Years of Speaking English. The regression on the best model was significant overall ($F[2,22] = 4.372$, $p < 0.05$, $r^2 = 0.219$). This shows that the model explains 21.9% of the variance in $Total\ Score_{CVAS,ARB}$. Flight hours was the only significant predictor in the best model (β = 0.014, t = 2.744, p < 0.05).

AIC generated few models whose delta AICc values were close to the best model (M0) above (see **Table B2**). A fully averaged model was examined to see whether there were other models worthy of consideration (**Table B3**). As Burham and Anderson (2004) noted, models that have delta AICc values below 2 are worthy of further investigation. The fully averaged model did not show any significant coefficient. Years of Speaking English was only significant in M1, M4, and M6. Flight Hours was consistently significant across all the models.

An outlier diagnosis based on studentized residuals was carried out, which identified three outliers. These were removed to examine the impact of their absence. During the assumption checks, Flight Hours was identified as having high GVIF value (2.35) and was dropped from the new model. After meeting the assumptions, AIC was performed to find the best new best model without the outliers. The new best model was significant (F[8,13] = 10.03, p < 0.001, $r^2 = 0.775$). The new best model contained three predictors: Familiarity with Aviation English (Good, Moderate, Poor, No Experience levels), Familiarity with Arabic (Moderate, Poor, No Experience levels), Years of Speaking English.

Three levels of Familiarity with Aviation English were significant: Moderate (β = -2.354, t = -4.563, p < 0.001), Poor (β = -2.127, t = -4.003, p < 0.01), No Experience (β = -3.099, t = -4.731, p < 0.001). One level of Familiarity with Arabic was significant: No Experience (β = 2.898, t = 3.397, p < 0.01). Years of Speaking English was significant (β = -0.199, t = -3.425, p < 0.01). In general, if a participant has less experience with AE, then they performed relatively poorly. Having no experience in Arabic had a positive coefficient, but no other levels were significant. Therefore, it is not known whether it can be said that having no experience in Arabic conferred any advantage.

It must be noted that the significance of the new best model did not change as a result of removing outliers. The difference was noted in adjusted R-squared (which improved) with other predictors introduced in the new best model that left out the outliers. In the best model that included outliers, only Flight Hours was found to be significant (a small positive coefficient). Therefore, Flight Hours can be said to be an important metric, which was one of the main points talked about in the CBE chapter of the thesis (i.e., it should not be completely discarded).

7.4.5 Effects of Demographics on the Total Score from NS with French Accent (CV)

The sum of the scores from the neutral scripts (NS) that specifically played French accent in both round 1 and 2 in ($Total\ Score_{CVNS,FRE}$) was regressed on the following predictors: Gender, Age, Language Background, Years of Speaking English, Familiarity with French. After meeting the assumption, AIC was performed on the global model to find the best model. The best model was significant overall (F[1,23] = 7.623, p < 0.05, $r^2 = 0.216$). This shows that the model explains 21.6% of the variance in $Total\ Score_{CVNS,FRE}$. The only significant predictor was Years of Speaking English (β = 0.165, t = 2.761, p < 0.05).

AIC generated models whose delta AICc values were close to that of the best model (M0) were also examined (see **Table B4**). The closest model was M1, which had a delta AICc value of

0.27. Akaike weight comparison revealed that M0 was 1.14 times better than M1, which means that M0 was the best representation of the true unknown model given the data. M1 contained the following predictors: Gender, Years of Speaking English. A detailed analysis of a fully averaged model revealed that Years of Speaking English was consistently significant across all weaker models (see **Table B5**). This shows that Years of Speaking English exerted influence on the listening test scores in this specific case.

Outliers and leverage diagnostics based on studentized residuals revealed two outliers, which were removed to assess the impact. The new model with no outliers was significant overall (F[2,20] = 13.02, p < 0.001, $r^2 = 0.522$). Two predictors were significant: Gender (Male) (β = 0.853, t = 3.363, p < 0.01), Years of Speaking English (β = 0.146, t = 3.406, p < 0.01). The significance did not change as a resulting of removing the outliers, but it did improve the adjusted R-squared. It is noteworthy that Years of Speaking English was still shown as significant.

7.4.6 Effects of Demographics on the Total Score from AS with French Accent (CV)

The sum of the scores from the aviation scripts (AS) that specifically played French accent in both round 1 and 2 in ($Total\ Score_{CVAS,FRE}$) was regressed on the following predictors: Gender, Age, Language Background, Years of Speaking English, Flight Rating, Familiarity with French, Familiarity with Aviation English. Flight Hours was dropped due to having a high GVIF value (4.67). After meeting the assumptions, AIC was performed on the global model to find the best model. The best model was significant overall (F[4,20] = 6.119, p < 0.01, $r^2 = 0.460$). This shows that the model explains 46.0% of the variance in $Total\ Score_{CVAS,FRE}$. Two levels in Familiarity with Aviation English were the only significant predictors: Poor (β = -2.600, t = -3.840, p < 0.01), No Experience (β = -3.067, t = -3.923, p < 0.001). The less experience in AE, the greater the decrease in the overall score.

Outlier and leverage diagnostics based on studentized residuals indicated that there were two outliers. The outliers were removed to assess their impacts. After meeting the assumptions, AIC was performed on the newly respecified model without the outliers to find the best model. The new best model was significant overall (F[5,17] = 11.85, p < 0.001, $r^2 = 0.712$). Two predictors were significant: Familiarity with Aviation English (Moderate, Poor, No Experience), Language Background (NNES).

Three levels in Familiarity with Aviation English were significant: Moderate ($\beta = -1.652$, $t = -4.274$, $p < 0.001$), Poor ($\beta = -2.218$, $t = -5.121$, $p < 0.001$), No Experience ($\beta = -2.305$, $t = -4.184$, $p < 0.001$). Language Background (NNES) was significant as well ($\beta = -0.954$, $t = -3.139$, $p < 0.01$).

In conclusion (for this section), the noticeable difference was twofold: 1) Improvement in the adjusted R-squared in the new best model, compared to the best model with all outliers included in; 2) Inclusion of the Language Background predictor. Hence, these were model fit outliers. But it is noteworthy that they did not change the hypothesis testing. The significance, regardless of whether the outliers were included or not, was still held. In simple words, being more familiar with AE is advantageous when confronting accented speech.

7.5 Limitations

This study, overall, shares the same limitations in Study 4. One important aspect that seems more prominent in this study is the rigorous identification of outliers. A question that arises in doing so is whether this step can be said to be necessary. The author's formal opinion is that it depends on the goal of the study. But the more insightful opinion is that of whether identifying outliers would help in addressing important insights not found by "not doing so". That is, this is a type of question that can be directly answered by a researcher's curiosity to learn more about the data at hand. In this study, the author (by looking back at the outliers) perhaps "over-analyzed" in a strict (or agonizing) manner more than it was necessary, and perhaps, it would have been sufficient to just state the finding in a simple manner – the best models (or analysis discontinued due to violation of assumptions, non-findings) for each regression result.

The author chose to present outliers and explain them, which was a recommendation according to Aguinis et al. (2013). Arguments for including or excluding outliers were also read from Osborne and Overbay (2004). However, the final decision is to be made by the researcher.

With respect to the confines of the research study design and the sample size, the null models for the Arabic accents may be a natural result. It does not mean that it is insignificant, therefore a look into other best models was necessary. An important thing to note is that AIC is not to be understood as just a tool for choosing the best model and "move on", but rather a method to quantify a level of uncertainty that might have been subsumed by the models. The other best models were thus given equally thoughtful interpretation to make an inference on which predictors could have been distantly influential. From this researcher's perspective, it is more important to take note of various results and document them than choose a model and interpret that "as the only valid answer". By describing the presence of uncertainty, it enriches the discussion the regression wisdom, that is, on the meanings of the best and alternative models in the face of uncertainty.

7.6 Condusion

In this chapter, multiple linear regression on the predictors (demographic information) with respect to the total scores from each accent was carried out. The purpose was to see if any of these factors (F.ESP, F.ARB, F.FRE) predicted total scores to some degree. The results were reported in two separate cases: With or Without Outliers. The results varied in the form of differing best models, therefore a full coverage of alternative best models and discussion on those were documented.

In the Human Voice Group, three predictors were found to be included in the best models: Age, F.FRE (Poor), and FLR (SP, PPL, CPL). There were no repeating predictors. There were two null models for the total scores in Arabic. The best model for the total scores in Spanish AS was not significant. On the other hand, in the Computer Voice Group, best models for each of ESP, ARB, FRE contained several predictors some of which reappeared in other best models (ex. LBG, YSE, GEN, F.AE).

It is important to discuss what these significant predictors mean in the real world, where more complex interactions take place. Age may symbolize overall experience, whether it be life experience that encompasses various subsets of experiences (language acquisition experience, communication experience etc.). Assuming that an individual's vocabulary knowledge expands as he or she learns more intricate features of a language, then it is no surprise that they would find themselves understanding an accented speech relatively easily. Having more knowledge of a foreign language may be a benefit. However, the possibility that confusion can arise while simultaneously appropriating attention and energy to two languages with different phonology cannot be ignored. In that case, it could make their listening performance worse.

Flight rating is a proxy for flight experiences. It could be argued that flight rating encompasses KSAs not included in flight hours. Each flight rating has clear competency requirements further categorized into performance criteria that are explicitly written, all of which must be successfully demonstrated by a trainee in front of an instructor. Having a higher rating (or flight license) is, as found in this study, beneficial.

In the Computer Voice Group, LBG seemed to be influencing total scores a lot. Being NNES is certainly a disadvantage when it comes to pilot-ATC communication because there are grammatical or lexical hurdles to overcome. Being familiar with AE helps a lot, that is, being able to use standard phraseology appropriately and understanding elliptic instructions can prevent confusion. Having more years of speaking English in general was positive, but since it only measures colloquial English, there is bound to be a limit on how much this predictor can confer advantages in reducing communication errors.

Chapter 8. Study 6: Post Survey Results

8.1 Introduction

In each Voice Group, the post surveys were completed by the participants who finished all the rounds in the study. The survey asked two questions divided by the Script Type: "Overall, how did you find the speech in the pilot-ATC conversations (neutral/non-aviation scripts) in terms of accent and the way of talking?" The variable to be measured was the post survey score, which was a Likert scale from 1 to 5, where 1 was "very unnatural" and 5 was "very natural". The research question for this study was as follows: Is there a difference between the post survey scores of each Script Type between the Voice Groups?

The post survey was made to gain some insights on how each participant felt about the test constructs overall. A blank text box was provided to let them write their opinions. There were no missing responses for the surveys.

8.2 Procedure

To analyze the data, participants' post survey scores were tabulated in a separate excel sheet. Some participants provided two separate scores for the accent and the way of talking, which was supposed to be a single score. In this case, an average was taken between the two. Participants' opinions were also noted as well. Once the scores were all added into the sheet, assumptions for the independent samples t-test were checked. However, normality was violated. Therefore, a Mann-Whitney U test was conducted instead.

8.3 Results

A summary of descriptive statistics of the results is shown below in **Table 26**.

Table 26

Summary of Descriptive Statistics for Post Surveys

Descriptive Statistics

	HVG AS	CVG AS	HVG NS	CVG NS
Valid	25	25	25	25
Missing	0	0	0	0

Descriptive Statistics

	HVG AS	CVG AS	HVG NS	CVG NS
Median	3.000	3.000	3.000	3.000
Mean	3.240	2.540	3.260	2.540
Std. Deviation	1.052	0.815	1.012	0.912
Minimum	1.000	1.000	1.000	1.000
Maximum	5.000	4.000	5.000	4.500

Note. HVG (Human Voice Group). CVG (Computer Voice Group), AS (Aviation Script), NS (Neutral Script).

A Mann-Whitney U test was conducted to compare the post survey scores between the same script types in each Voice Group (HVG AS vs. CVG AS, HVG NS vs CVG NS). Scores for the AS were higher in the Human Voice Group than the Computer Voice Group (W= 197.5, p-value < 0.01, r = -0.368), meaning the participants found the AS more natural in general. In relation to NS, the scores were higher in the Human Voice Group as well (W= 188, p-value < 0.05, r = -0.398). This indicated that the NS used in the Human Voice Group was more natural than in the Computer Voice Group.

8.4 Discussion

There were differences in the participants' experiences with the listening tests. In the Computer Voice Group, participants reported finding the accents unnatural most of the time (i.e. thick accents) and they had difficulty with deciphering what words were said. For AS, some participants found that AS was natural and understandable due to the familiar contexts comparable to real life flight experiences. For NS, the comments were mostly negative due to lack of contexts, but a few found the accents "believable". Some expressed difficulty with aviation terminologies which they were not familiar with before. There was a comment on accents being "over-exaggerated", which could mean that the accents were thicker than what was expected by them.

The reason as to why the accents were unnatural or thicker than usual is dependent on the nature of TTS and where it is provided. Every TTS is different in terms of sound quality, intonation, level of accents etc. Perhaps this is due to how they are coded by voice providers, the availability of linguistic corpora, or unique expectations they held regarding a particular language (and therefore its accents). It is rather premature to expect or hold an assumption that "TTS or A.I-backed computer voices resemble perfectly to real human voices". The accent strength of the computer voices used in this thesis might have been stronger than that of the human voices. Also, the configuration of the accent depends on what regional variation of accent is programmed. For example, if a programmer finds a particular accent of a language spoken in an administrative zone within a country that is not well-known and codes that accent to his or her own TTS system, would it be fair to say that "that accent is the standard of that language"?

On the other hand, in the Human Voice Group, participants expressed familiarity with the contexts for AS, mentioning that speech was clear and had well-represented real-life pilot-ATC conversations. In contrast to the participants in the Computer Voice Group, there were some positive comments on the accents being natural. However, accents still interfered with understanding both scripts and some clung onto callsigns of the aircraft or used their own guesses to fill in the blanks. Some commented on the uneven balance of pauses. This was unavoidable as the author had no knowledge of what recording equipment was used by each voice participants or what their natural speech rate was.

8.4. Limitation

On the post survey form, the author noticed that some participants misunderstood the question and provided two separate scores for the accent and way of talking for each Voice Group. The author originally intended and expected a single, comprehensive score (post survey score), which most participants were able to provide. However, there is a possibility that the conjunction 'and' would indicate a separate score needed to be provided. To make it more specific, a future study could include a clearer sentence for the question, stating clearly "provide two separate scores for the accent and way of talking". This way, scores can be recorded separately for accent and naturalness of the sound (for both voices).

8.5 Conclusion

Overall, the post survey scores were higher across both Script Type in the Human Voice Group. The participants in both Voice Groups relied on contexts when solving the AS portion of the tests but had more difficulty deducing answers due to lack of contexts in the NS portion. In both Voice Groups, they expressed that the foreign accents made the process of deducing an answer more difficult regardless of the script types. The implication of this finding is that accent training can be beneficial to pilots who are starting out for the first time, as they have not encountered what deleterious effect accents could have. Gaining earlier exposure to the accents of the ICAO's six official languages (French, Spanish, Arabic, Chinese, Russian; except English) would at least help them become familiar with their impacts on flight safety.

Chapter 9. Overall Discussion and Limitations

In this chapter, each of the research questions established in the Chapter 1 (1.1 Objectives & Overall Research Questions) is answered.

9.1 Study 1: Overall Discussion

Study 1. What is the relationship between CBE and Aviation English?

A) How is traditional pilot training different from CBE?

B) How is AE taught and evaluated?

C) What could be said about AE in relation to CBE?

Study 1 was a literature review that encompassed a wide range of contemporary literature that spanned from psycholinguistics to aviation-related research articles. This provided a thorough overview of the discussions pertaining to AE training such as how to set up an experimental procedure for language-related studies, the history of AE and CBE, and how these two are interrelated to one another.

For **A)**, traditional pilot training is different from CBE in general. The former places emphasis on flight hours as a main indicator of flight experience. However, this is not the right way to describe a trainee's overall skillset. CBE, on the other hand, is consultative and collaborative, meaning that needs of trainees are incorporated and reflected in a training curriculum. It clearly establishes a set of competencies that describe an attainable KSA, followed by performance criteria for each of the elements in KSA. Hours are also part of the bigger picture; that is, hours are not completely discarded but are an important denominator when it comes to evaluating a trainee's performance.

For **B)**, AE is taught in a variety of ways. It could be synchronous (offline, classroom) or asynchronous (online) depending on how a training curriculum is set up and the qualifications or interests of an instructor. An instructor who emphasizes importance on reducing communication error types is more likely to focus on the relationship between speaker and hearer and will likely be attentive in pointing out factual errors, inappropriate jargons, or background static noises. If an instructor has operational knowledge (by working previously at the airport or as an ATCO), then a more robust or advanced training curriculum may be provided. In such case, various scenarios for listening skills will be created by instructors if some trainees, for example, need more practice in decoding METAR codes or traffic information (ex. runways, vectors). Instructors may also ask "from this scenario, what is happening?" or "what does SHRA mean?" to a trainee who just finished listening to that particular script. Their performance is evaluated with ICAO LPR scale.

For **C)**, AE and CBE are not mutually exclusive. In other words, they are related. This is largely because one of the training aspects of both align themselves very closely in philosophical sense: The needs of a trainee. One of the discussions on AE was cultural background of a trainee and how this could interfere with level of understanding in an interlocution. There is no doubt that a trainee's cultural background must be acknowledged by an instructor if a successful completion of a pilot training is to be expected. It is up to the instructor to create training materials that reflect this. Also, communication should be two-way, not in one direction (the latter is a defining characteristic of a traditional classroom in which a teacher is the top of the command, and the students are mere listeners). Going back to the traditional pilot training, do "hours" indicate how communication took place between the instructor and the student? It's a number, not a description of the many trials and errors that each pilot trainee endured. Only through the knowing of a trainee's needs resulting from the accumulation of communication can an instructor track detailed progress of the trainee.

9.2 Study 2: Overall Discussion

Study 2. What can be gleaned from the literature on AE, Accent, or other disciplines that concern accents?

A) Is there sufficient literature on the impacts of foreign accent on pilot-ATC communication?
B) What lessons can be learned from other disciplines that have investigated foreign accents?

Study 2 was another literature review, but in the form of a content review of articles that were selected based on relevancy (to the thesis). The contents of each article were tabulated with their associated field of study, study topics, methods and tasks. This way of organizing information helped the author to create an idea for setting up an overall experimental procedure for the thesis. It also provided additional insights on accents in general, and once again confirmed that accents are frowned upon (that is, not received well by the hearer in conversation).

For A), the answer is no. There is no sufficient literature that documents the effect a foreign accent has on pilots' understanding. There were several articles (see the starred articles in **Table 11** & **Table 12**) that conducted empirical studies using data, but the number of these articles do not go over a single digit. Thus, the author studied other fields that have investigated accents to learn more about how those fields set up an experimental procedure and what materials were used. The articles from such fields were not related to aviation in the slightest, but their experimental methods were a source of inspiration, nonetheless.

For B), there are three important takeaways to discuss regarding AE: 1) Information asymmetry; 2) Test constructs; 3) Adaptation to accents. Information asymmetry is bound to happen in a cockpit because pilots cannot read or see the wealth of traffic information that ATCO has. In reverse, the ATCO cannot see or feel what the pilots are going through at the moment. They must rely on radiotelephony communication to make decisions that align with flight safety and their intentions. This is very difficult to achieve. Some of the accidents related to language issues have been noted as the worst disasters in aviation (see **Table 5**). If foreign accents are involved in this picture, the boundary of that asymmetry is going to become larger.

Test constructs wildly vary when AE is tested. This is not a matter of small significance. The creation of test constructs should reflect the real-life communication that takes place between a pilot and ATCO. Using ESL-level teaching to train pilots is not sufficient. Therefore, an instructor should have a relevant experience (operational experience at an ATC tower, for example) when training language of pilot trainees.

Somewhat distantly related to the issue of test constructs is how experienced pilots and ATCOs view ICAO LPR. By scheming over the LPR and its contents, one might feel that its nuance comes across as seemingly punitive to those NNES pilots as if they are the leading cause of major accidents. It has been noted that LPR is neglecting the fact that NES also make mistakes and are not immune to the difficulties posed by standard phraseology. Blaming NNES who are not as fluent in English as the former is not the correct approach to solve language issues in aviation. The correct approach is to look at test constructs that simplify the complications caused by standard phraseology, regardless of whether the complications are felt by the NES or NNES. Standard phraseology affects both, not just NNES.

Finally, accent adaptation is possible. However, this comes with time and partly, it is the willingness (or attitude) of the hearer to adapt to difficult foreign accents. That willingness to adapt to foreign accents may be influenced further by a set of cognitive strategies used by the hearer. It could be a wealth of vocabulary knowledge that enables the hearer to match the accented word with its exact match. It could also be a web of vocabulary knowledge and semantic skills that allow for the ease of extracting accurate information out of accented speech. In the context of aviation, a pilot who has flown to a very specific destination many times (Arabic regions, Hispanic regions, or Francophone regions) would have adapted to accented speech from those regions. Age might be an understandable proxy to this, but the other part (the willingness to adapt) could have more or less influence on the processing fluency than the former.

9.3 Study 3: Overall Discussion

Study 3. Are the scores in each Voice Group (CVG, HVG) clearly different?

A) Is there an improvement in the Voice Groups?

B) Does the Script Type and Rounds influence the scores in each Voice Group?

C) What are the implications for using computer voice and human voice in accent-listening training?

A) In the Human Voice Group, participants had more improvement in their scores from the portion of the tests that played AS. On the other hand, in the Computer Voice Group, there is evidence that scores from NS increased from round 1 to 2, but not for the AS. The improvement of the total score from Round 1 to 2 was seen only in the Computer Voice Group, not in the Human Voice Group.

B) When the impacts of Rounds and Script Types were examined on the scores in the Human Voice Group, there was no effect of Rounds on the scores, but Script Types were influential. The post-hoc comparison showed two important results: 1) No evidence of meaningful change between the scores from the NS portion of the tests between Round 1 and 2; 2) The change of the scores from the AS between Rounds 1 and 2, however, was clearly noticeable.

In the Computer Voice Group, only Rounds were influential. One meaningful post-hoc finding showed that the change in the scores from the NS portion of the tests between Round 1 and 2 was clearly noticeable.

9.4 Study 4 & 5: Overall Discussion

Study 4. Does demographic information of participants have any influence on the total scores in NS and AS in each Voice Group?

A) Which of the following demographic information (predictors) of the participants (Age, Gender, Years of Speaking English, Flight Hours, Flight Ratings, Language Background, Familiarity with Spanish, Arabic, French, and lastly Familiarity with Aviation English) can be said to be influential to the total scores in NS and AS? Do some of the predictors above belong to a best model equation?

Study 5. In relation to Study 4 above, does having familiarity with specific languages (ESP, ARB, FRE) influence the scores?

In terms of the influence of demographic information on the total scores, the significant predictors that appeared in the best model equations varied. It is therefore not appropriate to generalize which predictors could be construed the most influential among all and across all analyses.

A caution is needed in interpreting the results of the linear regression outputs, as there were outliers and cases of collinearity among predictors which were eliminated to conduct further analyses in their meanings. Outliers were identified and labeled if possible. Because the overall dataset was complex in structure, AIC was used to find the most parsimonious model. Detailed results with and without outliers, along with other AIC models worthy of consideration, were shown and discussed as well in Chapter 6 and 7. For the complete collection of the significant predictors for Study 4 and 5, see **Table *27***.

Table 27

Table of the Overall Summary of Regression Analyses

Regression Analyses		Influential Predictors ($p < 0.05$)
Human Voice Group	**By Script Types (Study 4)**	
	Total Score in HVNS	Without Outliers: LBG
	Total Score in HVAS	With Outliers: FLR (PPL) Without Outliers: FLR (SP, PPL CPL), F.FR (Moderate, Poor)
	By Accents (Study 5)	
	Total Score in HVNS (ES)	Without Outliers: Age
	Total Score in HVNS (AR)	Null model.
	Total Score in HVNS (FR)	Without Outliers: F.FR (Poor)
	Total Score in HVAS (ES)	Not significant with or without outliers.
	Total Score in HVAS (AR)	Null model.
	Total Score in HVAS (FR)	With Outliers: FLR (SP, PPL, CPL)
Computer Voice Group	**By Script Types (Study 4)**	
	Total Score in CVNS	With Outliers: F.ES (No Experience), LBG Without Outliers: F.ES (No Experience), LBG
	Total Score in CVAS	With Outliers: F.AE (Moderate, Poor, No Experience) Without Outliers: F.AE (Good, Moderate, Poor, No Experience), F.AR (Moderate, No Experience)
	By Accents (Study 5)	
	Total Score in CVNS (ES)	With Outliers & Influential Point: LBG Without Influential Point: LBG Without Outliers: GEN, LBG
	Total Score in CVNS (AR)	With Outliers: LBG Without Outliers: LBG
	Total Score in CVNS (FR)	With Outliers: YSE Without Outliers: GEN, YSE
	Total Score in CVAS (ES)	With Outliers: YSE
	Total Score in CVAS (AR)	With Outliers: FH Without Outliers: F.AE (Moderate, Poor, No Experience), F. AR (No Experience), YSE
	Total Score in CVAS (FR)	With Outliers: F.AE (Poor, No Experience) Without Outliers: F.AE (Moderate, Poor, No Experience), LBG

Note. Only significant predictors from the **best AIC models (M0)** are shown for the two cases: With Outliers, Without Outliers. Levels under the categorical predictors which were significant are shown in brackets. If one of the cases is missing in the third column, it was dropped because it violated either one of the regression assumptions or were not significant.

HV = Human Voice. CV = Computer Voice. NS = Neutral Script. AS = Aviation Script. LBG = Language Background (NNES). FLR (Flight Rating; SP = Student Pilot's License, PPL = Private Pilot's License, CPL = Commercial Pilot's License). YSE = Years of Speaking English. F.ES/AR/FR (Familiarity with Spanish, Arabic, French). FH = Flight Hours. GEN = Gender.

The influence of a person's language background (whether a person is NNES or NES) on the score cannot go unnoticed or overlooked. A participant who identified as NNES was more likely to be disadvantaged when completing the listening tests (i.e., they performed poorly). This is not surprising as there was a similar finding in Tiewtrakul & Fletcher (2010). A NNES would have smaller vocabularies and weaker understanding of English grammar that could have interfered with the understanding. This implies that NNES had more trouble understanding scripts which were embedded with foreign accents.

Being familiar with AE helped in scoring better. The scores were low when they had less experience with AE. It is important to remember the definition of AE: It is an ENGSP that is coded characterized by ellipsis. The purpose of AE is to convey information in a timely manner by eliminating the need to be verbose (or repetition). Having the knowledge of how AE is used, and how it should be interpreted once received, can be of immense help in navigating the challenges posed by the presence of foreign accents.

Having a higher flight rating (SP, PPL, CPL) does help in improving the score. Flight ratings are a proxy for flight hours and thereby flight experience. Having more flight experience means that a pilot has communicated with ATCOs many times, which builds familiarity on the rules and nuances that accompany the interaction. It is no surprise that having more experience in flight operations would increase the knowledge in dealing with complex communication between a pilot and ATC. However, a caution is needed when interpreting this because not many participants held flight ratings. This may imply that having familiarity with AE is still more important when it comes to actual flight operation than flight ratings. This resonates with the author's point of view that even if a pilot is more experienced, he or she is not immune to the communication complexities posed by standard phraseology.

It remains inconclusive about the impact of how long a person has spoken English. Although speaking English daily for a month or a year may improve one's English skills, it cannot be said decisively to lead to an increase in the listening test scores. Arabic pronunciation is vastly different from that of the Germanic languages where English belongs. It could be that the sudden exposure to computerized Arabic accent might have come across as a shock. Or they just had less exposure to Arabic language in general, hence making them feel more nervous when dealing with Arabic voices even if the participants themselves were fluent in English. In brief, it shows that having more years of speaking English might be helpful in understanding the basic premise of the conversation, but it is not enough in successfully navigating the complexities posed by foreign accents in standard phraseology.

The influence of age was not as clear. Age might play a role in how much of the information can be retained throughout the flight. For instance, accents can impose more cognitive load to process by reducing signal strength for older adults (Cristia et al., 2012, p. 8). However, what if the older adults have considerable exposure (or experience) in speaking, listening, or even writing Spanish? If adults, whether young or old, are well-versed in or have taken courses in Spanish, then recalling knowledge about Spanish would not be too challenging. As mentioned in the conclusion of Study 2, age may play a role, but perhaps it is the willingness to adapt to foreign accents – the attitude (KSA) – that could have played a more major role in this case.

For accents (ESP, ARB, FRE), it is difficult to say that having familiarity with any of the languages could confer significant or satisfactory advantage to overcome the impacts of accents. Only some of the models have the accents as predictors, and some of them had positive or negative coefficients. The directionality of the coefficients was not as clear compared to that of Familiarity with AE (i.e., the more experience in AE, the less negative the coefficients, and thus better the score). However, this reveals a rather interesting insight regarding the definition of accent. Accents are *varied,* and there is no single accent that can be said to represent 'the main dialect' of a language. For example, if a pilot is fluent in English and understands Texan accent well, but not New York accent, is it appropriate to conclude that the pilot in question understands all types of English accent (including the New York accent)? Will the pilot fare any better in listening to Cockney accent based on his familiarity with Texan accent? This could explain the random directionality reported in the test statistics: Being familiar with a language alone does not mean that one can completely understand various accents that come with that language. And certainly, being "good" at listening to one type of accent does not mean that the same advantage will be attained for a different accent.

Flight Hours was shown to be significant in only one best model. A caution needs to be taken regarding the interpretation of this predictor because the majority of the participants did not have enough flight hours at the time of recruitment. However, it is important to note that having more flight hours can marginally improve the overall score (i.e., a positive coefficient).

Gender was shown to be significant in only two best models for NS in the Computer Voice Group. The directionality of the coefficients of this predictor was different between whether the scripts were read in computerized Spanish (negative) or French accent (positive). For Spanish accent, male participants tended perform poorly than female participants. For French accent, male participants performed better than female participants. From this finding, it cannot be generalized that one of the two genders has performed significantly better than other.

9.7 Limitations

This thesis contributes a wealth of knowledge into AE research in terms of methodological approach such as testing A.I-backed TTS in training accent listening skills. However, it must be noted that the results outlined in the thesis should be understood in the confines of the overall research design and the external factors that could not be controlled by the author.

First Limitation. When the author started gathering participants, there was no funding. Therefore, there was no financial support for the duration of the entire thesis. This caused a limitation on the number of participants to the study. To correct for small sample size, various measures have been applied to the studies in the thesis such as using AICc for computation, checking statistical assumptions, reaching out to participants again for more information. The number of voice-over volunteers (for the Human Voice Group) and computer voices (for the Computer Voice Group), however, stayed the same as they were needed to make the scripts.

Second Limitation. It arises from the voice over volunteers for the Human Voice Group. Most of the participants were able to converse in English and it was very difficult to find native speakers whose English level was low, but a strength of accent was high. Also, it must be reminded that accents are not standardized in any language, as there are internal varieties within the same language. For instance, there is French Creole spoken in parts of the Caribbean or in the Pacific. Another example would be the Appalachian accent or Texan accent in American English. Depending on who is recruited for the voiceover volunteers for the Human Voice Group, controlling for the variance caused by the accents from different geographical regions was not possible. Not only that, what significantly delayed the time was the late responses (or no responses at all) from the volunteers, most of whom never sent any response back to the author even after multiple attempts to communicate. The computer voices, which the author bought from a TTS provider, had several errors embedded in the recordings upon downloading which took a lot of time to fix.

The author expected some anomalies in the obtained recordings because of the belief that not every recording was perfect. When every voice over volunteer had a different recording device and was recording in an unknown place or environment, an anomaly embedded in the recordings was bound to be present (ex. background noise, someone chatting in the background, or simply just a mysterious noise). The same can be said to the computer voices as well, as TTS is not perfect and cannot truly represent human voices with natural pauses, intonations, or phonology in general.

Third Limitation. The participants in the experimental studies were mostly students who were either enrolled in the aviation program or had familiarity with pilot-ATC communication. Participants were beginning to learn about the fundamentals of aviation and not many had flight licenses. It is unclear whether they would continue their studies in aviation or possibly change their major in the future. Not only that, when students were completing their listening tests, the author assumed that they would be in good condition to do the listening tests. But this is only an assumption; perhaps some participants were dealing with outside noises when doing the tests or they were pressured by an external factor such as having a series of mid-terms next week. These could potentially influence their performance on the listening tests, hence causing outliers. The outliers were accounted for in this thesis.

Conduding Remarks. What reminded the author was the frustration felt by researchers who were in a similar field that involved AeE language testing and analysis (Sullivan & Girginer, 2002; Alderson, 2010; Hamzah & Fook Fei, 2018). Responses to the recruitment posters were very low to none. The most effective method was cold emailing, which allowed the author to move from one person to another as they referred to someone who would be interested in volunteering. A word of advice to future researchers in this field would be that unless there is a pre-selected voice over volunteers or participants to the actual studies, hurdles described above will not disappear. Recruitment for the actual study was also a challenge, as participants just stopped responding to author's emails or stopped the tests in the middle with no prior notifications.

It is also reminded that this study was carried out in Canada, where English and French are officially recognized as the primary languages for communication. Future research could potentially investigate how Arabic fluency could help in extracting accurate information out of accented speech. Also, future research could include a larger sample size as it can detect other predictor variables that could be of significance. In conclusion, the findings of the studies included in this thesis should be taken carefully in relation to the limitations discussed above.

9.8 Recommendations for Future Work

The thesis presents a wealth of information about how to create testing material (audios) that has foreign accents embedded and a procedure to test listening comprehension of participants. From the making of these, the author gathered some important knowledge that might be worthy of discussion as these could provide some useful insights for future researchers in this area. The recommendations for future work are divided into following questions: I, II, III.

I. How to create a good speech material for AE training and assessment.

To create a good speech material for this purpose, the first step for a researcher is to learn about the grammatical nature of AE. AE is elliptic, which means that it is a codified language with an aim to convey a concise and clear message (or instruction) to the pilots (from ATCs). Being verbose or repetitive can confuse pilots, therefore pronunciation of each word contained in a sentence is very important.

The second step is to create a list of scripts to be recorded. The pre-requisite for this step is that the researcher must gather publications that contain examples of actual pilot-ATC communication. Other publications include ICAO manuals (Doc. 9835) or circulars (Cir. 323) that contain helpful information about the history of AE, ICAO LPR, considerations when training pilots on AE (qualification of the examiners, cultural backgrounds of interlocutors, setting up asynchronous or synchronous methods of teaching etc.). A publication example would be NAV Canada's IFR or VFR phraseology books, which contain many example instructions that an ATCO gives to a pilot.

The third step is the writing of the scripts. It is helpful to include the location of where the airplane is departing from or landing at. It is also helpful to decide how lengthy a script needs to be. A script that is filled with too many numbers or technical jargons will confuse trainees from the start. Thus, a more gradual approach whereby participants can get some time to adjust to the introduction of jargons and numbers is recommended. In the example of this thesis, the author wrote the scripts with imagination that was carefully integrated with realism. That is, the scripts contained an imaginary scenario but overall were not considered a "far stretch" from reality, as those were embedded with the author's learned knowledge of operation (of flights) by reading publications relevant to pilot-ATC communication. A fact-check of the scripts by an expert in flight operation may upgrade the quality.

The fourth step is to record the scripts. The author recruited 16 voiceover volunteers for the Human Voice Group. The other 16 were recorded with computer voices (Computer Voice Group). When recruiting human voiceover volunteers, a caution needs to be made: Not all have standardized accent. As reinforced elsewhere, accents are not standardized but vary regionally (and within a region, vary in strength). It is very difficult to recruit a voiceover volunteer who has the exact same Appalachian accent as another. Unless there is a strict protocol whereby only persons with Appalachian accent are recruited, this could mean more expenditure in terms of financial resources. And lastly, the researcher is then tasked with a difficult task of defining what centrally constitutes an Appalachian accent (and hopefully, the researcher in question is well-versed or an expert in comparative linguistics or phonology at this point). The further the scope is narrowed to control every aspect of the research; the more difficulties there will be. A careful balance between time and cost must be prioritized in this case.

Depending on the purpose of the studies, the recruitment strategy of voiceover volunteers (or the use of computer-generated voices), the number (or the length) of unique scripts to be written will be different. The intricate details of the scripts such as the inclusion of numbers, codes, jargons, scenarios will be also different. If the researcher needs help in revising the scripts, then it would be a wise decision to gain some suggestions or comments from experts who know well the operational or background knowledge of pilot-ATC communication.

II. How to design new experiments to test the training transfer from learning computer generated speech to understanding human AE speech

This thesis did not investigate the effect of training transfer from computer voice to human voice (an important question when thinking about whether computer voices can serve as pilot training materials). If the researcher wants to test the training transfer, then the following approach can be used as an example:

1. Have a participant complete a pre-test in which human voices are used to read the scripts.
    a. Train the participant in either computer or human voice.
2. Let the same participant complete a post-test in which human voices are used instead.
3. Compare the score in the pre- and post-test.

In this design, the listening test comprehension score for participants trained in CV vs. HV can be compared for HV pre- and post-test. Human voices are used in real pilot-ATC communication. The results of this test would help in answering whether there is evidence of the effect of training transfer.

The author would like to leave a helpful note regarding the above example. This is more related to the statistical interpretation of a significant finding. Suppose that a researcher who carried out the study above sees a significant finding that the participants' scores increased equally from pre- to post-test for the computer and human voice training, seeing the evidence of training transfer. The researcher concludes that the computer voices are an effective way to train pilots regarding AE. A flight instructor who is pressured to look for new ways to train pilots at the time might see this as a quick fix to the bottleneck in his or her training method.

The author thinks that one must carefully write the finding so as not to over-generalize. This is largely because computer-generated voices vary in terms of phonology. In other words, depending on who codes the TTS (and thereby configuring phonology along the way), the phonemic variation will be evident. A flight instructor who uses computer voices created by one company to train for accent listening skills versus another flight instructor who uses a different product created by another company may not see the same training transfer effect. There are too many variables to control for the pronunciation, rate of speech, and strength of accent, and lastly the issue of how scripts are written by the researchers or flight instructors.

Simply speaking, when writing a conclusion that is based on a direct comparison of data for the industry where upholding safety is high stakes (that is, safety as a concept is directly related to a dire consequence where loss of life is not avoidable), a researcher must be very careful. The author is of the opinion that training ab initio pilots with computer-generated accents may be helpful, but this material must not be accepted as an absolute substitute to the human voices. Just like flight hours is not completely discarded in CBE (discussed in Chapter 2), computer voice is an alternative tool whose importance is certainly not more than its counterpart. A good reason for flight instructors to keep computer voices in the training package may be because of its ease of access and relatively low cost in replacing older recordings. For instance, recruiting human voiceover volunteers for recording every new and unique script might not be possible for some flight instructors due to limited financial resources, so computer voices (if there is a consistent result of training transfer that conveys it is a good alternative tool) might be an ideal alternative. Future research (preferably with experts from various fields) might look at how to arrange a testing procedure that maximizes the utility value of the computer voices while still employing human voices.

The direction of the technological advancement in computer voices is also worthy of discussion. Briefly speaking, as the natural language processing and TTS-related technology improve, the author thinks that the pronunciations of computer voices may become more natural. Becoming more natural means that there will be less phonemic variations between human and computer voice.

Although it is not known how fast this will be achieved in the realm of TTS research, there is less doubt on TTS being less "robotic sounding" in the future. The question of how computer voices might be perceived by then is an interdisciplinary matter that cannot be decided by a single researcher and is a very difficult question to answer.

III. How to isolate and test the effects of the two aspects of learning (only learning accent vs. only learning AE knowledge).

This is more advanced research that allows one to observe how the listening test scores of the participants change depending on the types of training materials they are exposed to. For example:

1. Divide the participant group into three separate groups: A, B, C.
2. Design a pre- and post-test that utilized all training materials: Generic Accent Material (GAM), Aviation English without Accents (AEwA), Aviation English with Accents (AEA). In each test, a participant is tested with each training material.
3. Have a training phase where participants in A, B, and C use only one of the training materials: GAM, AEwA, AEA.
4. Compare the scores in each training material between those of pre- and post-test.

This question would answer the question of whether an improvement of the score could be seen from pre- to post-test for each participant, based on their exposure to one of the training materials. Thus, this enables isolating and testing the effects of the learning of accent vs. learning of AE knowledge possible.

Chapter 10. Condusion

This thesis focused on whether computer voices generated from A.I-backed text-to-speech system (TTS) can confer any benefit compared to human voices in terms of training. To determine this, the following studies have been conducted: (1) Comparison of the differences in the performances of the participants in the Human Voice Group and Computer Voice Group based on Rounds and the Script Types; (2) Investigation of the effects of demographics on the scores categorized by the Script Types and Accents. For (1), paired-samples t-tests were conducted along with two-way repeated measures ANOVA. For (2), multiple linear regressions were conducted with Akaike Information Criterion along with the computation of the generated models' fully averaged models, delta AICcs, Akaike Weights, and Evidence Ratios. The following demographic profiles have also been confirmed to influence the performance of the listening tests: Language Background, Flight Ratings, Familiarity with Aviation English, and Years of Speaking English.

The thesis also covered a detailed overview of competency-based education, history of Aviation English since World War I and II. For the literature review, literature from other academic disciplines was summarized and noted to delineate the methodological approaches that could be used to develop and enhance that of the thesis. This was largely because the author found no prior study on accent testing in the field of aviation. The results of the literature reviews found that the research on how to train and test accent-embedded instructions in aviation is scant, and not well-documented. Not only that, there was no emphasis on the use of TTS to train pilots' skills in extracting accurate meanings out of the accented speech. This study contributes to this gap that TTS (computer voices) can perhaps be used to train them in this regard.

The results of the thesis, as aforementioned elsewhere, should not be construed as advocating the use of TTS over human voices. Instead, it should be understood in the sense that using TTS may help in enriching a training curriculum by strategizing the purpose of usage. It is recommended that a flight instructor who is looking to include both human and computer accents in training start thinking about which virtual scenarios could be best served by these two accent types. It would be helpful to know the background of trainees as well, as it was shown in the thesis that some of their prior experiences may well determine the quality of their performances.

Technology evolves at a rapid pace. It is unknown whether language-related technology such as natural language processing combined with TTS could be used in a cockpit to streamline complex operations. Although the ways in which instructors apply TTS to train pilots will vary depending on the specification of each curriculum, it is a valuable technological resource whose influence cannot go unnoticed in the day and era of artificial intelligence.

References

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods*, *16*(2), 270–301. https://doi.org/10.1177/1094428112470848

Alderson, J. C. (2010). A survey of aviation English tests. *Language Testing*, *27*(1), 51–72. https://doi.org/10.1177/0265532209347196

Aviation Safety Network. (n.d.). *Boeing 747-249F* [ASN Aviation Safety Database]. Language/Communication Problems. https://aviation-safety.net/database/record.php?id=19890219-0

Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *The Journal of the Acoustical Society of America*, *137*(4), 2015–2024. https://doi.org/10.1121/1.4916265

Baugh, B., & Stolzer, A. (2018). Language-Related Communications Challenges in General Aviation Operations and Pilot Training. *International Journal of Aviation, Aeronautics, and Aerospace*. https://doi.org/10.15394/ijaaa.2018.1271

Borowska, A. (2017). Аеронавігаційна англійська: Аналіз окремих комунікативних стратегій, що використовуються персоналом-носіями англійської мови під час спілкування з персоналом с 4-м рівнем володіння англійською мовою. *Proceedings of the National Aviation University*, *71*(2), 140–147. https://doi.org/10.18372/2306-1472.71.11759

Borowska, A. P. (2017). *Avialinguistics: The study of language for aviation purposes*. Peter Lang.

Borowska, A. P. (2018). A Multilingual Speaker in Global Aviation Communication. *Applied Linguistics Papers*, *2/2018*(25), 11–19. https://doi.org/10.32612/uw.25449354.2018.2.pp.11-19

Burnham, K. P., & Anderson, D. R. (Eds.). (2004). *Model Selection and Multimodel Inference*. Springer New York. https://doi.org/10.1007/b97636

Cambridge Dictionary. (n.d.). Aeronautical. In *Cambridge Dictionary*.

CCAA. (2018). *Labour market information report*. CCAA.

Clark, B. (2017). *Aviation english research project: Data analysis findings and best practice recommendations* [Independent research report]. CAA.

Coertze, S., Conradie, S., Burger, C. R., & Huddlestone, K. (2014). Aviation English in South African airspace. *Stellenbosch Papers in Linguistics Plus*, *42*(0), 41. https://doi.org/10.5842/42-0-167

Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. MIT Press.

Cristia, A., Seidl, A., Vaughn, C., Schmale, R., Bradlow, A., & Floccia, C. (2012). Linguistic Processing of Accented Speech Across the Lifespan. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00479

Defalque, H. (2017). Competency-based training and assessment. *ICAO*.

Dissanayaka, Y. H. P. S. A. Y., Molesworth, B. R. C., & Estival, D. (2023). Miscommunication in Commercial Aviation: The Role of Accent, Speech Rate, Information Density, and Politeness Markers. *The International Journal of Aerospace Psychology*, *33*(1), 79–97. https://doi.org/10.1080/24721840.2022.2154672

Dragojevic, M., & Giles, H. (2016). I Don't Like You Because You're Hard to Understand: The Role of Processing Fluency in the Language Attitudes Process: Processing Fluency and

Language Attitudes. *Human Communication Research*, *42*(3), 396–420. https://doi.org/10.1111/hcre.12079

Emery, H. J. (2014). Developments in LSP Testing 30 Years On? The Case of Aviation English. *Language Assessment Quarterly*, *11*(2), 198–215. https://doi.org/10.1080/15434303.2014.894516

Eurocontrol. (2021). *Engligh Language Proficiency for Aeronautical Communication ELPAC Test Taker Preparation—Pilots*. Eurocontrol.

FAA. (2006). *Advisory Circular*. Federal Aviation Administration. https://www.icao.int/safety/airnavigation/OPS/CabinSafety/Cabin%20Safety%20Library/FAA%20AC%20120-54a%20-%20Advanced%20Qualification%20Program.pdf

FAA. (2023, March 7). *Boeing 747-206B and Boeing 747-121* [Transport Airplane Accident Library]. https://www.faa.gov/lessons_learned/transport_airplane/accidents/PH-BUF

Finland Safety Investigation Authority. (n.d.). *Risk of runway collision between two airlines at helsinki-vantaa aerodrome on 29 december 2011* (Accident Investigation L2012-01). Safety Investigation Authority, Finland.

Franks, P., Hay, S., & Mavin, T. (2014). Can competency-based training fly?: An overview of key issues for *ab initio* pilot training. *International Journal of Training Research*, *12*(2), 132–147. https://doi.org/10.1080/14480220.2014.11082036

Friginal, E., Mathews, E., & Roberts, J. (2020). *English in Global Aviation: Context, research, and pedagogy*. Bloomsbury Academic.

Gander, P. H., Mulrine, H. M., Van Den Berg, M. J., Smith, A. A. T., Signal, T. L., Wu, L. J., & Belenky, G. (2015). Effects of sleep/wake history and circadian phase on proposed pilot fatigue safety performance indicators. *Journal of Sleep Research*, *24*(1), 110–119. https://doi.org/10.1111/jsr.12197

Hamzah, H., & Fook Fei, W. (2018). Miscommunication in Pilot-controller Interaction. *3L The Southeast Asian Journal of English Language Studies*, *24*(4), 199–213. https://doi.org/10.17576/3L-2018-2404-15

Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When One Person's Mistake Is Another's Standard Usage: The Effect of Foreign Accent on Syntactic Processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887. https://doi.org/10.1162/jocn_a_00103

Hazrati, A. (2015). Intercultural Communication and Discourse Analysis: The Case of Aviation English. *Procedia - Social and Behavioral Sciences*, *192*, 244–251. https://doi.org/10.1016/j.sbspro.2015.06.035

Hellenic Republic Ministry of Transport & Communications. (2006). *Helios airways flight HCY522 being 737-31s at grammatiko, hellas on 14 august 2005*. Air Accident Investigation & Aviation Safety Board (AAIASB).

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297–307. https://doi.org/10.1093/biomet/76.2.297

ICAO. (2010, March 1). *Welcome to the Next Generation of Aviation Next Generation of Aviation Professionals Symposium Workshop*. https://www.icao.int/Meetings/AMC/NGAP/Documentation/MAF-NGAP%20Welcome.pdf

ICAO. (n.d.). *The History of ICAO and the Chicago Convention*. https://www.icao.int/about-icao/history/pages/default.aspx

ICAO 32nd Session of the Assembly. (n.d.). RESOLUTIONS ADOPTED AT THE 32ND SESSION OF THE ASSEMBLY. *ICAO*.

ICAO 40th Session of the Assembly. (2019). *Addressing the shortage of pilots and aviation professionals—A global approach to a global problem*. ICAO.

ICAO Air Transport Bureau. (2023, January 27). *Effects of Novel Coronavirus (COVID-19) on Civil Aviation: Economic Impact Analysis*.

ICAO Circular 323, & ICAEA. (2009). *Guidelines for Aviation English Training Programmes*. International Civil Aviation Organization.

ICAO Doc. 9835. (2010). *Manual on the implementation of ICAO language proficiency requirements* (2. ed). International Civil Aviation Organization.

ICAO Doc. 9868. (2016). *Procedures for Air Navigation Services—Training* (2nd ed). International Civil Aviation Organization. https://www.icao.int/mid/documents/2018/cbt%20atco%20and%20atsep%20wksp/doc%209868_en.pdf

ICAO Doc. 9995 (Ed.). (2013). *Manual of evidence-based training* (1st ed). International Civil Aviation Organization.

ICAO RASG-APAC/1. (2011). *Update on ICAO's competency-based approach to training and assessment*. ICAO. https://www.icao.int/APAC/Documents/rasg_apac/rasg_apac/IP08%20%20ICAO%20CBT.pdf

Joyner, C. W. (1995). The DACUM technique and competency-based education. In *Challenge and Opportunity: Canada's Community Colleges at the Crossroads* (pp. 243–255). UBC Press.

Kearns, S. K., Mavin, T. J., & Hodge, S. (2017). *Competency-Based Education in Aviation: Exploring Alternate Training Pathways* (1st ed.). Routledge. https://doi.org/10.4324/9781315563220

Kim, H., & Elder, C. (2015). Interrogating the construct of aviation English: Feedback from test takers in Korea. *Language Testing*, *32*(2), 129–149. https://doi.org/10.1177/0265532214544394

Knoch, U. (2014). Using subject specialists to validate an ESP rating scale: The case of the International Civil Aviation Organization (ICAO) rating scale. *English for Specific Purposes*, *33*, 77–86. https://doi.org/10.1016/j.esp.2013.08.002

Kraśnicka, I. (2016). English with Flying Colors: The Aviation English and the International Civil Aviation Organization. *Studies in Logic, Grammar and Rhetoric*, *45*(1), 111–124. https://doi.org/10.1515/slgr-2016-0019

Larandang, B., Thamrin, N. S., Naniwarsih, A., Omolu, F. A., & Nadrun, N. (2023). The top-down strategies in an EFL class of the vocational high school: An experimental study on enhancing students' listening comprehension skills. *Journal on English as a Foreign Language*, *13*(2), 399–426. https://doi.org/10.23971/jefl.v13i2.6269

Lev-Ari, S. (2015). Comprehending non-native speakers: Theory and evidence for adjustment in manner of processing. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01546

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. https://doi.org/10.1016/j.jesp.2010.05.025

Madlock-Brown, C., Wilkens, K., Weiskopf, N., Cesare, N., Bhattacharyya, S., Riches, N. O., Espinoza, J., Dorr, D., Goetz, K., Phuong, J., Sule, A., Kharrazi, H., Liu, F., Lemon, C., & Adams, W. G. (2022). Clinical, social, and policy factors in COVID-19 cases and deaths: Methodological considerations for feature selection and modeling in county-level analyses. *BMC Public Health*, *22*(1), 747. https://doi.org/10.1186/s12889-022-13168-y

Mansikka, H., Harris, D., & Virtanen, K. (2017). An Input–Process–Output Model of Pilot Core Competencies. *Aviation Psychology and Applied Human Factors*, *7*(2), 78–85. https://doi.org/10.1027/2192-0923/a000120

McLaughlin, D. J., Baese-Berk, M. M., Bent, T., Borrie, S. A., & Van Engen, K. J. (2018). Coping with adversity: Individual differences in the perception of noisy and accented speech. *Attention, Perception, & Psychophysics*, *80*(6), 1559–1570. https://doi.org/10.3758/s13414-018-1537-4

Ministry of Civil Aviation. (n.d.). *Report of court of inquiry on mid-air collision between saudi arabian boeing 747 and kazakhstan il-76 on 12th november, 1996 near delhi-india (charkhi-dadri, haryana)* [Accident Investigation].

Moskalenko, O. I., & Didenko, O. V. (2018). A COMPUTER TOOL FOR TRAINING PILOTS' LISTENING SKILLS IN AVIATION ENGLISH. *Information Technologies and Learning Tools*, *67*(5), 187. https://doi.org/10.33407/itlt.v67i5.2434

National Transportation Safety Board. (1991). *Aircraft accident report. Avianca, the airline of colombia boeing 707-321b, hk 2016 fuel exhaustion cove neck, new york january 25 1990* [Accident Investigation].

NAV Canada. (2022). *IFR Phraseology*. NAV Canada.

Netherlands Aviation Safety Board. (n.d.). *Final report and comments of the Netherlands Aviation Safety Board of the investigation into the accident with the collision of KLM flight 4805, Boeing 747-206B, PH-BUG and Pan American flight 1736, Boeing 747-121, N736PA at Tenerife airport, Spain on 27 March 1977* [Accident Investigation]. Netherlands Aviation Safety Board.

Nkamwesiga, J., Korennoy, F., Lumu, P., Nsamba, P., Mwiine, F. N., Roesel, K., Wieland, B., Perez, A., Kiara, H., & Muhanguzi, D. (2022). Spatio-temporal cluster analysis and transmission drivers for Peste des Petits Ruminants in Uganda. *Transboundary and Emerging Diseases*, *69*(5). https://doi.org/10.1111/tbed.14499

Osborne, J. W., & Overbay, A. (2019). *The power of outliers (and why researchers should ALWAYS check for them)*. https://doi.org/10.7275/QF69-7K43

Park, M. (2018). Innovative assessment of aviation English in a virtual world: Windows into cognitive and metacognitive strategies. *ReCALL*, *30*(2), 196–213. https://doi.org/10.1017/S0958344017000362

Parohinog, D., & Meesri, C. (2015). ICAO-Based Needs Assessment in Thailand's Aviation Industry: A Basis for Designing a Blended Learning Program. *Procedia - Social and Behavioral Sciences*, *208*, 263–268. https://doi.org/10.1016/j.sbspro.2015.11.202

Seiler, W. (2009). English as a lingua franca in aviation. *English Today*, *25*(2), 43–48. https://doi.org/10.1017/S0266078409000182

Sgro, H. J. A. (2019). *Supporting Canada's Flight Schools*.

Shum, K. K., Au, T. K., Romo, L. F., & Jun, S.-A. (2021). Learning Challenging L2 Sounds Via Computer Training: High-Variability Perceptual Training for Children and Adults. *Language Learning and Development*, *17*(4), 327–342. https://doi.org/10.1080/15475441.2021.1876699

Statistics Canada. (2023). Monthly civil aviation statistics, March 2023. *Statistics Canada*, *11*.

Sullivan, P., & Girginer, H. (2002). The use of discourse analysis to enhance ESP teacher knowledge: An example using aviation English. *English for Specific Purposes*, *21*(4), 397–404. https://doi.org/10.1016/S0889-4906(01)00042-4

Sumalinog, G. G. (2018). Listening anxiety in focus: The case of the foreign university students in the philippines. *International Journal of Current Research*, *10*(7), 71366–71371.

Symonds, M. R. E., & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, *65*(1), 13–21. https://doi.org/10.1007/s00265-010-1037-6

Tajima, A. (2004). Fatal miscommunication: English in aviation safety. *World Englishes*, *23*(3), 451–470. https://doi.org/10.1111/j.0883-2919.2004.00368.x

Thomas, M., & Richards, C. (2015). Determining Readiness for Solo Flight Training: Toward Defining the Elements of Competency. *Aviation Psychology and Applied Human Factors*, *5*(2), 114–124. https://doi.org/10.1027/2192-0923/a000084

Tiewtrakul, T., & Fletcher, S. R. (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. https://doi.org/10.1080/00140130903470033

Todd, M. A., & Thomas, M. J. W. (2013). Experience, Competence, or Syllabus? Influences on Flight Hours at Licensing of Commercial Pilots. *The International Journal of Aviation Psychology*, *23*(2), 169–180. https://doi.org/10.1080/10508414.2013.772851

Transport Canada. (n.d.-a). *Standard 421—Flight Crew Permits, Licences and Ratings—Canadian Aviation Regulations (CARs)* [Canadian Aviation Regulations]. Canadian Aviation Regulations (SOR/96-433). Retrieved June 28, 2023, from https://tc.canada.ca/en/corporate-services/acts-regulations/list-regulations/canadian-aviation-regulations-sor-96-433/standards/standard-421-flight-crew-permits-licences-ratings-canadian-aviation-regulations-cars#421_06

Transport Canada. (2019). *Guide to a formal aviation language proficiency demonstration* [Government Website]. https://tc.canada.ca/en/aviation/licensing-pilots-personnel/getting-aviation-language-proficiency-demonstration/guide-formal-aviation-language-proficiency-demonstration

Transport Canada. (n.d.b). *Chapter 1—AQP Introduction and Overview* [Government Website]. Development and Implementation of an Advanced Qualification Program (AQP). https://tc.canada.ca/en/aviation/commercial-air-services/approved-check-pilot-acp-advanced-qualification-program-aqp/development-implementation-advanced-qualification-program-aqp/11-introduction

Trippe, J., & Baese-Berk, M. (2019). A prosodic profile of American Aviation English. *English for Specific Purposes*, *53*, 30–46. https://doi.org/10.1016/j.esp.2018.08.006

Truong, L. T., Nguyen, H. T. T., Nguyen, H. D., & Vu, H. V. (2019). Pedestrian overpass use and its relationships with digital and social distractions, and overpass characteristics. *Accident Analysis & Prevention*, *131*, 234–238. https://doi.org/10.1016/j.aap.2019.07.004

Valenta, V. (2018). Effects of Airline Industry Growth on Pilot Training. *MAD - Magazine of Aviation Development*, *6*(4), 52–56. https://doi.org/10.14311/MAD.2018.04.06

Wu, Q., Molesworth, B. R. C., & Estival, D. (2019). An Investigation into the Factors that Affect Miscommunication between Pilots and Air Traffic Controllers in Commercial Aviation. *The International Journal of Aerospace Psychology*, *29*(1–2), 53–63. https://doi.org/10.1080/24721840.2019.1604138

Zhao, K., Guo, X., & Gao, X. (2017). Learning English to fly: A study of Chinese cargo airline pilots' learning engagement: Understanding Chinese pilots' English learning and use informs efforts to improve their command of English and ensure aviation safety. *English Today*, *33*(4), 5–11. https://doi.org/10.1017/S0266078417000219

Appendix A

Information Table for Akaike information Criterion

Table A1

Table of Abbreviations used for Multiple Linear Regression

CV	Computer Voice Group	NES	Native English Speaker
HV	Human Voice Group	NNES	Non-Native English Speaker
NS	Neutral Script	FH	Flight Hours
AS	Aviation Script	GEN	Gender
AICc	Corrected Akaike Information Criterion for small samples	YSE	Years of Speaking English
ES, ESP	Spanish	LBG	Language Background
AR, ARB	Arabic	F.FRE/ARB/ESP/AE	Familiarity with French, Arabic, Spanish, Aviation English
FR, FRE	French	aGVIF	"Adjusted" Generalized Variance Inflation Factor
AE	Aviation English	Log.Lik	Log Likelihood
w_i	Akaike Weight(s)	$\Delta AICc$	Delta AICc
ER_i	Evidence Ratio	$Total\ Score_{NS,ES\ (or\ AR,FR)}$	Participant's total score from Neutral Scripts that only played Spanish (or Arabic, French) accents
$Total\ Score_{AS,ES\ (or\ AR,FR)}$	Participant's total score from Aviation Scripts that only played Spanish (or Arabic, French) accents		

Table A1 shows the abbreviations of the words commonly used in the multiple linear regressions analyses in the thesis. Some of these abbreviations such as Log.Lik, w_i, ER_i, ΔAICc are related to Akaike Information Criterion (AIC). AIC generates a list of models based on the AIC score from a global model that is complex (has all necessary predictors fitted in at first). A model with the smallest AIC score is the most parsimonious (i.e., the best) model that are closer to a representation of a true unknown model. AICc is defined by Hurvich and Tsai (1989) as the corrected version of the AIC for small samples. Equations for these are shown in the Table A2 below.

According to Snipes and Taylor (2014), when the best model (the most parsimonious model) is selected, hypothesis testing can be performed on it. Also, when the best model is chosen, other generated models "can be effectively ignored" (Snipes & Taylor, 2014). The definition of the best model is based on the concept of parsimony, which is balancing the drawbacks of underfitting and overfitting with many predictor variables in the model equation (Burnham & Anderson, 2004). In this sense, the best model can be said to be the closest approximation of the true unknown model of the data being analyzed, and that also mathematically balances bias which stem from the aforementioned drawbacks.

Table A2

List of Equations for AIC

AIC Score	$AIC = 2K - 2\log(L(\hat{\theta}\|y))$	K = number of estimable parameters $\left(L(\hat{\theta}\|y)\right)$ = log-likelihood function
AICc	$AICc = AIC + \frac{2K(K+1)}{n-K-1}$	n = sample size (Other parameters are the same as above)
ΔAICc	$\Delta AICc = AICc(i) - AICc_{min}$	AICc(i) = individual AICc score for each model $AICc_{min}$ = minimum AICc score of the models
Akaike Weight	$w_i = \frac{\exp(-\frac{1}{2}\Delta AICc_i)}{\sum_{r=1}^{R}\exp(-\frac{1}{2}AICc_R)}$	R = number of generated models r = the model in question
Evidence Ratio	$ER_i = \frac{w_{best}}{w_i}$	w_{best} = the best model's weight w_i = the nest best model's weight
Note. The list of equations shown on this table were adapted from Snipes and Taylor (2014).		

Akaike weights, which quantify model selection uncertainty (Burnham & Anderson, 2004) – were used to calculate evidence ratios. Evidence ratio helps in comparing the level of support each model has based on the given data (Burnham & Anderson, 2004). In brief, evidence ratio quantifies how many times better the best model is compared to the next best model.

Appendix B

Supplementary Tables for the Regression Outputs in the Voice Groups

This section lists all the supplementary tables that have been used to make judgments on selecting the best model for each regression analyses completed in both Voice Group (CV, HV). These tables are known as model selection tables, which contain all the models generated by the AIC. These tables helped in comparing and investigating which model equation was the most parsimonious given test statistics (log likelihood, AICc, ΔAICc, Akaike Weights, and Evidence Ratios).

According to Burnham and Anderson (2004), the "next best models" may contain important inferential insights based on evidence ratios and confidence intervals. To follow this best practice in handling information-theoretic approach, two tables were created for each case of with and without outliers: 1) If the ΔAICc values of the next best models were close to each other, including that of the best model (a difference somewhere between 0 and 1), then a model selection table was created to compare the models; 2) A table that showed coefficients and confidence intervals of a fully averaged model was created to provide the totality of the evidence. In terms of showing which models to display in the tables, the models whose ΔAICc values lied somewhere between 1 and 2 were chosen as those had "substantial support and should receive consideration in making inferences (Burnham & Anderson, 2004). The model selection tables are organized by their Voice Group (CV, HV), Script Types (NS, AS), and Accents (ESP, FRE, ARB), followed by a dash with the case of 'with outlier' or 'without outlier' (or influential point, leverage).

Computer Voice Group

Table B1

Model Selection Table (CV NS ESP – Without Influential Point)

Models	df	Log.Lik	AICc	$\Delta AICc$	w_i	ER_i
M0 (best model)	5	-22.221	57.8	0.00	0.225	1
M1	6	-20.756	58.5	0.68	0.160	1.41
M2	6	-21.312	59.6	1.79	0.092	2.45

Models:

M0: $TS_{CVNS,ESP} = \beta_0 + \beta_1 Age + \sum_e \beta_e(F.ESP) + \sum_l \beta_l(LBG) + \varepsilon$

M1: $TS_{CVNS,ESP} = \beta_0 + \sum_e \beta_e(F.ESP) + \sum_g \beta_g(GEN) + \sum_l \beta_l(LBG) + \varepsilon$

M2: $TS_{CVNS,ESP} = \beta_0 + \beta_1 YSE + \sum_e \beta_e(F.ESP) + \sum_l \beta_l(LBG) + \varepsilon$

Table B2

Model Selection Table (CV AS ARB – With Outliers)

Models	df	Log.Lik	AICc	$\Delta AICc$	w_i	ER_i
M0 (best model)	4	-42.851	95.7	0.00	0.112	1
M1	5	-41.356	95.9	0.17	0.103	1.09
M2	5	-41.399	96.0	0.25	0.099	1.13
M3	3	-44.613	96.4	0.67	0.080	1.4
M4	5	-41.969	97.1	1.39	0.056	2
M5	4	-43.721	97.4	1.74	0.047	2.4
M6	6	-40.490	97.6	1.94	0.042	2.7

Models:
M0: $TS_{CVAS,ARB} = \beta_0 + \beta_1 FH + \beta_2 YSE + \varepsilon$
M1: $TS_{CVAS,ARB} = \beta_0 + \beta_1 Age + \beta_2 FH + \beta_3 YSE + \varepsilon$
M2: $TS_{CVAS,ARB} = \beta_0 + \beta_1 FH + \sum_g \beta_g (GEN) + \beta_2 YSE + \varepsilon$
M3: $TS_{CVAS,ARB} = \beta_0 + \beta_1 FH + \varepsilon$
M4: $TS_{CVAS,ARB} = \beta_0 + \beta_1 FH + \sum_l \beta_l (LBG) + \beta_2 YSE + \varepsilon$
M5: $TS_{CVAS,ARB} = \beta_0 + \beta_1 FH + \sum_g \beta_g (GEN) + \varepsilon$
M6: $TS_{CVAS,ARB} = \beta_0 + \beta_1 Age + \beta_2 FH + \sum_g \beta_g (GEN) + \beta_2 YSE + \varepsilon$

Table B3

Average Model (CV AS ARB – With Outliers)

Models	AICc	$\Delta AICc$	*Age*	*Flight Hours*	*GEN* [b]	*LBG* [c]	*YSE* [d]
M0	95.7	0.00	–	[0.003,0.025]	–	–	[-0.407, 0.026]
M1	95.9	0.17	[-0.303, 0.036]	[0.003, 0.024]	–	–	[-0.442, -0.013]
M2	96.0	0.25	–	[0.004, 0.025]	[-0.263, 2.057]	–	[-0.422, 0.001]
M3	96.4	0.67	–	[0.001, 0.022]	–	–	–
M4	97.1	1.39	–	[0.002, 0.023]	–	[-2.115, 0.536]	[-0.463, -0.008]
M5	97.4	1.74	–	[0.001, 0.022]	[-0.474, 1.988]	–	–
M6	97.6	1.94	[-0.280, 0.072]	[0.004, 0.024]	[-0.512, 1.892]	–	[-0.448, -0.021]
Model Averaging [a]			-0.037	0.01	0.255	-0.141	-0.126

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Note. 95% confidence interval for coefficients (in square brackets).

[a] Coefficients from a fully averaged model. Only the coefficients (rounded if possible) for the predictors mentioned in the table are provided.
[b] Gender (Male).
[c] Language Background (NNES).
[d] Years of Speaking English.

Table B4

Model Selection Table (CV NS FRE – With Outliers)

Models	df	Log.Lik	AICc	$\Delta AICc$	w_i	ER_i
M0 (best model)	3	-30.439	68.0	0.00	0.214	1
M1	4	-29.143	68.3	0.27	0.188	1.14
M2	4	-29.558	69.1	1.10	0.124	1.73
M3	4	-29.675	69.3	1.33	0.110	1.95

Models:
M0: $TS_{CVNS,FRE} = \beta_0 + \beta_1 YSE + \varepsilon$
M1: $TS_{CVNS,FRE} = \beta_0 + \sum_g \beta_g(GEN) + \beta_1 YSE + \varepsilon$
M2: $TS_{CVNS,FRE} = \beta_0 + \beta_1 Age + \beta_2 YSE + \varepsilon$
M3: $TS_{CVNS,FRE} = \beta_0 + + \sum_l \beta_l(LBG) + \beta_1 YSE + \varepsilon$

Table B5

Average Model (CV NS FRE – With Outliers)

Models	AICc	$\Delta AICc$	*Age*	*Gender* [b]	*LBG* [c]	*YSE* [d]
M0	68.0	0.00	–	–	–	[0.041, 0.288]
M1	68.3	0.27	–	[-0.175, 1.208]	–	[0.034, 0.276]
M2	69.1	1.10	[-0.166, 0.040]	–	–	[0.021, 0.272]
M3	69.3	1.33	–	–	[-1.197, 0.330]	[0.001, 0.269]
Model Averaging [a]			-0.017	0.199	-0.151	0.121

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Note. 95% confidence interval for coefficients (in square brackets).

[a] Coefficients from a fully averaged model. Only the coefficients (rounded if possible) for the predictors mentioned in the table are provided.
[b] Gender (Male)
[c] Language Background (NNES)
[d] Years of Speaking English

Human Voice Group

Table B6

Model Selection Table (HV NS ESP – With Outliers)

Models	df	Log.Lik	AICc	$\Delta AICc$	w_i	ER_i
M0	3	-23.518	54.2	0.00	0.160	1
M1	2	-24.877	54.3	0.12	0.150	1.07
M2	3	-23.991	55.1	0.95	0.100	1.6
M3	5	-21.284	55.7	1.55	0.074	2.16
M4	4	-22.965	55.9	1.75	0.067	2.39

Models:
M0: $TS_{HVNS,ESP} = \beta_0 + \beta_1 Age + \varepsilon$
M1: $TS_{HVNS,ESP} = \beta_0 + \varepsilon$
M2: $TS_{HVNS,ESP} = \beta_0 + \beta_1 YSE + \varepsilon$
M3: $TS_{HVNS,ESP} = \beta_0 + \beta_1 Age + \sum_e \beta_e (F.ESP) + \varepsilon$
M4: $TS_{HVNS,ESP} = \beta_0 + \beta_1 Age + \beta_2 YSE + \varepsilon$

Table B7

Average Model (HV NS ESP – With Outliers)

Models	AICc	$\Delta AICc$	*Age*	*YSE* [b]	*F.ESP* [c]
M0	54.2	0.00	[-0.019, 0.159]	–	–
M1	54.3	0.12	–	–	–
M2	55.1	0.95	–	[-0.017, 0.074]	–
M3	55.7	1.55	[-0.002, 0.174]	–	[-1.371, 1.527] [d], [-1.920, 0.770] [e]
M4	55.9	1.75	[-0.031, 0.152]	[-0.024, 0.068]	–
Model Averaging [a]			0.036	0.009	0.032 [d] -0.097 [e]

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Note. 95% confidence interval for coefficients (in square brackets).

[a] Coefficients from a fully averaged model. Only the coefficients (rounded if possible) for the predictors mentioned in the table are provided.
[b] Years of Speaking English
[c] Familiarity with Spanish
[d] Familiarity with Spanish (Poor)
[e] Familiarity with Spanish (No Experience)

Table B8

Model Selection Table (HV NS ESP – Without Outliers)

Models	df	Log.Lik	AICc	$\Delta AICc$	w_i	ER_i
M0	5	-10.152	33.8	0.00	0.266	1
M1	6	-8.316	33.9	0.05	0.260	1.02
M2	6	-9.167	35.6	1.75	0.111	2.40

Models:
M0: $TS_{HVNS,ESP} = \beta_0 + \beta_1 Age + \sum_e \beta_e (F.ESP) + \varepsilon$
M1: $TS_{HVNS,ESP} = \beta_0 + \beta_1 Age + \sum_e \beta_e (F.ESP) + \sum_l \beta_l (LBG) + \varepsilon$
M2: $TS_{HVNS,ESP} = \beta_0 + \beta_1 Age + \sum_e \beta_e (F.ESP) + \sum_g \beta_g (GEN) + \varepsilon$

Table B9

Average Model (HV NS ESP – Without Outliers)

Models	AICc	$\Delta AICc$	*Age*	*GEN* [b]	*LBG* [c]	*F.ESP* [d]
M0	33.8	0.00	[0.033, 0.156]	-	–	[-0.916, 1.037] [e] [-1.521, 0.306] [f]
M1	33.9	0.05	[0.030, 0.148]	-	[-0.058, 0.670]	[-0.938, 0.928] [e] [-1.647, 0.130] [f]
M2	35.6	1.75	[0.032, 0.154]	[-0.152, 0.614]	–	[-0.806, 1.163] [e] [-1.439, 0.386] [f]
Model Averaging [a]			0.082 *	0.045	0.116	0.053 [e] -0.528 [f]

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Note. 95% confidence interval for coefficients (in square brackets).

[a] Coefficients from a fully averaged model. Only the coefficients (rounded if possible) for the predictors mentioned in the table are provided.
[b] Gender (Male)
[c] Language Background (NNES)
[d] Familiarity with Spanish.
[e] Familiarity with Spanish (Poor)
[f] Familiarity with Spanish (No Experience)

Table B10

Model Selection Table (HV AS ESP – Without Outlier)

Models	df	Log.Lik	AICc	$\Delta AICc$	w_i	ER_i
M0	3	-38.175	83.6	0.00	0.217	1
M1	2	-39.569	83.7	0.16	0.200	1.09

Models:
M0: $TS_{HVAS,ESP} = \beta_0 + \sum_l \beta_l(LBG) + \varepsilon$
M1: $TS_{HVAS,ESP} = \beta_0 + \varepsilon$

Table B11

Average Model (HV AS ESP – Without Outlier)

Models	AICc	$\Delta AICc$	LBG [b]
M0	83.6	0.00	[-2.124, 0.060]
M1	83.7	0.16	–
Model Averaging [a]			-0.378

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Note. 95% confidence interval for coefficients (in square brackets).

[a] Coefficients from a fully averaged model. Only the coefficients (rounded if possible) for the predictors mentioned in the table are provided.
[b] Language Background (NNES)

Table B12

Model Selection Table (HV NS ARB – Without Outliers)

Models	df	Log.Lik	AICc	$\Delta AICc$	w_i	ER_i
M0	2	-17.841	40.3	0.00	0.218	1
M1	3	-16.959	41.2	0.90	0.139	1.57
M2	3	-17.052	41.4	1.09	0.126	1.73

Models:
M0: $TS_{HVNS,ARB} = \beta_0 + \varepsilon$
M1: $TS_{HVNS,ARB} = \beta_0 + \sum_g \beta_g(GEN) + \varepsilon$
M2: $TS_{HVNS,ARB} = \beta_0 + \sum_l \beta_l(LBG) + \varepsilon$

Table B13

Average Model (HV NS ARB – Without Outliers)

Models	AICc	$\Delta AICc$	LBG [b]	GEN [c]
M0	40.3	0.00	-	-
M1	41.2	0.90	-	[-0.809, 0.189]
M2	41.4	1.09	[-0.732, 0.191]	-
Model Averaging [a]			-0.091	-0.091

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Note. 95% confidence interval for coefficients (in square brackets).

[b] Language Background (NNES)
[c] Gender (Male)

Table B14

Model Selection Table (HV NS FRE – Without Outliers)

Models	df	Log.Lik	AICc	$\Delta AICc$	w_i	ER_i
M0	5	–13.371	40.3	0.00	0.192	1
M1	6	–11.539	40.3	0.06	0.187	1.03
M2	6	–12.033	41.3	1.05	0.114	1.69
M3	3	–17.234	41.7	1.46	0.093	2.08

Models:
M0: $TS_{HVNS,FRE} = \beta_0 + \sum_f \beta_f (F.FRE) + \varepsilon$
M1: $TS_{HVNS,FRE} = \beta_0 + \sum_f \beta_f (F.FRE) + \sum_g \beta_g (GEN) + \varepsilon$
M2: $TS_{HVNS,FRE} = \beta_0 + \sum_f \beta_f (F.FRE) + \sum_l \beta_l (LBG) + \varepsilon$
M3: $TS_{HVNS,FRE} = \beta_0 + \sum_l \beta_l (LBG) + \varepsilon$

Table B15

Average Model (HV NS FRE – Without Outliers)

Models	AICc	$\Delta AICc$	*F.FRE* [b]	*GEN* [c]	*LBG* [d]
M0	40.3	0.00	[–1.078, 0.378] [e] [–1.543, –0.407] [f] [–1.041, 0.126] [g]	–	–
M1	40.3	0.06	[–1.094, 0.297] [e] [–1.585, –0.492] [f] [–1.045, 0.069] [g]	[–0.799, 0.070]	–
M2	41.3	1.05	[–1.020, 0.401] [e] [–1.427, –0.263] [f] [–0.933, 0.246] [g]	–	[–0.737, 0.125]
M3	41.7	1.46	–	–	[–0.959, –0.029]
Model Averaging [a]			–0.268 [e] –0.716 [f] –0.322 [g]	–0.113	–0.146

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Note. 95% confidence interval for coefficients (in square brackets).

[a] Coefficients from a fully averaged model. Only the coefficients (rounded if possible) for the predictors mentioned in the table are provided.
[b] Familiarity with French

[c] Gender (Male)
[d] Language Background (NNES)
[e] Familiarity with French (Moderate)
[f] Familiarity with French (Poor)
[g] Familiarity with French (No Experience)

Appendix C

Arrangement of Script Types (NS, AS) and Accents in HV and CV

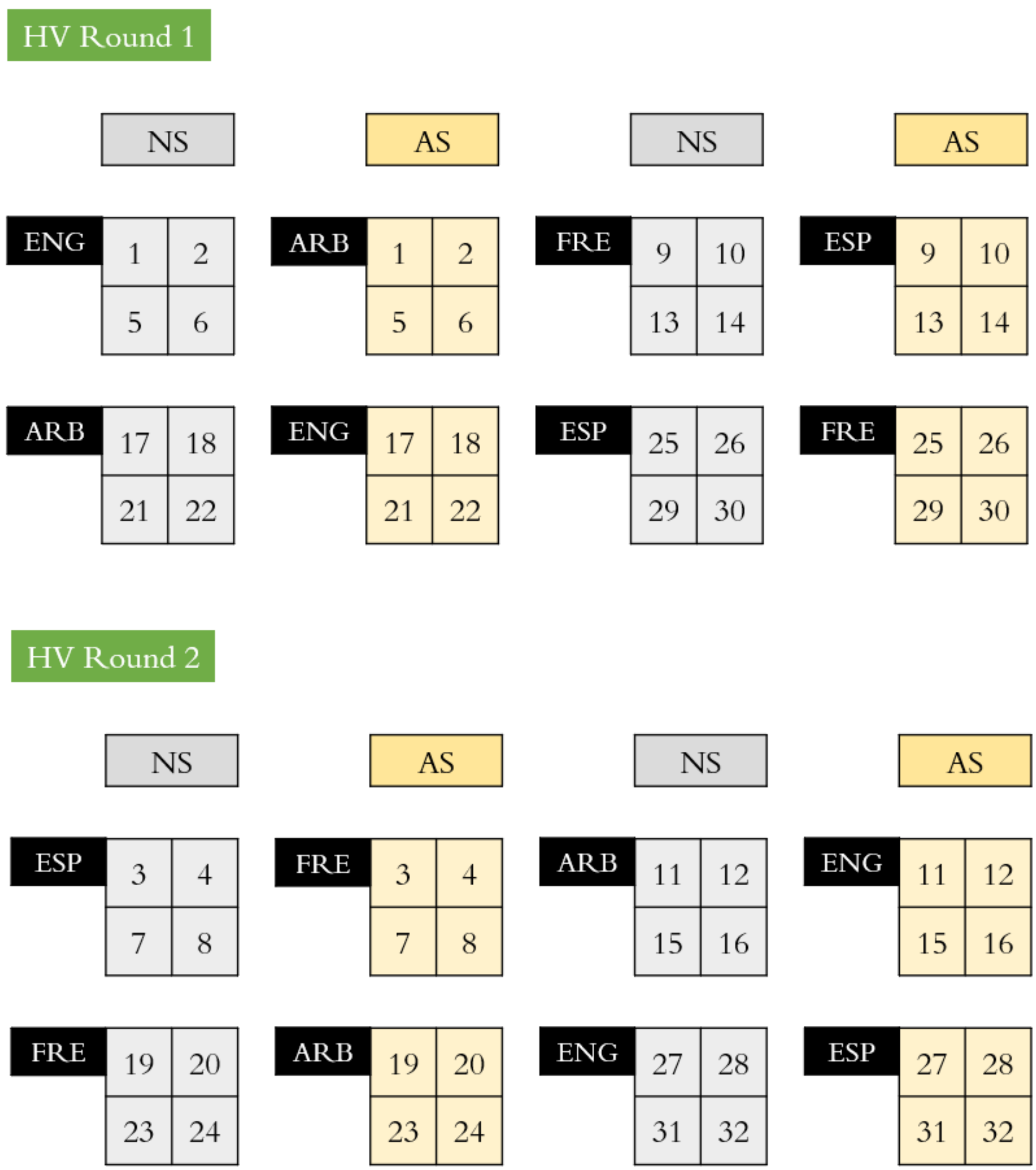

Figure C1. The arrangement of foreign accents (black boxes) and script types (gray and yellow boxes above; NS, AS) in the Human Voice Group Round 1 and 2. Only the human voices were played in this group.

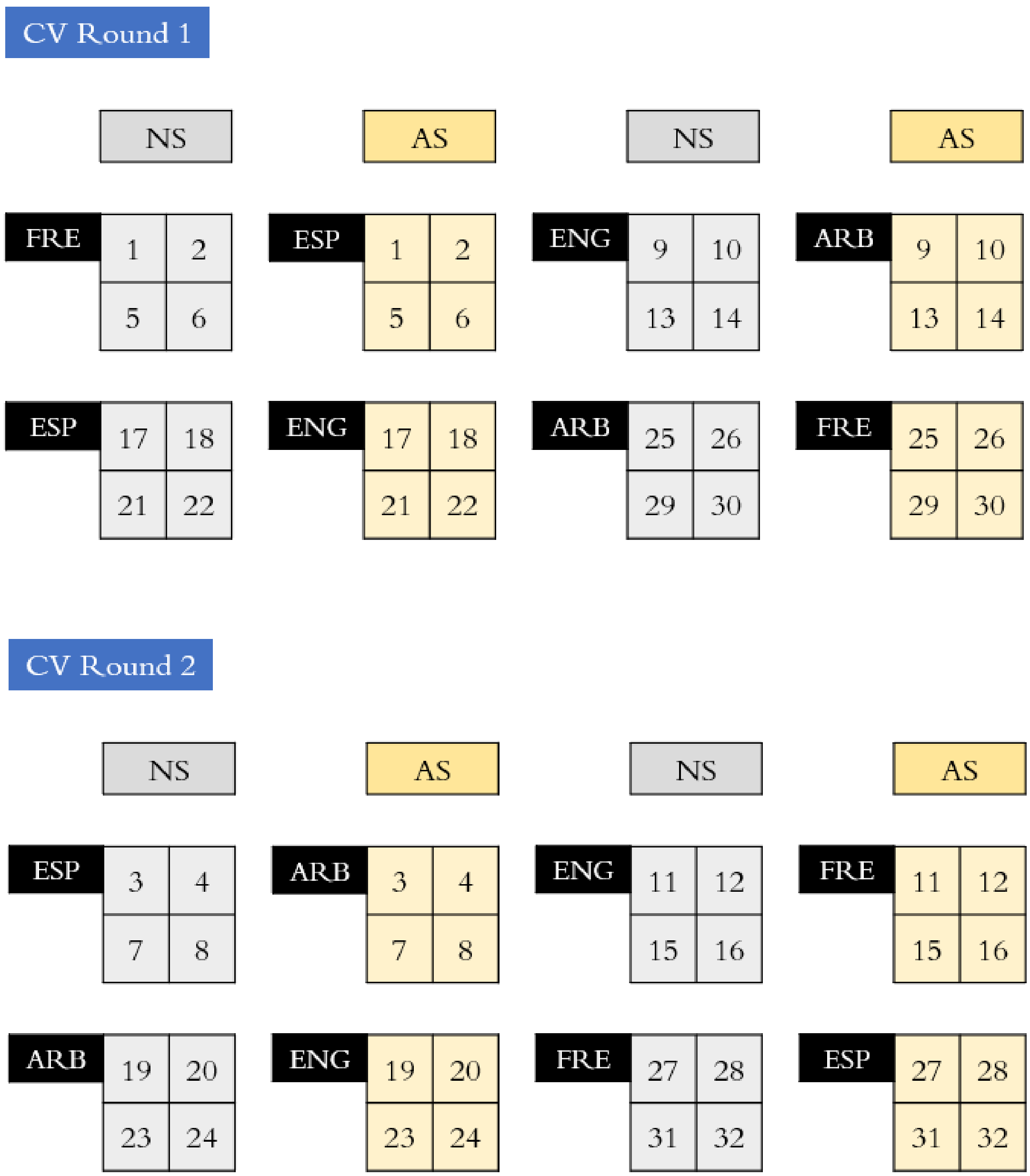

Figure C2. The arrangement of foreign accents (black boxes) and script types (gray and yellow boxes; NS, AS) in the Computer Voice Group Round 1 and 2. Only the computer voices (TTS) were played in this group.

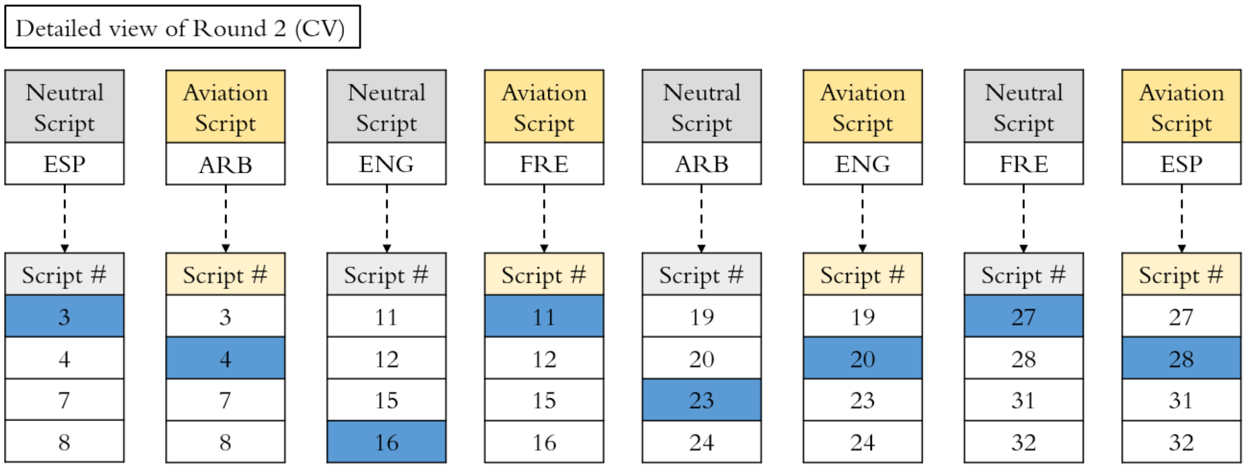

Figure C3. Sample scripts randomly chosen by Qualtrics to play for each participant in CV Round 2.

Due to a lack of specific function in Qualtrics that allowed arranging the script types and foreign accents in the format the author preferred, the author created a Python program that randomized the arrangement of accents for each test in the Voice Groups. The resulting arrangement (**Figure C1** and **Figure C2**) was then programmed into the tests on the platform. The reason for this step was because Qualtrics would sometimes show two foreign accents clustering together in succession (ex. French accent for NS and the same accent for AS). As can be seen in both Figure C1 and C2, in the arrangement of accent for CV Round 2 (ESP – ARB – ENG – FRE – ARB – ENG – FRE – ESP), there was no single accent shown twice in succession.

Qualtrics chose a random script in each block that consisted of four scripts. This overall format prevented the use of the same scripts in Round 1 and 2. The Script Types also alternated in the following fashion: NS – AS – NS – AS – NS - AS – NS - AS.

The highlighted cells in the **Figure C3** shows the random selection of the scripts to play by Qualtrics. For each participant who completed a test, the author made sure to program in the function that all scripts have equally likely chance to be selected.

Appendix D

Preliminary (or Demographic) Survey

1. Type your ID:

2. Age:

3. Gender: [ ] Male [ ] Female [ ] Non-binary/third-gender [ ] Prefer not to say [ ] Prefer to self-identify

4. Place of Birth (Country):

5. Field of Study and Year:

6. Occupation:

7. What is your flight rating or certificate?

Certificate	Rating (if applicable)
None	
Student Pilot (SP)	
Private Pilot (PPL)	
Commercial Pilot (CPL)	

8. How long (in **hours**) have you been flying?

9. How long (in **years**) have you been speaking English?

10. How often do you speak English at home?
    A. Always (100%; I do not speak my native language at home)
    B. Often (75%; I speak both, but I speak English more)
    C. Neutral (50%; I speak half and half)
    D. Seldom (25%; I speak both, but I speak my native language more)
    E. Never (0%; I do not speak English at home)

11. What language background do you come from? (If non-native English speaker, please specify what is your native language)
    A. Non-native English speaker ( )
    B. Native English speaker

12. What other languages do you speak other than your native language?

13. For Spanish, how would you rate your fluency? (Circle below)

1 – No experience 2 – Poor 3 – Moderate 4 – Good 5 – Excellent

14. For Arabic, how would you rate your fluency? (Circle below)

1 – No experience 2 – Poor 3 – Moderate 4 – Good 5 – Excellent

15. For French, how would you rate your fluency? (Circle below)

1 – No experience 2 – Poor 3 – Moderate 4 – Good 5 – Excellent

16. How would you rate your familiarity with Aviation English?

1 – No experience 2 – Poor 3 – Moderate 4 – Good 5 – Excellent

Appendix E

Aviation Scripts (Human & Computer Voice)

Author: Hyun Su (Winfred) Seong

Introduction

All airline names and the contents of the scripts are fictional and unique. Scripts are inspired by examples online or actual live recordings, VFR and IFR phraseology from NAV Canada, and Airplane Flying Handbook from Federal Aviation Administration. Geographic names such as a city or town names (ex. Montreal Centre) are used and runways are also included respectively. There are a mix of VFR and IFR scenarios. The aircraft callsign is fictional but may contain various homophones that have similar phonetics.

Scripts contain a mix of standard phraseology and plain English. Some contain speech disfluencies, (ex. "uhh"), repeats (ex. "it seems like… it seems like"), pauses (ex. "…") as they are found in real-life conversation.

Set Up

The first table below is shown with each script. It is a background information table. It is only used by the author to remember which script belonged to which phase of the flight and location. It is not shown to participants.

Phase		Location	

- **Phase** refers to the phases of flight.
- **Location** is the airport (usually an ATS unit) in question.
- Anything else not mentioned in the table above will be used for making M/C questions to be answered by the participants. They are but not limited to:

Flight designator & number (callsign)	Weather Information (PIREP, METAR, SPECI)	Runway Conditions, Number, and Direction
Passenger Information	Causes of Emergency	Frequency
Flight Crew Information	Read back Procedure	IPA Phonetics
Altitude, Vector, Airspeed	Transponder	Flight Procedure (recovery, approach, landing, takeoff etc)

Next, the table in which a script is written is structured as follows:

ID	Gen.	Script #
ATC	OF	
Pilot	YM	
ATC	OF	
Question Cues		

- **ID** refers to the two roles: Pilot or ATC.
- **Gen.** refers to the gender pairs, which are equally represented in the scripts (total: 32)
    - YM, OM (young male, old male) x 8

    - YF, OF (young female, old female) x 8
    - YF, OM (young female, old male) x 8
    - YM, OF (young male, old female) x 8
- **Script#** enumerates the specific number of the script.

References used to write the scripts

- VFR Phraseology (NAV Canada)
- IFR Phraseology (NAV Canada)
- LiveATC.net and conversation videos on YouTube
- Airplane Flying Handbook (FAA)
- Canadian Aviation Regulations (laws-lois.justice.gc.ca)
- TC AIM (2021)
- FAA AIM (U.S.)
- FAA Pilot Glossary
- Skybrary
- EASA (EU)
- RadioReference.com (for some airport frequencies; ex. Toronto)
- Airport Diagrams (for finding runways)
- Botch & Johnson (2008). Alcohol-Related Aviation Accidents Involving Pilots With Previous Alcohol Offenses. FAA

Acknowledgement

I would like to sincerely thank Brad Moncion for helpful suggestions on the scripts and questions.

For Human Voices

Script 1 – Animal on Runway

Phase	Departure	Location	TPA Tower

ID	Gen.	Script 1
ATC	OF	Albatross one-two-three, wind two-six-zero at ten, runway two-eight, cleared for takeoff.
Pilot	YM	Unable, there is an animal crossing the runway. It seems like… it seems like a fox or something? Don't know what it is… uhh… and it just stopped moving, embedded on the runway.
ATC	OF	Albatross one-two-three, roger. Take-off clearance cancelled. Hold position. We will send the wildlife team up there.

Script 2 – Mayday (Fire on Engine)

Phase	Cruise	Location	YQB Terminal

ID	Gen.	Script 2
Pilot	YF	MAYDAY MAYDAY MAYDAY, Lucan Air two-three-four-five, left engine fire. Unable to maintain one-one-thousand. Descending to eight thousand.
ATC	OM	Lucan Air two-three-four-five, Quebec Terminal, radar identified. Roger Mayday. Descend eight thousand. When able, state intentions.
Pilot	YF	Lucan Air two-three-four-five, roger.

Script 3 – Pilot Confused Instruction (Frequency)

Phase	Departure	Location	YYZ North Tower

ID	Gen.	Script 3
ATC	OF	Parthenon niner-two-seven-five, runway one-five right, taxi Juliet, contact Departure one-two-seven-five-seven, holding short.
Pilot	YF	Roger. Runway one-five right, taxi Juliet, contact Tower one-one-eight-three-five as usual, holding short, Parthenon niner-two-seven-five.

ATC	OF	Parthenon niner-two-seven-five, negative. Contact Tower one-two-seven-five-seven.

Script 4 – Callsign Confusion

Phase	Cruise	Location	YYZ Tower

ID	Gen.	Script 4
Pilot	YM	Toronto Tower, Chrome one-niner-two-seven would like to cancel flight following.
ATC	OM	Uhh… Coulomb… one-niner-seven-two, roger. There are other similar sounding aircraft approaching the airport in all directions. Squawk VFR, frequency change approved.
Pilot	YM	Negative. That's not us. We are Chrome one-niner-two-seven.

Script 5 – Bird and Taxiing

Phase	Ground (Taxi in)	Location	YGK Ground

ID	Gen.	Script 5
ATC	YM	Sutton one-three-two-five, Kingston Ground, hold short runway one-niner on Bravo. A herd of bison moving and impeding the taxiway.
Pilot	OF	YGK Ground, Sutton one-three-two-five, requesting progressive taxi to apron five.
ATC	YM	Sutton one-three-two-five, negative. Wait until the birds are gone and… wait until further advised.

Script 6 – Pan-Pan (Altimeter Lost)

Phase	Departure	Location	TPA Tower

ID	Gen.	Script 6
Pilot	OM	PAN-PAN, PAN-PAN, PAN-PAN. IPT. We have lost the altitude meter. We would like to request immediate landing on runway two-five.
ATC	YF	IPD, Tampa Bay Tower, wind two-seven-zero at thirteen gusting nineteen, you are cleared to land on runway two-five.
Pilot	OM	Cleared to land two-five, IPT.

Script 7 – Noticing Another Accident (Fire)

Phase	Approach	Location	YYZ Terminal

ID	Gen.	Script 7
ATC	YF	Rigel one-one-niner-two, Toronto Terminal, orbit north of Centennial Park Conservatory, right turns, continue orbiting until we provide further instructions.
Pilot	OF	Rigel one-one-niner-two, right turns, north of Centennial Park Conservatory. Continue orbiting. May I ask why?
ATC	YF	Fire on the runway. We're having fire trucks roll out now. Never thought that this would occur during my shift. Well, at least, nobody's hurt so that's good!

Script 8 – Detailed METAR

Phase	Arrival	Location	LHR Director

ID	Gen.	Script 8
Pilot	OM	Heathrow Director, Northwest one-nine-two, I would like to divert to another airport. Request METAR.
ATC	YM	Heathrow Director, read you two by three, background static. Heathrow METAR at 0630 Zulu, wind 120 at 26 knots, visibility less than 3 statute miles, light drizzle, ceiling six thousand two hundred broken, towering cumulus, temperature 21, dewpoint 19, QNH 1025 hectopascals.
Pilot	OM	Roger, Heathrow Director, Northwest one-nine-two.

Script 9 – Too Many Scheduled Arrivals

Phase	Approach	Location	YHZ Terminal

ID	Gen.	Script 9
ATC	OF	Aqua Heavy, Halifax Terminal. Turn left heading to two-five-zero, cleared the visual approach.
Pilot	YM	Aqua Heavy, heading two-five-zero, roger.

ATC	OF	Aqua Heavy, you need to make that turn immediately because others in the queue are requesting descent but can't due to you and other arrivals occupying the lower altitudes.

Script 10 – Wind

Phase	Departure	Location	YYZ Tower

ID	Gen.	Script 10
Pilot	YF	Toronto Tower, Marianas zero-seven-three, holding short of runway one-five left on victor. Waiting for takeoff clearance.
ATC	OM	Marianas, wind one-one-six at seven, variable between zero-six-zero varying one-two-five. Runway one-five left, cleared for takeoff.
Pilot	YF	Runway three-three right, cleared for takeoff, Marianas zero-seven-three.

Script 11 – Extreme Weather (Dust)

Phase	Arrival	Location	YEG Tower

ID	Gen.	Script 11
ATC	OF	Rain Air one-four-eight, Edmonton Tower, runway three-zero, wind one-three-four at five knots, altimeter setting three-zero-two-niner, widespread dust moving southeast. Cleared downwind, report final.
Pilot	YF	Rain Air one-four-eight, cleared downwind, report final. Uhh… sorry I couldn't quite catch the weather info. Could you repeat again?
ATC	OF	Rain Air, no problem. Dust moving southeast. Expect stronger and abrupt headwind when entering base leg.

Script 12 – Weather and Traffic (Mayday)

Phase	Cruise	Location	YFD Tower

ID	Gen.	Script 12
Pilot	YM	MAYDAY MAYDAY MAYDAY, Brantford Tower, Octopus two-three-six-eight. We lost weather radar and don't have visual on weather and looking for traffic to change the altitude.

ATC	OM	Octopus two-three-six-eight, Brantford Tower, roger. Wind three-zero-five at fifteen. Thunderstorm at two o'clock. A commuter liner at two thousand feet, southbound 737, three thousand feet.
Pilot	YM	Uhh… thunderstorm at two o'clock and… traffic not in sight, but we'll let you know when we see it, Octopus two-three-six-eight.

Script 13 – Unnecessary Greetings

Phase	Ground (Taxi in)	Location	DFW Arrival

ID	Gen.	Script 13
ATC	YM	Golf Mike Whiskey, greetings and welcome to Dallas! And that's the new 777X isn't it? What a beautiful aircraft! Anyhow… next available left, contact ground on two-one-decimal-six-five.
Pilot	OF	Thanks, it's been a long flight. Uhh… can you repeat frequency again? Sorry, there is a background static noise on the radio and I can't hear you very well.
ATC	YM	Correction. Contact Ground on two-one-decimal-eight-five.

Script 14 – Redirect due to Extreme Weather

Phase	Cruise	Location	YHM Tower

ID	Gen.	Script 14
Pilot	OM	Hamilton Tower, Vein one-three-two-eight, I would like to update my flight plan and divert to another airport.
ATC	YF	Vein one-three-two-eight, radar contact, roger. Just a quick question: How much fuel is left and where do you want to go?
Pilot	OM	Ahh… about forty-five minutes. Looking out the window I see ice pellets nearby Hamilton. Request vectors to YYZ and weather over enroute.

Script 15 – Blizzard and the Pilot

Phase	Ground (Taxi in)	Location	YXU Tower

ID	Gen.	Script 15
ATC	YF	Flour Air three-one-niner-four, London Tower, exit on Charlie, taxi via India and hold short of Bravo. Contact Ground two-one-decimal-niner when off.

Pilot	OF	Two-one-decimal-niner when off, Flour Air three-one-niner-four.
ATC	YF	Flour Air three-one-niner-four. Correction. Exit via Charlie, hold short of India. Low visibility due to blizzard. We just sent the snow removal team to the runway. Contact Ground when the visibility recovers.

Script 16 – Extended METAR Information

Phase	Approach	Location	YSB Terminal

ID	Gen.	Script 16
Pilot	OM	Sudbury Terminal, Cedar Air seven-two-zero-eight, inbound from the north to land. Is runway three-zero available? Also, requesting the latest weather in Sudbury.
ATC	YM	Cedar Air seven-two-zero-eight, stand by… Sudbury weather at 0250 Zulu, wind 270 at twelve gusting eighteen, visibility less than three quarters of a statute miles, intensifying squall with mist, sky overcast, temperature 9, dewpoint 6, altimeter three-zero-zero-nine inches.
Pilot	OM	Cedar Air seven-two-zero-eight, roger.

Script 17 – Contact Ground

Phase	Ground (Taxi in)	Location	IND Tower

ID	Gen.	Script 17
ATC	OF	Chili Air eight-one-two-five Super, Indianapolis Tower, do not exit on Hotel. Exit on Juliet.
Pilot	YM	Got it, Chili Air eighty-one-twenty-five Super.
ATC	OF	Chili Air eight-one-two-five Super, Contact Ground two-one-decimal-eight for taxi.

Script 18 – Landing Gear Malfunction

Phase	Approach	Location	LAX Approach

ID	Gen.	Script 18
Pilot	YF	SoCal Approach, Soul Air two-five-five, we got a landing gear problem and we request time to run the checklist.

ATC	OM	Soul Air two-five-five, roger. Maintain five thousand. Continue downwind. Traffic at three o'clock, six miles, southwest bound.
Pilot	YF	We're looking, Soul Air two-five-five.

Script 19 – Brief Off the Radar Moment

Phase	Descent	Location	STH Tower

ID	Gen.	Script 19
ATC	OF	Caret Air one-two-niner-seven, Strathmore Tower, radar contact lost. Can you recycle transponder?
Pilot	YF	I can't hear you very well… I will do that... Strathmore Tower, this is Caret Air one-two-niner-seven. Confirm radar contact?
ATC	OF	Radar contact confirmed, Caret Air one-two-niner-seven.

Script 20 – FO Incapacitated

Phase	Landing	Location	CDG Tower

ID	Gen.	Script 20
Pilot	YM	Paris Tower, this is Crux Air zero-niner-zero inbound to land from south. My first officer is incapacitated. He is bleeding heavily and… is losing consciousness and murmuring something...
ATC	OM	Crux Air zero-niner-zero, Paris Tower, roger. Cleared to left base two-seven right.
Pilot	YM	Cleared left base runway two-four right, Crux Air zero-niner-zero.

Script 21 – Fire Electrical

Phase	Descent	Location	YZF Tower

ID	Gen.	Script 21
ATC	YM	Pharaoh Air one-two-five-seven, Yellowknife Tower. Just to be clear, is the fire electrical in nature?
Pilot	OF	Yes. I can see the smoke from… from… the cabin area and I believe so. It is coming into the cockpit now… We're running out of time as well as the fuel. Only fifteen minutes, Pharaoh Air one-two-five-seven.

ATC	YM	Pharaoh Air one-two-five-seven, roger. Runway three-two centre is open and no traffic there so you are cleared to land.

Script 22 – Lost Airspeed Indicator

Phase	Departure	Location	ATL Tower

ID	Gen.	Script 22
Pilot	OM	MAYDAY MAYDAY MAYDAY Lapps Air two-four-eight-eight has lost airspeed indicator and we are losing altitude. We're… trying to keep the nose down.
ATC	YF	Lapps Air two-four-eight-eight, Atlanta Tower, I understand that you lost the airspeed indicator and your altitude is falling rapidly?
Pilot	OM	Affirm. We're reducing power.

Script 23 – Emergency with Another Aircraft

Phase	Approach	Location	GRU Tower

ID	Gen.	Script 23
ATC	YF	Pear Air three-six-nine, Sao Paulo Tower. Expect an extended left downwind. There's an emergency inbound and expect forty-five minute delay.
Pilot	OF	Roger, Pear Air three-six-nine.
ATC	YF	Pear Air three-six-nine, correction, expect an hour delay.

Script 24 – Weather Display Lost

Phase	Cruise	Location	YHZ Tower

ID	Gen.	Script 24
Pilot	OM	Halifax Tower, Palate Air six-eight-five. We lost the weather radar and we don't have weather information. Request the latest weather.
ATC	YM	Palate Air six-eight-five, Halifax Tower, roger. Halifax Special Weather Report at 0900 Zulu, wind one-five-zero at eight with peaks up to fifteen. Visibility three-quarters of a mile. Funnel cloud at seven o'clock.
Pilot	OM	Roger, Palate Air six-eight-five.

Script 25 – Weather and Flight Update

Phase	Cruise	Location	PBH Tower

ID	Gen.	Script 25
ATC	OF	Planar Air one-five-niner, Paro Tower. There is a tilting cumulonimbus rapidly developing at two o'clock. Wind gust five-five at forty, variable between twenty and ninety. What's your intention?
Pilot	YM	Planar Air one-five-niner, I would like to update ETA due to weather. ETA thirty minutes.
ATC	OF	Planar Air one-five-niner, roger.

Script 26 – PIREP #1

Phase	Cruise	Location	YKF Tower

ID	Gen.	Script 26
Pilot	YF	Waterloo Radio, Razor Air three-zero-seven with PIREP.
ATC	OM	Razor Air three-zero-seven, go ahead.
Pilot	YF	Razor Air three-zero-seven, seven miles west of the Cambridge VOR, temperature twenty-six degrees, heavy rain, cumulonimbus clouds at one-five thousand feet, high turbulence.

Script 27 – Climb Correction

Phase	Climb	Location	CTK Tower

ID	Gen.	Script 27
ATC	OF	Hill Air one-six-six, Ingersoll Tower. Climb and maintain eight thousand five hundred. Turn left, heading one-four-zero. Traffic below at seven thousand five hundred.
Pilot	YF	Maintain eight thousand, turn left, heading two-four-zero, Hill Air one-six-six.
ATC	OF	Correction. Climb and maintain nine thousand, turn right, heading one-six-zero.

Script 28 – Practice Area

Phase	Descent	Location	YXU Tower

ID	Gen.	Script 28
Pilot	YM	Whale Air eight-two-five, approaching practice area from the west, four thousand feet over Fanshawe Golf Course, planning to operate south of Thames River to the Robins Hill Road.
ATC	OM	Whale Air eight-two-five, London Tower. I couldn't hear the last part. Say again?
Pilot	YM	Whale Air eight-two-five planning to operate south of Thames River to the Robins Hill Road.

Script 29 – Cabin Depressurization

Phase	Cruise	Location	YYZ Centre

ID	Gen.	Script 29
ATC	YM	Beat Air three-seven-three, radar contact at two-two-thousand. Toronto Centre. I understand that you have a medical emergency on board due to cabin depressurization. Cleared for right turn heading one-three-five and descend to ten-thousand immediately.
Pilot	OF	Cleared for right turn… and… and… descend to… ten-thousand, Beat Air three-seven… three.
ATC	YM	Beat Air three-seven-three. Just a quick question: How many souls on board?

Script 30 – Chemical Leak

Phase	Cruise	Location	YQK Tower

ID	Gen.	Script 30
Pilot	OM	MAYDAY MAYDAY MAYDAY Kenora Tower, Yves Air two-four-four. We have an emergency regarding chemical substance leaking in the cabin area. Some passengers are ill. We don't know where it's coming from.
ATC	YF	Yves Air two-four-four, radar contact, roger. The closest airport for diversion is Red Lake. Do you want vectors to that airport?
Pilot	OM	We request vectors to Red Lake and weather as well, Yves Air 244.

Script 31 – Routing Change

Phase	Climb	Location	YUL Tower

ID	Gen.	Script 31
ATC	YF	Aero Axel six-eight-niner, Montreal Tower. I understand that you are ready to change your direct course due to towering cumulonimbus and low visibility, advise ready to copy clearance to Quebec City.
Pilot	OF	Montreal Tower, Aero Axel six-eight-niner is ready to copy.
ATC	YF	Aero Axel six-eight-niner is cleared to the Quebec City Airport via present position direct and maintain flight level two-six-zero.

Script 32 – Normal Operation

Phase	Cruise	Location	JAX Approach

ID	Gen.	Script 32
Pilot	OM	Jacksonville Approach, Basses Air seven-two-five. Inbound from southeast, altitude six thousand three hundred. Requesting to land on runway two-six.
ATC	YM	Basses Air seven-two-five. Cleared for visual approach. Descend and maintain three thousand four hundred until established on final.
Pilot	OM	Roger, Basses Air seven-two-five.

For Computer Voices

Script 1 - Stall

Phase	Cruise	Location	YYZ Tower

ID	Gen.	Script 1
ATC	OF	Beys Air one-two-seven, Toronto Tower, fly heading two-seven-zero, climb six thousand.
Pilot	YM	Beys Air one-two-seven, roger, heading two-seven-zero… STALL! STALL! STALL!
ATC	OF	Beys Air one-two-seven, do you require assistance?

Script 2 – Circling

Phase	Approach	Location	YEG Approach

ID	Gen.	Script 2
Pilot	YF	Edmonton Approach, Air Orange three-zero-four-zero, inbound from southwest.
ATC	OM	Air Orange three-zero-four-zero, reduce speed to minimum approach speed. Cleared to ILS runway one-six approach circling for runway one-two.
Pilot	YF	Roger, reduce speed to one-niner-zero, cleared to ILS runway one-two approach circling for runway two-six, Air Orange three-zero-four-zero.

Script 3 – Follow 747

Phase	Cruise	Location	YXU

ID	Gen.	Script 3
ATC	OF	Mayflower two-three-seven-niner, London Tower, radar contact. Number two for approach. You are following 747 at your four o'clock, four miles at two thousand three hundred feet. Do you have it in sight?
Pilot	YF	Mayflower two-three… 737 what…? Say again? Sorry I am a student pilot who is doing a solo for the first time.
ATC	OF	Correction. You are following 747 at five o'clock, four miles at two thousand five hundred feet. Read back.

Script 4 – Plume of Smoke

Phase	Approach	Location	YHM Approach

ID	Gen.	Script 4
Pilot	YM	Hamilton Approach, Lion Air two-five-seven-six. We spotted a plume of smoke over the runway two-four and the visibility is less than a half mile.
ATC	OM	Lion Air two-five-seven-six, in the event of missed approach, turn right heading two-four-zero and maintain four thousand. Expect vectors for the ILS runway one-two.
Pilot	YM	Turn right heading two-four-zero, maintain four thousand. Expect vectors for runway one-two, roger, Lion Air two-five-seven-six.

Script 5 – Traffic blurred by Cumulonimbus Cloud

Phase	Cruise	Location	YQR Centre

ID	Gen.	Script 5
ATC	YM	Kangaroo Air five-zero-five-eight, Regina Centre, traffic at six o'clock four miles, northbound 737 at nine thousand feet descending.
Pilot	OF	Kangaroo Air five-zero-five-eight in IMC, traffic not in sight due to cumulonimbus clouds in the said direction. We'll alert you if it's in sight.
ATC	YM	Kangaroo Air five-zero-five-eight, roger.

Script 6 – Altitude Deviation during Landing

Phase	Approach	Location	YZF Tower

ID	Gen.	Script 6
Pilot	OM	Yellowknife Tower, Air Plato two-seven-eight-niner, inbound from northwest. Is runway three-four available for landing?
ATC	YF	Air Plato two-seven-eight-niner, confirm maintaining three two thousand four hundred feet?
Pilot	OM	Tower, Air Plato two-seven-eight-niner correcting immediately.

Script 7 – Take-Off (Hold Position)

Phase	Departure	Location	YVR Tower

ID	Gen.	Script 7
ATC	YF	Air Okra one-two-five-three, Vancouver Tower, turn left heading two-seven-zero, cleared for take off runway zero-eight left. Climb to flight level four thousand feet.
Pilot	OF	Cleared for take off runway one-five left, climb to four-thousand, Air Okra one-two-five-three.
ATC	YF	Air Okra one-two-five-three, hold position.

Script 8 – Request Option

Phase	Descent	Location	YYG Tower

ID	Gen.	Script 8
Pilot	OM	Charlottetown Tower, Paprika six-zero-eight-four, downwind two-eight, request the option.
ATC	YM	Paprika zero-six-eight-four, cleared for the option runway two-eight.
Pilot	OM	Roger, cleared for the option runway one-five. Paprika six-zero-eight-four.

Script 9 – Request RVR

Phase	Cruise	Location	YFB Tower

ID	Gen.	Script 9
ATC	OF	Roncesvalles two-four-two-three, Iqaluit Tower. Wind three-one-zero at seventeen, visibility less than three-quarters.
Pilot	YM	Roger. Request RVR for one-six, Roncesvalles two-four-two-three. Over.
ATC	OF	Roncesvalles, Iqaluit Tower, RVR four thousand. Squall in the vicinity.

Script 10 – FO Pregnant on Airplane

Phase	Cruise	Location	TPA Tower

ID	Gen.	Script 10

Pilot	YF	TPA Tower, Great Lakes Air four-six-five-three, the female first officer is pregnant and she is gasping and… showing a sign of vomiting.
ATC	OM	Great Lakes Air four-six five-three, roger. Fly heading zero-six-five and maintain present altitude. Do you require assistance when you land?
Pilot	YF	Fly heading zero-six-three, maintain present altitude. Yes, we would like an ambulance when we land. Great Lakes Air four-six-five-three.

Script 11 – Maintain Critical Distance

Phase	Approach	Location	YZF Terminal

ID	Gen.	Script 11
ATC	OF	Keywest Air three-seven-six-five, Yellowknife Approach, you need to reduce your speed to one-eight-zero knots!
Pilot	YF	Roger, reducing speed to one-eight-zero knots immediately, Keywest Air three-seven-six-five.
ATC	OF	Keywest Air three-seven-six-five, maintain current speed.

Script 12 – Blowing Snow (with altitude block)

Phase	Cruise	Location	YMM Tower

ID	Gen.	Script 12
Pilot	YM	Fort McMurray Tower, Giraffe Air two-five-niner-eight Alpha Heavy, request weather information and altitude block at the airport.
ATC	OM	Giraffe Air two-five-niner Alpha Heavy, at 1417 Zulu, wind 240 at 39 gusting to 43 knots, visibility less than three-quarters of a statute mile, blowing snow, temperature minus 19. Maintain block three thousand two hundred to three thousand five hundred.
Pilot	YM	Maintain block three thousand and to… three thousand five hundred, Giraffe Air two-five-niner-eight Alpha Heavy.

Script 13 – Maintain Visual Separation with Runway Confusion

Phase	Approach	Location	IAD Terminal

ID	Gen.	Script 13

ATC	YM	Langley Air two-four-six-four, Dulles Approach, cleared visual approach runway one-niner center, maintain visual separation from 737 on approach to runway one-niner right.
Pilot	OF	Langley Air two-four-six-four roger, cleared visual approach runway one-niner right, maintaining visual separation from 737 on approach to runway one-niner center.
ATC	YM	Langley Air two-four-six-four, I say again, cleared visual approach runway one niner center and maintain visual separation on one-niner right!

Script 14 – Direction to Next Available Airport

Phase	Cruise	Location	PBI Tower

ID	Gen.	Script 14
Pilot	OM	Palm Beach International, Write Air six-four-seven at three thousand nine hundred feet. We would like to update our flight plan and request vector to Fort Lauderdale.
ATC	YF	Write Air six forty-seven, roger. Cleared to the Pahokee V-O-R. Passing four thousand, cleared direct La Belle V-O-R on course.
Pilot	OM	Cleared to the Pahokee V-O-R one-one-five-four. Passing four thousand, cleared direct La Belle V-O-R on course, Write Air six forty-seven.

Script 15 – Off Course

Phase	Cruise	Location	YVR Center

ID	Gen.	Script 15
ATC	YF	Cubik Air six-zero-two, Vancouver Center, we show you off course, confirm direct Vernon.
Pilot	OF	Kelowna Tower, Cubik Air six-zero-two, requesting left turn heading 150 to avoid a large… towering cloud.
ATC	YF	Cubik Air six-zero-two, left turn approved, report new heading 150 and when able proceed direct Kelowna NDB on course.

Script 16 – PIREP #2

Phase	Cruise	Location	YBR Centre

ID	Gen.	Script 16
Pilot	OM	Brandon Centre, Coral Air two-three-six-niner, I would like to report weather over Nelson.
ATC	YM	Coral Air two-three-six-niner, go ahead.
Pilot	OM	A large nimbostratus with showers and thunderstorm at twelve o'clock. A long line of altostratus connected to the top of the nimbostratus at eight thousand. Turbulence is severe. Will provide updates to my ETA.

Script 17 – Alcohol Abuse

Phase	Ground (before departure)	Location	YYZ Ground

ID	Gen.	Script 17
ATC	OF	Air Toronto one-four-two-six, Toronto Ground, cross runway one-five left, continue Romeo, left on Bravo, and hold short of Hotel.
Pilot	YM	Toronto Ground, we would like to report an incident with the… captain. I can smell alcohol from his breath and he keeps dosing off so we need to return to the terminal immediately.
ATC	OF	Air Toronto, roger. I will taxi you back to the apron now. Hold up for just a second… Turn right on Hotel, right on Alpha, keep going straight and enter Alpha Juliet for Terminal 3 at your discretion.

Script 18 – Mayday (Engine Failure)

Phase	Climb	Location	SFO Tower

ID	Gen.	Script 18
Pilot	YF	Ocean Air twenty sixty-nine, we need to return immediately to San Francisco due to engine failure.
ATC	OM	Ocean Air twenty sixty-nine, SFO Tower, roger. VMC. Make a left base and you are cleared for visual approach runway two-eight right.
Pilot	YF	Affirmative.

Script 19 – Midair Collision

Phase	Cruise (in emergency)	Location	YQB Tower

ID	Gen.	Script 19
ATC	OF	Bowl Air six twenty, radar contact lost, recycle transponder. We can't see you on the radar… no altitude, heading, and airspeed.
Pilot	YF	MAYDAY MAYDAY MAYDAY Bowl Air six twenty. We collided with another aircraft and we lost the fuel gauge indicators for both engines. Continuing to descend three thousand… five hundred feet.
ATC	OF	Bowl Air six twenty, we have crews coming, cleared to land on runway zero-six.

Script 20 – Tail Strike

Phase	Climb	Location	YXU Tower

ID	Gen.	Script 20
Pilot	YM	London Tower, Coconut Air four-six-one-five, climbing twelve thousand, we felt something bumping in the tail area when we were taking off and now there's a mechanical problem.
ATC	OM	Coconut air four six-one-five, maintain four thousand five hundred, what's the nature of the emergency?
Pilot	YM	MAYDAY MAYDAY MAYDAY Coconut Air four-six-one-five, we have a fire on the tail and the normal brakes won't work… Trying to fly as stable as possible but it keeps turning left! We'll need a long runway and have the trucks roll as well.

Script 21 – Immediate Departure

Phase	Departure	Location	BFS Ground

ID	Gen.	Script 21
ATC	YM	Rhode Island Air two-two-five-seven, Belfast Ground, are you able an immediate departure?
Pilot	OF	Affirmative.
ATC	YM	Roger. Winds at Belfast are two-six-zero at eleven (knots), sky clear. Cleared for immediate takeoff runway three-five. Climb and maintain four thousand five hundred. Leaving four thousand five hundred, fly heading two-six-zero.

Script 22 – Towering Cumulonimbus before Departure

Phase	Departure	Location	YQT Ground

ID	Gen.	Script 22
Pilot	OM	Thunderbay Ground, Blue Lagoon six-zero-two-four, expediting to runway two-five, taxiing via Bravo.
ATC	YF	Blue Lagoon six-oh-two-four, hold short of the runway. METAR at 0700 Zulu, wind three-zero-zero at twenty four (knots), heavy rain, overcast, towering cumulonimbus, visibility less than three quarters of a statute mile to the northwest.
Pilot	OM	Roger, Blue Lagoon six-zero-two-four.

Script 23 – Student Pilot's Confusion

Phase	Approach	Location	YGK Tower

ID	Gen.	Script 23
ATC	YF	Black Forest one-five-eight-three, Kingston Tower, turn right heading one-six-five, intercept the localizer, descend and maintain two thousand five hundred now.
Pilot	OF	Turn right heading one-six-five, uhh… I am not sure what's going on but… I was cleared to descend to one thousand twenty by another ATC before you. I am confused… and I am a student pilot.
ATC	YF	Black Forest Jet one-five-eight-three, you need to comply with my instruction now because the airport is becoming congested.

Script 24 – Student Solo with Altitude Problem

Phase	Cruise (Emergency)	Location	YZR Tower

ID	Gen.	Script 24
Pilot	OM	Nimbus two-five-seven-zero Alpha, engine failure. Altitude decreasing significantly, and I am a student solo. I am being pushed by strong gust. I need help please.
ATC	YM	Nimbus two-five-seven-zero, Sarnia Tower, squawk ident if able. The help is on the way. Just a quick question: How many souls onboard?
Pilot	OM	I need to land right now… Altitude five hundred decreasing rapidly. I am going to have to land it on an open rice paddy field, eleven o'clock from my position. Can't talk much longer!

Script 25 – Confusing Approach Instruction

Phase	Approach	Location	YUL Tower

ID	Gen.	Script 25
ATC	OF	Pineapple Air one-niner-two-three, Montreal Tower, make a left base turn and descend to one thousand eight hundred now. Cleared visual approach for runway 24 Right. Wind one-eight-zero at thirteen knots, gusting to twenty-four.
Pilot	YM	Make a left base turn and descend one-eight… Sorry I was not able to copy the descending instruction. Speak slower. Pineapple Air one-niner-two-three.
ATC	OF	Make a left base turn and descend to one thousand eight hundred.

Script 26 – Unfamiliar Area

Phase	Cruise	Location	YYB Tower

ID	Gen.	Script 26
Pilot	YF	North Bay Tower, Burrow Air six-two-one Heavy, request traffic information over Ashburton. Unfamiliar with the area.
ATC	OM	Burrow Air six-two-one Heavy, you have traffic… 737 three o'clock six miles at two thousand, and 747 at seven o'clock ten miles at three thousand feet.
Pilot	YF	737 at three thousand and 747 at two thousand, roger, Burrow Air six-two-one Heavy.

Script 27 – Verify Enroute Traffic

Phase	Cruise	Location	YXU Tower

ID	Gen.	Script 27
ATC	OF	Scent Air two-one-six-four Heavy, London Tower, verify your position.
Pilot	YF	Enroute traffic, Scent Air two-one-six-four Heavy, overhead Lucan, two-thousand-nine-hundred feet, direct to London. Request heading two-eight-five.
ATC	OF	Request approved. Heading two-eight-five.

Script 28 – Low on Oxygen

Phase	Cruise (nearing approach)	Location	YYZ Approach

ID	Gen.	Script 28
Pilot	YM	Toronto Approach, this is Doughs Air four-two-niner-one with information Papa, inbound from south to land. Uhh… we are running low on oxygen and I'm wearing one.. I need… to find somewhere to land.
ATC	OM	Doughs Air four-two-niner-one I need you to descend to nine-thousand feet. Fly heading two-five-zero.
Pilot	YM	Yes, yes… Doughs Air four… uhh… two… niner-one turning right and descend to uhh …two-nine thousand feet.

Script 29 – Unruly Passenger

Phase	Approach	Location	YYZ Terminal

ID	Gen.	Script 29
ATC	YM	Leopard Air two-zero-seven-niner, Toronto Terminal, radar contact seven miles southeast of the airport. I understand that you are declaring an emergency regarding unruly passenger?
Pilot	OF	Affirm, he also had a pocketknife in his right hand. Captain is stabbed in the forehead, right leg, and left hand. We need clearance to land immediately.
ATC	YM	Leopard Air two-zero-seven-niner, you are cleared left base runway two-three. We'll have police and the ambulance ready for you on the runway.

Script 30 – Airplane Crash Site

Phase	Approach	Location	YVR Tower

ID	Gen.	Script 30
Pilot	OM	Ahhhh… Vancouver Tower, this is Libre Air two forty-nine, I just spotted an airplane crashing on the highway ninety-nine and there are two civilian cars turned upside down.

ATC	YF	Libre Air two forty-nine, roger. I have dispatched a helicopter, officers, medical team, and firefighters on the way. Remain outside the control zone until further advised.
Pilot	OM	Libre Air six twenty-five, wilco.

Script 31 – Speak Slower

Phase	Approach	Location	YYZ Approach

ID	Gen.	Script 31
ATC	YF	Lace Air two-six-seven-three, Toronto Approach, emergency inbound runway three-three left. Now expect runway three-three right. Maintain four thousand two hundred.
Pilot	OF	Toronto Approach, I missed the last part. Speak slower, Lace Air two-six-seven-three.
ATC	YF	Lace Air two-six-seven-three, maintain four thousand two hundred until established, cleared visual approach runway three-three right.

Script 32 – Passenger with Cardiac Arrest

Phase	Cruise	Location	YEG Centre

ID	Gen.	Script 32
Pilot	OM	Edmonton Centre, Vein Air triple three, inbound from north. I have a passenger onboard suffering from cardiac arrest. Request latest winds and ceiling in Edmonton?
ATC	YM	Vein three-three-three, roger. Winds at Edmonton, 1200 Zulu, two-five-zero at five knots, scattered at 10000 feet. ILS runway one-two in use.
Pilot	OM	Roger. We would like to divert to Edmonton, requesting vectors for the ILS runway one-two. Also requesting EMS on landing. Vein Air three-three-three.

Appendix F

Neutral Scripts & Questions (Human & Computer Voice)

Author: Hyun Su (Winfred) Seong

Introduction

This is a list of semantically anomalous (context-free or neutral) scripts. It uses Harvard Sentences, which contains a total of 720 sentences, divided into Lists (ex. 1,2,3… 72). Each List contains ten sentences. For this thesis, the author has chosen List 1 to 20.

Each sentence in the range of the List has been selected in the following manner:

1. Pick the sentence #1 from List 1, #2 from List 2, #3 from List 3.
2. Put those sentences into a table.
3. Attach a gender pair.

Fill-in-the-blanks Questions

This is the type of the questions a study participant will encounter. Each question will have four blanks to be filled by the participant (0.25 mark for each blank).

Human Voice

Gen.	Script 1
OF	The birch canoe slid on the smooth planks.
YM	The boy was there when the sun rose.
OF	The small pup gnawed a hole in the sock.
Question Cues	▪ The [?] [?] slid on the [?] [?]. ▪ The [?] was [?] the [?] [?]. ▪ The small [?] [?] a [?] in the [?].

Gen.	Script 2
YF	Hoist the load to your left shoulder.
OM	A king ruled the state in the early days.
YF	The frosty air passed through the coat.
Question Cues	▪ [?] the [?] to your [?] [?]. ▪ A [?] [?] the [?] in the [?] days. ▪ The [?] [?] passed [?] the [?].

Gen.	Script 3
OF	We talked of the slide show in the circus.
YF	A yacht slid around the point into the bay.
OF	The Navy attacked the big task force.
Question Cues	▪ We [?] of the [?] [?] in the [?]. ▪ A [?] [?] around the [?] into the [?]. ▪ The [?] attacked the [?] [?] [?].

Gen.	Script 4
YM	The slush lay deep along the street.
OM	Glue the sheet to the dark blue background.
YM	A rod is used to catch pink salmon.
Question Cues	▪ The [?] [?] [?] along the [?]. ▪ [?] the [?] to the [?] [?] background. ▪ A [?] is used to [?] [?] [?].

Gen.	Script 5
YM	The fish twisted and turned on the bent hook.
OF	Take the winding path to reach the lake.
YM	The ship was torn apart on the sharp reef.
Question Cues	▪ The fish [?] and [?] on the [?] [?]. ▪ Take the [?] [?] to [?] the [?].

	▪ The [?] was [?] apart on the [?] [?].

Gen.	Script 6
OM	The crooked maze failed to fool the mouse.
YF	Use a pencil to write the first draft.
OM	The two met while playing on the sand.
Question Cues	▪ The [?] [?] failed to [?] the [?]. ▪ Use a [?] to [?] the [?] [?]. ▪ The [?] [?] while [?] on the [?].

Gen.	Script 7
YF	See the cat glaring at the scared mouse.
OF	A wisp of cloud hung in the blue air.
YF	It's easy to tell the depth of a well.
Question Cues	▪ [?] the [?] [?] the scared [?]. ▪ A [?] of [?] [?] in the blue [?]. ▪ It's [?] to [?] the [?] of a [?].

Gen.	Script 8
OM	The source of the huge river is the clear spring.
YM	Press the pants and sew a button on the vest.
OM	Note closely the size of the gas tank.
Question Cues	▪ The [?] of the [?] [?] is the [?] spring. ▪ Press the [?] and [?] a [?] on the [?]. ▪ [?] [?] the size of the [?] [?].

Gen.	Script 9
OF	Sickness kept him home the third week.
YM	Adding fast leads to wrong sums.
OF	He ran half way to the hardware store.
Question Cues	▪ [?] [?] him home the [?] [?]. ▪ [?] fast [?] to [?] [?]. ▪ He [?] [?] to the [?] [?].

Gen.	Script 10
YF	The ink stain dried on the finished page.
OM	There are more than two factors here.
YF	A pound of sugar costs more than eggs.
Question Cues	▪ The [?] [?] dried on the [?] [?]. ▪ There [?] more [?] two [?] [?]. ▪ A [?] of [?] [?] more than [?].

Gen.	Script 11
OF	These days a chicken leg is a rare dish.
YF	Kick the ball straight and follow through.
OF	The swan dive was far short of perfect.
Question Cues	▪ These days a [?] [?] is a [?] [?]. ▪ Kick the [?] [?] and [?] [?]. ▪ The [?] [?] was [?] [?] of perfect.

Gen.	Script 12
YM	Wipe the grease off his dirty face.
OM	The wide road shimmered in the hot sun.
YM	The show was a flop from the very start.
Question Cues	▪ [?] the [?] off his [?] [?]. ▪ The [?] road [?] in the [?] [?]. ▪ The [?] was a [?] from the [?] [?].

Gen.	Script 13
YM	The clock struck to mark the third period.
OF	The walled town was seized without a fight.
YM	The hat brim was wide and too droopy.
Question Cues	▪ The [?] [?] to mark the [?] [?]. ▪ The [?] [?] was [?] [?] a fight. ▪ The [?] [?] was [?] and too [?].

Gen.	Script 14
OM	The fin was sharp and cut the clear water.
YF	Rice is often served in round bowls.
OM	Help the woman get back to her feet.
Question Cues	▪ The [?] was [?] and [?] the [?] water. ▪ [?] is often [?] in [?] [?]. ▪ [?] the [?] get [?] to her [?].

Gen.	Script 15
YF	The beauty of the view stunned the young boy.
OF	Mend the coat before you go out.
YF	The lazy cow lay in the cool grass.
Question Cues	▪ The [?] of the [?] [?] the [?] boy. ▪ [?] the [?] [?] you go [?]. ▪ The [?] cow [?] in the [?] [?].

Gen.	Script 16
OM	A saw is a tool used for making boards.

YM	A small creek cut across the field.
OM	The lease ran out in sixteen weeks.
Question Cues	▪ A [?] is a [?] used for [?] [?]. ▪ A [?] [?] [?] across the [?]. ▪ The [?] [?] out in [?] [?].

Gen.	Script 17
OF	The lawyer tried to lose his case.
YM	The play seems dull and quite stupid.
OF	The juice of lemons makes fine punch.
Question Cues	▪ The [?] tried [?] [?] his [?]. ▪ The [?] seems [?] and [?] [?]. ▪ The [?] of [?] makes [?] [?].

Gen.	Script 18
YF	A pot of tea helps to pass the evening.
OM	Two blue fish swam in the tank.
YF	The wrist was badly strained and hung limp.
Question Cues	▪ A [?] of [?] helps to [?] the [?]. ▪ [?] [?] fish [?] in the [?]. ▪ The [?] was badly [?] and [?] [?].

Gen.	Script 19
OF	Lift the square stone over the fence.
YF	The wagon moved on well oiled wheels.
OF	Cars and busses stalled in snow drifts.
Question Cues	▪ [?] the [?] [?] over the [?]. ▪ The [?] moved [?] well [?] [?]. ▪ [?] and [?] [?] in snow [?].

Gen.	Script 20
YM	A tame squirrel makes a nice pet.
OM	The grass curled around the fence post.
YM	Bail the boat, to stop it from sinking.
Question Cues	▪ A [?] [?] makes a [?] [?]. ▪ The [?] [?] around the [?] [?]. ▪ [?] the [?], to [?] it from [?].

Gen.	Script 21
YM	The box was thrown beside the parked truck.
OF	Smoky fires lack flame and heat.
YM	Her purse was full of useless trash.

Question Cues	▪ The [?] was [?] beside the [?] [?]. ▪ Smoky [?] [?] [?] and [?]. ▪ Her [?] was [?] of [?] [?].

Gen.	Script 22
OM	The stray cat gave birth to kittens.
YF	The rope will bind the seven books at once.
OM	March the soldiers past the next hill.
Question Cues	▪ The [?] [?] gave [?] to [?]. ▪ The [?] will [?] the seven [?] at [?]. ▪ [?] the [?] [?] the next [?].

Gen.	Script 23
YF	The set of china hit, the floor with a crash.
OF	The horn of the car woke the sleeping cop.
YF	Cut the pie into large parts.
Question Cues	▪ The [?] of [?] [?], the [?] with a crash. ▪ The [?] of the car [?] the [?] [?]. ▪ [?] the [?] into [?] [?].

Gen.	Script 24
OM	The term ended in late June that year.
YM	The hogs were fed chopped corn and garbage.
OM	The soft cushion broke the man's fall.
Question Cues	▪ The [?] ended in [?] [?] that [?]. ▪ The [?] were [?] [?] [?] and garbage. ▪ The [?] [?] [?] the man's [?].

Gen.	Script 25
OF	The colt reared and threw the tall rider.
YM	The young girl gave no clear response.
OF	Hop over the fence and plunge in.
Question Cues	▪ The [?] [?] and threw the [?] [?]. ▪ The [?] [?] gave no [?] [?]. ▪ [?] over the [?] and [?] [?].

Gen.	Script 26
YF	A cup of sugar makes sweet fudge.
OM	This is a grand season for hikes on the road.
YF	The heart beat strongly and with firm strokes.
Question Cues	▪ A [?] of [?] makes [?] [?]. ▪ This is a [?] [?] for [?] on the [?].

	▪ The [?] [?] strongly and with [?] [?].

Gen.	Script 27
OF	Men strive but seldom get rich.
YF	A tusk is used to make costly gifts.
OF	Four hours of steady work faced us.
Question Cues	▪ [?] [?] but [?] [?] rich. ▪ A [?] is used to [?] [?] [?]. ▪ [?] [?] of [?] work [?] us.

Gen.	Script 28
YM	The salt breeze came across from the sea.
OM	It snowed, rained, and hailed the same morning.
YM	The meal was cooked before the bell rang.
Question Cues	▪ The [?] [?] came [?] from the [?]. ▪ It [?], [?], and [?] the same [?]. ▪ The [?] was [?] before the [?] [?].

Gen.	Script 29
YM	The friendly gang left the drug store.
OF	Place a rosebush near the porch steps.
YM	The dune rose from the edge of the water.
Question Cues	▪ The [?] [?] left the [?] [?]. ▪ [?] a [?] near the [?] [?]. ▪ The [?] [?] from the [?] of the [?].

Gen.	Script 30
OM	The pearl was worn in a thin silver ring.
YF	Always close the barn door tight.
OM	Ten pins were set in order.
Question Cues	▪ The [?] was [?] in a [?] [?] ring. ▪ Always [?] the [?] [?] [?]. ▪ [?] [?] were [?] in [?].

Gen.	Script 31
YF	Large size in stockings is hard to sell.
OF	The girl at the booth sold fifty bonds.
YF	Read verse out loud for pleasure.
Question Cues	▪ [?] size in [?] is [?] to [?]. ▪ The [?] at the [?] [?] fifty [?]. ▪ [?] [?] [?] loud for [?].

Gen.	Script 32
OM	What joy there is in living.
YM	Mesh wire keeps chicks inside.
OM	Both lost their lives in the raging storm.
Question Cues	▪ [?] [?] there [?] in [?]. ▪ [?] [?] [?] [?] inside. ▪ Both [?] [?] lives in the [?] [?].

Computer Voice

Gen.	Script 1
OF	Those words were the cue for the actor to leave.
YM	The fruit peel was cut in thick slices.
OF	He lay prone and hardly moved a limb.
Question Cues	▪ Those [?] were the [?] for the [?] to [?]. ▪ The [?] [?] was cut in [?] [?]. ▪ He [?] [?] and [?] moved a [?].

Gen.	Script 2
YF	The bill as paid every third week.
OM	Oak is strong and also gives shade.
YF	The bark of the pine tree was shiny and dark.
Question Cues	▪ The [?] as [?] every [?] [?]. ▪ [?] is [?] [?] also gives [?]. ▪ The [?] of the [?] tree [?] shiny and [?].

Gen.	Script 3
OF	Type out three lists of orders.
YF	A cramp is no small danger on a swim.
OF	The young kid jumped the rusty gate.
Question Cues	▪ [?] out [?] [?] of [?]. ▪ A [?] is [?] [?] danger on a [?]. ▪ The young [?] [?] the [?] [?].

Gen.	Script 4
YM	The empty flask stood on the tin tray.
OM	The jacket hung on the back of the wide chair.
YM	Steam hissed from the broken valve.
Question Cues	▪ The [?] [?] stood on the [?] [?]. ▪ The [?] [?] on the back of the [?] [?]. ▪ [?] [?] from the [?] [?].

Gen.	Script 5
YM	Acid burns holes in wool cloth.
OF	The fruit of a fig tree is apple-shaped.
YM	Cats and dogs each hate the other.
Question Cues	▪ [?] [?] holes in [?] [?]. ▪ The [?] of a [?] tree is [?]-[?].

	▪ [?] and [?] [?] [?] the other.

Gen.	Script 6
OM	Leaves turn brown and yellow in the fall.
YF	The harder he tried the less he got done.
OM	He said the same phrase thirty times.
Question Cues	▪ [?] [?] [?] and [?] in the fall. ▪ The [?] [?] tried the [?] [?] got done. ▪ He [?] the [?] [?] [?] times.

Gen.	Script 7
YF	Guess the results from the first scores.
OF	A speedy man can beat this track mark.
YF	At that high level the air is pure.
Question Cues	▪ [?] the [?] from the [?] [?]. ▪ A [?] man [?] [?] this [?] mark. ▪ At [?] high [?] the [?] is [?].

Gen.	Script 8
OM	The child almost hurt the small dog.
YM	Fairy tales should be fun to write.
OM	Corn cobs can be used to kindle a fire.
Question Cues	▪ The [?] almost [?] the [?] [?]. ▪ [?] [?] should [?] fun to [?]. ▪ [?] [?] can be used to [?] a [?].

Gen.	Script 9
OF	The pipe began to rust while new.
YM	The pennant waved when the wind blew.
OF	The boss ran the show with a watchful eye.
Question Cues	▪ The [?] began to [?] [?] [?]. ▪ The [?] [?] when the [?] [?]. ▪ The [?] ran the [?] with a [?] [?].

Gen.	Script 10
YF	Pluck the bright rose without leaves.
OM	A salt pickle tastes fine with ham.
YF	He broke a new shoelace that day.
Question Cues	▪ Pluck [?] [?] [?] [?] gloves. ▪ A [?] [?] [?] fine with [?]. ▪ [?] [?] a [?] shoelace that [?].

Gen.	Script 11
OF	Drop the two when you add the figures.
YF	There was a sound of dry leaves outside.
OF	Eight miles of woodland burned to waste.
Question Cues	▪ [?] the [?] when you [?] the [?]. ▪ There [?] a [?] of [?] leaves [?]. ▪ [?] [?] of woodland [?] to [?].

Gen.	Script 12
YM	Where were they when the noise started.
OM	Open the crate but don't break the glass.
YM	Split the log with a quick, sharp blow.
Question Cues	▪ [?] [?] they when the [?] [?]. ▪ [?] the [?] but don't [?] the [?]. ▪ Split the [?] with a [?], [?] [?].

Gen.	Script 13
YM	The cup cracked and spilled its contents.
OF	Two plus seven is less than ten.
YM	The just claim got the right verdict.
Question Cues	▪ The [?] [?] and [?] its [?]. ▪ [?] [?] [?] is less than [?]. ▪ The just [?] got [?] [?] [?].

Gen.	Script 14
OM	The coffee stand is too high for the couch.
YF	A filing case is now hard to buy.
OM	The sky that morning was clear and bright blue.
Question Cues	▪ The [?] [?] is too [?] for the [?]. ▪ A [?] [?] is now [?] to [?]. ▪ The [?] that [?] was [?] and [?] blue.

Gen.	Script 15
YF	The third act was dull and tired the players.
OF	The paper box is full of thumb tacks.
YF	Add the sum to the product of these three.
Question Cues	▪ The [?] [?] was [?] and [?] the players. ▪ The [?] [?] is full of [?] [?]. ▪ [?] the [?] to the [?] of these [?].

Gen.	Script 16
OM	Burn peat after the logs give out.

YM	Paste can cleanse the most dirty brass.
OM	The glow deepened in the eyes of the sweet girl.
Question Cues	▪ [?] [?] after the [?] give [?]. ▪ [?] can [?] the most [?] [?]. ▪ The [?] [?] in the [?] of the [?] girl.

Gen.	Script 17
OF	These thistles bend in a high wind.
YM	The urge to write short stories is rare.
OF	An abrupt start does not win the prize.
Question Cues	▪ These [?] [?] in a [?] [?]. ▪ [?] urge to [?] [?] stories is [?]. ▪ An [?] start [?] not [?] the [?].

Gen.	Script 18
YF	Torn scraps littered the stone floor.
OM	A young child should not suffer fright.
YF	Sell your gift to a buyer at a good gain.
Question Cues	▪ [?] [?] [?] the [?] floor. ▪ A [?] [?] should not [?] [?]. ▪ [?] your [?] to a [?] at a good [?].

Gen.	Script 19
OF	Thieves who rob friends deserve jail.
YF	He ordered peach pie with ice cream.
OF	The slang word for raw whiskey is booze.
Question Cues	▪ [?] who [?] friends [?] [?]. ▪ He [?] [?] [?] with [?] cream. ▪ The [?] word for [?] [?] is [?].

Gen.	Script 20
YM	Bring your problems to the wise chief.
OM	Pure bred poodles have curls.
YM	The pencils have all been used.
Question Cues	▪ [?] your [?] to the [?] [?]. ▪ [?] [?] [?] have [?]. ▪ The [?] have [?] [?] [?].

Gen.	Script 21
YM	Wood is best for making toys and blocks.
OF	Sunday is the best part of the week.
YM	Add the column and put the sum here.

Question Cues	▪ [?] is best for [?] [?] and [?]. ▪ [?] is the [?] [?] of the [?]. ▪ [?] the [?] and put the [?] [?].

Gen.	Script 22
OM	The tongs lay beside the ice pail.
YF	The ripe taste of cheese improves with age.
OM	Weave the carpet on the right hand side.
Question Cues	▪ The [?] [?] beside the [?] [?]. ▪ The [?] [?] of [?] [?] with age. ▪ [?] the [?] on the [?] [?] side.

Gen.	Script 23
YF	It caught its hind paw in a rusty trap.
OF	Write a fond note to the friend you cherish.
YF	The tree top waved in a graceful way.
Question Cues	▪ It [?] its [?] paw in a [?] [?]. ▪ [?] a [?] [?] to the friend you [?]. ▪ The [?] top [?] in a [?] [?].

Gen.	Script 24
OM	The pirates seized the crew of the lost ship.
YM	The office paint was a dull sad tan.
OM	The doctor cured him with these pills.
Question Cues	▪ The [?] [?] the [?] of the lost [?]. ▪ The office [?] was a [?] [?] [?]. ▪ The [?] [?] him with [?] [?].

Gen.	Script 25
OF	We admire and love a good cook.
YM	The petals fall with the next puff of wind.
OF	Act on these orders with great speed.
Question Cues	▪ We [?] and [?] a [?] [?]. ▪ The [?] [?] with the next [?] of [?]. ▪ [?] on these [?] with [?] [?].

Gen.	Script 26
YF	Hemp is a weed found in parts of the tropics.
OM	The wharf could be seen at the farther shore.
YF	Clothes and lodging are free to new men.
Question Cues	▪ Hemp [?] a [?] found in [?] of the [?]. ▪ The [?] could be [?] at the [?] [?].

	▪ [?] and lodging [?] free to [?] [?].

Gen.	Script 27
OF	The spot on the blotter was made by green ink.
YF	We tried to replace the coin but failed.
OF	He knew the skill of the great young actress.
Question Cues	▪ The [?] on [?] blotter was [?] by [?] ink. ▪ [?] tried to [?] the coin but [?] [?]. ▪ He [?] the [?] of the [?] young [?].

Gen.	Script 28
YM	The new girl was fired today at noon.
OM	There the flood mark is ten inches.
YM	Bring your best compass to the third class.
Question Cues	▪ The new [?] was [?] [?] at [?]. ▪ [?] the [?] [?] is ten [?]. ▪ Bring your [?] [?] to the [?] [?].

Gen.	Script 29
YM	The hog crawled under the high fence.
OF	A lame back kept his score low.
YM	Feel the heat of the weak dying flame.
Question Cues	▪ The [?] [?] under the [?] [?]. ▪ A [?] [?] kept his [?] [?]. ▪ [?] the [?] of the weak [?] [?].

Gen.	Script 30
OM	We frown when events take a bad turn.
YF	Mud was spattered on the front of his white shirt.
OM	She sewed the torn coat quite neatly.
Question Cues	▪ We [?] when [?] take a [?] [?]. ▪ [?] was [?] on the [?] of his white [?]. ▪ She [?] the [?] [?] quite [?].

Gen.	Script 31
YF	A rag will soak up spilled water.
OF	They felt gay when the ship arrived in port.
YF	He carved a head from the round block of marble.
Question Cues	▪ A rag [?] [?] up [?] [?]. ▪ They [?] [?] when the [?] arrived in [?]. ▪ He [?] a [?] from the round [?] of [?].

Gen.	Script 32
OM	They could laugh although they were sad.
YM	Move the vat over the hot fire.
OM	We find joy in the simplest things.
Question Cues	▪ They [?] [?] [?] they were [?]. ▪ [?] [?] vat [?] the hot [?]. ▪ We [?] [?] in the [?] [?].

Appendix G

Aviation Script Questions (Human & Computer Voice)

Author: Hyun Su (Winfred) Seong

Questions for Human-Voiced Scripts

Script 1 – Animal on Runway

1. What is the aircraft callsign?
    a. Albatross 132
    b. Albatross 123
    c. Albatross 122
    d. Albatross 213
    e. Albatross 312
2. What was the mistake of the pilot when responding to the ATC?
    a. Did not clearly speak; committed many repetitions and stutter
    b. Did not clearly identify the species of the animal, which could help wildlife team
    c. Did not specify the animal's distance from the aircraft position
    d. Did not specify the flight designator and number
    e. B and C
3. As a pilot, what would be the most helpful info to provide for the ATC, based only on the conversation?
    a. The changes in weather that can significantly affect take-off.
    b. The direction the animal is travelling.
    c. The intention to change runways if the runway configuration is activated.
    d. A and C
    e. B and C

Script 2 – Mayday (Fire on Engine)

1. What was the callsign and the nature of the emergency?
    a. Lucan Air 3425; fire on left engine
    b. Lucan Air 2345; fire on left engine
    c. Lucan Air 2345; fire on right engine
    d. Lucan Air 3245; fire on right engine
    e. Lucan Air 2354; fire on right engine
2. Based on the scenario, the "MAYDAY" signifies which of the following?
    a. It is non-threatening declaration than SQUAWK 7500
    b. It is non-threatening declaration than PAN-PAN
    c. Aircraft is now given special privilege to update flight routes to maintain safety
    d. ATC makes this aircraft top priority
    e. A and D
3. Suppose in the scenario, the pilot turns off the master electrical switch to put out the fire while the airplane's altitude is decreasing. What is a likely consequence of this action?
    a. It can put out the fire rapidly as no electricity will flow to the engine.
    b. It can short circuit other electrical components by reset.
    c. It can help gain attention of the ATC and gain priority over radar.

    d. A and C
    e. B and C

Script 3 – Pilot Confused Instruction (Frequency)

1. ATC corrected pilot's frequency from [1] to [2]:
    a. 127.55 → 127.75
    b. 127.75 → 118.35
    c. 118.53 → 118.75
    d. 127.57 → 118.53
    e. 118.35 → 127.57
2. Based on the scenario, the pilot said "as usual" after mentioning a frequency she already knew, but later was corrected by the ATC. Thus, the pilot committed which of the following human factors?
    a. Spatial Disorientation
    b. Fatigue
    c. Confirmation Bias
    d. Contextual Bias
    e. Actor-Observer Bias
3. Suppose the pilot was a foreign pilot who is having trouble with numbers. Which standard phraseology could the pilot use to overcome this mishap?
    a. “Roger, contact Tower.”
    b. “Speak slower.”
    c. “Re-confirm once more.”
    d. “It is difficult to understand.”
    e. “Negative.”

Script 4 – Callsign Confusion

1. What was the correct callsign of the aircraft spoken by the pilot?
    a. Coulomb 1297
    b. Crumb 1927
    c. Crown 1297
    d. Chrome 1927
    e. Krome 1297
2. Based on the scenario, what could have been the reason as to why the ATC made a mistake in reading the callsign?
    a. Spatial disorientation
    b. Increase in workload
    c. Accumulating fatigue
    d. Congested airspace
    e. Communication breakdown

3. Suppose the ATC caught his own mistake and fixed the problem The ATC should say which of the following standard phraseology to confirm the cancellation of flight following?
    a. Radar service terminated
    b. Frequency change approved.
    c. Roger.
    d. All of the above are correct.
    e. None of the above.

Script 5 – Bird and Taxiing

1. What was the situation on the runway 19 Bravo?
    a. A herd of birds moving to the end of the runway.
    b. A herd of mammals moving to the north end of the apron.
    c. A herd of python moving to the south end of the taxiway.
    d. A herd of bison moving to the end of the taxiway.
    e. A herd of mycin moving to the end of the taxiway/runway.
2. Why is the pilot's readback incorrect in this case?
    a. The pilot did not acknowledge the ATC's instruction to hold short.
    b. The pilot did not report a confirmation for pre-flight checklist on initial contact.
    c. The pilot did not provide the aircraft's whereabouts with respect to animals.
    d. A and B
    e. All of the above.
3. Suppose the ATC finds out that another animal is now approaching the aircraft in question from the south. Which of the following standard phraseologies should alert the pilot?
    a. [direction of birds] [altitude] [species] [speed] [aircraft callsign] [ATS unit]
    b. [species] [speed] [altitude] [aircraft callsign] [ATS unit]
    c. [speed] [altitude] [species] [direction of birds] [ATS unit]
    d. [aircraft callsign] [ATS unit] [altitude] [species] [direction of birds] [speed]
    e. None of the above.

Script 6 – Pan-Pan (Altitude Meter Lost)

1. What was the instruction given by the ATC?
    a. 27013G19, cleared to land on RW 25
    b. 20719G13, cleared to land on RW 52
    c. 20713G19, cleared to land on RW 25
    d. 27019G13, cleared to land on RW 52
    e. 20713G19, cleared to land on RW 25
2. Based on the conversation, it seems that the ATC made a mistake in reading back a critical information. What could have the pilot done to correct ATC's mistake?
    a. ATC should have identified its ATS units or identifier.
    b. Pilot did not read back its current vector, VOR, position, or altitude.

c. ATC should have clarified what obstacles could be near the pilot, which could improve flight safety.
d. Pilot should have read back the aircraft callsign according to International Phonetic Alphabet pronunciation.
e. ATC should have read back for the wind speed from the pilot.

3. Suppose the airport was in the unfamiliar terrain for the pilot and clouds were interfering with visibility, forcing the pilot to initiate a dangerous unstable approach. Which of the following would be the most appropriate action by the pilot?
    a. Request to divert to another airport.
    b. Initiate contact with MAYDAY.
    c. Initiate contact with PAN-PAN.
    d. Ask the ATC to have ambulances ready.
    e. A, B, and C are all appropriate.

Script 7 – Noticing Another Accident (Fire)

1. What is the aircraft callsign?
    a. Risel 1192
    b. Rigel 1192
    c. Rijewel 1192
    d. Rizell 1192
    e. Rijol 1192
2. Based on the conversation, what was the instruction given to the pilot?
    a. Orbit, right turns, continue orbiting
    b. Orbit north, right turns, continue orbiting until further instructions
    c. Orbit north, left turns, continue orbiting until further instructions
    d. Orbit, right turns, continue orbiting until the fire has weakened
    e. Orbit north, right turns, continue orbiting, contact FIC for weather
3. In the scenario, the ATC uses a colloquial tone (i.e., normal English) during the conversation to explain the situation. What effect does this have on the nearby traffic at the airport?
    a. No effect; there is no danger to safety.
    b. The pilots on radio communication may be confused.
    c. No effect; pilots will continue to fly in their respective vectors.
    d. The ATC will follow up with instructions once the aircraft makes landing.
    e. The trucks will respond to ATC's instructions regarding the position of the accident.

Script 8 – METAR and S-Turn

1. What information is requested by the pilot?
    a. Terminal Aerodrome Forecast
    b. Special Weather Report
    c. Meteorological Aerodrome Report

d. Pilot Report (PIREP)
e. Meteorological Agency Terminal Weather Report

2. Based on the scenario, what was the weather and cloud ceiling?
   a. light drizzle, cloud ceiling at 6200 ft
   b. heavy drizzle, cloud ceiling at 6200 ft
   c. light drizzle, cloud ceiling at 6020 ft
   d. drizzle, cloud ceiling at 6000 ft
   e. light drizzle, cloud ceiling at 6020 ft
3. Suppose that there was a large cloud developing vertically ahead of the aircraft. The pilot notices that changing the (flight) path is likely. What standard phraseology is the most appropriate to use?
   a. "Heathrow Director, Northwest one-nine-two requesting deviations left/right of track due to weather ahead."
   b. "Pan-pan pan-pan pan-pan. Heathrow Director, Northwest one-nine-two requesting diversion to another airport due to severe weather."
   c. "Heathrow Director, Northwest one-nine-two requesting information on the altitude of the cloud."
   d. "Mayday mayday mayday. Heathrow Director, Northwest one-nine-two requesting to veer left/right of track due to severe weather."
   e. None of the above.

Script 9 – Too Many Scheduled Arrivals

1. What was the instruction from the ATC in the beginning?
   a. Turn left heading two-five-zero, cleared the visual approach.
   b. Turn right heading two-two-five-zero, cleared the visual approach.
   c. Turn left heading two-zero-five, cleared the visual approach.
   d. Turn right heading two-zero-two-five, cleared the visual approach.
   e. Turn left heading two-five-two-zero, cleared the visual approach.
2. Based on the scenario, it can be surmised that the airport is characterized by…?
   a. Low number of departures
   b. High number of departures
   c. High number of arrivals
   d. Understaffed ATC
   e. All of the above
3. In the scenario, the pilot is given approach clearance. What additional information does the pilot need in order to land the aircraft safely?
   a. Weather over the airport, especially over the runway.
   b. Communicate to the ATC of the possibility of planning missed approach.
   c. Current vectors.
   d. Runway number.
   e. All of the above.

Script 10 – Wind

1. What was the wind speed and direction?
    a. 116007KT 060V125
    b. 161007KT 060V125
    c. 116017KT 060V125
    d. 1611107G10KT 060V125
    e. 11607KT 060V125
2. What mistake did the pilot make in the conversation?
    a. Reading back incorrect fly heading
    b. Reading back incorrect runway number
    c. Taking off while not being given a permission to do so
    d. Responding with incorrect flight callsign number
    e. Ignoring parts of ATC's instructions
3. Suppose that, along with the previous weather conditions, the airport is now overcast at nine thousand feet at eight o'clock eastern time. Flight Information Centre (FIC) worries that this weather change is significant. FIC sees increasing squall, temperature eleven, dew point ten. The FIC issues which type of the weather report?
    a. Report Type: SPECI, Time: 0800Z, Weather: INCR SQ OVC090 10/11
    b. Report Type: TAF, Time: 0800Z, Weather: SQ OVC09 11/10
    c. Report Type: PIREP, Time: 0800Z, Weather: INCR SQ OVC0900 10/11
    d. Report Type: METAR, Time: 1200Z, Weather: HEAVY SQ OVC09 11/10
    e. Report Type: SPECI, Time: 1200Z, Weather: +SQ OVC090 11/10

Script 11 – Extreme Weather (Dust)

1. What was the callsign of the aircraft?
    a. Reign Air 184; 1430005KT
    b. Grain Air 148; 134050KT
    c. Reign Air 184; 13405KT
    d. Rain Air 148; 13405KT
    e. Rain Air 148; 1340050KT
2. Based on the scenario, there is one runway that is affected by the said weather. What is the runway and where is it (the weather) moving towards?
    a. RW 30; southeast
    b. RW 03; northwest
    c. RW 30; northwest
    d. RW 03; southwest
    e. RW 30; northeast
3. Suppose the weather in question has deteriorated. What is a possible outcome for exacerbated weather?
    a. Possibility of a "go around"
    b. Unstable air leading to rainy thunderstorm
    c. Ditching (emergency landing)

d. A and B
e. B and C

Script 12 – Weather and Traffic (Mayday)

1. What was the aircraft callsign and what information did the pilot request?
    a. Octopus 2668; weather radio malfunction and traffic information
    b. Octopus 2638; weather radio malfunction and visibility
    c. Octopus 2368; visibility and weather radar malfunction
    d. Octopus 2836; traffic information and weather radar malfunction
    e. Octopus 2368; weather radar malfunction and traffic information
2. What key information did the ATC provide in terms of traffic?
    a. A 737 at 3000 ft, A regional airliner at 2000 ft
    b. A 747 at 2000 ft, A regional airliner at 3000 ft
    c. A 737 at 3500 ft, A low cost carrier at 2000 ft
    d. A 747 at 3000 ft, A low cost carrier at 2000 ft
    e. A 737 at 2000 ft, A regional airliner at 3000 ft
3. Suppose the aircraft experiences a mechanic malfunction, resulting in the failure of two way radio communication. The aircraft should call which of the phraseology?
    a. SQUAWK 7700
    b. SQUAWK 7500
    c. SQUAWK 7600
    d. Pan-Pan
    e. none of the above.

Script 13 – Unnecessary Greetings

1. What frequency was given prior (before) and after the correction?
    a. 121.86 → 121.65
    b. 121.58→ 121.56
    c. 121.56 → 121.58
    d. 121.85 → 121.65
    e. 121.65 → 121.85
2. The beginning of the conversation initiated by the ATC does [X] standard phraseology rule because [Y]:
    a. [X] not violate; [Y] it is a common greeting
    b. [X] violate; [Y] it uses colloquial (plain conversational English) tones
    c. [X] not violate; [Y] it still contains aviation-specific vocabularies
    d. [X] violate; [Y] it is awkward and shows incompetency as an ATC
    e. [X] not violate; [Y] it does not contain harmful or misleading information
3. Suppose the conversation became more friendly and took more space in the radio frequency being used. This is a precarious situation because…?

a. It will interfere with situation awareness of other pilots
b. Passengers will be frustrated due to delay in arrival
c. Radio may malfunction if the frequency is occupied for too long
d. A and B
e. All of the above

Script 14 – Redirect due to Extreme Weather

1. What was the callsign of the aircraft?
    a. Ban 1328
    b. Vaughn 1328
    c. Been 1823
    d. Vein 1328
    e. Bane 1283
2. Based on the scenario, the temperature outside is [X]. The latent threat (or "hidden" threat) in this weather conditions may result in [Y].
    a. [X] Cold; [Y] decreasing altitude
    b. [X] Cold; [Y] unstable levelling of the wings
    c. [X] Below Freezing; [Y] confusion in ATC's understanding
    d. [X] Below Freezing; [Y] tail strike
    e. [X] Below Freezing; [Y] loss of thrust
3. What should the ATC include when giving a vector to the alternative airport mentioned in the communication?
    a. [aircraft ID] [Fly Heading xyz degree]
    b. [aircraft ID] [Turn Right/Left] or [Fly Heading xyz degree]
    c. [aircraft ID] [Turn Right/Left] or [Fly Heading xyz degree] [TAF] [ETA]
    d. [aircraft ID] [PIREP] [Obstacles]
    e. None of the above are standard phraseologies.

Script 15 – Blizzard and the Pilot

1. What was the aircraft callsign and Ground frequency?
    a. Flower Air 3194; contact Ground 121.90
    b. Flour Air 3194; contact Ground 121.90
    c. Floor Air 3194; contact Ground 121.90
    d. Flower Air 3149; contact Ground 112.90
    e. Flour Air 3149; contact Ground 112.09
2. Based on the scenario, the ATC told the pilot to hold short of [X] due to [Y] (weather code).
    a. [X] India; [Y] SHRA
    b. [X] Charlie; [Y] SHRA
    c. [X] India; [Y] SHRA BLZD
    d. [X] Charlie; [Y] BLZD

   e. [X] India; [Y] BLZD

3. Suppose the weather condition became significantly worse and the aircraft in question is unable to move. What could be the first step the ATC could do to minimize impacts of this particular weather for arriving planes?
   a. Issue SPECI
   b. Close the runway
   c. Ask them to divert to another airport
   d. Ask them to report PIREP
   e. All of the above

Script 16 – Extended METAR Information

1. What was the callsign and what wind speed was mentioned?
   a. Seeder Air 7028; 27012G18KT
   b. Cider Air 7208; 27012G18KT
   c. Seder Air 7028; 27012G18KT
   d. Cedar Air 7208; 27012G18KT
   e. Sweeter Air 7208; 27012G18KT
2. Based on the scenario, the sky is [W] outside and the visibility is [X]. Temperature is [Y] and the dewpoint is [Z].
   a. [W] Overcast [X] less than ¾ SM [Y] 6 [Z] 9
   b. [W] Scattered [X] ¾ SM [Y] 6 [Z] 9
   c. [W] Overcast [X] less than ¾ SM [Y] 9 [Z] 6
   d. [W] Broken [X] less than ¾ SM [Y] 9 [Z] 6
   e. [W] Overcast [X] ¾ SM [Y] 9 [Z] 6
3. Suppose there is another flight following this airplane and the pilot of that aircraft is looking for weather info as well. But the pilot sees the weather condition that is rapidly changing and different from the weather the preceding aircraft has experienced. The pilot believes that it will likely impact other aircraft in the vicinity. What could the pilot do in that specific case?
   a. Transmit radio to other aircraft and communicate the changes in observation.
   b. Communicate with the ATC of Pan-Pan and let him/her know the intention to divert, if it can be safely done so.
   c. Report PIREP
   d. B and C
   e. Do not send radio transmission as it will interfere with situation awareness of other pilots.

Script 17 – Debris on Runway

1. What is the aircraft callsign?
   a. Chile Air 8125 Super
   b. Chili Air 8125 Super
   c. Cheerly Air 8125 Super
   d. Cherry Air 8125 Super

    e. Chiri Air 8125 Super
2. What should have been the correct readback for the pilot?
    a. "Hold short of Hotel. Contact Ground two-one-one-decimal-eight for taxi"
    b. "Do not exit on Hotel. Exit Juliet. Contact one-one-two-decimal-eight for taxi"
    c. "Do not exit on Hotel. Exit Juliet. Contact Ground one-two-decimal-eight for taxi"
    d. "Exit Juliet. Contact Ground one-two-one-decimal-eight for taxi"
    e. "Hold short of Hotel. Exit Juliet. Contact Ground one-two-one-decimal-eight-zero for taxi"
3. Suppose the operation on the runway is taking longer time than expected. Other runways are busy. Another airplane is ready for landing soon, but currently on a downwind. What is one realistic option the ATC can do for this pilot who is on a downwind?
    a. Initiate runway configuration so that one runway is for takeoff, the other for landing
    b. Close the runway and divert the aircraft to the nearest airport to maintain safety
    c. Ask how much fuel is left and whether it can make it to another airport
    d. Tell the pilot to expect longer downwind until further advised
    e. B and C

Script 18 – Landing Gear Malfunction

1. Choose the correct statement that lists the aircraft callsign and its altitude.
    a. Sol Air 255; 5000 ft
    b. Sole Air 255; 5000 ft
    c. Seoul Air 255; 5000 ft
    d. Soul Air 255; 5000 ft
    e. Sul Air 255; 5000 ft
2. What key information could the ATC add so that the pilot can elevate his/her situation awareness?
    a. Arrival schedules to the designated runway
    b. The altitude of another aircraft
    c. METAR or TAF
    d. A, B and C
    e. A and B
3. Suppose ATC notices that the aircraft is nearby a wind turbine farm. When providing [X], the ATC may consider a factor such as [Y].
    a. [X] Landing clearance; [Y] Area of the farm
    b. [X] Approach clearance; [Y] Terrain of the farm
    c. [X] Altitude clearance; [Y] Height of the obstacle
    d. [X] Approach clearance; [Y] Height of the obstacle
    e. [X] Unstable Approach clearance; [Y] Height of the obstacle

Script 19 – Brief Off the Radar Moment

1. What was the aircraft callsign?
    a. Carut Air 1297
    b. Carat Air 1297
    c. Karat Air 1297
    d. Caret Air 1297
    e. Carib Air 1297
2. ATC mentions "recycle transponder", which means…?
    a. turn transponder off and on
    b. turn ignition switch off and on
    c. inspect comm selectors
    d. check ground radio
    e. look for a second transponder that can be turned on
3. Suppose that the radio is malfunctioning and sound cannot be heard. The last instruction from the ATC, before being disconnected, was "If comms are lost…". What does this mean?
    a. Two-way communication is lost.
    b. One-way communication is lost.
    c. ATC can still see the aircraft on radar, but not the altitude.
    d. ATC can hear what the pilot is saying, but the pilot can't hear the ATC.
    e. Pilot must tune into another frequency for emergency.

Script 20 - FO Incapacitated

1. What was the callsign of the aircraft?
    a. Cruze Air 090
    b. Cruise Air 090
    c. Crews Air 090
    d. Crus Air 090
    e. Crux Air 090
2. Based on the scenario, the ATC's instruction means which of the following?
    a. Cleared to land on the runway and report landing
    b. Descend and intercept localizer and maintain present heading
    c. Descend, intercept, and report final once entered the base leg
    d. Turn right to face the runway and maintain present vector until further advised
    e. Enter the base leg from the left and prepare to descend
3. Based on the conversation, the pilot made a mistake in reading back something. What should the pilot have done right after making a mistake?
    a. The pilot should have corrected the read back on runway number
    b. The pilot should have declared Pan-Pan
    c. The pilot should have declared Mayday
    d. The pilot should have used more clear vocabularies when describing the injury of FO
    e. A and B

Script 21 – Fire Electrical

1. What is the aircraft callsign?
   a. Sparrow Air 1257
   b. Farrow Air 1257
   c. Faro Air 1257
   d. Burrow Air 1257
   e. Pharaoh Air 1257
2. Based on the scenario, the source of the fire is [W]. The fuel remaining is [X] and the runway given is [Y] as there is no [Z].
   a. [W] Mechanical [X] 51 minutes; [Y] 23R [Z] Obstacle or Traffic
   b. [W] Electrical [X] 15 minutes; [Y] 32L [Z] Obstacle or Traffic
   c. [W] Both Mechanical and Electrical [X] 15 minutes [Y] 32C [Z] Traffic
   d. [W] Electrical [X] 15 minutes; [Y] 32C [Z] Traffic
   e. [W] Both Mechanical and Electrical [X] 15 minutes [Y] 23C [Z] Traffic
3. Suppose the smoke is affecting the visibility. The pilot decides to initiate ditching. What is the most crucial information the pilot needs to make this happen?
   a. Maintain and climb to eliminate the "nosedive"
   b. Immediately communicate for an alternative airport to land
   c. Request flight crew to help passengers not panic
   d. Find the largest open field anywhere to land
   e. Extend landing gears to avoid tail strike

Script 22 – Lost Airspeed Indicator

1. The callsign of the aircraft is [X]. The cause of this incident is called [Y].
   a. [X] Laps Air 2488 [Y] Loss of thrust
   b. [X] Lapps Air 2488 [Y] Loss of altitude
   c. [X] Lapse Air 2488 [Y] Spatial disorientation
   d. [X] Laughs Air 2488 [Y] Loss of altitude
   e. [X] Lavs Air 2488 [Y] Stall
2. The [1] is lost based on the conversation, which means that [2]…?
   a. [1] Power; [2] It is displaying an unreliable altitude.
   b. [1] Pitot Tubes; [2] It is displaying an unreliable air pressure.
   c. [1] Speed; [2] It is displaying an unreliable airspeed.
   d. [1] Speed; [2] It is displaying an unreliable heading.
   e. [1] Speed and Power; [2] Every display will be affected to some degree.
3. Suppose that the pilot was experiencing dizziness from unstable altitude control. This would affect which of the following?
   a. Spatial disorientation
   b. Vision
   c. Speech rate
   d. Situation Awareness
   e. All of the above

Script 23 – Emergency with Another Aircraft

1. What was the aircraft callsign?
    a. Pair Air 369
    b. Pear Air 369
    c. Fair Air 369
    d. Pare Air 369
    e. Bear Air 369
2. Based on the scenario, the ATC's instruction literally means which of the following?
    a. with a forty-five minute delay expected
    b. with an hour delay expected
    c. with an undetermined time of delay expected
    d. wait until further advised when airport officials solve problems
    e. wait until further advised when traffic congestion improves
3. Suppose the airplane – where the intoxicated passenger is onboard – is rounding out as fast as possible at night. The pilot has "red eyes". The pilot's vision could possibly be affected by which of the following?
    a. Laser point from the ground
    b. Lights from cockpit displays
    c. Diabetes
    d. Drowsiness
    e. All of the above

Script 24 – Weather Display Lost

1. What was the aircraft callsign?
    a. Palate Air 685
    b. Palace Air 685
    c. Pallet Air 685
    d. Ballad Air 685
    e. Ballet Air 685
2. Based on the conversation, the weather report is called which of the following:
    a. SPECI
    b. METAR
    c. TAF
    d. PIREP
    e. SPECIAL WX
3. Suppose the funnel cloud is accompanied by heavy showers and heavy rain where the aircraft is enroute. A proper weather code would look like which of the following?
    a. + SHTS
    b. + FCTS
    c. + SHRA

d. + RASH
e. SHRA

Script 25 – Weather and Flight Update

1. What was the aircraft callsign?
    a. Playner Air 159
    b. Planar Air 159
    c. Planner Air 159
    d. Plainer Air 159
    e. Flaner Air 159
2. A weather described in the communication has which of the following characteristics?
    a. A narrow, funnel cloud shape
    b. Tilting of the cloud
    c. Towering of the cloud
    d. A and C
    e. B and C
3. Suppose this aircraft is a lighter sport aircraft (LSA) and it is going through a valley located in a tall mountain. This is a precarious situation in terms of weather because…?
    a. Obstacles such as tall trees can impact visibility and altitude
    b. Mountain winds cause rapid sinking
    c. There is no safe landing zone if there is an emergency
    d. Radio communication errors exist in high mountains
    e. All of the above

Script 26 – PIREP #1

1. What was the callsign of the aircraft?
    a. Razor Air 307
    b. Razer Air 307
    c. Laser Air 307
    d. Raiser Air 307
    e. Lazer Air 307
2. What has been described by the pilot for reporting weather?
    a. Distance, Location, Temperature, Weather
    b. Ground Position, VOR, Final Destination, Altitude
    c. Previous Destination, Location, Weather, Altitude
    d. Estimated Time of Arrival, Location, Cloud Types, Turbulence
    e. None of the above.
3. Suppose the aircraft is a light sport airplane (LSA). What is NOT an influence to the safety of the flight pertaining to the scenario?
    a. Differential heating

b. Crosswinds
c. Terrain
d. Turbulence
e. All are relevant influencers.

Script 27 – Climb Correction

1. What is the callsign of the aircraft?
    a. Hele Air 166
    b. Hell Air 165
    c. Heal Air 156
    d. Hail Air 166
    e. Hill Air 166
2. What's the primary purpose of altitude assignment, as can be inferred from the scenario?
    a. To avoid potential obstacles on the ground
    b. To schedule different takeoff times for the airplanes in the airspace
    c. To make ATC's work easy in busy airspace
    d. To separate airplanes
    e. To allow airplanes to avoid potential extreme weather and obstacles
3. Based on the conversation, it seems that the pilot read back the wrong altitude assignment. But the ATC corrected it.
    a. 2000 ft
    b. 1000 ft
    c. 2500 ft
    d. 1500 ft
    e. 500 ft

Script 28 – Practice Area

1. What is the aircraft callsign?
    a. Wail Air 825
    b. Wale Air 825
    c. Where Air 825
    d. Whale Air 825
    e. Ware Air 825
2. Suppose the pilot is practicing a stabilized approach. What factors would the pilot consider?
    a. Airspeed
    b. Descent rate
    c. Lateral track
    d. A and B only
    e. All of the above

3. Suppose this is a student pilot doing solo but a day before he/she had two exams and two days before three exams. Which human factor would most likely result?
    a. Confidence
    b. Workload
    c. Human-Computer Interaction
    d. Thermal Sense
    e. None of the above are human factors

Script 29 – Cabin Depressurization

1. What is the aircraft callsign?
    a. Bit Air 373
    b. Beat Air 373
    c. Beet Air 373
    d. Brit Air 373
    e. Bee Air 373
2. Why is this considered a dangerous situation, according to the scenario?
    a. Pilots lose decision-making ability at higher altitude
    b. Pilots have hard time multi-tasking
    c. Pilots are not well-trained in English conversation and stutter
    d. Pilots experience drowsiness at higher altitude
    e. Pilots pretend to stay alert even when the fatigue kicks in
3. As the altitude increases, the amount of oxygen [X] and pressure [Y].
    a. [X] Increases; [Y] Increases
    b. [X] Increases; [Y] Remains Constant
    c. [X] Decreases; [Y] Increases
    d. [X] Decreases; [Y] Decreases
    e. [X] Remains Constant; [Y] Increases

Script 30 – Chemical Leak

1. What is the callsign of the aircraft?
    a. If Air 244
    b. Yves Air 244
    c. Ibs Air 244
    d. Ebbs Air 244
    e. Ives Air 244
2. What is MAYDAY used for?
    a. For life-threatening situations such as becoming lost, system failure
    b. For non life-threatening situations such as becoming lost, system failure
    c. When it is closer to the airport during cruise
    d. A and C

    e. A, B, and C are all correct as safety is top priority
3. Suppose that at Red Lake Airport, the weather there is characterized by broken clouds at 10000 ft, temperature 22 and dewpoint 10, and the wind speed is 280 at 10 knots. What would be the correct METAR format in proper order?
    a. 28010KT BKN100 22/10
    b. BKN100 22/10 28010KT
    c. 22/10 28010KT BKN100
    d. 280010KT BKN100 22/10
    e. 28010KT BKN010 22/10

Script 31 - Routing Change

1. What is the aircraft callsign?
    a. Aero Axil 689
    b. Aero Axe 689
    c. Aero Axel 689
    d. Aero Axol 689
    e. Aero Axer 689
2. The pilot decided to change direct course because… ?
    a. Adverse weather condition due to cloudy conditions
    b. Adverse weather condition due to low visibility
    c. Adverse weather condition due to stormy weather
    d. Adverse weather condition due to stratocumulus and low visibility
    e. Adverse weather condition due to cumulonimbus and low visibility
3. If the ATC wants to convey more information such as specific route changes via VOR, the correct order of such instruction is…?
    a. [aircraft ID] [Routing Change] or [Amendment to the Route]
    b. [Routing Change] or [Amendment to the Route] [ATS Unit]
    c. [aircraft ID] [VOR or VOR-DME]
    d. [aircraft ID] [ETA] [Routing Change] or [Amendment to the Route]
    e. [aircraft ID] [Amendment to the Route] [ETA] [VOR or VOR-DME] [Obstacles] [Runway Conditions]

Script 32 – Normal Operation

1. What is the aircraft callsign?
    a. Basus Air 725
    b. Beiges Air 725
    c. Bassex Air 725
    d. Basics Air 725
    e. Basses Air 725
2. Based on the scenario, what were the altitude changes?

a. 6000 ft → 3400 ft
b. 6300 ft → 3400 ft
c. 6030 ft → 3000 ft
d. 6300 ft → 3040 ft
e. 6000 ft → 3040 ft

3. Based on the conversation only, what is a fatal mistake in the ATC's approach clearance instruction?
    a. Not confirming the runway number requested by the pilot.
    b. Not confirming the weather condition over the runway.
    c. Not confirming the remaining traffic on the runway.
    d. Not confirming the runway conditions.
    e. No mistake from ATC as the pilot finally said "Roger", which means understood everything.

Questions for Computer-Voiced Scripts

Script 1 - Stall

1. What is the aircraft callsign?
    a. Beys Air 127
    b. Bas Air 127
    c. Beige Air 127
    d. Baize Air 127
    e. Vase Air 127
2. Which of the following is TRUE?
    a. Keep the present heading and airspeed constant
    b. Stall can occur at any speed, any altitude, at any thrust.
    c. Communicate whether immediate landing is possible after recovery
    d. Ask flight attendants to calm the passengers during and after the stall
    e. Low airspeed is the cause of the stall and adding power would help recover
3. Suppose the aircraft is going through a secondary stall. This means that the pilot tried abruptly to do which one of the following?
    a. Not focusing on the thrust power and misusing thrust
    b. Turning off the master electrical switch
    c. Quickly returning to the flight path
    d. Decreasing pitch angle
    e. Decreasing bank angle

Script 2 - Circling

1. The aircraft callsign is [X].
    a. [X] Air Orange 3400
    b. [X] Air Orange 0304
    c. [X] Air Orange 3044
    d. [X] Air Orange 3040
    e. [X] Air Orange 3004
2. What fatal mistake did the pilot make?
    a. Pilot was over-speeding.
    b. Pilot misunderstood the runway numbers and mixed them up.
    c. Pilot is running out of fuel and didn't communicate it properly.
    d. Pilot had not been listening attentively to ILS instructions.
    e. Pilot responded with wrong callsign again, leading to ATC's.
3. Suppose the pilot encountered a large variation in wind direction from the west and descent rate has been affected, causing excessive lateral deviations. This means that the pilot might carry out which of the following actions?
    a. Stabilized Approach
    b. Unstabilized Approach

c. Undershoot/Overshoot
d. Missed Approach
e. Circling

Script 3 – Follow 747

1. What is the aircraft callsign?
    a. Mayflower 23
    b. Mayflower 237
    c. Mayflower 2397
    d. Mayflower 2937
    e. Mayflower 2379
2. What was the correction?
    a. Aircraft type, Heading, and Altitude
    b. Heading and Altitude
    c. Aircraft type, Altitude
    d. Aircraft type, Heading, Distance
    e. Heading, Distance, Altitude
3. Suppose the pilot has finished many extracurricular activities at his university prior to flying and his muscle reflex is not great. Which human factors could likely influence student's actions?
    a. Fatigue
    b. Startle Response
    c. Thermal Feedback
    d. A and B
    e. A, B and C

Script 4 – Plume of Smoke

1. What is the callsign of the aircraft?
    a. Lion Air 2576
    b. Lion Air 2567
    c. Lion Air 2657
    d. Lion Air 2675
    e. Lion Air 2667
2. What was the heading, altitude, and runway assigned in the event of missed approach?
    a. Turn right heading 240, maintain 4000 ft, expect vectors for RW 21
    b. Turn left heading 240, maintain 4000 ft, expect vectors for RW 21
    c. Turn right heading 240, maintain 4500 ft, expect vectors for RW 21
    d. Turn left heading 240, maintain 4500 ft, expect vectors for RW 12
    e. Turn right heading 240, maintain 4000 ft, expect vectors for RW 12
3. Which of the following is FALSE?

a. ATC does not expect pilot to report PIREP for other aircraft at the airport
b. Pilot should maintain continuous communication with ATC
c. Pilot should not fly over the smoke as it can affect stability of the aircraft
d. ATC expects the pilot to continuously report smoke while flying over the runway
e. ATC expects the pilot to calculate and maintain critical distance from the smoke

Script 5 – Traffic blurred by Cumulonimbus Cloud

1. What is the aircraft callsign?
    a. Kangaroo Air 5850
    b. Kangaroo Air 5508
    c. Kangaroo Air 5805
    d. Kangaroo Air 5085
    e. Kangaroo Air 5058
2. What was the instruction given by the ATC?
    a. Traffic at 6, 3 miles, southbound 747, 9500 ft descending
    b. Traffic at 4, 6 miles, northbound 737, 9000 ft descending
    c. Traffic at 6, 4 miles, northbound 737, 9000 ft descending
    d. Traffic at 5, 6 miles southbound 747, 8000 ft descending
    e. Traffic at 6, 3 miles northbound 737, 9000 ft descending
3. Suppose this communication took place at night over a city where the pilot can see many bright lights. If the pilot happens to stare at a light source for too long, this could cause which of the following?
    a. Increased heartrate
    b. Heightened attention
    c. Confirmation Bias
    d. Visual Autokinesis
    e. All of the above

Script 6 – Altitude Deviation during Landing

1. What is the callsign of the aircraft?
    a. Air Plato 2799
    b. Air Plato 2879
    c. Air Plato 2978
    d. Air Plato 2789
    e. Air Plato 2897
2. Based on the scenario, why did the ATC ask to confirm whether the aircraft was maintaining a specific altitude?
    a. ATC noticed a vector deviation
    b. ATC noticed a heading deviation
    c. ATC noticed a change in the flight plan

    d. ATC noticed a change in severe weather
    e. ATC noticed an altitude deviation
3. Suppose the problem in the conversation was fixed. The ATC should follow up with the pilot which of the following instructions?
    a. runway clearance for landing
    b. weather condition on the runway
    c. reporting PIREP
    d. terminating radar service
    e. none of the above

Script 7 – Take-Off (Hold Position)

1. What is the aircraft callsign and runway given by the ATC?
    a. Air Okra 1253; 08L
    b. Air Orca 1253; 08L
    c. Air Oakra 1253 15L
    d. Air Occra 1235; 15L
    e. Air Oaxaca 1235; 08L
2. Based on the scenario, what mistake did the pilot make in read back?
    a. Did not request a special weather report for a pre-flight
    b. Did not complete a pre-flight checklist
    c. Did not read back the runway number
    d. Did not read back what ATS unit it was speaking to
    e. All of the above are correct since they are all related
3. Suppose during takeoff the ATC notices there is a fire on the right engine. What would be the proper phraseology from the ATC to notify the pilot?
    a. "Do not take off. Smoke on your right engine. Ambulance will be ready upon your return to the runway."
    b. "Do not take off. Smoke on your right engine. Do you need a vector to the closest airport?"
    c. "Takeoff clearance cancelled. We see smoke from your right engine. Do you require assistance?"
    d. "Takeoff clearance cancelled. We see smoke from your right engine. Ambulance is ready when you return and land."
    e. All of the above are reasonable standard phraseologies.

Script 8 – Request Option

1. What is the CORRECT callsign repeated by the pilot?
    a. Paprika 0684
    b. Paprika 6048
    c. Paprika 0648

    d. Paprika 6084
    e. Paprika 6408
2. What runway option did the ATC give?
    a. RW 25
    b. RW 28
    c. RW 15
    d. RW 06
    e. None of the above
3. Suppose the ATC was busy communicating with other pilots in the vicinity. If he could not give a timely instruction, the phraseology should be structured as which of the following?
    a. [ATS unit] [aircraft callsign] [divert to another runway] [reason]
    b. [aircraft callsign] [ATS unit] [expect longer downwind]
    c. [ATS unit] [aircraft callsign] [unable due to] [reason]
    d. [aircraft callsign] [ATS unit] [read back]
    e. [ATS unit] [aircraft callsign] [stay on the radio]

Script 9 – Request RVR

1. What is the aircraft callsign?
    a. Roncesvalles 2234
    b. Roncesvalles 2432
    c. Roncesvalles 2342
    d. Roncesvalles 2243
    e. Roncesvalles 2423
2. What was the weather information provided by the ATC regarding the said runway?
    a. R16/040, VCTS
    b. R16/0400, VCSQ
    c. R16/4400, VCSH
    d. R16/0400, VCTS
    e. R16/4000, VCSQ
3. Suppose the pilot confused the definition of runway visual range with that of visibility. What would be the rule of thumb for remembering the two concepts?
    a. RVR is measured in feet, Visibility is in statute mile
    b. RVR is measured in statute mile, Visibility in feet
    c. Both RVR and Visibility are measured in feet
    d. Both RVR and Visibility are measured in statute mile
    e. RVR is measured in kilometres, Visibility is in metres

Script 10 – FO Pregnant on Airplane

1. What is the callsign of the aircraft?
    a. Great Lakes Air 3456

    b. Great Lakes Air 4653
    c. Great Lakes Air 4656
    d. Great Lakes Air 4536
    e. Great Lakes Air 4346
2. Based on the conversation, what is actually happening?
    a. Captain confused about vector
    b. First officer being incapacitated
    c. One of the passenger is pregnant
    d. First officer is pregnant
    e. Captain is pregnant
3. The captain was not able to land on the runway and initiated a go-around unfortunately. Based on the conversation, what mistake did the captain make?
    a. Captain did not request correct weather information
    b. Captain did not request ambulance
    c. Captain read back incorrect fly heading
    d. Captain read back incorrect flight number and call sign
    e. C and D

Script 11 - Maintain Critical Distance

1. What is the aircraft callsign?
    a. Keywest Air 3675
    b. Keywest Air 3765
    c. Keywest Air 3657
    d. Keywest Air 3756
    e. Keywest Air 3567
2. In the scenario, it seems that there is an aircraft flying closely ahead. What would be the most appropriate standard phraseology for the ATC to tell the pilot to reduce speed?
    a. "Hey, decrease the speed please."
    b. "Can you reduce your speed to one-eight-zero knots immediately"
    c. "You are speeding on my radar. Reduce the speed to one-eight-zero knots for spacing"
    d. "Speed one-eight-zero knots for spacing"
    e. "Collision possible. Reduce speed to one-eight-zero knots immediately"
3. Suppose the aircraft in question has landed and is taxiing. Another aircraft right behind this aircraft is on final, trying to land on the same runway. The pilot of this later aircraft should wait until the preceding pilot has completed which of the following?
    a. The preceding aircraft has intercepted the localizer
    b. The preceding aircraft has communicated landing to the ATC
    c. The preceding aircraft is taxiing towards the exit
    d. The following aircraft knows the preceding aircraft's position and intention
    e. The following aircraft knows the preceding aircraft has landed

Script 12 – Blowing Snow (with altitude block)

1. What is the aircraft callsign?
    a. Giraffe Air 2958 AH
    b. Giraffe Air 2598 AH
    c. Giraffe Air 2895 AH
    d. Giraffe Air 2985 AH
    e. Giraffe Air 2859 AH
2. As can be inferred from this scenario. what could be the reason for establishing a block altitude?
    a. To avoid mid-air collision
    b. To avoid obstacles in the air (bird) or radio towers (ground)
    c. To identify the aircraft in communication on the radar
    d. To minimize impact of icing conditions or turbulence
    e. To prepare for emergency landing in case of weather-related emergency
3. What could possibly go wrong from this conversation?
    a. ATC did not identify the ATS unit
    b. Pilot did not clarify the aircraft callsign
    c. ATC did not provide traffic around the aircraft in question
    d. Pilot did not correctly read back the range of block of altitude
    e. ATC did not request estimated time of arrival for arrival due to his own personal matters

Script 13 – Maintain Visual Separation with Runway Confusion

1. What two runways were being used in the communication?
    a. Visual Approach 19R, Visual Separation 19L
    b. Visual Approach 19C, Visual Separation 19L
    c. Visual Approach 19C, Visual Separation 19R
    d. Visual Approach 19L, Visual Separation 19R
    e. None of the above
2. This situation could be most likely a precursor to which of the following accidents?
    a. CFIT
    b. Mid-air collision
    c. LOC-I
    d. Runway Excursion
    e. Overshoot
3. Suppose the pilot (in IFR flight) of the intended aircraft heard the last instruction but is unable to establish radio contact due to radio malfunction during the approach (i.e. Two-war radio communications failure). What good judgment should the pilot make?
    a. Initiate radio contact with any ATS or aircraft
    b. Continue on the extended downwind
    c. Squawk 7500
    d. Circle until the ATC establishes an emergency communication
    e. Request another airport to divert to

Script 14 - Direction to Next Available Airport

1. What was the callsign of the aircraft?
    a. Bright Air 647
    b. Fright Air 647
    c. Ripe Air 647
    d. Write Air 647
    e. Wrath Air 547
2. Based on the scenario, what is VOR?
    a. Visual Omni Directional Range
    b. VHF Omni Directional Range
    c. Visual Orientation Directional Range
    d. Visual Octagonal Directional Range
    e. Visual Optometric Directional Range
3. Suppose that the communication took place at night and pilots are busy studying charts looking for frequencies of VOR. What could be the *first* possible mistake that could occur during this situation?
    a. Increased workload due to stress
    b. Gradually losing situation awareness as they are looking back and forth at the cockpit and the chart
    c. Inadvertently changing direction due to spatial disorientation caused by reduced visibility
    d. Inputting wrong frequency
    e. All of the above

Script 15 – Off Course

1. What is the callsign of the aircraft?
    a. Cubik Air 602
    b. Cubeek Air 602
    c. Cuvrik Air 602
    d. Cubrik Air 602
    e. Koobik Air 602
2. Based on the scenario, the large cloud that is blocking the path is most likely [X] and it is known to cause [Y].
    a. [X] Cumulonimbus Cloud [Y] Thunderstorm
    b. [X] Stratocumulus Cloud [Y] Blizzard
    c. [X] Cirrus Cloud [Y] Strong updraft
    d. [X] Cumulus Cloud [Y] Hail
    e. [X] Nimbostratus Cloud [Y] Heavy squall

3. Suppose the pilot wanted to pass through the cloud because he/she thinks that it is not as large as she expected and ensures herself that it won't be dangerous. She decides to go through it. This action based on such thought is referred to as…?
    a. Tunnel vision
    b. Confirmation bias
    c. Heuristics
    d. Swiss Cheese Theory
    e. Error estimation

Script 16 – PIREP #2

1. What is the aircraft callsign?
    a. Koral Air 2366
    b. Chloral Air 2369
    c. Coral Air 2369
    d. Kural Air 2369
    e. Corel Air 2369
2. Based on the scenario, what information would the pilot be interested when he/she is trying to descent and preparing for precision landing?
    a. Debris on the runway
    b. RVR
    c. Closest location of the VOR/DME
    d. Whether de-icing service is available
    e. Whether emergency service is available
3. Suppose the pilot noticed ice pellets building up on the rim of the window. Which mechanical part of the aircraft, if impacted by ice, would impact airspeed and altitude?
    a. Wings
    b. Tail
    c. Pitot tubes
    d. Flaps
    e. All of the above

Script 17 – Alcohol Abuse

1. What is the aircraft callsign?
    a. Air Toronto 1264
    b. Air Toronto 1462
    c. Air Toronto 1624
    d. Air Toronto 1426
    e. Air Toronto 1246
2. Based on the scenario, the plane turned right on [W], right on [X], and continued straight and entered [Y].

   a. [W] Delta [X] Alpha [Y] Bravo
   b. [W] Hotel [X] Alpha [Y] Alpha Juliet
   c. [W] Bravo [X] Juliet [Y] Alpha
   d. [W] Hotel [X] Bravo [Y] Alpha Juliet
   e. [W] Hotel [X] Alpha [Y] Terminal 2
3. Suppose the captain was found to be heavily under the influence of a substance. Transport Canada investigated the incident and his/her license was suspended after. Turns out, there were more incidents like this in other airports as well. In aviation, which of the following analogy would best describe this situation?
   a. Glass Ceiling
   b. Nepotism
   c. Swiss Cheese
   d. Embezzlement
   e. Office Politics

Script 18 – Mayday (Engine Failure)

1. What is the aircraft callsign?
   a. Ocean Air 2609
   b. Ocean Air 2096
   c. Ocean Air 2906
   d. Ocean Air 2960
   e. Ocean Air 2069
2. Based on the scenario, the pilot should have done which of the following at the beginning of the transmission?
   a. Pan-Pan
   b. Mayday
   c. SQUAWK 7500
   d. SQUAWK 7600
   e. SQUAWK 7700
3. Suppose the pilot is having trouble in controlling the aircraft and as can be heard from the radio, the pilot just replied with "Affirmative". What should have been the correct standard phraseology for the read back?
   a. "Roger, visual runway two-eight right."
   b. "Got it, visual runway two-eight right."
   c. "Copy that, visual runway two-eight right."
   d. "Yes, visual runway two-eight right."
   e. None of the above as it is mutually understood that the aircraft is in emergency

Script 19 – Midair Collision

1. What is the aircraft callsign?

    a. Ball Air 620
    b. Bole Air 620
    c. Boll Air 620
    d. Bawl Air 620
    e. Bowl Air 620
2. Based on the scenario, the ATC requested to recycle the transponder. What if the ATC told the pilot to "Squawk ident"? What is the difference between these two?
    a. Squawk ident is used when ATC can see the target and wants to verify
    b. Squawk ident is used for non-threatening emergencies such as moderate turbulence
    c. Squawk ident is used for resetting altitude errors
    d. A and B
    e. A, B, and C
3. Suppose the aircraft crashed and didn't make it. Safety investigators are writing up a final report on what happened. Which of the following would NOT be included in the report?
    a. Identification and type of aircraft
    b. Weather conditions at the altitude at which midair collision occurred
    c. Injuries
    d. Degree of evasive actions from pilots
    e. All of the above are included

Script 20 - Tail Strike

1. What is the [X] aircraft callsign, [Y] the instruction given to the pilot after the bumping sound, and finally [Z] the nature of emergency based on the conversation?
    a. Coconut Air 4165, Maintain 4500, In-flight fire
    b. Coconut Air 4156, Maintain 4500, Smoke due to fire
    c. Coconut Air 4615, Maintain 4500, Tail Strike
    d. Coconut Air 4165, Maintain 4500, Broken brakes
    e. Coconut Air 4651, Maintain 4500, Broken vertical stabilizer
2. Based on the scenario, the pilot may experience which of the following as a result of the aircraft keep turning left?
    a. Rapid eye movement
    b. Frustration
    c. Spatial Disorientation
    d. Sweating
    e. All of the above
3. Suppose the runway was slippery due to snow buildup and is contaminated. The pilots should think of which factors upon landing (or flare)?
    a. Braking efficiency
    b. Touchdown within a prescribed zone
    c. Possibility of go around
    d. A and B
    e. All of the above

Script 21 – Immediate Departure

1. What was the callsign of the aircraft?
   a. Rhode Island Air 2527
   b. Rhode Island Air 2257
   c. Rhode Island Air 2725
   d. Rhode Island Air 2557
   e. Rhode Island Air 2752
2. Based on the scenario, the ATC gave a conditional clearance. What was the runway and clearance altitude, and heading?
   a. RW 35; 4500 ft; heading 265
   b. RW 53; 4500 ft; heading 265
   c. RW 35; 4500 ft; heading 260
   d. RW 53; 4000 ft; heading 260
   e. RW 35; 4000 ft; heading 265
3. Suppose the immediate departure resulted in the pilot feeling pressured to complete the climb prematurely, overlooking some details on the pre-flight checklist. The pilot thinks that everything is in order and is confident that the flight will be conducted safely. This is an example of…?
   a. Confirmation bias
   b. Heuristics
   c. Overconfidence bias
   d. Rationalization
   e. Optimization

Script 22 – Towering Cumulonimbus before Departure

1. What is the aircraft callsign?
   a. Blue Lagoon 6024
   b. Blue Lagoon 6204
   c. Blue Lagoon 6420
   d. Blue Lagoon 6402
   e. Blue Lagoon 6240
2. Based on the scenario, what were the elements of METAR provided to the pilot?
   a. 30024KT, Heavy rain, M¾ SM
   b. 30024KT, Thunderstorm, ¾ SM, Cumulonimbus
   c. 30024KT, Heavy drizzle, ¾ SM, Overcast
   d. 35024KT, Heavy rain, M¾ SM, Cumulonimbus
   e. 35024KT, Thunderstorm, ¾ SM, Overcast, Cumulonimbus
3. Suppose the ATC, due to traffic congestion at the airport, gave clearance for takeoff to the airplane. The said cloud was moving towards the plane. Unfortunately, the plane was enshrouded by the cloud within a matter of minutes. What could possibly happen to the aircraft?
   a. There will be ice buildup around the wings and engines

b. Not all cumulonimbus clouds are stormy, but pilots will still be on the lookout
c. Low visibility is not safe so the pilots will find ways to return to the runway
d. Strong wind updraft and downdraft will cause turbulence
e. All of the above

Script 23 – Student Pilot's Confusion

1. What is the aircraft callsign?
    a. Black Forest 1538
    b. Black Forest 1583
    c. Black Forest 1853
    d. Black Forest 1835
    e. Black Forest 1385
2. What was the reason as to why the ATC tried to stop the pilot from descending further?
    a. Traffic congestion
    b. Misunderstanding of instructions
    c. Another plane at 2500 ft
    d. A and B
    e. A and C
3. Suppose the student pilot had a lavish party the night before the flight, where alcohol was involved. However, he/she did not tell the flight instructor. During the landing flare, the pilot misunderstands a very important landing instruction from the ATC by doing everything opposite, although he didn't mean to. This is a primary example of…?
    a. Individualism
    b. Communication breakdown
    c. Intentional disobedience
    d. Confirmation bias
    e. All of the above

Script 24 - Student Solo with Altitude Problem

1. What is the aircraft callsign?
    a. Nimbus 2570 A
    b. Nimbus 2750 A
    c. Nimbus 2507 A
    d. Nimbus 2057 A
    e. Nimbus 2705 A
2. Based on the scenario, what should have been declared by the pilot?
    a. Pan-Pan
    b. Mayday
    c. Squawk 7600
    d. Squawk 7500

    e. All of the above, at this moment, are reasonable declarations or transponder modes

3. Unfortunately, there was no mention of a location or whereabouts he/she was flying. Not even terrain information was provided. This is troublesome for the ATC because…?
    a. ATC cannot suggest other suitable areas or airports for landing
    b. ATC cannot figure out the name of the pilot and flight crew on board
    c. ATC does not know whether the pilot is held as a captive by a terrorist
    d. ATC may have trouble reporting to higher authorities
    e. None of the above as the pilot mentioned that he/she will land on an open field

Script 25 – Confusing Approach Instruction

1. What is the aircraft callsign?
    a. Pineapple Air 1923
    b. Pineapple Air 1293
    c. Pineapple Air 1932
    d. Pineapple Air 1239
    e. Pineapple Air 1392
2. Based on the scenario, what was the wind speed given and what runway was it for?
    a. 18013G25KT; Runway 06L
    b. 18013G25KT; Runway 06R
    c. 18013G25KT; Runway 24L
    d. 18013G24KT; Runway 24R
    e. 18013G24KT; Runway 24C
3. Suppose that the pilot was experiencing fatigue due to having long-haul flight yesterday, hence being confused. What remedies could she use to combat fatigue?
    a. Drink caffeine if need be
    b. Drink plenty of water
    c. Get a prescription for a pill that reduces fatigue beforehand
    d. Fill the sleep bank
    e. All of the above

Script 26 – Unfamiliar Area

1. What is the callsign of the aircraft?
    a. Borough Air 621H
    b. Burro Air 621H
    c. Borrow Air 621H
    d. Bureau Air 621H
    e. Burrow Air 621H
2. Based on the scenario, the aircraft in conversation is surrounded by [W] at [X], and [Y] at [Z].
    a. [W] 737 [X] 3000 ft [Y] 747 [Z] 2000 ft
    b. [W] 737 [X] 3500 ft [Y] 747 [Z] 2500 ft

    c. [W] 737 [X] 2000 ft [Y] 747 [Z] 3000 ft
    d. [W] 737 [X] 2500 ft [Y] 747 [Z] 3000 ft
    e. [W] 737 [X] 2000 ft [Y] 747 [Z] 2000 ft
3. Suppose the traffic was not in the control zone the ATC was responsible for. What standard phraseology would the ATC use to indicate this after giving altitude information?
    a. "Unable to confirm"
    b. "Confirm"
    c. "Stand by"
    d. "Unverified"
    e. "Correction"

Script 27 – Verify Enroute Traffic

1. What is the aircraft callsign?
    a. Cents Air 2164H
    b. Scent Air 2164H
    c. Sent Air 2164H
    d. Zent Air 2164H
    e. Xent Ait 2164H
2. Based on the scenario, what information did the pilot give?
    a. [location] [altitude] [destination]
    b. [location] [airspeed] [weather] [destination] [heading]
    c. [location] [altitude] [destination] [weather]
    d. [location] [altitude] [type of flight] [time] [heading]
    e. [altitude] [overhead route] [type of flight] [time] [heading]
3. Suppose the pilot finds there is a thunderstorm brewing nearby London Airport. The pilot feels this could be threatening. What could the pilot do to alert other pilots flying in the area?
    a. Report METAR
    b. Report PIREP
    c. Report TAF
    d. A and B
    e. A, B, and C

Script 28 – Low on Oxygen

1. What is the aircraft callsign?
    a. Toes Air 4291
    b. Doze Air 4291
    c. Docks Air 4291
    d. Doughs Air 4291
    e. Downs Air 4291

2. Based on the scenario, the pilot is suffering from [X]. This may have caused the pilot to confuse or forget [Y] and [Z].
    a. [X] Delusion [Y] Airspeed [Z] Heading
    b. [X] Spatial disorientation [Y] Flight number [Z] Altitude
    c. [X] Increased workload [Y] Airspeed [Z] Altitude
    d. [X] Hypoxia [Y] Altitude [Z] Heading
    e. [X] Pre-Cardiac Symptoms [Y] Altitude [Z] Heading
3. If the pilot believes that he is not going to make it to the destination airport (that he is aware that losing consciousness soon is highly probable), then he has the option to do which of the following?
    a. Forced landing
    b. Precautionary landing
    c. Ditching
    d. Tell the ATC to have ambulance ready
    e. Declare MAYDAY

Script 29 - Unruly Passenger

1. Identify the aircraft callsign and choose all those that describe the injuries sustained by the captain.
    a. Leopard Air 2079 - stabbed in the right arm, right leg, left hand
    b. Leopard Air 2709 - stabbed in the neck, right leg, right hand
    c. Leopard Air 2970 - stabbed in the forehead, left leg, left hand
    d. Leopard Air 2907 - stabbed in the forehead, left leg, right hand
    e. Leopard Air 2079 - stabbed in the forehead, right leg, left hand
2. Based on the scenario, this is an example of which of the following?
    a. Pan Pan
    b. Mayday
    c. Indictable Offence
    d. A and C
    e. B and C
3. Suppose the passenger is tied to a seat finally and the injured pilot is the captain; the first officer is not injured. The first responsibility of the first officer is to:
    a. Consult the checklist
    b. Communicate with the ATC to request police services upon landing
    c. Land the aircraft safely
    d. Calm down the unruly passenger by talking
    e. Ban the passenger from flying the pilot's airline

Script 30 - Airplane Crash Site

1. What is the situation of the accident?

   a. Two civilian cars damaged
   b. Smoke rising from the crash site
   c. Six civilians dead
   d. A and B
   e. A and C
2. Based on the scenario, the ATC called [U], [V], [W], [X] on the way. ATC also issued the pilot to [Y].
   a. [U] Police Officer [V] Trucks [W] Ambulance [X] Safety Personnel [Y] Circle
   b. [U] Security Guards [V] Firefighters [W] Ambulance [X] Police [Y] Circle
   c. [U] Airport officials [V] Firefighters [W] Ambulance [X] Police [Y] Remain outside
   d. [U] Firefighter [V] Airport Manager [W] Ambulance [X] Helicopter [Y] Remain outside
   e. [U] Police Officer [V] Helicopter [W] Ambulance [X] Firefighters [Y] Remain outside
3. At the end, the pilot landed the aircraft too early and it nearly caused a collision with another aircraft departing. This is because the pilot made which mistake in the conversation?
   a. he replied with a wrong aircraft number
   b. he did not identify the aircraft name correctly
   c. he would have probably feared that he would be late to the airport
   d. A and B
   e. None of the above.

Script 31 – Speak Slower

1. What is the aircraft callsign?
   a. Less Air 2673
   b. Glass Air 2673
   c. Raise Air 2673
   d. Lycee Air 2673
   e. Lace Air 2673
2. Based on the scenario, the switch of the runway was from [X] to [Y] and the altitude until established was [Z].
   a. [X] 33R [Y] 33L [Z] 4500
   b. [X] 33L [Y] 33R [Z] 4000
   c. [X] 33R [Y] 33L [Z] 4200
   d. [X] 33L [Y] 33R [Z] 4200
   e. [X] 33R [Y] 33L [Z] 4000
3. Suppose the pilot landed and was questioned by a supervisor on her slurred speech during communication. She had not slept well for two days straight. She has another flight in 10 hours. What remedy would work for her situation?
   a. A prescription of a drug
   b. A cup of coffee
   c. Eat sugary foods

d. Fill sleep bank
e. None of the above

Script 32 – Passenger with Cardiac Arrest

1. What is the callsign of the aircraft?
    a. Von Air 333
    b. Vein Air 333
    c. Fan Air 333
    d. Band Air 333
    e. Vaughn Air 333
2. What was the weather and runway information at Edmonton?
    a. 20505KT, ILS RW 12
    b. 20550KT, ILS RW 21
    c. 25005KT, ILS RW 12
    d. 25005KT, ILS RW 21
    e. 25055KT, ILS RW 12
3. Suppose the pilot thinks that transponder would be more appropriate to alert ATCs. He/she pressed Squawk 7700, what does this mean?
    a. Hijacking of the aircraft
    b. Emergency
    c. Display malfunction
    d. Radio communication failure
    e. All of the above can capture the ATC's attention

Appendix H

Department of Systems Design Engineering, University of Waterloo

PARTICIPANTS NEEDED FOR RESEARCH IN *AVIATION ENGLISH*

We are looking for volunteers to take part in a study of Aviation English. The study involves participants to listen to: a) various pilot and air traffic controller conversations and; b) random sentences. And answer multiple-choice and fill-in-the-blanks questions (total of 48 questions) with their personal computers.

Expectations

- Participant will get links that allow them to access to two listening tests (Round 1 and 2; 24 questions in each).
- The listening tests will take a total of 1 hour and 30 minutes. No time limit.
- Participants will complete a preliminary (demographic) survey before starting.
- No identifiable information about participants will be published.
- In appreciation for your time, you will receive up to $20 Tim Hortons digital gift card**.**

Eligibility Criteria

- Participants can be either:
  - Undergraduate students
    - Enrolled in the aviation program.
    - May have a minimum exposure to piloting an aircraft **or** no experience.
    - May have licenses (PPL, CPL, ATPL).
    - Enjoy listening to pilot-air traffic controller conversation on YouTube.
  - Graduate students
    - Conducting research in aviation-related studies.
    - May have a minimum exposure to piloting an aircraft **or** no experience.
    - May have licenses (PPL, CPL, ATPL).
    - Enjoy listening to pilot-air traffic controller conversation on YouTube.
- Participants should have normal hearing (i.e. no hearing issues).
- Participants must be at least eighteen years old.

The study is being carried out by student investigator Hyun Su Seong who is pursuing his PhD in Systems Design Engineering. The faculty supervisors for this experiment are Dr. Shi Cao (SYDE) and Dr. Suzanne Kearns (ENV). This study has been reviewed by and received ethics clearance from the University of Waterloo Research Ethics committee. Everything related to the experiment would be provided by the research team administering the experiment. If you are interested in participating or have questions, please contact **Hyun Su Seong** at **hs2seong@uwaterloo.ca**