12-2020

The Effect of Intensive Auditory Cues on Flight Passengers' Safe Behaviors

Tianhua Li

The Effect of Intensive Auditory Cues on Flight Passengers' Safe Behaviors

by

Tianhua Li

Bachelor of Engineering
Civil Aviation Maintenance Engineering
Nanjing University of Aeronautics and Astronautics
2015

Master of Science
Aeronautics
Embry-Riddle Aeronautical University
2017

A dissertation
submitted to the College of Aeronautics at
Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Aviation Sciences

Melbourne, Florida
December 2020

We the undersigned committee hereby recommend that the attached document be accepted as fulfilling in part the requirements for the degree of Doctor of Philosophy of Aviation Science.

"The Effect of Intensive Auditory Cues on Flight Passengers' Safe Behaviors"

by

Tianhua Li.

Deborah S. Carstens, Ph.D.
Professor
College of Aeronautics
Major Advisor

Meredith Carroll, Ph.D.
Associate Professor
College of Aeronautics
Committee Member

John Deaton, Ph.D.
Professor
College of Aeronautics
Committee Member

Brooke Wheeler, Ph.D.
Assistant Professor
College of Aeronautics
Committee Member

William Gabrenya, Ph.D.
Professor Emeritus
School of Psychology
Committee Member

Ulreen Jones-McKinney, Ph.D.
Assistant Professor and Dean
College of Aeronautics

Abstract

TITLE: The Effect of Intensive Auditory Cues on Flight Passengers' Safe Behaviors and Attitudes

AUTHOR: Tianhua Li

MAJOR ADVISOR: Deborah S. Carstens, Ph.D.

When passengers do not follow in-flight announcements, injuries can occur. It is critical to draw passengers' attention and help them be aware of the importance of the instructions provided during in-flight announcements. Passengers' attention can be improved by providing intensive auditory cues before announcements. The auditory cues with intensive acoustic parameters increase the hearers' perceived level of urgency, and then, they are more likely to perform safer behaviors. Also, when the intensity level increases to an intermediate level, the stimulated persons should have better performance. However, when it increases excessively, their performance may be impaired.

The study examined the effect of the intensity level of auditory cues on passengers' performance mediated by perceptions, arousal levels, and attitudes. Auditory cues included five levels, one baseline, and four intensive levels. Performance referred to passengers' response time to instructions. Perceptions included the perceived level of urgency and the perceived level of risk. Arousal levels were determined by analyzing heart rate, heart rate variability, and skin conductance. Attitudes were defined by how annoyed passengers were with the cues.

The findings demonstrated that auditory cues with a moderately high-intensity level significantly increased passengers' performance and an exceedingly high level of cues impaired their performance. The intensity level had a positive effect on perceptions and attitudes. Although perceptions and attitudes were not linearly correlated with performance, they formed a potential inverted-U relationship with the performance. Arousal levels were not affected by the intensity level and did not have an effect on performance, possibly because passengers' physiological responses were not sensitive in a short period.

Table of Contents

List of Figures

List of Tables

Acknowledgments

This dissertation could not be finished without the help and support from my department and many professors. As the ultimate challenge of student lives, the dissertation has never been easy. It demands the student’s hard work and efforts, but financial and academic supports are imperative. Initially, I want to thank the College of Aeronautics for purchasing aircraft seats and providing a laboratory room for researching. I also want to thank the facility department for installing the seats in a safe and functional way.

For professors, I am grateful for the help of my dissertation advisor, Dr. Deborah S. Carstens. She dedicated hundreds of hours to review and edit my dissertation. Also, she provided me with devices to support my experiments. In addition, I appreciate Dr. Carstens for hiring me to work on an FAA grant, which offers me scholarships to pay my tuition. Another professor that I want to thank is Dr. Meredith Carroll because she provided me with equipment to measure physiological signals and introduced relevant theories to consolidate my study. She helped me resolve physiological data collection kindly and in patience. Without the devices, experiments could not be conducted. I also want to thank Dr. William Gabrenya and Dr. John Deaton for their support in statistics. I am grateful for Dr. Brooke Wheeler’s guidance of grammars and format and her dedication and preciseness when designing the experiment, determining lab setups (e.g., seat pitch), and build a simulated cabin, make curtains, and guide the facilities’ work. I also want to express appreciation for Dr. Michael Gallo for his help with my topic selection and also the manual that he created for dissertations. Last, but not least, I would like to thank Maria as she taught me to use the devices and software.

Dedication

I am dedicating this dissertation to my parents for their financial support. It reduced my burden and paved the way for me to take the program and accomplish my dissertation. Also, financial support enabled me to work on my doctorate degree energetically and dedicatedly.

I am also dedicating my dissertation to my girlfriend, Xin (Holly). Her encouragement and belief in me was imperative. Therefore, I had adequate confidence to finish the dissertation.

Chapter 1

Introduction

Background and Purpose

Background. Every year, many passengers are injured in commercial aircraft, which could be serious and even fatal, because they do not follow the instructions, especially the fasten-seat-belt instruction. There can be two common reasons for a passenger not to follow instructions. One condition is that passengers ignore the announcements. The other is that some passengers do not realize the importance of compliance though they notice the instructions. Therefore, it is necessary to make flight passengers attend to announcements and follow the instructions. Intensive auditory cues can help solve both problems. For attention, when the auditory cue is intensive, the stimulation becomes stronger to draw more attention to the cue. To make passengers follow the instructions, Wilde's (1982a) risk homeostasis theory (RHT) provides an approach. According to the RHT, human interventions and education do not help due to the homeostasis caused by the negative feedback; instead, only the factors that can change the target level of risk (TLR), which is the only element that is not in the closed-loop, can alter individuals' behaviors. However, it is not practical to change a passenger's TLR unless they can be awarded for compliance or fined for not following the instructions. In this case, changing the perceived level of risk (PLR) could be a feasible method. If intensive auditory cues induce higher PLRs, passengers will tend to make cautious behaviors to lower the risk that they are taking. In contrast, the actual level of risk (ALR) does not increase. Although Wilde believed a change in the PLR would not have lasting effects on car drivers' behaviors because they will be influenced by the negative feedback, the feedback that drivers receive is different from passengers. Therefore, if flight passengers can perceive a higher risk level than the ALR, they will be more likely to perform safe behaviors.

Hellier and Edworthy (1989), Edworthy, Loxley, and Dennis (1991), and Hellier, Edworthy, and Dennis (1993) studied the effects of acoustic parameters on perceptions of audio warnings' urgency. By modifying these parameters, audio warnings can be intensive and make people believe the situation is urgent and serious. For example, the change in the volume, frequency, or inter-onset intervals (IOI) of a warning can lead to other perceived levels of urgency (PLUs; Wang, Guo, Ma, & Li, 2016). Therefore, these intensive auditory cues may be appropriate stimuli to raise passengers' PLR. Also, they can draw passengers' attention better than typical auditory cues or no cues. Although it has been demonstrated that intensive audio warnings can convey a higher level of urgency, the researchers did not examine the effect of auditory cues. The subjects in past studies were mainly operators, but it is unknown whether the findings apply to passengers who do not have tasks. These gaps can be filled by conducting the current study.

Moreover, past studies did not reveal if a person with a higher PLU can have different performance. Based on the arousal theory, it is believed that the relationship between arousal and performance is consistent with Yerkes and Dodson's law (Duffy, 1957). An intensive auditory cue is very likely to increase the arousal level. The auditory cue with different acoustic parameters can result in different PLUs, so it can be assumed that intensive auditory cues can affect their arousal levels and further affect performance. In the current study, the arousal theory will be tested on passengers instead of operators.

Purpose of study. The main objective of the study was to test the influence of the intensity level of auditory cues on flight passengers' performance regarding compliance with safety instructions. Physiological signals that were used to identify arousal levels, perceptions, and attitudes (i.e., annoyance with these intensive auditory cues) were treated as mediating variables (MVs). Therefore, there was a primary purpose, which was to discover the effect of the independent variable (IV) on the dependent variable (DV), and two secondary purposes, which were to identify the effect of IV on MVs and to test the mediating effect of MVs between the IV and DV. The primary purpose of the study was to examine the effects of intensity levels of auditory cues (set A) on flight passengers' performance (set E). There were also two secondary purposes. One was to test the effects of intensity level (set A) on arousal levels (set B), perceptions (set C), and attitude (set D). The other secondary purpose was to study the mediating effects of arousal levels (set B), perceptions (set C), and attitudes (set D) on the relationship between intensity levels (set A) and performance (set E). The structure of the model is displayed in Figure 1.1.

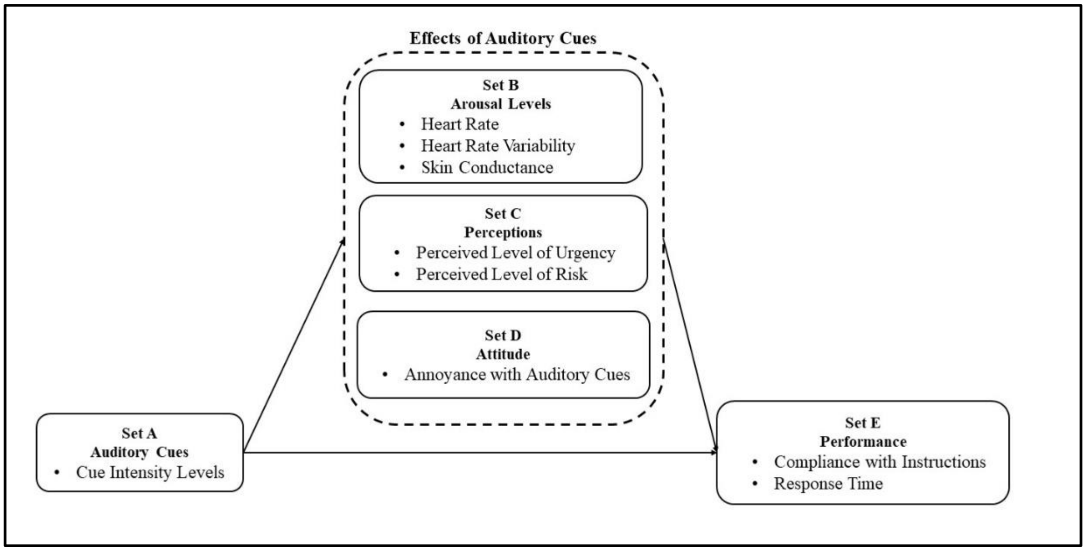

Figure 1.1. **Model of hypothesized relationships among interested factors.**

In the context of the study, the auditory cue was defined as a sound that draws passengers' attention before an in-flight announcement. The intensity level was defined as the extent to which the auditory cue with different sets of acoustic parameters can strike flight passengers. There were five intensity levels with variations of four acoustic parameters, which can impact passengers differently. The acoustic parameter was the characteristic of an auditory cue, including the frequency (0.5 and 1.0 kHz), waveform (sine wave and triangle wave), number of pulses (1 and 2 pulses), and length of each pulse (600 and 1,200 milliseconds). The performance was defined as compliance with instructions and response time.

Compliance meant whether or not the participant followed the instructions, and the response time was how many milliseconds it took the participant to initiate the action that was required by the instructions. The arousal level referred to physiological signals that reflected an individual's activation, and signals included the heart rate (HR), heart rate variability (HRV), and skin conductance. HR and skin conductance referred to the elevation instead of actual values. The perceptions referred to an individual's subjective judgments of the auditory cues. They included the PLU and the PLR. The attitude was annoyance with intensive auditory cues. The variables are summarized in Table 1.1.

Table 1.1

Variables and Sets

Category	Variables
Set A = Auditory Cues	X_1 = Cue Intensity levels
Set B = Arousal Levels	M_1 = Heart Rate
	M_2 = Heart Rate Variability
	M_3 = Skin Conductance
Set C = Perceptions	M_4 = Perceived Level of Urgency
	M_5 = Perceived Level of Risk
Set D = Attitude toward Auditory Cues	M_6 = Annoyance with Auditory Cues
Set E = Performance	Y_1 = Compliance with Instructions
	Y_2 = Response Time

Note. X = independent variable. M = mediating variable. Y = dependent variable.

Definition of Terms

Acoustic parameter. The acoustic parameter was the characteristic of sound. In the current study, acoustic parameters were operationally defined as the frequency, waveform, length, and number of pulses.

Arousal. Arousal was a physiological and behavioral concept, which exists in a continuum and can energize and direct behaviors (Duffy, 1957; Ribble, 2010). Its level can be influenced by the stimulation, baseline level, and stimulus sensitivity (Ribble, 2010). In the context of the study, arousal was defined as the activation generated by the stimulation of intensive auditory cues.

Arousal level. The arousal level refers to "how calming or soothing, versus how exciting or agitating, a particular stimulus or event is perceived to be" (Arousal level, 2012, p. 15). It was operationally defined as the physiological signals that reflect the level of the individual's arousal. The signals included HR (i.e., HR elevation), HRV, and skin conductance (i.e., skin conductance elevation). This is similar to the definition that Duffy (1957) provided to describe the concept of the arousal level, which was "variations in the arousal or excitation of the individual as a whole, as indicated roughly by any one of a number of physiological measures" rather than "the activation pattern in the EEG" (p. 265).

Attitude. Attitude is defined as a "mindset or perspective that determines behavior and conduct" (Attitude, 2013, p. 18). In the current study, attitudes referred to passengers' annoyance with auditory cues, which will be collected with a questionnaire.

Auditory cue. An auditory cue refers to "a sound signal that represents an incoming sign received through the ears, causing the brain to hear" ("Sensory cue," 2018). In the context of the study, an auditory cue meant a sound that was played before an in-flight announcement to draw passengers' attention.

Compliance. Compliance refers to "an overt, public action performed in accordance with a request from an external source," which "can be from another person(s) or from an object" (Mead, 2007, p.161). In the context of the study, compliance referred to an individual's response to safety instructions in announcements. There were two levels, complying and not complying.

Intensity level. The auditory cue's intensity level was the extent to which the auditory cue was striking to alert flight passengers. It was designed by altering four acoustic parameters, one at a time. They were frequency, waveform, number of pulses, and length of pulses.

Intensive auditory cue. The intensive auditory cue was defined as the auditory cue with a higher intensity level. The features of intensive auditory cues in the current study were a higher frequency, a more striking waveform, a longer length, and more pulses.

Perceptions. Perceptions are measurements of an individual's feelings about the auditory cues. The measurements included the PLU and PLR.

Perceived level of risk. In the context of the study, the PLR referred to how the individual thinks the situation is risky. It was measured to examine the RHT and risk compensation theory (RCT). It was measured by employing bipolar adjective scales.

Perceived level of urgency. The PLU was how urgent the individual believes the situation was. It was collected to determine the effect of acoustic factors on intensity levels. This factor was acquired by utilizing bipolar adjective scales.

Performance. Performance is defined as how well an individual can conduct the tasks. It included compliance with instructions and response time. The variables in this set were combined after in exchange for the practical data analyses.

Response time. The response time referred to the duration from the point that the action-related keywords in the instructions (e.g., keep your seatbelts fastened) were given to the point at which the individual initiated the responses.

Research Questions and Hypotheses

Research questions. Based on the purpose of the study, there were three research questions, which are shown as follows:

RQ1: What is the effect of auditory cues (set A) on passengers' performance (set E)?

RQ2: What is the effect of auditory cue intensity levels (set A) on arousal levels, perceptions, and attitudes (sets B, C, and D)?

RQ3: If the auditory cues (set A) have a zero-order relationship with flight passengers' performance (set E), to what extent is the relationship mediated by arousal levels, perceptions, and attitude (sets B, C, and D) variables?

Research hypotheses. As physiological signals were variable and sensitive to tiny stimuli, they were hard to control and predict. Therefore, the hypotheses were not made with respect to the research questions about physiological measures. According to the research questions, the hypotheses in this study were based on a hypothesized model (Figure 1.1) as follows:

H1: Auditory cues will have an effect on passengers' performance.

H2: Auditory cues will have effects on passengers' perceptions and attitudes.

H3: If the auditory cues have a zero-order relationship with flight passengers' performance, the relationship will be mediated by at least one of the perceptions and attitude variables.

Study Design

The study utilized a repeated-measures design. The IV was the intensity level of auditory cues and included five conditions. There were one baseline auditory cue and four intensive cues by altering one of the four parameters at one time: frequency, waveform, number of pulses, and length of each pulse. Each participant was measured in each of the five conditions, and each auditory cue could be compared to one another within subjects. The DV was the performance of participants, which included whether or not a participant complied with instructions and response time if participants followed it. Also, participants' HR elevation, HRV, and skin conductance elevation were analyzed to obtain their arousal levels. The participants reported their PLUs and PLRs after hearing each auditory cue to confirm the accuracy of the physiological signals obtained, and they also identified the annoyance with each auditory cue.

The session lasted for an hour, and the simulation took 40 minutes. Before the simulation started, participants were directed to sign informed consent forms and read instructions. Participants were then instructed to wear the Equivital vest that collected physiological data. The first six minutes of the simulation were for participants to relax so that their resting levels of each physiological signal could be measured. At the 6:00 minute, the first announcement was given. After that, other announcements were played at 10:30, 15:00, 19:30, 24:00, 28:30, 33:00, 37:30 minutes. Each announcement was approximately 10-second long. The interval between the two adjacent announcements allowed participants to recover from the previous auditory cue. Announcements could be action-required announcements or general announcements. Action-required announcements instructed participants to (a) fasten seat belts, (b) lower tray tables, (c) raise tray tables, and (d) adjust seatbacks, respectively. Auditory cues were played before each announcement in a counterbalanced order. After the simulation, participants were asked to hear auditory cues another time and reported their PLRs, PLUs, and annoyance with auditory cues using an adjective bipolar item. Each scale included two reverse-scored items, and each item had two extreme adjectives with nine boxes in between. The participant checked any box that he or she believed was appropriate to describe feelings about the auditory cue.

Significance of the Study

Every year, more than 50 flight passengers are injured because they do not follow instructions to fasten seatbelts when encountering turbulence (Hiatt, as cited in Davies, 2013). If passengers were to pay more attention to the announcements and were to be more motivated to comply with safety instructions, they would tend to keep themselves away from potential risks. It is suggested that audio warnings with different parameters can lead to different levels of individuals' perceived urgency (Edworthy et al., 1991; Hellier & Edworthy, 1989; Hellier et al., 1993). Therefore, an intensive auditory cue may also be an excellent method to draw passengers' attention and make them perceive a higher level of urgency and probably a higher level of risk. According to Wilde's (1982a) RHT, individuals tend to alter their actions to keep their PLR close to the TLR. In this case, perceiving a high level of risk can result in cautious behaviors, further facilitating their compliance. This study determined if intensive auditory cues increased passengers' PLRs and helped them perform safer behaviors. Also, as auditory cues are designed to be intensive, they can likely cause individuals' arousal levels to increase. According to the arousal theory, an increase in the arousal level can initially improve and then impair individuals' performance (Duffy, 1957). In this sense, the current study discovered the relationship between passengers' performance and arousal levels induced by intensive auditory cues. Additionally, the study recognized individuals' annoyance with intensive auditory cues. The results can be generalized to young and educated passengers with normal hearing abilities in the United States and may be able to be generalized to other passengers.

Study Limitations and Delimitations

Limitations.

1. ***Sample size.*** In the current study, the convenience sampling strategy was applied. Participants volunteered to take part in the study. If this study is replicated with a larger sample, results could be different, and non-significant results may show significant effects.
2. ***Demographics.*** As the sample was not randomly selected, the demographics might not represent the target population very firmly. The accessible population will be the students, faculty, and staff at the Florida Institute of Technology (FL Tech). Considering the demographics of flight passengers in the United States, the sample might not be representative of the target population. Therefore, the results might not be applicable to all the passengers onboard U.S. commercial aircraft. It is possible that population validity was impaired,

and population generalizability was limited. If the study is replicated with another sample, the results could be different.

3. ***Authenticity of responses.*** Because of the convenience sampling strategy, participants were volunteers, and they might not perform appropriately to provide decent data. Participants might not care about the validity of the findings and only sought to participate to receive rewards or extra credits. As it was a simulation study, the environment was different from a real cabin. The simulated cabin could not simulate the turbulence, so the participants did not feel the shaking of the fuselage. Therefore, participants may not have been as concerned about their safety as they might have been on an actual flight, and the arousal levels might not be as high as they might be in a real cabin. In addition to the mental difference, passengers on a real-world shaking aircraft may have higher arousal levels. In this case, the results from the current study might deviate from passenger behaviors and performance in the real-world. Despite this, every effort was made to create a realistic cabin.

Delimitations.

1. ***Auditory cue design.*** The auditory cues with different intensity levels that were used in the current study were created by manipulating one acoustic parameter at each time. By doing this, the number of auditory

cues was considerably reduced, which made the study practical. Nevertheless, it was not feasible to distinguish the effect of intensity level from the effect of the change in auditory cues.

2. ***Laboratory experiment.*** The experiment was conducted in a lab setting to control extraneous variables and increase internal validity. However, due to the different environment, the ecological generalizability as a part of external validity might be undermined. The outcomes may not be able to accurately reflect the passengers' performance on flights.
3. ***Data collection method.*** The instrument that was used in the current study, including the physiological data collecting devices and psychological responses collecting scales, was valid and reliable. However, the studies in which other devices or scales are applied could lead to different results. A replication that is conducted with different instruments may show different results. Also, physiological signals were not reliable. The signals could vary between-subjects and within-subjects. In this case, the results regarding arousal levels could not be applied to practice or be referenced by other studies.

Chapter 2

Review of Related Literature

Introduction

The importance of passengers' compliance with in-flight safety instructions is illustrated. Although it is critical to follow the instructions, some passengers do not comply due to three possible reasons. The first reason is that passengers do not notice the presence of announcements. However, the intensive auditory cues can increase passengers' arousal levels, and an individual's attention can be drawn to the stimulation that elevates the arousal level the most (Eysenck, 1982). The second reason is that passengers do not want to follow instructions because it is uncomfortable, or they are tired of following instructions. A third reason is that they do not see other passengers following the safety instructions. For the last two reasons, based on the risk homeostasis theory (RHT), if passengers perceive higher levels of risk, they are more likely to follow instructions from flight attendants and the captain. This is the reason why the RHT is also introduced. However, perceiving an extremely high level of urgency may not be beneficial with respect to passengers' compliance because it may increase passengers' arousal levels exceedingly. According to the arousal theory, if the arousal level is beyond the optimal level, the performance will be impaired. The arousal theory is discussed next. One way to increase the perceived level of risk (PLR) and perceived level of urgency (PLU) is to raise the intensity of the auditory cues, so it is necessary to introduce the influence of acoustic parameters on individuals' perceived urgency. In addition to theories, several studies are reviewed to support the theories and provide practical and effective methods to design the current study. The studies are about the RHT, the arousal theory, and the acoustic parameters of audio warnings. Finally, the theories and study reviews are summarized, and the implications of the current study are presented.

Overview of Underlying Theory

Importance of passengers' compliance. From 1980 to 2008, two-thirds of the passengers who were fatally injured during turbulence overlooked illuminated seatbelt signs and did not fasten their seatbelts (Davies, 2013). It is apparent that passenger's compliance with safety instructions is critical as the aviation industry strives to achieve zero accidents. It is not merely closely related to their own safety but also other passengers' safety. The purpose of aircraft seat belts is to protect passengers from being thrown out of seats when the aircraft encounters turbulence or has a hull breach ("In-Flight Seat," 2017; Scales, 2018). If a passenger does not wear the seatbelt, sudden turbulence can lift the passenger out of the seat and throw the passenger into hard surfaces (e.g., armrest) or even adjacent passengers (Toohill, 2015). Most of the time, turbulence is not serious, so passengers learn that fasten-seatbelt instructions are simply routines, and they usually place little importance on those announcements. However, when they realize the turbulence is severe, it is too late.

On December 28, 1997, United Airlines flight 826 was on its route from Tokyo to Honolulu. About one and a half hours after taking off, the aircraft suffered from severe turbulence. In only six seconds, the aircraft was carrying a load of -0.8 G, which was strong enough to eject occupants out of seats if they were not fastening seat belts. Although the seat belt signs were turned on, many passengers did not follow the instructions. In this accident, one passenger was killed, and 74 passengers received injuries ("1 Dead, Scores", 1997).

Similarly, on May 26, 2013, the Singapore Airlines Flight SQ 308 was flying from Singapore to London. When the turbulence occurred, passengers were having meals. Although passengers were instructed to fasten seatbelts, and the seatbelt lights were on, some passengers did not choose to comply with the instructions. Suddenly, the aircraft dropped 65 feet, and 11 passengers were injured (Moran, 2013).

In less than two months, two passengers aboard OZ 214 received fatal injuries because they did not follow instructions as well. On July 6, an Asiana Airlines flight OZ 214 was landing at San Francisco International Airport (SFO), and the aircraft hit a seawall at the airport due to the pilots' mistakes. In terms of the National Transportation Safety Board's (NTSB) report, two passengers did not wear seatbelts though they were asked to do so before landing; as a result, they were thrown out of the cabin during the impact and were fatally injured. Aarons

(2014) suggested the casualty could have been avoided if these passengers followed the instructions and fastened seatbelts.

In addition to the events that are introduced above, there are still many injuries due to not fastening seatbelts during turbulence, which were stated by Scales (2018) and AirSafe.com (Turbulence Accidents That Killed Airline Passengers, 2012). The airlines, aircraft types, numbers of injuries of these accidents are shown in Table 2.1. Reportedly, among non-catastrophic commercial aircraft accidents, the injuries of passengers and crew members were mainly due to the turbulence (“1 Dead, Scores”, 1997). An effective way to avoid this kind of injury is to comply with in-flight safety instructions. There are three major reasons why passengers do not follow the instructions. One reason is they do not notice the presence of the requirements, which are usually spread out via flight announcements. Another reason is that they do not want to follow the instructions. However, they are aware, probably due to the lack of motivation because it is not comfortable to fasten seatbelts, or the instructions are too frequent and boring. It is possible that they want to exchange precautions of unlikely dangers for comfort, or perhaps they merely want to be maverick. The third reason is they follow others who do not respond to instructions. The given reasons suggest the necessity of making passengers aware of important announcements and propelling them to comply with instructions. To accomplish the former, auditory cues are efficient to draw passengers' attention; as for the latter, intensive auditory cues could make passengers realize the significance of following instructions.

Table 2.1

Injuries Caused by Turbulence

Year	Carrier	Aircraft Type	Region	Fatal Injuries	Non-fatal Injuries
1980	Indian Airlines	B 737	India	2	0
1982	China Airlines	B 747	Hong Kong	2	0
1990	Eastern Air Lines	DC 9	Florida	1	25
1996	Air France	B 747	Burkina Faso	1	2

Risk homeostasis theory and risk compensation theory. The RHT and risk compensation theory (RCT) can be applied to help passengers recognize the importance of following safety instructions. Therefore, in this section, RHT and RCT are introduced. Compared to RCT, RHT is more popular but also controversial.

Mechanism of RHT. RHT is a population-level closed-loop process, which means the theory is applied to the population in general instead of an individual person. It contains six major elements, which are (a) target level of risk (TLR), (b) PLR, (c) desired adjustment, (d) adjustment action (e) resulting accident loss, and (f) lagged feedback (Hoyes, Neville, & Taylor, 1996; Wilde, 1982a). Five of the elements are included in the closed loop, whereas the TLR is not included because the input of the TLR is not affected by any other element in the closed loop (Wilde, 2014). In a closed loop, the changes in elements can be mitigated by the negative feedback of the system. The process is shown in Figure 2.1, which is adapted from "Target risk 3: Risk in everyday life," by G. J. S. Wilde, 2014. Wilde illustrated the process of the RHT with respect to vehicle operators that will be referred to as drivers in this section. In short, when operating a vehicle, the driver accepts a certain level of risk (i.e., TLR) to gain benefits and also perceives the current level of risk (i.e., PLR). Then, the driver compares the TLR with the PLR repeatedly and chooses the next actions accordingly to minimize the discrepancy. The action of each driver contributes to the accidental loss, and the information about the accident rate will come back to the drivers through some factors (e.g., the mass media, witnessing) in the long term. At this moment, when the driver perceives the level of risk, the newly-updated information is taken into consideration as well. Additionally, it is necessary to clarify the definition of the risk. Most people misinterpret the concept of risk as a probability; however, the term "risk" is supposed to be defined as the product of the event probability and its corresponding cost (Haight, 1986). In other words, when an individual is expecting or perceiving the level of risk, he or she does not simply consider what the outcome is or how likely the outcome would happen. Instead, the individual thinks about these two factors collectively. For example, there are two options: 10% chance of losing $1,000 and 0.1% chance of losing $1,000,000. Although the first option (10% ×

$1,000 = $100) has a higher probability of losing money, the second option (0.1% × $1,000,000 = $1,000) is riskier.

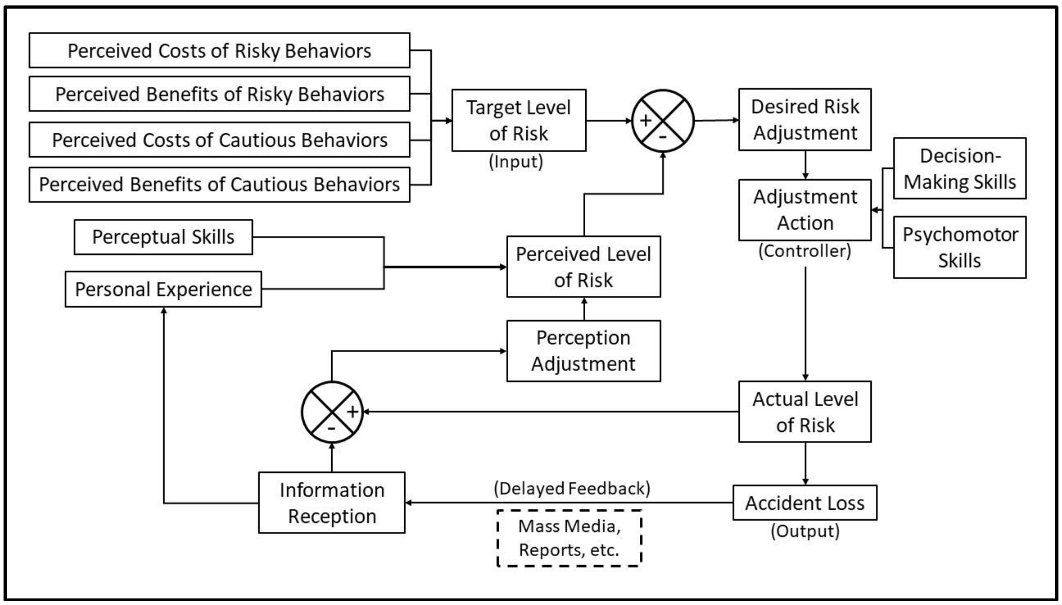

Figure 2.1. **RHT flow diagram.**

TLR is an individual's accepted level of risk, which is usually acquired intuitively; namely, it is "what feels right" (p.32) (Wilde, 2014). TLR is dependent on the expected benefits and costs of different actions. When performing a task, an individual usually unconsciously considers four motivating factors: (a) expected benefits acquired from risky behaviors, (b) expected losses due to risky behaviors, (c) expected benefits acquired from safe behaviors, and (d) expected losses due to safe behaviors, and they can be related to various areas (e.g., economy, culture, society) (Wilde, 1988; Wilde, 2014). By comprehensively weighing these factors, a

TLR is developed (Wilde, 2014). In addition to drivers, which have been interpreted by Wilde, these factors are also applicable to flight passengers. For example, regarding compliance with fasten seatbelt instructions, the expected benefits of complying with the direction can be a safe flight, whereas anticipated costs may be an uncomfortable feeling. On the contrary, the prospected benefits of not following the instruction are passengers having the ability to more easily move around their seats, and the expected costs may be injuries caused by sudden turbulence. Moreover, the factors can be long-term, short-term, or momentary (Wilde, 1982a; Wilde 1988). For commercial flight passengers following instructions, a long-term factor may be the obesity, whereas a short-term factor can be the back pain, and a momentary factor would be the need to go to the restroom.

PLR is a subjective level of risk, and the process is usually automatic as well (Wilde, 2014). Wilde identified three sources of the perception of risk. The first source is past experience. If the individual knows the task is difficult and dangerous, he or she will perceive a higher level of risk. Another source is the assessment of current situations. The third source is the extent to which the individual has confidence in the ability to deal with the situation. According to the sources, it can be implicated that flight passengers' PLR can be determined by (a) past flight experience as passengers, (b) the assessment of current situations in the cabin, and (c) the confidence in dealing with potential emergencies. The individuals who are in the same cabin do not necessarily have the same perception of the risk.

Therefore, there is always a discrepancy between the subjective level of risk (i.e., PLR) and the objective level of risk. It depends on the individual's perceptual skills.

In addition to expecting and perceiving the level of risk, individuals are also comparing the TLR with the PLR automatically, though sometimes consciously (Wilde, 1982a). Also, individuals come up with the desired adjustments to minimize the difference between TLR and PLR (Wilde, 1988; Wilde, 2014). If the PLR is higher than TLR, the individual will seek more cautious behaviors to lower the PLR; if the PLR is lower than the TLR, he or she will pursue riskier behaviors intuitively to meet the expectancy. For example, if a passenger feels the flight is smooth and safe, the passenger will perceive a relatively low level of risk and perform risky behaviors, such as walking around the cabin. By contrast, if the aircraft is shaking along with the captain's safety briefings, the passenger is more likely to fasten their seatbelts as tight as possible and pay full attention to the presented instructions. Once the difference is less than the just-noticeable difference (JND), the individual will remain at the current risk level of behaviors until it exceeds the JND (Wilde, 2014). Although people attempt to minimize the difference between the TLR and the PLR, what they want to eliminate is the difference between the TLR and the actual level of risk (ALR). However, as stated above, there is a discrepancy between the subjective level and the objective level. The minimal difference between the TLR and the PLR does not mean the equivalence between the TLR and the ALR. This difference is called the steady-state error (Wilde, 1982a). In other words, sometimes people think they are taking the risk at their expected levels; actually, they are receiving higher or lower levels of risk depending on their perceptual skills. From the perspective of the population, it is uncertain whether or not the steady-state error exists (Wilde, 1988).

After having the desired adjustment (i.e., a higher level of risk, the same level of risk, or a lower level of risk), the individual takes the adjustment action. At this step, the individual selected an appropriate way to perform the subsequent behavior to reach the TLR. Although the individual has an estimation of the direction and strength of the adjustment, it is possible that the person underestimates or overestimates the skills, and the difference between the TLR and the PLR still exists. According to Wilde (2014), for drivers, skills include decision-making skills and vehicle handling skills. For example, if a person is driving on a nearly empty highway, the driver will unintentionally behave more riskily to keep the PLR close to the TLR. However, if the driver has poor decision-making skills and overestimates the driving skill, the driver may choose to accelerate in exchange for time. In this case, the ALR exceeds the TLR. Therefore, aside from perceptual skills, decision-making skills, and psychomotor skills determine the difference between the TLR and the ALR as well (Wilde, 1988). As for commercial aircraft passengers, they may overestimate their abilities to fasten their seatbelts if needed (e.g., severe turbulence) before it is too late so that they choose not to follow the safety instructions in exchange for comfort.

Individual behavior can have an influence on the accident loss. It does not matter whether they underestimate or overestimate their skills because the ALR they are taking will be reflected on the accident loss. Wilde (1988) defined accident loss as the product of the frequency of accidents and accidents related costs. It can be analyzed from three different perspectives, which are spatial, temporal, and per capita. The spatial accident loss refers to the accident loss per a certain unit of distance, and the temporal accident loss refers to the accident loss within a certain duration. The accident loss for a person on average is measured per capita. Among them, only temporal accident loss is compliant with the RHT, and it is also the outcome of the closed-loop (Wilde, Claxton-Oldfield, & Platenius, 1985).

The accident loss will be acquired by individuals through lagged feedback (Wilde, 2014). This information will be taken into consideration the next time drivers perceive the current level of risk. Therefore, the fact that drivers overestimate their driving skills and choose to speed results in some accidents. Other drivers hear about these accidents through some means (e.g., from TV news, from the newspaper, from accident reports, from witnesses, on-site), making them reconsider their driving skills. In this case, they may have more accurate PLR and take more appropriate actions. The temporal accident loss will finally remain at a certain level. That is the complete process of the RHT closed-loop. Similarly, for those flight passengers who overestimate their abilities to fasten seatbelts in an emergency, they may hear about what happened on flight SQ 308 and decide to behave in a relatively safe manner.

Methods to reduce accident loss. Reducing the accident loss is a collective desire of society. A conventional approach is to apply human interventions, which contain two major types. The first type is the environment consisting of both road and car designers. Road designers aim to make the road error-tolerant (e.g., wide lanes, clear lines, emergency stopping lanes on both sides, road reflectors). Car designers also attempt to increase cars' safety, such as airbags, air curtains, energy-absorbing bumpers, and assistive warning systems. Another type is regulation. Law enforcement requires drivers to drive below the speed limits and not text while driving to prevent drivers from performing risky behaviors. Car manufacturers are mandated to meet safety standards and are encouraged to equip vehicles with safety-related technology. However, according to the RHT, all of these efforts are unavailing. Although the level of risk and the accident loss will decrease temporarily when the intervention is carried out, once people realize the ALR became low, they will make riskier behaviors to achieve the TLR (Wilde, 1982a; Wilde, 1988). Therefore, regarding a long-term influence on the population accident loss, the interventions are not sufficient.

Nevertheless, there are many opponents of the RHT, especially those who work for safety agencies, because their efforts are denied by the RHT. O'Neill and

Williams (1998) attempted to refute the theory and proved that the accident loss could be lowered through interventions. They pointed out that the per capita death rate did not significantly change in 1987 compared to 1927, but individuals were driving considerably more in 1987. For this reason, O'Neill and Williams claimed that the motor vehicle was safer than 60 years ago. However, Wilde (1988) had already solved this controversy. He stated interventions can indeed reduce spatial accident loss and may lead to some fluctuations in temporal and per capita accident loss. Still, after a long time, the temporal accident loss and the per capita accident loss increase back to the original level. This occurs because when people notice the spatial accident loss decreased, they will spend more time driving (Wilde, 2014). Although the fatality per kilometer drops, the driving distance per time unit raises accordingly, and the slope of these two variables was near -1 (Wilde, 1982b). Therefore, in the long run, interventions and innovations can change the spatial accident loss but not the temporal accident loss (Wilde, 1988). That was the reason why Wilde (1988) and Wilde (1989) emphasized the distinction between three types of accident losses and the concept of homeostasis (i.e., equilibrium instead of constant). Moreover, there was another flaw in O'Neill and Williams' debate. They talked about the motor vehicle death rate, whereas Wilde (1988) clarified the RHT dealt with the accident loss (i.e., the product of accident frequency and corresponding losses).

The second possible method is to enhance individuals' skills. Three skills that have been mentioned in the RHT can affect drivers' behaviors, which are perceptual skills, decision-making skills, and psychomotor skills (Wilde, 1982a). They can be increased through training and education. Perceptual skills influence the perception of risk (Wilde, 2014). The primary purpose of increasing perceptual skills is to eliminate the difference between the subjective level of risk (i.e., PLR) and the objective level of risk (i.e., ALR). In this case, the drivers are less likely to underestimate the risk they are taking, so they may have relatively cautious behaviors. Admittedly, it is also possible that the drivers overestimated the level of risk initially and have a lower PLR. The other two skills, decision-making skills and psychomotor skills, have an influence on adjustment actions (Wilde, 2014). By enhancing these skills, it can ensure that drivers can understand their capabilities better and make the most appropriate decisions correspondingly. Similarly, after training or education, their PLR may become higher or lower. Nonetheless, it is also not effective to increase any skill regarding the aggregate accident loss. The elements in the RHT process that are affected by these three skills are the PLR and the adjustment action. These two elements are in a closed loop, so the effects of any input should be offset after the long term because of the negative feedback. Therefore, efforts to increase individuals' skills do not have lasting effects on accident loss (Wilde, 1982a).

Another method that may be effective in reducing the accident loss would be increasing the PLR. As aforementioned, there are three sources of the PLR. These are the past experience, the assessment of current situations, and the confidence in coping skills. Among them, past experience cannot be changed via intentional interventions. Even if the driver is told that dangerous driving behaviors may lead to accidents, the experience shows it is safe to perform risky behaviors because the driver has not experienced an accident. The extent to which an individual is confident in the coping skills can be improved by safety training. However, regarding the PLR, the increase in confidence is bad for safety. When a driver believes he or she can deal with the current situations, the driver's PLR goes down, so the driver will behave more riskily to reach the TLR. Although it is also possible to impair their confidence by offering safety education, which informs drivers of the actual difficulty in coping with the situations, it may lead to worse performance with respect to the self-efficacy theory. According to Bandura (1977), self-efficacy refers to the extent to which an individual believes he or she can perform the task successfully. It determines whether or not the individual performs coping behaviors and how hard and persistent the individual will be. As for the assessment of current situations, there are several practical ways to reduce accident loss. Trimpop (1996) mentioned three situations in which people can have higher PLRs. The first one is people realize the real dangers. For example, most drivers do not know the danger of speeding partly because they did not witness or hear about the disastrous outcomes. If they are informed that when the speed is beyond 100 mph, the survival rate after tires blowing out is minimal, they will reassess the risk in which they are involved. The second situation is the economic fluctuations. Typically, the extra costs due to speeding (e.g., more fuel consumption, tickets, costs of repairing vehicles) are affordable, so some drivers decide to take more risks. However, if the individual or even the whole society encounters the economic issue, the PLR, when speeding, would become considerably higher than usual as the extra costs could bring a serious outcome to daily life. In this case, people will drive very carefully to prevent the PLR from being too high. The third situation is the individuals are provided with monetary incentives. If drivers are awarded a certain amount of money for zero accidents, they incline to drive safely. Although it is similar to a method to change the TLR, Trimpop (1996) considered this as changing the PLR. He also mentioned a problem of this method, which is it does not change the subjective level of risk. In other words, if there is no more incentive, their behaviors will be risky again. Nevertheless, Wilde (1982a) asserted that the PLR is in the closed-loop. It is admitted that the change in PLR will have a positive effect on safety in the short term. Still, because drivers will finally receive the information about the accidental loss and have lower PLRs, the temporal accident loss will rise to the previous level at last (Wilde, 1988).

TLR is the only element that is outside the closed-loop (Wilde, 1982a). According to Wilde's flow diagram, the TLR is not influenced by the feedback. In other words, only TLR can change the temporal accident loss entirely and have a lasting effect (Wilde, 1982a; Wilde, 1982b). Based on the four motivating factors that determine the TLR, Wilde (1982a) provided four techniques to lower the TLR, including (a) reducing the benefits of risky behaviors, (b) increasing the losses of risky behaviors, (c) reducing the losses of cautious behaviors, and (d) increasing the benefits of cautious behaviors. Wilde (1988) addressed that it is necessary to enhance safety from these perspectives, and the interventions that do not influence the TLR will not result in any outcomes in the long term.

Risk compensation theory. As mentioned before, many scholars argued against the RHT. The RHT emphasizes that it is the TLR rather than the environmental risk that has a significant influence on the accident loss (Hoyes et al., 1996). However, there are a large number of people who are making efforts to lower the environmental risk factor. Because if they accept the RHT, they have to admit that their contributions are meaningless. Also, O'Neill and Williams (1998) asserted that the RHT is merely a hypothesis but not a theory. Nevertheless, risk RCT is widely accepted. The RHT originated from the RCT (Trimpop, 1996). Wilde (1982a) stated that the word "homeostasis" is more appropriate than "compensation." The reason is that a single individual will not compensate for risk; rather, the individual makes behaviors according to the desired adjustment so that the per capita risk stays at the same level (Wilde, 1988). The difference between the RHT and the RCT is that the RHT states that the compensation is complete, whereas the RCT has partial compensation (Trimpop, 1996). Specifically, both theories support when there is an intervention or another kind of input to the system, individuals' behaviors will shift to the opposite side to somewhat offset the influence of the stimuli. The difference is that, regarding the RHT, the accident loss will recover back to the original level in the long term if it does not change the TLR, whereas, based on the RCT, the accident loss will not change to the desired extent. In this case, it does not matter whether the RHT or the RCT is correct. Compensation after interventions is needed in the current study.

Connection with the current study. The purpose of the current study is to motivate flight passengers to follow the safety instructions, such as fastening seatbelts and raising tray tables. Simply improving passengers' safety from the perspective of engineering is not helpful. If passengers are provided with a seemingly safer cabin environment, they may take more risk to achieve the TLR in exchange for benefits, such as comfort. According to Wilde (1982a), it is effective to change passengers' TLR to make them perform more cautious behaviors. However, the long-term TLR remains constant during a flight, and the short-term TLR is not likely to change within a flight either. Although a passenger's momentary TLR can change, the difference is too tiny to alter their behaviors. Unless passengers can be awarded for fastening seatbelts or be fined for not raising tray tables, the change in the TLR cannot have a considerable influence on passengers' behaviors.

Nevertheless, it is also possible to induce passengers to follow instructions by improving the PLR. As mentioned above, three factors affect a passenger's PLR: (a) past flight experience as a passenger, (b) the assessment of current situations in the cabin, and (c) the confidence in dealing with potential emergencies. The likelihood that their past flight experience and confidence in coping skills will change within a flight is minimal. Despite this, it is possible that passengers evaluate the situations differently under the influence of some stimuli, such as auditory cues. If passengers believe the announcement following the auditory cue is serious, their PLR will increase. As the TLR will not change significantly, the amount of risk that passengers want to take should decrease to remain the homeostasis. Under this circumstance, passengers are more likely to listen to the announcement and comply with instructions unintentionally. Therefore, if the auditory cue is striking and intensive, it may enhance passengers' safety awareness and minimize the presence of injuries.

Admittedly, the PLR is in the closed-loop, and any effect can be eliminated after individuals keep receiving feedbacks for a long time (Wilde, 1982a). There are some differences between vehicle drivers and aircraft passengers. First, drivers are fully aware of the surrounding situations. They can perceive the level of risk very accurately regarding the objective level of risk. Unlike drivers, flight passengers are passive, and they have limited sources to acquire information about real situations. For this reason, it is possible to increase passenger's PLR while keeping the ALR at the previous level. Second, the urgency is usually perceived unintentionally when people hear auditory cues. Although they may know it is simply an announcement, they may still perceive a relatively high level of risk when an intensive auditory cue suddenly appears. Third, flight passengers have different feedback because they are more apt to receive their feedback from mass media versus the first-hand witness. They are less likely to witness passengers receive injuries from severe turbulence. On the contrary, due to the number of car accidents, drivers have more opportunities to witness accidents. Whereas, passengers' injuries are less likely to be witnessed but more likely to be reported by mass media, and it has a stronger impact. Therefore, the feedback that is received when the PLR exceeds the TLR might not work effectively or efficiently for passengers in the cabin. In this sense, it could be feasible to change passengers' PLR by using intensive auditory cues and then increase the discrepancy between the ALR and the PLR. As a result, they will choose to behave cautiously and follow safety instructions.

Arousal theory. Nevertheless, it is not necessarily beneficial to keep increasing the intensity level of cues regarding the arousal theory. The arousal theory is a theory about internal motivation (Ribble, 2010). It is mainly about the relationship between an individual's arousal and his or her performance (Ribble, 2010). The relationship is congruent with the Yerkes-Dodson Law, which can be illuminated with an inverted U-shape curve (Figure 2.2). Before the arousal level surpasses the optimal level of arousal (OLA), which is illustrated as the summit of the curve, the arousal level has a positive effect on performance; when it exceeds the OLA, then there is an inverse relationship between the arousal level and performance (Ribble, 2010).

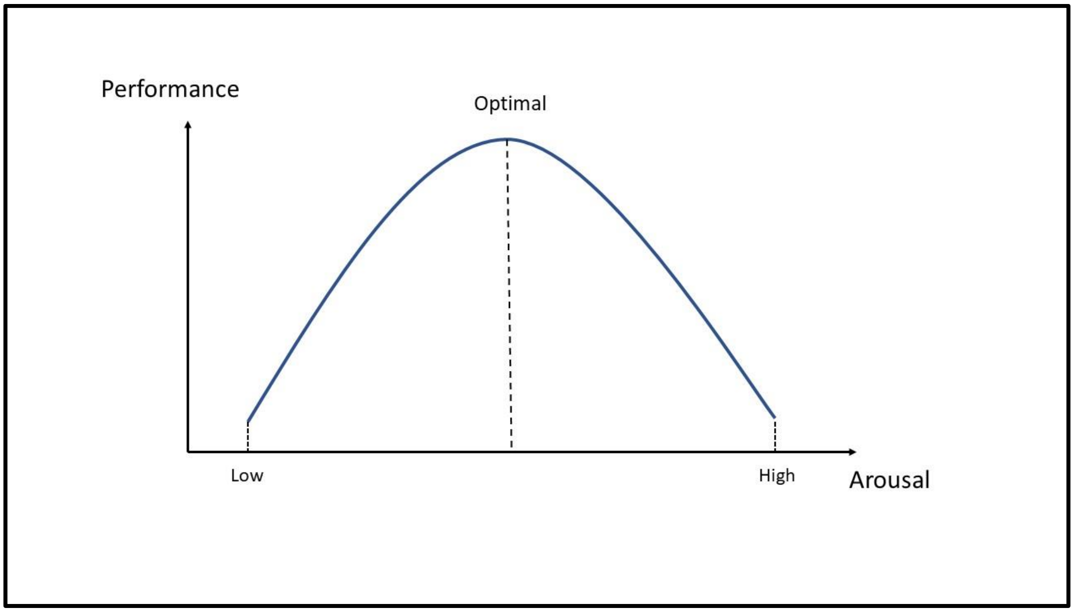

Figure 2.2. **Yerkes-Dodson law.**

Arousal. According to Ribble (2010), arousal is considered a drive, which motivates and directs individuals' behaviors. In the concept of arousal theory, arousal is the excitation of the whole individual rather than the activation pattern of electroencephalograph (EEG) (Duffy, 1957). Arousal includes psychological arousal and physiological arousal. From the psychological perspective, the arousal is generated by the autonomic nervous system (ANS); from the physiological aspect, arousal is controlled by the cerebral cortex and reticular arousal system (RAS) (Ribble, 2010). Based on Berlyne (1960), RAS is a nervous system that relates the most to attend. It reaches the whole cortex via multiple sensory tracts and collateral fibers, and it provides a path for excitation to travel from stimulation receptors to the cortex. When the receptor is stimulated, the information about the location and the quality of stimulations is sent via RAS and stored in the cortical projection area. The activation of the diffuse projection system makes the entire cortex alert by delivering impulses to it. After the cortex is alert, it transmits the information to motor functions via the RAS accordingly.

Arousal is reflected by the whole individual instead of a single system in the body (Duffy, 1957). It is a continuous variable varying from an extremely low level (e.g., sleeping) to an extremely high level (i.e., frantic excitation) around a central tendency, resting level of arousal (RLA), which is variable among individuals, to describe the temporal mental status (Berlyne, 1960; Duffy, 1957; Ribble, 2010). When the arousal is low, the individual is bored or relaxed, whereas the individual with a high arousal level shows anxiety or excitation. At any time, the arousal level is determined by (a) the stimulus level at present and in the near past, (b) RLAs, and (c) the sensitivity to the stimuli (Ribble, 2010). The stimulus could be a drug, hormone, physical exercise, and some motivations, and the effects of stimuli can be accumulated and last for a certain period (Duffy, 1957; Ribble, 2010).

Relationship between arousal and performance. The arousal theory is widely considered to originate from the Yerkes-Dodson Law (Ribble, 2010). It was established by Yerkes and Dodson (1908). They used the electric shock as a stimulus to force mice to learn to distinguish the white box from the black box. The findings showed that when the intensity of the stimulation is weak or too strong, mice learned slower compared to moderate stimuli. Also, the relationship between the strength of the electric shock and the learning speed is dependent on the difficulty of learning. The details of this study will be described later in the chapter. Afterward, Duffy (1957) stated the best degree of activation is moderate, and the relationship between activation and performance complies with an inverted U-shape. When an individual is in a state of over motivation (i.e., high level of motivation), the response is disorganized; when the individual has drowsiness or fatigue, which is caused by a low level of arousal, the impairment of performance appears as well (Duffy, 1957). Arousal is a concept that is similar to the motivation (i.e., electric shock) in the Yerkes and Dodson’s (1908) study, so the motivation in the Yerkes-Dodson Law can be replaced with the arousal (Ribble, 2010). By modifying Yerkes and Dodson’s findings, Daniel Berlyne developed the arousal theory and introduced the concept of the OLA, and he also maintained it was also applied to interest, curiosity, pleasure (Ribble, 2010; Strohminger, 2014). When an individual is conducting a task, a moderately difficult task can increase the arousal level, and an appropriate level of arousal further enhances the performance;

however, if the person is stimulated exceedingly and has a considerably high arousal level, it will lead to the impairment of the performance (Ribble, 2010). In addition to the general performance, Berlyne (1960) addressed that the degree of alertness also becomes the highest at the moderate arousal level.

Similar to the risk, the arousal level also stays in the state of homeostasis because the individual has the best performance and feeling at the OLA (Ribble, 2010). When the arousal level is low, the individual has the motivation to increase the arousal level toward the OLA (Ribble, 2010; Strohminger, 2014). This process is called the arousal boost (Berlyne, 1972). If the arousal level rises above the OLA, the individual attempts to decrease the arousal level, which is called the arousal jag (Berlyne, 1960). Both of these processes can enhance the degree of pleasure, and it is called arousal boost-jag (Berlyne, 1972). Moreover, from the homeostatic perspective, it can be explained that the characteristics of people can be affected by the RLA (Eysenck, 1982). Those who usually have low RLA have more tolerance of arousing things (e.g., noise, crowd, excitation); as for the individuals who frequently have high RLA tend to stay alone and avoid this kind of activities to prevent the arousal level from exceeding the OLA (Ribble, 2010). Nevertheless, it is also possible that a low-RLA individual has been immersed in an arousing condition for a period and would like to keep peace in a certain amount of time (Ribble, 2010).

Arousal and attention. Sheridan (2007) defined attention as “the focusing of sensory, motor, and/or mental resources on aspects of the environment to acquire knowledge” (p. 16). In the context of the arousal theory, Berlyne (1960) explained the attention from two different perspectives. One explanation is the extent to which the information can be transmitted from the external environment and internal stimuli to motor functions. Another one is the extent to which the capacity to send the information is occupied by the acquired information. From either perspective, when the demand capacity exceeds the actual capacity, some information cannot be processed. Berlyne asserted it is possible that the information from some sources is preserved, and the transmission of the stimuli from other sources is prohibited by the efferent fibers of the central nervous system. Otherwise, a certain part of the information from each source is transmitted, whereas other parts are not processed, but more errors may be made in this circumstance. Therefore, some behaviors are dependent on one stimulus, and some responses may be determined by partial stimulations from multiple sources. Among the stimuli, novel, variating, and complex stimuli are more likely to be attended to, such as an intensive auditory cue in the cabin.

As Nä tänen stated in 1973, a high arousal level does not lead to the impairment of performance directly (as cited in Eysenck, 1982). Arousal is generated by some stimuli which do not come from the main tasks most of the time, and attention tends to be paid to the source that increases the arousal level the most.

If the arousal level exceeds the OLA, too much attention is drawn to the source of stimuli, so insufficient attention can be paid to the main tasks. Therefore, individuals have bad performance in this situation. For example, flight passengers are aroused by intensive auditory cues, but their tasks are to follow the instructions. If the stimulus of an auditory cue is too strong, it will distract passengers' attention from their primary tasks, and it will impair their performance regarding safety instructions.

Measurements. Arousal level can be measured subjectively and objectively. The subjective measurement is simply to employ a self-report survey. Perala and Sterling (2007) stated acquiring the physiological information (e.g., stress, workload) using a self-report survey is convenient and cost-efficient, but it may not be accurate as participants report it based on their feelings. Also, it was always reported after the experiment, so participants have to recall their feelings. Therefore, Perala and Sterling suggested checking self-reported data with physiological measurements (e.g., galvanic skin response). By contrast, the objective measurement methods are various. In general, the information about the arousal level is usually acquired by measuring the individual's responses. Duffy (1957) pointed out two dimensions of responses, which are the direction and the intensity. The direction means the individual chooses to make this response instead of alternatives, and the response is positive or negative but not the opposite. The intensity, on the other hand, refers to the degree of arousal. These two dimensions are not correlated to each other, so they are always measured independently (Duffy, 1957). In the current study, only the intensity (i.e., arousal) is considered, so the direction of responses will not be discussed. The measurements include (a) the skin conductance, (b) muscle tension, (c) EEG, (d) the heart rate (HR), (e) the respiration rate, (f) the blood pressure, and so on (Berlyne, 1960; Duffy, 1957; Strohminger, 2014; Wilde, 2014). Although cortisol, epinephrine, salivary amylase, and some other physiological feedbacks are all the signs of the changes in the arousal level, they cannot provide continuous readings, and they are expensive to collect (Perala & Sterling, 2007; Strohminger, 2014).

Among these measurements, skin conductance is one of the most popular methods. Skin conductance is usually collected using galvanic skin response (GSR). When there is a change in the arousal level, sweat glands are stimulated, and the amount of sweat in sweat ducts and its concentration of electrolytes change accordingly (Galvanic skin response, 2009; Galvanic skin response, 2015; Perala & Sterling, 2007). Two electrodes are attached to the skin and measure the momentary changes in electric skin resistance to acquire the information about emotional arousal (Galvanic skin response, 2009; Galvanic skin response, 2012; Galvanic skin response, 2015; Perala & Sterling, 2007). When the arousal level increases, the skin conductance rises accordingly, and the GSR scores go higher (Berlyne, 1960). However, it is possible that a considerable change in the GSR appears with a trivial stimulus, or the GSR does not sensitively respond to the consecutive stimuli correspondingly (Duffy, 1957). As for other methods, EEG is also a common approach to detect the arousal level as well. The changes in α-wave rhythm can reflect intensive RAS activation (Berlyne, 1960; Duffy, 1957). Muscle tension, as an indicator of the skeletal-muscle functioning, has a direct relationship with the arousal level, but it may be different among individuals under the same stimulus (Berlyne, 1960; Duffy, 1957). As the arousal level increases, the HR, respiration rate, and blood pressure increase (Berlyne, 1960).

Connection with the current study. As stated in the previous section, it may be practical to use intensive auditory cues to increase the level of risk that flight passengers perceive. When the PLR is beyond the TLR, passengers choose cautious behaviors to reduce the risk that they are taking, so they tend to follow the instructions. Moreover, an increase in arousal caused by a stimulus can help the individual pay full attention to the related information and ignore others (Ribble, 2010). In this case, the auditory cues should be as striking as possible. Nevertheless, according to the arousal theory, the increase in the intensity of auditory cues does not necessarily always have positive effects on passengers' performance. If the arousal level exceeds the OLA, passengers will have worse performance, though it is not very likely for passengers to exceed the OLA in the cabin. Although they know the instruction is important, they may have a bad decision-making ability and perform worse than usual (Ribble, 2010).

Based on the statement of Ribble (2010), the determinants of the temporal arousal level contain (a) the stimulus level at present and in the near past, (b) RLA, and (c) the sensitivity to the stimuli. For an individual, the RLA and the sensitivity are thought to be constant within a flight, so the arousal level is simply a function of the stimulus strength. A sudden intensive sound, such as an auditory cue, is a strong stimulation, and it can lead to a rise in flight passengers' arousal levels. According to the arousal theory, the relationship between the arousal level and the performance follows an inverted U-shape curve. If the arousal level increases above the OLA, an exceeding amount of attention is paid to the auditory cue. Regarding the compliance with safety instructions, performing safe behaviors (e.g., fasten seatbelts, raise tray tables) are the tasks, whereas listening to instructions is not one of them. When the information about non-task stimulations occupies a majority of the information-transmitting resource, the organism may have an insufficient capacity to process the task-related information, so the passenger's performance of following instructions would be impaired. On the contrary, an auditory cue with an appropriate intensity provides a moderate stimulus, and passengers' arousal level increases to an optimal degree. Under this circumstance, they are more likely to comply with safety instructions. Therefore, it is necessary to determine whether or not auditory cues with a higher level of stimuli have a negative influence on cabin safety. If it follows the Yerkes-Dodson Law, it is critical to recognize the OLA to ensure passengers' performance.

Audio warning design.

Audio warning. As aforementioned, an intensive auditory cue is key to control flight passengers' behavioral patterns and optimize their performance. Therefore, it is necessary to understand the effect of sound parameters on passengers' attention, perceived urgency, and performance. There is a large amount of literature that is related to audio warnings, but they are not the same as auditory cues. Generally, conventional audio warnings include merely warning sounds, which can draw people's attention and make them notice abnormalities; whereas, the auditory cue in the context of this study is a sound that is played before an announcement. The purpose of the auditory cue is to help people be ready to listen to the announcement before it starts, so they will less likely miss the beginning of the announcement. The auditory icon, which is also being studied by researchers, refers to an informative warning sound. It can help people know which part is abnormal, so they do not have to scan the system to detect the issues, but auditory icons do not sound urgent (Belz, Robinson, & Casali, 1999). Training is always mandatory for operators to know what a specific sound means. The diagrams of the audio warning, the auditory cue, and the auditory icon are shown in Figure 2.3. For the first one, it is simply a warning sound. It can be long or short, but it does not contain any information. The second one, which is the auditory cue, is a short sound to draw people's attention to the following announcements. The third one, however, includes both an attention-drawing sound and the information.

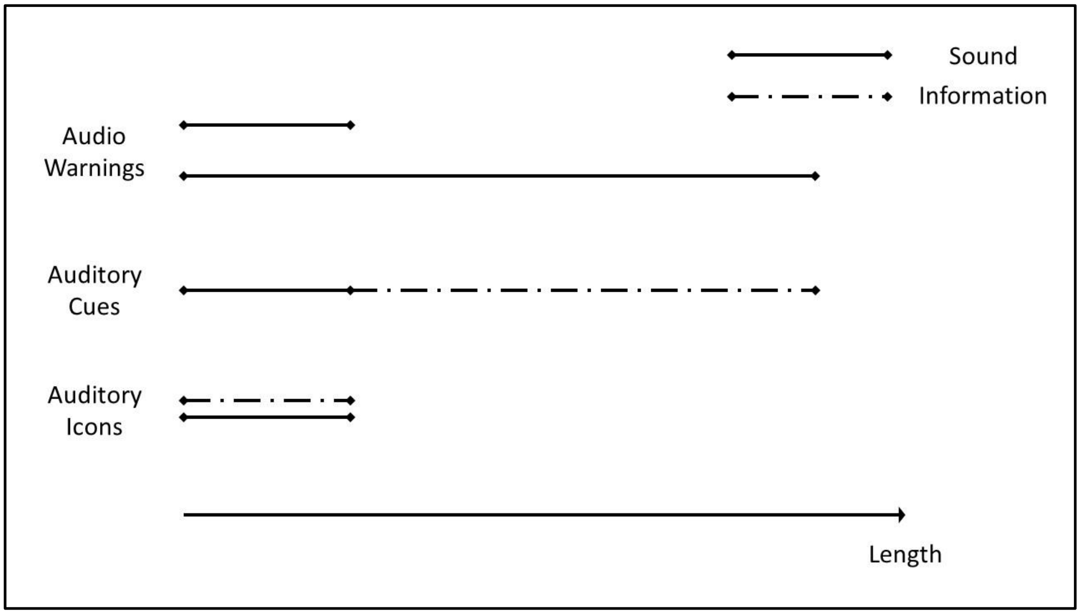

Figure 2.3. **Diagrams of auditory signals.**

Auditory cue versus visual cue. The reason why only the auditory cue but not the visual cue is researched in this study is that auditory sense is a major source to acquire warning information, and humans can automatically detect and process the warnings (Patterson & Mayfield, 1990). Besides, the audio warning is omnidirectional so that passengers do not have to look in a specific direction to be aware of the warnings (Sirkka, Fagerlönn, Lindberg, & Frimalm, 2014). Because passengers are a group of untrained individuals who are usually not ready for anomalies during the whole flight, it is not feasible to require and expect them to be prepared and have decent performance regarding responses to instructions. Moreover, auditory stimulations are stronger than visual stimulations, so they are more likely to generate responses (Berlyne, 1960).

Audio warning design. Hellier and Edworthy (1989) and Hellier, Edworthy, and Dennis (1993) applied Stevens' Power Law to quantitatively determine the effect of acoustic parameters on passengers' perceived urgency. Stevens' Power Law stated that the relationship between subjective parameters and objective parameters is compliant with the equation, $S = k \cdot O^{m}$, where S = the subjective value, O = the objective value, m = the ability to change the subjective parameters, and k = a constant. Hellier and Edworthy (1999) concluded the results of these two studies and showed the order of effects of different parameters (Table 2.2). It can be seen that the speed has the most significant effect, whereas the inharmonicity has a minimal effect.

Table 2.2

Effects of Acoustic Parameters

Order	Acoustic Parameters	Exponent
1	Speed	1.35
2	Repetition	0.50
3	Length	0.49
4	Frequency	0.38
5	Inharmonicity	0.12

Note. Speed = pulse rate. Repetition = the number of two-pulse units.

Nevertheless, the effects of various factors are acquired using different units. For example, the unit of length is the millisecond, but the unit of frequency is hertz. The exponents are based on different standards. If the unit of frequency were changed from hertz to kilohertz, the exponent would be 4.17 instead of 0.38, and it would become the most effective factor. Fortunately, Edworthy, Loxley, and Dennis (1991) integrated multiple parameters and compared the effects among combinations of seven factors. The characteristics of the warning sound that was considered the most urgent included (a) standard envelope, (b) random harmonic regularity, (c) short pulse-to-pulse interval (150 ms), (d) regular rhythm, (e) high average frequency (600 Hz), (f) long pitch range (300 Hz), and (g) random pitch contour. On the contrary, the combination that conveyed the least urgency were (a) slow offset, (b) regular harmonic regularity, (c) long pulse-to-pulse interval (550 ms), (d) slowing rhythm, (e) low average frequency (290 Hz), (f) short pitch range (75 Hz), and (g) downward pitch contour. Nevertheless, the lowest average pitch was 175 Hz, and the shortest pitch range was 50 Hz. In other words, the least urgent combination should be the sound with 175-Hz average frequency and 50-Hz pitch range, but the researchers did not test this combination. That was why the sound that was considered the least urgent did not have the lowest average frequency and the shortest pitch range. This implies that these two parameters do not have substantial effects on perceived urgency.

Wang, Guo, Ma, & Li (2016) demonstrated the relationship between pilots' perceived urgency and their reaction time, where the reaction time is an aspect of performance. In other words, if flight passengers' perceived urgency can be raised by changing the acoustic parameters of warning sounds, they can have better performance. Now the influence of acoustic parameters on the perceived urgency is apparent, so it is necessary to decide the appropriate ranges for each parameter on commercial flights. Patterson (1982) established guidelines for audio warning systems for cabin crews on commercial aircraft. Although they are for audio warnings, not auditory cues, and they are made for pilots instead of passengers, some of them can be referenced when designing auditory cues. In the guidelines, Patterson discussed the volume, the pitch, the length, and the frequency.

Initially, Patterson (1982) suggested that the volume of warning sounds should be appropriate, which is between the minimum volume that can draw people's attention and the maximum volume that may be annoying. The optimal volume should be higher than the masked threshold by between 15 and 30 dB. If the volume is too low, the warning sound may be covered by the background noise, and the crew is less likely to attach enough importance to the warning; if it is too high, the warning sound would be annoying and bad for communications, and the crew may have to cancel the warnings first rather than deal with the issue (Patterson, 1982). In the case that the background noise is loud, increasing the volume from the masked threshold by 25 dB would be enough. However, the masked threshold is inconsistent, and it varies across the frequency. For example, Patterson calculated the appropriate volumes for the warnings on the Boeing 727. At around 2 kHz, the volume should be between 90 and 100 dB, whereas 60 to 70 dB would be high enough when the frequency is at 5 kHz. When the background noise is stationary, the masked threshold can be predicted using the theory of auditory masking, which is very complex, so the process is not introduced. The results that Patterson calculated will be used in the current study.

Moreover, if the warning sound is played at the maximum volume from the beginning, it will be too startling. In this case, people always make involuntary and not appropriate responses to the warnings with this kind of characteristics, so an onset is necessary (Patterson, 1982). The onset refers to the period in which the sound from the beginning to the maximum or 90% of the maximum volume; similarly, the offset refers to the period in which the volume decreases from the maximum to the end (Edworthy et al., 1991). Admittedly, a short onset can convey the urgency and severity very effectively, and a rapid change in the volume can draw attention. Still, the onset should not be too sharp. Otherwise, there would be very little distinction from the sound without an onset, which could impair pilots' cognitive skills. Therefore, Patterson (1982) asserted that the onset rate should be between 1 and 10 dB/ms. This statement about the onset is consistent with the arousal theory. As for the offset, it can be simply symmetrical to the onset.

The optimal lengths of the onset and the offset that Patterson (1982) mentioned are 25 ms. Also, the duration between the onset and the offset should be long enough for people to detect it under the influence of the background noise (Patterson, 1982). The integration time of the auditory system is usually less than 50 ms, and the length of the whole pulse, including the onset, the offset, and the steady-state part, needs to be at least twice the length of the integration time

(Patterson, 1982). However, if the length of the pulse is too long, such as more than 150 ms, it will increase the length of the burst (i.e., the combination of all the pulses), negatively affect communication, and limit the variations of the warnings. Therefore, Patterson addressed the length of the steady-state part should be around 100 ms.

Furthermore, according to Patterson (1982), in the cockpit, the background noise, especially when cruising, that is below 0.5 kHz is louder than the noise above 0.5 kHz, so the low-frequency warning needs to be at a high volume to ensure that pilots can notice it. Typically, the warning sound below 0.5 kHz has to be at least 85 dB, but being immersed in a loud warning sound for a long-time impairs the hearing ability. However, the warning sound does not have to be louder than the noise. Instead, it can be at the frequency at which the noise has a low volume, so the frequency of warning sounds can be over 0.5 kHz (Patterson, 1982). Also, concerning the ability to hear high-frequency sound decreases as the person is growing old, the frequency should not exceed 5 kHz so that all of the people can be warned. Although people can easily hear a 5-kHz sound, it is still harmful to hearing (Patterson, 1982). Therefore, Patterson concluded that the range from 0.5 kHz to 5 kHz is acceptable, and 1 to 4 kHz would be better.

Connection with the current study. According to previous studies, a significant part of the information about the effect of acoustic parameters on people’s perceived urgency and the performance can be acquired. Although the parameters were researched from different perspectives among the studies, the results appeared to be consistent. It can be concluded that (a) the frequency, (b) the volume, (c) the number of pulses, (d) the interval between pulses, (e) the number of repetitions, (f) the length, (g) the irregularity, (h) the rhythm, (i) the pitch, and (j) the inharmonicity have effects. Therefore, these parameters can be considered when designing auditory cues in the current study.

Moreover, Patterson (1982) established guidelines for cockpit audio warning systems. Admittedly, these guidelines are for the cockpit, but it applies to the cabin considering the purpose, the environment, and human nature. For this reason, the range of the frequency, the range of the volume, the pitch, and the length of the auditory cues that will be used in this study also need to reference the guidelines.

Review of Past Research Studies

After understanding the theories related to this study, it is also necessary to review relevant studies previously conducted. By doing this, the findings and outcomes of different research can be acquired. Also, the design and procedure of other experiments can be referenced when designing the current study.

Risk homeostasis theory. The RHT is a population-level theory, which can only be verified in a society with millions of interactive people; also, it requires several months or years to achieve the status of homeostasis (Hoyes et al., 1996; Wilde, 1982a). In other words, there are three necessary conditions for the RHT, which are (a) a large population, (b) an extended period, and (c) the interaction within the society. By contrast, in a lab, the number of individuals and the amount of time is limited, and researchers usually conduct separate observations (Wilde et al., 1985). Despite this, Hoyes et al. (1996) addressed if the necessary feedback is timely, it is also possible to identify short-term behavioral compensations in a lab. Therefore, in a simulated environment, only the compensation but not the homeostasis can be studied.

Wilde et al. (1985) designed a study to test individuals' risk compensation behaviors in a controlled situation. In the study, participants were provided with a box, which included a stimulus light, a response button, and a rest button. The light would be turned on automatically, and participants were requested to push the response button within 800 ms to cancel the light. If it was at or beyond 800 ms, there would be a probability that the participant received a financial penalty (i.e., ¢ 10); if the response was made less than 800 ms, they would be awarded. The amount of the award was dependent on the extent to which the response time was close to 800 ms, and it was from ¢1 to ¢ 5. After the participant canceled the light, the light was turned on again in a random period between 700 and 1,600 ms. The participants consisted of 110 male college students, and they were divided into two groups. Both groups were given a 10-trial practice session followed by an additional 150 trials. Participants received feedback from the system for some of the trials so that they could know their actual response time, the amount of money earned or lost for the current trial, and their aggregate earnings. The settings of these 150 trials are shown in Table 2.3. Afterward, both groups were given 25 additional trials with feedback to test their timing skills. In this session, the closer the response time was to 800 ms, the more awards they could gain. However, they would still receive awards even if the response time was longer than 800 ms. Each participant was given $2 initially, and they could keep the money that they gained. Moreover, participants completed a questionnaire, which was to test their risk-taking tendency. The top 20 participants who scored the highest were considered risk seekers, whereas the 20 participants who had the lowest scores were classified as risk avoiders.

Table 2.3

The Probability of Loss and Feedback for 150 Trials

	Feedback	Probability of Loss	
		Group 1	Group 2
First 50 Trials	Yes	0.3	0.7
Second 50 Trials	Yes	0.7	0.3
Third 50 Trials	No	0.7	0.3

The results of Wilde et al.'s (1985) study showed that risk seekers made responses longer than 1,000 ms significantly more frequently than risk avoiders (p = .0340). However, regarding the times exceeding 700, 800, and 900 ms, there were no significant differences between risk seekers and risk avoiders. The average response times for the two groups were not considerably different either. Also, the mean response time was significantly higher when the probability of loss was 0.3 (M = 682 ms) in comparison to 0.7 (M = 653 ms) with p = .0002. The difference in response time between participants with and without feedback was not significant.

Interestingly, when participants were provided with feedback, if they did not receive penalties on the preceding trial, risk seekers were more likely to take risks and respond longer than 800 ms on the next trial. When no feedback was given, participants tended to make cautious responses. As for the earnings, no significant result was found regarding the risk-taking tendency (p = .2875). However, there were significant differences in earnings between with and without feedback (p = .0111). Similarly, the earnings were different between 0.3 probability of loss and 0.7 (p = .0123). It was apparently not compliant with the RHT. According to the RHT, the earnings should be the same across different probability of loss. Despite this, the result did not contradict the RCT. Additionally, researchers studied the effect of skill on performance. The skill was scored using the difference between their response times and 800 ms on the 25 skill-testing trials. Risk seekers had almost the same skills as risk avoiders. Also, no significant relationship was found between skill and either response times or earnings. This finding confirmed Wilde's (1982a) statement that increasing individuals' psychomotor skills do not affect the outcome lastingly. Moreover, participants made cautious responses when the probability of loss was high, and the responses became risky when the probability of loss was low. It indicated risk compensation could happen in a simulated environment within a short period. Besides, if they took the low-loss-probability session at first, they would perform more riskily in the subsequent high-loss-probability session.

Concerning the current study, it was found by Wilde et al. (1985) that when the probability of loss was high, individuals tended to take risks to a lesser extent. Therefore, it can be expected that flight passengers would be concerned more about their safety if they feel the probability of abnormality is high. In other words, if the crew can make passengers perceive a higher level of risk (e.g., make announcements after intense auditory cues), passengers would perform more cautiously. It supports the hypothesis. However, the result of Wilde et al.'s study also showed that if risk-takers did not fail on the first trial, they would perform more riskily on the subsequent trial. If it is applied to the cabin, it is possible that if the passengers do not receive injuries the first time they hear the intense auditory cue, risk-takers will not follow the instructions the next time the intense auditory cue is played. A difference between the experimental results and the real situation in a cabin is that participants could receive feedback instantly, knowing the assessment of their performance and outcomes. In contrast, flight passengers can simply know the outcome, which is whether or not they are injured, but not the discrepancy between their performance and the demanding performance, namely the adjustment actions they need to take. The result also demonstrated that when there was no feedback, people were more likely to behave more cautiously. It indicated that passengers, who only receive limited feedback, are not very likely to have risky behaviors though they are not injured before. Moreover, the finding that there was no effect of skills on the earnings implied that the implementation of passenger training is not helpful to reduce the causalities.

Hoyes et al. (1996) conducted a within-subjects study to test the validity of RHT. They recruited 70 participants. Each participant was requested to drive on a simulator in two different sessions. In the first session, frontal cars were moving at lower speeds, which were between 20 and 30 mph. In the second session, which was treated as a riskier situation, the speeds of the cars in front of the simulated driving position were between 30 and 40 mph. The researchers focused on (a) overtakes, (b) collisions, (c) driving behaviors, and (d) speeds. For overtakes, a risky overtake refers to passing when an oncoming car is in the driver's vision. Collisions included the collisions with other cars and the collisions with curbs. Driving behaviors were the distance from the car ahead and the position regarding the line and the curb.

One of the findings of Hoyes et al.'s (1996) study was when the environmental risk was high, risky overtakes were significantly fewer than a low-risk condition ($t = 2.00$, $p < .05$). Similarly, aborted passes were substantially more frequent in the second session ($t = 3.15$, $p < .01$). However, the number of successful overtakes were significantly more in the low-risk condition ($t = 1.96$, p

< .05), and there were considerably fewer collisions with other cars ($t = 1.70$, $p < .05$). The total number of overtakes and the number of collisions with the curb were not significantly different between the two conditions. Moreover, as for behaviors, participants were closer to frontal cars when they were driving in a riskier condition ($t = 2.67$, $p < .01$). It apparently contradicted the RHT. Theoretically, they should increase the distance from the cars in front of them to make behavioral compensation, but their performance was the opposite of this theory. Hoyes et al. explained that they did not need to overtake cars in the high-risk condition, so it was not necessary for them to keep enough space for the maneuver. The difference in road position (i.e., the distance from lines and curbs) was not significant. Besides, the mean speed was significantly higher when the environmental risk was high ($t = 8.98$, $p < .0001$). Higher speed was considered a riskier behavior, and it was not compliant with the RHT as well. However, other cars were moving at between 30 and 40 mph in the high-risk session, whereas the speeds were only between 20 and 30 mph in the other session. There was no wonder that the mean speed was higher in the second session. Also, the mean speed in the low-risk condition was nearly 35 mph, which was higher than other cars' speeds, but it was merely 40 mph in the condition of high environmental risk. It could imply that participants behaved more riskily in the low-risk situation considering the relative speed. The gas pedal position was not significantly different between the two conditions. In contrast, the brake pedal use was

significantly more frequent in the high-risk condition ($t = 4.53$, $p < .0001$) partly because they had to reduce the speed when aborting an overtake. It was also found that the mean steering wheel position differed significantly ($t = - 3.97$, $p < .0001$). The steering wheel was manipulated more frequently in the low-risk condition partly because participants made more overtakes.

First, the results of Hoyes et al.'s (1996) simulation also supported that when the environment is risky, individuals are more likely to make cautious behaviors and abort risky behaviors, which paved the way for the current study. Despite this, the difference in accident rates between the safe condition and the risky condition was not significant, which demonstrated the intervention was not effective. Although the study did not present the difference in the frequency of accidents between safe and unsafe situations, flight passengers cannot receive information timely about the injury rate in an emergency, and it will inhibit them from knowing the ALR. Furthermore, it can be implicated from the Hoyes et al.'s study that behavioral compensations could happen and be detected in a short-term laboratory condition. Considering fewer collisions in the high-risk environment, behaviors were over-compensated. As the RHT states the initial compensation can be perfect, imperfect surplus, or imperfect deficit, the results did not show incongruence with the theory (Hoyes et al., 1996). Additionally, the interval between the two sessions was only 10 minutes, but they were still able to identify the changes in the risk and make compensations. It indicates the variation of the environmental risk in addition to the feedback, leads to the changes in the PLR and thus brings out behavioral compensation instantly. An intensive auditory cue may be able to make passengers perceive a higher level of environmental risk. However, due to the inaccurate estimate of the ALR, the compensation may not be precise regarding the difference from the TLR. It is compliant with the RCT. Furthermore, the discrepancy between the PLR and the ALR is mainly determined by feedback. Individuals keep altering their behaviors to reach the TLR according to feedback so that the accident loss appears to be homeostatic. As the feedback is always delayed, there are fluctuations after the environmental risk changes. This explains the RHT. In this case, if the feedback is not delayed, risk compensation can be detected, and the compensations should be nearly perfect rather than surplus or deficit. As for the situation in the cabin, flight passengers can easily notice the changes in the environment but have extremely delayed feedback on the average rate of injuries. Therefore, behavioral compensation can be made if passengers perceive dangers by hearing intensive auditory cues, whereas the homeostasis may not be achieved.

Arousal theory.

Yerkes and Dodson's study. It is widely accepted that the arousal theory is on the base of the curvilinear relationship between arousal and behaviors, which originates from Yerkes-Dodson Law (Ribble, 2010). Therefore, it is necessary to introduce the study "the dancing mouse," in which Robert Yerkes and John Dodson discovered the relationship. Yerkes and Dodson (1908) intended to help mice learn to enter the white box. They prepared two boxes in the experiment area. One box was covered with white cards, and the other box was covered with black ones. Cards were switched between two boxes no longer than four sessions. The boxes had the tops opened so that the light was not blocked by the tops. Only one of the 40 mice sampled was involved in the experiment in each session, and each mouse attended only one session each day. After the mouse entered the experiment area, it encountered the two boxes. If the mouse chose the white box, it would be fine; whereas, if it entered the other one, it would be punished with an electric shock. The independent variables (IVs) included the strength of the light (i.e., the difficulty of learning) and the intensity of the stimulation (i.e., the circuit of the electric shock). Each IV had three levels (i.e., low, moderate, and high). The dependent variable (DV) was the number of times the mouse tried until it entered the white box in three successive sessions, which means the mouse learned that it should enter the white box. The scores were analyzed using descriptive statistics.

In the first experiment of the Yerkes and Dodson (1908) study, the light strength was moderate. It was found that when the intensity of the electric shock was moderate, it took mice the least amount of time to acquire the habit. Weak and strong electric shocks had similar results, and they led to poor performance. In the second experiment, the brightness was high, so the difficulty decreased in comparison to the first experiment. As the intensity of the stimulation rose, mice performed better and learned in the shortest time when the electric shock was the highest. Researchers believed the difference between the two experiments was due to the difficulty, so they conducted another experiment to perform further testing. In the third one, the light was the darkest, which means it is the most difficult condition for mice to acquire habits. The results showed mice had the best performance with moderate electric shocks, and it confirms their previous conclusion and the assumption. Except for the moderate level, mice's behaviors were worse with strong stimulations than weak stimulations.

Yerkes and Dodson (1908) made several conclusions, which can be summarized in two statements. The first statement is that the learning speed increases as the stimulation intensity is enhanced; after a certain level, the performance is impaired and keeps decreasing as the stimulation intensity increases, and strong stimulations could be worse than weak ones. Another statement is the relationship between the intensity of the stimulation, and the learning speed is affected by the learning difficulty. The more difficult the task is, the lower level the optimal intensity is at.

Although Yerkes and Dodson's (1908) findings are considered the basis of the arousal theory and some other similar theories, they are not flawless. For example, Eysenck (1982) commented that this study, as well as related studies that focused on the Yerkes-Dodson Law, usually included three or at most four levels. As Duffy (1957) maintained, the arousal level is continuous; more levels are necessary to see the pattern of the relationship. Moreover, it may be correct for

Yerkes and Dodson to conclude that the strong stimulation led to even worse results than the weak stimulation. However, if different levels of stimulations are employed, the difference in the performance between the strong and the weak stimuli may be changed. Additionally, the effect of the difficulty on the arousal is valid only when the arousal is caused by aversive stimulations. If it is the incentive that leads to an increase in the arousal, the effect does not exist (Eysenck, 1982). Despite this, an intensive auditory cue is an aversive stimulation, so it should be compliant with the statement of Yerkes and Dodson.

Effects of music on driving performance and arousal levels. Ünal, de Waard, Epstude, and Steg (2013) studied the influence of music on drivers' arousal levels and performance. Participants were instructed to follow the simulated car in front of them on a two-lane road, and they were told to drive at the same speed as the frontal car and keep a close and safe distance. Also, the other lane is always occupied by the oncoming traffic to avoid participants from passing the car ahead of them. The frontal car was moving at a random speed, which was between 60 and 80 kph, and the speed changed after 10 to 40 seconds. There were two IVs, including a within-subjects variable and a between-subjects variable. The within-subjects variable was whether or not participants drove with music. As for the music that was being played, participants had the opportunity to select their own playlists, so participants were familiar with and liked the music they listened to while driving. Besides, Ünal et al. employed a questionnaire to ensure that they liked the music. The between-subjects variable was the level of the volume, which included loud (85 dB) and moderate (70 dB). The DVs consisted of two categories, which were the performance and the arousal level. The performance was acquired by measuring (a) the delay in response, (b) the accuracy of car-following, and (c) the lane-keeping performance. The delay in response means how long it takes the participant to change the speed after the speed of the simulated frontal car changes. The accuracy of car-following was the correlation coefficient between the participant’s driving speed and the speed of the car in front, where 1 indicates perfectly accurate, and 0 means not accurate at all. The lane-keeping performance refers to the standard deviation of lateral positioning (SDLP). A low score corresponds to good performance, whereas a high score reflects atrocious performance. The arousal level was measured objectively and subjectively. Objective measurements were comprised of (a) EEG, (b) the mean HR, and (c) the HR variability. For both the EEG and the HR, there were resting measures, which were taken prior to and after the simulation, and task measures, which were conducted during the simulation. Also, researchers directed participants to self-report their deactivation to measure their arousal levels subjectively. The self-reported factors were boredom, tiredness, sleepiness, and energy level. The results indicated that the higher the score, the higher the level of deactivation, and the lower the level of arousal. The researchers made three hypotheses based on the findings of previous studies. The first one was that there was no adverse effect of the presence of music on task performance. The second one stated (a) that the music could increase the arousal levels and (b) that the music with a high volume could increase the arousal levels. The third hypothesis stated that the heart rate variability (HRV) would be higher without music than with music.

Ünal et al. (2013) discovered a significant main effect of the volume on the accuracy of car-following ($F = 2.77$, $p < .05$). It showed participants had better performance when the music was loud, which demonstrated loud music did not impair but improve the performance on following the lead car. Although the omnibus main effect of the presence of music was also significant, when researchers analyzed the data for every 5-minute segment, no effect of the music presence on the accuracy of car-following was significant. Therefore, the presence of music did not influence the performance regarding following the lead car. As for the delay in response, there was a significant main effect of the presence of music, which demonstrated that participants responded to the change in the speed of the frontal car faster when listening to music in comparison to not listening to music ($F = 3.30$, $p < .01$). Also, the main effect of the volume was significant as well ($F = 5.28$, $p < .05$), showing participants could make responses faster when the music was loud. For the other measurement on the performance, which is the SDLP, it showed a significant main effect of the presence of music ($F = 3.3$, $p < .05$), but the main effect of the loudness was not significant. No interaction effect was identified for either performance measurement. Therefore, the results were consistent with the first hypothesis.

Moreover, Ünal et al. (2013) also reported significant main effects on arousal levels. The main effect of the presence of music on deactivation was significant ($F = 55.33$, $p < .001$). The arousal level was shown to be higher (i.e., a lower level of deactivation) when the music was being played. The volume, however, did not have a significant main effect on the self-reported deactivation. As for objective measurements, the HR was significantly higher with music than with no music ($F = 5.12$, $p < .05$), but there was no significant difference in the mean HR between 70 and 85 dB. According to the results of arousal levels, the hypothesis 2a was retained, but the results rejected the hypothesis 2b. The HRV did not differ significantly. It rejected the third hypothesis. It indicated that participants had relatively the same mental efforts when driving with music versus driving without music.

Although the presence of music or the volume did not significantly affect some measurements of the performance or the arousal level, they showed significant effects on the performance and the arousal level in general. The findings are somewhat consistent with the arousal theory. The music, as a stimulus, raised the arousal level, which further improved the performance. Nevertheless, it did not reflect a decrease in the performance as the arousal level increased. There are two plausible explanations. One was that the strength of the stimulations was not high enough. If the volume was at a high level, such as 120 dB, it might negatively affect the performance, but it would very likely impair participants' hearing abilities as well. The other explanation was insufficient levels of the stimulations. The optimal level of arousal was between 70 and 85 dB, so the researchers only detected an increase in the performance but failed to identify how the performance changed from 70 to 85 dB. In relation to the current study, it can be known that the HR is a decent approach to measure the arousal level. Also, it is beneficial to acquire self-reported scores to confirm with objective measurements. Concerning the shortness of the Ünal et al.'s (2013) study, it would be better to set more levels of stimulations in the current study design, and the range needs to be wide enough.

Effects of illuminance on performance and arousal levels. Smolders and de Kort (2014) conducted a study about the illuminance levels. The design was similar to the Ünal et al.'s (2013) study. There were two IVs, which were the illuminance and the mental condition. Illuminance included two levels: 200 lx and 1000 lx, and the mental condition could be the fatigued condition or the normal condition, which served as the control group. It was a within-subjects design, so 29 participants experienced all the experimental conditions. Each participant took part in four sessions at the same time on four separate days. Each session included (a) a 7-min baseline measurement, (b) a 29-min mental condition manipulation, (c) a short questionnaire, and (d) a 30-min task measurement with corresponding light conditions. After the fourth session, participants were requested to report their characteristics. There were three tasks in this experiment, including the auditory psychomotor vigilance task (PVT), the auditory Go-NoGo task, and the 2-Back task. In the PVT, participants need to press the spacebar when they hear a tone to identify their sustained attention. The tone is played at 400 Hz, and the next tone is played between 1 and 9 seconds after the spacebar is pressed. The auditory Go-NoGo task is similar to the PVT. During this task, the tones are played at 400 Hz and 600 Hz, and participants should press the spacebar only when they hear 400-Hz tones, which appear with a probability of 80%. The intervals between the two tones are the same as the PVT. The purpose of this task is to acquire information about participants' inhibitory capacity. The 2-Back task is employed to measure working memory and executive functioning. Several characters are shown to participants one by one; they need to press the spacebar when the same character appears twice in a row. During the baseline measurement, participants simply performed these tasks for a short time. To manipulate mental conditions, participants were instructed to engage in complex tasks (i.e., the Multi-Attribute Task Battery and the Stroop color-naming task with incongruent-color ink) to have a fatigued condition. Participants were also asked to conduct relaxing tasks (i.e., watch a movie, read magazines, and the Stroop color-naming task with congruent-color ink) to avoid a fatigued condition. During the task measurement, participants were exposed to different light conditions, which could be 200 or 1000 lx. There were two identical measurements within the task measurement duration. Each measurement consisted of (a) a 1-min rest period, (b) a 5-min auditory PVT, (c) a 3-min auditory Go-NoGo task, (d) a 2-min 2-Back task, (e) a short questionnaire asking about the sleepiness, the vitality, the tension, the positive effect, and the negative effect, and (f) questions about the subjective state self-control, the evaluation of lighting condition and the environment, the time of going to sleep the night before, the time of awakening, and the time spent outside. The questions were asked to identify the influence of extraneous variables. The one-day session lasted approximately 75 minutes. The DVs were in three categories, including the experience, task performance, and the autonomic arousal level. The measurements in the experience category were alertness, mood, self-control capacity, appraisal, and beliefs. The measurements on the task performance were the response times for three main tasks. The physiological arousal levels of ANS were acquired by measuring the HR, the HRV, and skin conductance.

Smolders and de Kort (2014) performed linear mixed model analyses to identify the significance of differences. Initially, the researchers tested the differences in baseline measurements. No significant difference in the arousal level was detected, but there were some differences in the self-report scores (e.g., the vitality, the positive effect) and the task performance (i.e., interaction effects on the Go-NoGo task and the 2-Back task). For this reason, they decided to consider the baseline measurements as covariates. Afterward, the effects of fatigued manipulation were checked, and it showed the most of the subjective measures (i.e., sleepiness, vitality, tension, and positive effects) differed significantly between the fatigued manipulation and the control manipulation. In contrast, there was no significant difference in negative effects. The effects caused by the fatigue induction were demonstrated to still exist during the 30-min task measurement. Because participants had not been exposed to experimental lighting conditions at the moment of this measurement, the lighting manipulation did not show any effect. As for measurements during the treatment, when the illuminance level was at 1,000 lx, participants reported significantly lower sleepiness ($F = 5.46$, $p = .02$), higher vitality ($F = 10.25$, $p < .01$), and higher happiness ($F = 14.94$, $p < .01$). However, participants tended to have worse performance with a strong lighting condition. They had significantly longer reaction time on the auditory Go-NoGo task ($F = 6.85$, $p < 0.1$) and lower accuracy on the 2-Back task ($F = 8.71$, $p < .01$), and there was no other significant difference between the two lighting conditions with respect to the performance. Also, their skin conductance levels were higher in the lighting condition of 1,000 lx ($F = 7.45$, $p < .01$), which indicated higher arousal levels, but the HR and the HRV did not have significant differences. Regarding the mental conditions, when participants suffered from fatigue, they had longer reaction time on the PVT ($F = 11.61$, $p < .01$), the Go-NoGo task ($F = 14.62$, $p < .01$), and the 2-Back task ($F = 5.46$, $p = .02$). The differences in the accuracy of the tasks and arousal levels were not significant.

Smolders and de Kort's (2014) study was about the effect of lighting conditions and mental conditions, which was distinct from the current study, but the DVs (i.e., the experience, the task performance, and the ANS arousal levels) were related. In general, the lighting condition was a stimulus, which could generate a high level of arousal, and fatigue was considered an inhibitory factor because it mitigated the degree to which the arousal level increased. It can be learned from the results that a stronger lighting condition increased participants' arousal levels somewhat and decreased the performance. Also, a worse mental condition descended arousal levels and lengthened the reaction time. From the perspective of the arousal theory, when the arousal levels of participants dropped due to fatigue, their performance was worse. It demonstrates that an increase in the arousal level before the optimal level can lead to an increase in performance. Besides, the 1,000-lx lighting condition was a strong stimulus, which increases the arousal levels. It was possible that the arousal level exceeded the optimal level, and the impairment of the performance appeared.

Audio warnings. Many studies, such as Hellier and Edworthy (1989), Edworthy et al. (1991), and Hellier et al. (1993), have been conducted to identify the effects of acoustic parameters on individuals' perceived urgency, which can further affect the perceived risk and the arousal level. However, a considerable part of them was conducted nearly 30 years ago, and it is relatively less valuable to review these studies. Therefore, only recent studies about audio warnings will be reviewed in this part, but the results and methods of measurements in old studies are also briefly introduced.

Audio warnings for pilots. Wang et al. (2016) researched the influence of audio warning parameters on individuals' performance on receiving information. This study is being reviewed because the purpose was to modify the audio warning systems in the cockpit, where the environment is similar to the current research. Wang et al. tested two levels of the volume (i.e., 65 and 75 dB), three levels of the frequency (i.e., 700, 1,200, and 1,700 Hz), and four levels of the inter-onset interval (IOI; i.e., 100, 150, 300, and 600 ms), which is the interval between the onset of two pulses. Based on these levels of IVs, 24 combinations (2 × 3 × 4) of warning tones were created using the software Cool-Edit Pro 2.0. Every combination was paired with each of the other 23 combinations. For example, in one pair, one combination was the sound with 65 dB, 700 Hz, and 100-ms IOI, and it was followed by another sound, such as 65 dB, 700 Hz, and 150-ms IOI. The group of these two sounds was treated as a pair. When taking the order into consideration, there were 552 (A^2_{24}) possible combinations. A total of 13 participants were told to press the button as soon as possible when they heard the warning sound, and after each pair had been played, participants needed to compare the urgency of two tones. During the whole experiment, each participant needed to complete 552 pairs. In other words, one participant heard warning sounds 1,104 times and listened to each sound 46 times. By using this design, each pair was played twice in different positions so that the order effect could be offset. The experiment lasted for approximately two hours for each participant.

Wang et al. (2016) ran an analysis of variance (ANOVA) to analyze the effect of each parameter, and all of the factors were shown to have significant main effects on the perceived urgency. For volume, participants perceived a higher level of urgency when it was 75 dB in comparison to 65 dB ($F = 288.943$, $p < .001$). As for the frequency, it can be acquired that as the frequency increased, the level of perceived urgency increased accordingly ($F = 70.922$, $p < .001$). Also, when the IOI became longer, the urgency that participants perceived went down ($F = 236.176$, $p < .001$). However, Wang et al. did not report the post-hoc results, so it is unknown whether or not the differences between every two levels were significant. Moreover, they also tested the effects on reaction time. The volume and the IOI appeared to have significant effects, whereas the frequency did not ($p = .127$). When the volume was at 75 dB, participants had significantly shorter reaction time than 65 dB ($F = 5.597$, $p = .030$); as for the IOI, the reaction time increased as a result of the increase in the IOI ($F = 47.573$, $p < .001$), and the increase rate were lower when it exceeded 300 ms. Still, the researchers did not provide the post-hoc results. Additionally, they also detected a significant interaction effect between the volume and the IOI ($F = 14.240$, $p = .004$), but no detail was given. Another finding of Wang et al.'s study is that there was a significant correlation between two DVs, which are the perceived urgency and the reaction time ($r = .611$, $p < .01$), which demonstrated the individual's performance is partially dependent on the level of perceived urgency.

Although the levels of some parameters that Wang et al. (2016) selected were not within the ranges that Patterson (1982) suggested, the information about the effects of factors on the perceived urgency and the reaction time was valid. The results imply that the volume, the frequency, and the IOI are all appropriate parameters to be tested. In addition to the results, the design can be referenced as well.

Auditory confirmation signals. Although Bodendörfer, Kortekaas, Weingarten, and Schlittmeier's (2015) study was for auditory confirmation signals instead of audio warnings, the design of auditory tones and the process can be used for reference. Therefore, this study was reviewed. Bodendörfer et al. tested four parameters, including the frequency, the number of pulses, the pulse-to-pulse time, and the frequency ratio. The pulse-to-pulse time is the interval between two pulses, which is similar to the IOI. The frequency-ratio refers to the ratio of the frequency of the first pulse to the frequency of the second one. The frequencies were 393, 524, 655 Hz, respectively, so the frequency-ratio was 1, 0.8 (524/655), and 1.33 (524/393). There could be 1, 2, or 3 pulses with either 50, 150, or 300-ms intervals. The parameters of each combination were shown in Table 2.4, and the compositions of sounds are shown in Table 2.5. Therefore, there were nine combinations of parameters in total. Based on these combinations, 36 pairs ($C^2{}_9$)

were created. To counterbalance, each sound was played as the first stimulus four times and as the second one another four times. All the sounds were played at the volume of 60 dB. Bodendörfer et al. recruited 31 participants, and the whole session took around 15 minutes.

Table 2.4

Bodendörfer et al.'s (2015) Sound Parameters

Variation	Frequency (Hz)	Number of Pulses	Pulse-to-Pulse Time (ms)	Frequency-Ratio
Baseline	524	2	50	1
1	655	2	50	1
2	393	2	50	1
3	524	1	NA	NA
4	524	3	50	1
5	524	2	150	1
6	524	2	300	1
7	524	2	50	0.80
8	524	2	50	1.33

Table 2.5

Bodendörfer et al.'s (2015) Sound Design

Variation	Pulse 1 (Hz)	Interval 1 (ms)	Pulse 2 (Hz)	Interval 2 (ms)	Pulse (Hz)
Baseline	524	50	524		
1	655	50	655		
2	393	50	393		
3	524				
4	524	50	524	50	524
5	524	150	524		
6	524	300	524		
7	524	50	655		
8	524	50	393		

As the purpose of Bodendörfer et al.'s (2015) study was to identify the appropriate confirmation signal, which was different from the current study, the results are not discussed. One thing that can be learned from this study is that the volume can be set constant. Considering the environment of the cabin, it may not be practical to test various volumes. Also, Bodendörfer et al. did not create all the possible combinations of all the parameter levels. They simply established a baseline and changed one factor each time. The advantage is that the number of combinations can be much fewer than the way that Wang et al. (2016) designed. The current study is a simulation of a short-period commercial flight, which can only have a limited number of auditory cues, so it is not feasible to include too many combinations. However, a disadvantage is that it is unable to test the interaction effect.

Annoyance caused by audio warnings. When designing audio warnings, another factor that needs to be considered is the listeners' annoyance with the sound, which is also included in the current study. Sirkka et al. (2014) researched the warning sounds for industrial control rooms, and they also considered the annoyance level. There were 14 industrial control room operators who participated in the study. They were instructed to hear four sets of warning sounds and rate the level of urgency that they perceived and the level of annoyance that they felt on stepless scales. Also, the order of their perceived urgency was compared with the order of the designed urgency. These four sets contained two baseline sets and two designed sound sets, and all of them had three levels of urgency. Baseline 1 was created by the system manufacturer, and the three levels were (a) four 580-ms pulses with 580-ms intervals, (b) one 2,600-ms pulse, and (c) two 250-ms interpolating pulses. Baseline 2 included five 140-ms pulses with 80-ms intervals (the length of the last interval was 50 ms). For the two designed sounds, the compositions of three levels from low to high urgency were the same: (a) a 500-ms pulse, (b) two pulses with a total length of 1,200 ms, and (c) three pulses with a total duration of 2,300 ms. The frequencies were shown in Table 2.6.

Table 2.6

Frequencies of Pulses

		Frequencies (Hz)				
Baseline 1		Not Given				
Baseline 2	Low Urgency	300				
	Medium Urgency	300	900	3,450		
	High Urgency	300	2,450	2,550	3,450	3,513
Design 1	Low Urgency	261.63 (C4)				
	Medium Urgency	261.63 (C4)	293.66 (D4)			
	High Urgency	261.63 (C4)	293.66 (D4)	329.63 (E4)		
Design 2	Low Urgency	261.63 (C4)				
	Medium Urgency	261.63 (C4)	349.23 (F4)			
	High Urgency	261.63 (C4)	349.23 (F4)	392.00 (G4)		

Sirkka et al. (2014) discovered a significant main effect on the perceived urgency level, and each level in two designed warnings was significantly different from the other two levels. For example, in Design 1, the medium level of urgency differed from the low and high levels of urgency. In Baseline 2, the high level was not significantly higher than the medium level. Although the differences among the levels in Baseline 1 was significant, the low-level sound was considered more urgent than the medium level. The *F*-values and *p*-values were not provided. As for the degree of annoyance, the order was consistent with the designed levels of urgency. In other words, participants were annoyed the most with the sound with a high designed level of urgency, whereas they were least annoyed with a low-urgency-level sound. Moreover, Baseline 1 and Baseline 2 were scored more annoying than those two designed sounds. However, it was unknown whether or not the differences were significant because the researchers did not mention the results of the inferential statistics. Additionally, Sirkka et al. addressed that people probably tended to be more annoyed with long-duration sounds, but it was not confirmed with the inferential statistics.

When designing the current study, it should be practical to measure participants' annoyance with stepless scales, such as bipolar scales, because Sirkka et al. (2014) validated the usability of stepless scales. Also, to reduce the level of annoyance, it may be helpful to use lower frequencies and shorter durations. For the perceived urgency, it showed that the sound with more pulses and higher frequencies conveyed higher levels of urgency. Therefore, these two factors can be taken into consideration when designing intensive auditory cues in the current study.

Summary of auditory parameters effects. For pilots, the increase in (a) the volume, (b) the frequency, and (c) the interval between pulses can enhance the urgency that they perceive (Wang et al., 2016). Also, Wang et al. addressed that a high volume and a short interval can shorten pilots' reaction times. Whereas, the change in the frequency does not influence the reaction time.

For the workers in industrial control rooms, Sirkka et al. (2014) found that more pulses with higher frequencies where the differences can be identified can increase the perceived urgency. Specifically, the orders in which warnings generate the urgency from low to high are (a) C4 (i.e., 261.63 Hz), C4 + D4 (i.e., 293.66 Hz), and C4 + D4 + E4 (i.e., 329.63 Hz) or (b) C4, C4 + F4 (i.e., 349.23 Hz), and C4 + F4 + G4 (i.e., 392.00 Hz). When using these warnings, workers can also distinguish the urgency of audio warnings successfully. However, if the frequency is too high, people are annoyed with the sound though it can convey the urgency better (Sirkka et al., 2014).

In general, Hellier and Edworthy (1989) discovered the positive effects of the repetition and the length on the perceived urgency and the inverse relationship between the interval and the urgency. Among them, the repetition and the interval have strong effects, whereas the influence of the length on the perceived urgency is relatively weak (Hellier & Edworthy, 1989). Despite this, as stated before, the comparison among different types of acoustic parameters is suspicious. Similarly, Hellier et al. (1993) tested the speed, repetition, frequency, and inharmonicity. The effect of the frequency is higher than the speed and repetition, and the inharmonicity has little effect. Moreover, Suied, Susini, and McAdams (2008) measured the effects of acoustic parameters subjectively based on the reaction time and objectively based on the perceived urgency. They found the IOI had significant influences on the reaction time and urgency. Also, individuals have shorter reaction times and higher urgency when the warning sounds are irregular (Suied et al., 2008).

Summary of urgency measurements. Various types of measurements have been implemented by researchers to acquire participants' perceived urgency. It is beneficial to review their measurements to identify a method that best fits the current study. Hellier and Edworthy (1989) instructed participants to rank and rate the urgency levels of sounds. When ranking the urgency levels, participants were asked to hear three sounds each time, and they selected the most urgent and the second most urgent sound until each sound had been compared with every other sound. Afterward, participants were directed to imagine the most urgent sound and consider it as 100 and also consider the most non-urgent sound as 0. Each participant needed to assign a number between 0 and 100 to describe each sound. The methods that Edworthy et al. (1991) employed were similar to Hellier and Edworthy (1989). The difference was that participants ranked four sounds each time. Initially, they selected the most urgent one among four sounds. After that, they listened to the remaining three sounds and chose the most urgent one, and they listened to the rest two sounds and made the decision. The rating methods were the same in these two studies. Hellier et al. (1993) used a different way to measure PLUs. Participants were told to draw a line according to the perceived urgency. If the participant perceived a higher level of urgency, he or she drew a long line; whereas, the line should be short when the sound was considered low urgency. This method was adapted from the study that was conducted by Mashour and Hosman in 1968 (as cited in Hellier et al., 1993).

As for recent studies, Suied et al. (2008) instructed participants to rate the subjective urgency on continuous scales. The left side of the scale referred to not urgent at all, and the right side was labeled very urgent. Similarly, in Sirkka et al.'s (2014) study, participants rated the level of urgency on a "stepless scale" from extremely low to extremely high (p. 539). Wang et al. (2016) created 552 (A^2_{24}) comparative pairs with 24 sounds. Participants needed to compare between the two sounds in each pair. The urgency level of each sound was then scored according to the results of the comparisons.

In general, to rank the urgency level, researchers tended to ask participants to compare one sound with each other (Edworthy et al., 1991; Hellier & Edworthy, 1989). To rate the urgency level, methods included (a) bipolar scales, (b) scoring from 0 to 100, (c) drawing lines, and (d) comparing within each pair and scoring accordingly (Edworthy et al., 1991; Hellier et al., 1993; Hellier & Edworthy, 1989;

Sirkka et al., 2014; Suied et al., 2008; Wang et al., 2016). Among them, the most popular and recent method is the bipolar scale.

Summary and Study Implications

Summary. Annually, a considerable number of flight passengers experience various injuries, and many of the passengers were injured due to a failure to exhibit safe behaviors, especially fastening seatbelts. Although most injuries are not serious, fatalities are not rare either. Generally, there are two ways to tell passengers to fasten their seatbelts. The most direct way is a lighted fasten-seatbelt sign accompanied by a soft sound. Another way is for flight attendants or pilots to make announcements to instruct passengers to fasten their seatbelts. However, most passengers ignore them. Even if passengers notice the sign or the announcement, they may not perform the activity because they think of it as an unnecessary routine and do not place importance on it. A potential reason for this is that most passengers have not encountered severe turbulence and do not realize the consequence of not wearing seatbelts. In other words, they underestimate the necessity of following instructions and the levels of risk they are taking when not complying with announcements. Wilde's (1982a) RHT may provide a feasible way to solve this problem. According to the RHT, each individual adjusts behaviors to make his or her PLR as close as possible to the TLR. Flight passengers usually perceive a relatively low level of risk because they have not experienced severe turbulence, so they perform normally or even take some risk to achieve the TLR.

However, if passengers can perceive a high level of risk, which can be higher than the ALR, passengers would perform safely to lower the risk level. It is possible that a seemingly unsafe cabin environment would increase passengers' PLR, but it may reduce the number of passengers. However, if the visual or the audio signals are striking, the PLR can be increased unintentionally. Admittedly, the visual sign is not intrusive, and it can stay illuminated for a long time, so every time passengers look at the signs, they are reminded to fasten their seatbelts. Nevertheless, if passengers do not look at the signs, they cannot notice them. By contrast, the auditory signal is omnidirectional, and passengers can detect the signals without special attention (Patterson & Mayfield, 1990; Sirkka et al., 2014). Also, the stimulation caused by auditory cues is stronger than visual cues (Berlyne, 1960). Therefore, it would be better to modify auditory signals rather than visual ones to increase passengers' PLR.

Typically, an in-flight announcement is led by an auditory cue. In comparison to the announcement, it is much easier to make the auditory cue striking to draw passengers' attention and increase their PLUs. Several past studies and guidelines demonstrated the effects of some acoustic parameters on people's perceived urgency, which can very likely increase people's arousal levels and PLR. The parameters included (a) the volume, (b) the frequency, (c) the number of pulses, (d) the length of each pulse, (e) the length of each interval, (f) the onset and the offset, (g) the pitch, (h) the rhythm, (i) the inharmonicity, etc. Not all of the parameters are appropriate for the current study. For example, according to Patterson's (1982) guidelines, the volume of aircraft system sounds should be at least 15 dB and no more than 30 dB higher than the masked threshold. Considering the background noise, such as the engine noise, the volume should be around 90 dB. If the volume is treated as a factor, the sound with a high level of volume (i.e., significantly higher than 90 dB) may impair people's hearing abilities. Therefore, only some of the parameters can be examined in the current study to test if intensive audio warnings can increase PLR and lead to safe behaviors.

Nevertheless, it is not beneficial to make the auditory cues extremely intensive. One reason, as stated above, is that it may harm passengers' hearing. Another reason is if the auditory cue is too striking, it may increase passengers' arousal levels exceedingly, and it could impair their performance regardless of passengers wanting to follow instructions. Although it is not possible in most cases, it is still beneficial to consider this issue. The third reason is that high-intensity sounds are annoying. For the first reason, Patterson (1982) established guidelines for cockpit system designs, which should apply to the cabin. If the designed auditory cues are within the ranges that Patterson provided, the auditory cues are not likely to hurt passengers' ears. As for the second reason, according to the arousal theory, when the arousal level is increased to a level below the OLA, it can enhance people's performance; whereas, if the arousal level is beyond the optimal level, the stimulation may draw too much attention to the stimulation, and individuals perform worse on the main tasks (Eysenck, 1982). In the cabin, an intensive auditory cue may make passengers focus on the auditory cue instead of focusing on the instructions. Regarding the third reason, if the sounds have many pulses or high frequencies, they are more likely to be considered annoying (Sirkka et al., 2014). Based on these reasons, it is also necessary to test if passengers have a bad performance or have annoyance when the intense levels of stimulations created by auditory cues are extremely high. If the results are compliant with the Yerkes-Dodson Law, which means a strong stimulation can impair the performance, it is necessary to detect the optimal level to design the most appropriate auditory cue. Also, it is needed to consider if passengers are annoyed with the auditory cue that can lead to the OLA.

Study implications. Past studies about the RHT, such as Wilde et al. (1985) and Hoyes et al. (1996), validated the RHT in the laboratory environment, which has a limited number of people, limited interaction, limited time, and limited feedback. In actuality, their results supported the RCT rather than the RHT because they simply identified the compensation but not the homeostasis considering the short periods of the experiments. Despite this, it is enough for the current study. If passengers can make compensatory behaviors, it should demonstrate the effects of intensive auditory cues. The studies about the arousal theory provided several possible methods to measure physiological arousal levels and subjective arousal levels. Also, as Eysenck (1982) stated, they usually had insufficient levels of stimulation. For example, Ünal et al. (2013) and Smolders and de Kort (2014) only detected one direction that the performance shifted in each analysis. Therefore, in the current study, it is necessary to set up more intense levels of auditory cues with various parameters. Wang et al. (2016) used the matrix of different levels of acoustic parameters and created 552 auditory combinations, which were too many to be included in a simulation. Fortunately, Bodendörfer et al. (2015) employed a new way that they changed one parameter each time so that only nine sounds were containing three levels for four parameters.

Nonetheless, all these studies did not include flight passengers and auditory cues. They were mainly about operators instead of the individuals who did not have tasks. Passengers are usually not prepared for performing tasks, and few of them have received training. It makes the current study different from others. Moreover, for operators, their arousal levels are relatively high because they are engaging in the tasks. By contrast, passengers are relaxed and doing their favorite things or work, which are not related to following instructions. For this reason, audio warnings can make operators' arousal levels over the OLA easily as their arousal levels are already high. In contrast, the same stimulation may not make passengers' arousal levels reach that level. This is another reason why this study is necessary. Although the current study is distinct from other relevant studies, the procedures, the measurements, and the findings of other research can be referenced when designing the current study.

Chapter 3

Methodology

Population and Sample

Population. The target population was flight passengers in the United States. In 2018, there were 1,011,100,232 passengers who took flights in the United States, and it was the first time that the number of passengers exceeded one billion (Bureau of Transportation Statistics, 2019). Among them, 777,919,130 passengers flew domestically, and 233,181,102 were international travelers. The accessible population was the students, faculty, and staff at FL Tech, who (a) were 18 years or older, (b) were proficient in English, (c) had normal hearing ability, and (d) have been passengers on an aircraft at least once. Therefore, the target population was further defined as adult flight passengers who have proficiency in English and normal hearing abilities in the United States. Although the accessible population was not completely representative of the target population, the accessible population could demonstrate the effect of manipulations within the target population to some degree.

Sample. The primary sampling strategy was a convenience sampling approach. Those people who met the requirements and volunteered to participate were recruited. The study was posted on the Sona system to recruit participants. The system was open to students from the College of Psychology & Liberal Arts at FL Tech. Students received extra credits after participating in the study. Also, an email was sent to students through FITFORUM. A secondary sampling strategy was snowball sampling, where participants were encouraged to introduce the experiment to acquaintances within the campus community to increase the number of participants.

Power analysis. Three factors, including α, the effect size, and the desired power, were utilized to determine the minimum sample size (Cohen, 1992a). Cohen (1992b) asserted that committing a Type I error (i.e., a false positive claim) was more serious than a Type II error (i.e., a false negative claim), so it is practical to make the probability of a Type II error four times a Type I error to reflect their corresponding importance. Because the convention for α-level is set at .05, the β-value can be .20; therefore, the desired power was set at .80. The effect size was unknown, so a medium effect size was used to calculate the minimum sample size.

Power analyses needed to be conducted to determine the minimum sample sizes regarding separate tests. For the effect of auditory cues on physiological parameters, an analysis of variance (ANOVA) was run with one group and five measurements. The effect size was set at medium, which is .25 (Faul, Erdfelder, Buchner, & Lang, 2013). The correlation among repeated measures was assumed to be 0. The non-sphericity correction should not be less than .75 in a repeated-measures ANOVA based on one convention ("How to use," n.d.). The result showed 43 participants are needed as the minimum. The power analysis results for effects of auditory cues on response time and annoyance were supposed to be the same as physiological measures because the difference between variables did not change the minimum sample size.

For the analysis of the relationship between two continuous variables, including (a) the relationship between arousal levels and perceptions, (b) the relationship between perceptions and attitude, (c) the relationship between arousal levels and attitude, (d) the relationship between arousal levels and response time, (e) the relationship between perceptions and response time, and (f) the relationship between attitude and response time, a repeated-measures correlation (RMCORR) was run. According to Bakdash and Marusich's (2017) statement, approximately 25 participants were needed to reach a power of 0.80 with a medium effect size.

Compliance is dichotomous, and the within-subjects effect will be analyzed. Therefore, Cochran's Q tests and McNemar tests will be run to detect the effect of A on compliance; also, binary logistic regression will be conducted to identify the effect of B and the effect of C on compliance. For the power analysis prior to the McNemar tests, it will be a two-tailed test because it is unknown which auditory cue will have larger influences. The odds ratio and the proportion of discordant pairs cannot be estimated, and the minimum sample size is largely variable as a result of tiny changes in these parameters, especially the odds ratio. In this case, the power analysis is not meaningful, so a proper minimum sample size has not been determined. The minimum sample size for the binary logistic regression tests was relatively hard to determine because many parameters cannot be expected, such as

Pr(Y = 1|X = 1) H0 (i.e., the probability of Y = 1 when the X is one *SD* above the mean) and R^2 of other predictors. In this case, all the parameters are filled by default (odds ratio = 1.3, Pr(Y = 1|X = 1) H0 = .2, R^2 of other predictors = 0, normal distribution for X, $\mu_X = 0$, and $\sigma_Y = 1$), then 568 participants are needed to make the power at .80. However, by simply increasing the odds ratio to 2.1, it will be necessary to recruit 78 participants to acquire a power of .80. However, in the primary analysis, two DVs were integrated into one continuous variable. The power analysis results in relation to the dichotomous DV, which are discussed in this paragraph, may not be included.

In conclusion, if the effect size of the effects on continuous variables is at or above the medium sample size, 43 participants will be enough regardless of the power analyses with undecidable parameters. The results are summarized in Table 3.1. As for the dichotomous DVs, the default values cannot lead to high power. However, if the odds ratio is slightly high than the default values (at around 3), 43 participants are enough to keep the power above .80. Therefore, slightly more than 43 participants will be recruited in case of missing data.

Table 3.1

Summary of Power Analyses

Analyses	Tests	Minimum Sample Size	Estimated Power
A on B	ANOVA	43	.80
A on C	ANOVA	43	.80
A on D	ANOVA	43	.80
A on Compliance (E)	McNemar	Not Determined	
A on Response Time (E)	ANOVA	43	.80
B and C	RMCORR	25	.80
B and D	RMCORR	25	.80
B on Compliance (E)	Binary Logistic Regression	Not Determined	
B on Response Time (E)	RMCORR	25	.80
C and D	RMCORR	25	.80
C on Compliance (E)	Binary Logistic Regression	Not Determined	
C on Response Time (E)	RMCORR	25	.80
D on E	RMCORR	25	.80

Instrumentation

Tone generator. The purpose of a tone generator was to create tones with different acoustic parameters. The NCH Tone Generator was the software that was used in this study. It generated tones with various frequencies and waveforms.

Tones. There were five different tones. Each tone varied in four acoustic parameters consisting of the frequency, the waveform, the number of pulses, and the length of each pulse. The frequency could be 0.5 and 1.0 kHz. The waveform included the sine wave and the triangle wave. Some tones had one pulse, whereas others had two pulses. The volume of each tone was set at 90 dB, which was close to the suggested volume of auditory warnings, 90-100 dB when it was below 2kHz, in the cockpit (Patterson, 1982). For each pulse, the length could be 600 ms and 1200 ms. The ratio of the onset (i.e., 0% to 100% volume) to the main part to the offset (i.e., 100% to 0% volume) was 1:4:1. For example, if the length of the pulse was 600 ms, the length of the main part was 400 ms, and the lengths of the onset and the offset were both 100 ms. When there were two pulses, the inter-onset interval (IOI) was 300 ms. The tone with 0.5 kHz, a sine wave, one pulse, and 600-ms length was the most moderate, so it was used as a baseline. Every time the designed urgency level increases, one parameter was changed from low to high (Table 3.2). Therefore, the tone with 1 kHz, a triangle wave, two pulses, and the 1200-ms length was the most intensive. Bodendörfer, Kortekaas, Weingarten, and Schlittmeier (2015) applied the same method to develop different stimulation levels.

The raw tones were made with the NCH Tone Generator. It produced tones at the frequency of 0.5 or 1.0 kHz in either the sine or triangle waveform. Then, the raw tones were edited using Cool Edit Pro 2.1 to create onsets and offsets and change the length, number of pulses, volume, and the IOI.

Table 3.2

Intensive Auditory Cues

Auditory Cue Intensity	Frequency	Waveform	Number of Pulses	Length of Pulse
Baseline	0.5 kHz	Sine wave	1	600 ms
Intensity 1	1.0 kHz	Sine wave	1	600 ms
Intensity 2	1.0 kHz	Triangle wave	1	600 ms
Intensity 3	1.0 kHz	Triangle wave	2	600 ms
Intensity 4	1.0 kHz	Triangle wave	2	1,200 ms

Sound meter. A sound level meter was used to measure the volume of tones, announcements, and background noise. It ensured that the volumes of tones were within the ranges of 90 to 100 dB when the frequency was below 2 kHz that Patterson (1982) suggested. Also, it could make sure that the announcements and the background noise were at an appropriate level.

Aircraft seats. There were four rows of aircraft seats with three seats in each row. Each seat was equipped with a seatbelt and a tray table. However, there was no tray table in front of the first row, so only the back three rows were utilized in the study.

Scales. Several scales were prepared to collect subjective information (Appendix A). All the items that were included in the scales are bipolar. Each item has two opposite adjectives, and nine boxes between two ends for each item; this was a flexible method to identify individuals' attitudes and was easier for participants to fill out than a Likert-scale item (Ary, Jacobs, Sorensen, & Asghar, 2010). Participants were instructed to mark a cross on each line according to their feelings. It was then transformed into a value between 1 and 9.

Instruction. Instructions were provided to participants before the simulation. They informed participants of the directions and requirements (Appendix B). Participants were asked to behave as if they were passengers on a commercial flight. Besides, they were not allowed to talk, sleep, listen to music, use cell phones (except for reading e-books), or leave their seats. They were recommended to read magazines that were in the seat pockets in front of them. They were also told that there would be several announcements, and they needed to listen to and adhere to these.

Sensors. The Equivital EQ02 LifeMonitor was used to collect physiological data. It included a sensor electronics module (SEM) and a vest. The SEM can transmit the live data to a tablet via Bluetooth and can also store and transfer the data via a cable after the experiment. The vest can be harnessed on the chest and hold the SEM as well as several sensors, which were connected to the SEM. In this study, only the Electrocardiograph (ECG) and galvanic skin response (GSR) data were collected to measure the heart rate (HR), heart rate variability (HRV), and skin conductance. From the ECG, the HR based on each pulse can be acquired by calculating the R-R interval, and it was an accurate way to collect this information. In the context of the study, HR means the elevation in HR instead of pure HR. As for the GSR, two electrodes were placed on the participant's index and middle fingers on the left hand with adhesive, and it also referred to the elevation.

Camera. A camera was used to record the simulations. In this case, no experimenter was needed to collect data in the lab room. The experimenter was able to monitor participants' compliance and measure their response time after the simulation based on the recordings. The camera was connected to a cell phone, and monitoring occurred to view the simulation in real-time. The experimenter was able to annotate any observations or intervene if necessary.

Procedures

Research methodology. The study employed a repeated-measures design. Each participant experienced each of the auditory cues, and they were measured in each of the conditions. A benefit is that the influence of individual differences can be eliminated (Ary et al., 2010). Various levels in the independent variable (IV) were compared to each other within subjects. The resting levels of physiological signals were different among individuals, and the responses to stimulations were various as well, so the comparison within subjects was appropriate because it could lead to more convincing results than the between-subjects comparison. Moreover, because there were multiple levels in the auditory cues, the within-subjects design can result in a relatively high power with the same sample size (Ary et al., 2010). However, when applying the within-subjects design, participants received different treatments, and this might result in the carry-over effect (Ary et al., 2010). To minimize the effect, the order of auditory cues was counterbalanced to reduce the impact. Also, the event rate was designed to be 50% to mitigate this effect.

Human subjects research. The current study included human subjects. The process was reviewed by the FL Tech Institutional Review Board (IRB) to make sure the experiment was ethical (Appendix C). Monetary incentives in the form of a raffle prize were used to recruit participants. Therefore, only participants that won the prizes received rewards. Moreover, the experiment brought no more harm or risk than everyday life to participants, and participants were allowed to quit without penalty.

However, there were two minor concerns. One was the auditory cue because it was loud. In this experiment, the volume of an auditory cue was over 90 dB or longer than 3 seconds, and it was measured using a calibrated sound meter and checked before simulations. The National Institute of Occupational Safety (NIOSH) stated that for a 90 dB sound, the exposure limit is two hours and 31 minutes (NIOSH, 1998). Similarly, according to Title 29 of the electronic Code of Federal Regulations (29 e-CFR) part 1910, § 1910.95, the Occupational Safety and Health Administration (OSHA) regulated that the exposure to a 90-dB noise for eight hours is allowed. Therefore, the auditory cue would not harm the participants' hearing abilities. Another concern was face recognition because the simulation was video recorded, but the experimenters did not have access to face-recognition software and did not need this information. Therefore, the researchers did not seek out facial recognition of participants. The video recordings were entered and stored on a password-protected computer. Only researchers directly involved in this study had access to video recordings. They were stored in a locked office for three years after any publication, and then will be shredded and destroyed.

Description of IVs and DVs. There was one IV, which was the intensity level of auditory cues. The intensity level included five levels, which were four intensive tones and one baseline tone (Table 3.2). Also, there were six mediating variables (MVs), which were the variables that were functions of the IV and affect DVs. They are (a) HR, (b) skin conductance, (c) perceived level of urgency (PLU), (d) perceived level of risk (PLR), and (e) annoyance. The HR, HRV, and skin conductance served as physiological measures of arousal level. The HR was the frequency of heartbeats, and the HRV refers to the extent to which the HR varies. GSR was measured to gather information about the skin conductance. As there were confounding variables and hard to predict, these physiological measures were treated as exploratory MVs. Moreover, participants were instructed to self-report their PLUs and PLRs for each cue using bipolar scales, which were then transformed into values from 1 to 9 (Appendix A). Annoyance data were also collected because some tones might make people uncomfortable. If passengers on commercial flights are annoyed with the auditory cue sounds, they may wear earbuds to avoid hearing the sounds resulting in a safety briefing not being heard. Similar to emergency vehicles, although their sirens can successfully result in drivers' and pedestrians' attention, most people do not like to hear the warning sound.

There are two DVs, including (a) compliance with instructions and (b) response time. Compliance was defined as either a participant complying or not complying. If the participant did not respond to the instructions, it was considered as not complying. The response time was the duration from the point that keywords in the instruction were given to the point that participants initiate responses, which ranged from 0 to 10 seconds.

Pilot study. A pilot study was run before the data collection. It was to see the participant's opinions about the study. Also, it tested if the noise outside the laboratory room affected the auditory effects in the experimental environment.

Study implementation. Before the simulation starts, participants signed the informed consent form (Appendix D) and read the instructions. Afterward, participants were directed to don the vest and sensors devices, and they sat on aircraft seats located in a dedicated lab space belonging to the College of Aeronautics. Only one participant was in part of each session, so there was no social influence effect in groups. There were eight announcements, four of which were action-required, and the others were general announcements. Action-required announcements instructed passengers to fasten seatbelts, lower tray tables, close tray tables, and adjust seatbacks, respectively. The general announcements were about the emergency exit, no-smoking requirement, meal service, and destination weather, and participants did not need to make responses to them. The script of announcements is shown in Appendix E. The purpose of general announcements was to relax participants, and they did not learn that every announcement would request them to make an action, and it made the simulation more realistic. In this case, the event rate was 50% (4/8). Also, the general announcement did not have an intensive auditory cue but simply a baseline auditory cue (i.e., 0.5 kHz, sine wave, one pulse, and 600 ms). The order, contents, and instructions of announcements are shown in Table 3.3. Each action-required announcement followed a random intensive auditory cue. The order of auditory cues was counterbalanced based on the reduced Latin square to offset the carry-over effect. Auditory cues had the same volume, which was 90 dB. An engine sound was played at 75 dB as background noise. Intensive auditory cues included four levels, and parameters are presented in Table 3.2. As for general announcements, baseline auditory cues will lead them.

Table 3.3

Announcements

Order	Announcement Type	Content	Instructions
1	General	Emergency Exit	---------
2	General	No Smoking	---------
3	Action-Required	Turbulence	Fasten Seatbelts
4	Action-Required	Turbulence	Close Tray Tables
5	Action-Required	Meal Service	Lower Tray Tables
6	General	Food	---------
7	General	Destination Weather	---------
8	Action-Required	Descending	Adjust Seat Backs

The simulation lasted for 40 minutes. Participants were asked to remain seated on aircraft seats during the entire study session. During the study, participants performed a secondary task consisting of solving Sudoku puzzles, but they were not allowed to talk, sleep, use cell phones, listen to music, or leave their seats. Announcements were played at 6:00, 10:30, 15:00, 19:30, 24:00, 28:30, 33:00, 37:30 minutes, respectively. The first six minutes were for participants to relax in order to measure the resting levels of physiological parameters. Moreover, the physiological reading before each announcement was used as a baseline for the following announcement. Before each auditory cue, participants were given four minutes to recover from the previous stimulation and action. For HR, a normal adult's heart recovery rate should be above 12 beats per minute (bpm) (Cole, Blackstone, Pashkow, Snader, & Lauer 1999). The elevation of HR during the simulation was not expected to be over 48 beats per minute, so four minutes would be enough for subjects' hearts to recover. For example, Walther et al. (2018) allowed participants two minutes to recover from an increase in HR caused by low-intensity tasks. Despite the absence of evidence, four minutes were also believed to be enough for a GSR reading to drop back to the resting level. An experimenter who acted as a flight attendant walked around the cabin and made sure participants had lowered their tray tables between the fifth and the sixth announcement. If a participant did not lower the tray table 10 seconds after the instructions, the experimenter asked the person to do so. In this case, the participant received dual stimulations (i.e., the presence of an auditory cue and announcement and the presence of the experimenter). The experimenter walked in the cabin softly and spoke very gently to minimize the effect of the uninterested stimulation. A several-second delay in participants' recovery after the fourth announcement is expected for those who do not comply with the instruction. Despite this, the interval between the two stimulations was still long enough for people to recover. After the simulation, each auditory cue without an announcement was played in a random order to participants at another time, and they were requested to complete three items for each tone. These items are adjective-bipolar and include (a) how do you describe your sense of urgency in the situation after hearing this auditory cue, (b) what extent did you perceive the risk in the situation after hearing the auditory cue, and (c) how annoying did the auditory cue seem to you. The demographics, including gender and age, were collected as well. Moreover, the simulation was recorded to check participants' behaviors and collect compliance information. Hellier, Edworthy, and Dennis (1993) suggested that in addition to psychological measures, it was also necessary to measure subjects' behavioral responses, such as response time. Therefore, objective measures were made, and the data were acquired by watching the video recordings of simulations.

Data collection. There were four categories of data, including physiological, perceptions, performance, and attitude data. The physiological data were collected during the simulation, whereas the performance information was acquired after the simulation based on the recordings. Perceptions and attitude data were collected after the simulation but still in the lab. Also, the participant numbers in each session will be noted.

Physiological data. Participants' physiological information, which included the HR, HRV, and the GSR, was measured during the entire 40-minute session. It was broken down into nine segments (Figure 3.1). The first four minutes (from 00:00 to 04:00) were assigned in the first segment, which was for participants to relax and get used to the experimental setting, such as aircraft seats and engine noise. The following eight segments corresponded to eight announcements, respectively. Each segment had four and a half minutes and contained three parts, which were 2 mins before (two minutes), stimulation (30 seconds), and 2 mins after (two minutes). The 2-mins-before part was the baseline of the segment. During the stimulation part, auditory cues and announcements were given during the first 10 seconds, and participants needed to initiate responses in the second 10 seconds and finish the movements within another 10 seconds. The next two minutes were for participants to relax and recover so that their arousal levels could stay at the resting levels. The physiological data were collected separately in different parts.

Psychological data. Mental data included participants' perceived urgency and perceived risk. They were collected using bipolar scales. After the simulation, participants were asked to hear the auditory cues without announcements another time in random order. They needed to draw a proper check on each item. The graphics were then transformed into values from one to nine.

Performance. The performance data were compliance and response time. The sessions were recorded during simulations. After that, the recordings were watched by the experimenter, and data were acquired.

Attitude. Attitude data, which was the annoyance with the auditory cues, was similar to mental data. The data were collected using the bipolar scale after each session. The attitude item was put on the same questionnaire following the items that collected mental data for each auditory cue.

Questionnaire. Participants were asked about their demographical information (Appendix F). The first section included (a) gender, (b) age, and (c) hearing ability. The next section contained items on (a) proficiency in English, (b) nationality, (c) first language, (d) nation of high school, and (e) nation of college attended for the participant's bachelor's degree. The third section was about flight experience, including (a) whether or not a participant has been a passenger on commercial flights, (b) frequency of flights, (c) the compliance with fasten-seatbelt instruction, (d) compliance with airplane-mode instruction, and (e) attention to pre-flight safety briefings.

Debriefing. In the debriefing, participants were informed of the true purpose of the study. Also, they were asked about their attitudes toward the simulation (Appendix G). It included boredom, anger (in the questionnaire that collected the attitude, participants were asked how mad they were at the simulation instead of angry), and interest.

Time	Segments	Parts	Lengths
00:00.0			
00:30.0			
01:00.0			
01:30.0	1st Segment (4:00)	Relax	4:00
02:00.0			
02:30.0			
03:00.0			
03:30.0			
04:00.0			
04:30.0		2 min Before	2:00
05:00.0			
05:30.0			
06:00.0	2nd Segment (4:30)	Stimulation	0:30
06:30.0			
07:00.0		2 min After	2:00
07:30.0			
08:00.0			
08:30.0			
09:00.0		2 min Before	2:00
09:30.0			
10:00.0			
10:30.0	3rd Segment (4:30)	Stimulation	0:30
11:00.0			
11:30.0		2 min After	2:00
12:00.0			
12:30.0			
13:00.0			
13:30.0		2 min Before	2:00
14:00.0			
14:30.0			
15:00.0	4th Segment (4:30)	Stimulation	0:30
15:30.0			
16:00.0		2 min After	2:00
16:30.0			
17:00.0			
17:30.0			
18:00.0		2 min Before	2:00
18:30.0			
19:00.0			
19:30.0	5th Segment (4:30)	Stimulation	0:30
20:00.0			
20:30.0		2 min After	2:00
21:00.0			
21:30.0			

Time	Segments	Parts	Lengths
22:00.0			
22:30.0		2 min Before	2:00
23:00.0			
23:30.0			
24:00.0	6th Segment (4:30)	Stimulation	0:30
24:30.0			
25:00.0		2 min After	2:00
25:30.0			
26:00.0			
26:30.0			
27:00.0		2 min Before	2:00
27:30.0			
28:00.0			
28:30.0	7th Segment (4:30)	Stimulation	0:30
29:00.0			
29:30.0		2 min After	2:00
30:00.0			
30:30.0			
31:00.0			
31:30.0		2 min Before	2:00
32:00.0			
32:30.0			
33:00.0	8th Segment (4:30)	Stimulation	0:30
33:30.0			
34:00.0		2 min After	2:00
34:30.0			
35:00.0			
35:30.0			
36:00.0		2 min Before	2:00
36:30.0			
37:00.0			
37:30.0	9th Segment (4:30)	Stimulation	0:30
38:00.0			
38:30.0		2 min After	2:00
39:00.0			
39:30.0			

***Figure 3.1.* Session components.**

Threats to internal validity. Campbell and Stanley (1963) introduced eight types of threats to internal validity, which are (a) history, (b) maturation, (c) testing, (d) instrumentation, (e) statistical regression, (f) selection, (g) experimental mortality, and (h) selection-maturation interaction. Additionally, Ary et al. (2010) introduced more types of threats besides the eight types that Campbell and Stanley mentioned. They are (a) experimenter effect, (b) subject effects, and (c) diffusion. Besides the threats that Campbell and Stanley and Ary et al. stated, there is another type of threat, which is the location.

History. As this was a repeated-measures design, a specific event might have happened between two measures, and it might lead to the effects of other factors on the results than treatments. It was considered a history threat (Campbell & Stanley, 1963). For example, between two measures, a participant read the news that several passengers are injured on a flight because they did not fasten their seat belts. The participant might choose to follow the instructions faster and feel more nervous. The effects were not caused by intensive auditory cues but an event during the simulation. To eliminate this effect, participants had limited use of their cell phones. Also, the simulation duration was only 30 minutes, so it was unlikely that participants became aware of an event that could affect their arousal levels and behaviors.

Maturation. According to Campbell and Stanley (1963), maturation refers to the condition in which there are some specific changes within subjects between measures, which can result in effects on measurements. The change can be caused from growing older, hungrier, more tired, and so on. For example, during the simulation, if a participant was hungry and suffered from fatigue, the participant might have a slower response time even if the auditory cues were more intensive. Therefore, the simulation lasted for only 30 minutes, and the changes within each individual were not likely to be significant.

Testing effect. Campbell and Stanley (1963) stated the threat of testing means the previous test has an effect on following tests. For example, the participant understood that the first instruction asked him or her to fasten their seat belt, so the participant is ready to fasten the seat belt when the second announcement was played. In this case, the response time would be shorter because of the testing effect instead of the effect of intensive auditory cues. To minimize the effect of this threat, two successive announcements will not instruct participants to do the same action.

Instrumentation. When the study is under the influence of the threat of instrumentation, there is a change in the measuring instrument or experimenters (Campbell & Stanley, 1963). For example, the device that measured the GSR was not reliable. The readings might vary under equivalent stimuli. Therefore, in the current study, devices (e.g., stopwatch) were calibrated before sessions.

Statistical regression. If a group of participants with extreme scores is selected, their scores can only vary in one direction (Campbell & Stanley, 1963).

For example, if participants were selected because of their extremely poor performance in the pre-assessment, their scores must be better than or equivalent to the previous scores. The cluster sampling strategy was not applied in this study, so it was not likely that a group that had extreme scores were selected. Also, the order of auditory cues was counterbalanced, and the effects of statistical regression threat, if it exists, could be offset.

Selection bias. A threat of selection will appear if participants are recruited as groups, and there is already a difference in the DV between two groups before the treatment is deployed (Campbell & Stanley, 1963). For example, one group of participants were the students from the College of Aeronautics, whereas the other group is not in the field of aviation. The first group was more likely to comply with instructions because they understood the importance of fastening seat belts from a safety perspective. However, repeated-measures design was applied in the current study, so it should not be a significant problem.

Experimental mortality. This threat means participants quit the experiment so that the data are missing (Campbell & Stanley, 1963). For example, some participants might have claustrophobia, and they could not stay in the simulated cabin for a long time, so they quit the simulations. To prevent a threat of experimental mortality from occurring, the number of participants that were recruited was several more than the minimum sample size.

Selection-maturation interaction. In quasi-experimental studies with multiple groups, it is possible that these groups have different maturation rates (Campbell & Stanley, 1963). For example, it was easier for one group to be tired than the other group, so this group had relatively poor performance. In this study, the effects were compared within subjects, and the order of auditory cues was counterbalanced. Therefore, the threat of selection-maturation interaction should not have negative effects on internal validity.

Experimenter effect. It refers to there being multiple experimenters in the study, and they have various characteristics, so they have different interactions with participants, which have distinct influences (Ary et al., 2010). For example, in the current study, an experimenter who acted as a flight attendant walked around and made sure the participant had lowered the tray table before the instructions of raising a tray table. It was possible that one experimenter unintentionally revealed that there was a raise tray table instruction soon so that participants were ready for the next announcement. A solution that was implemented was for the acted flight attendant to be the same person in all the sessions and say the same words to participants.

Subject effects. The threat of subject effects are the threats related to participants' attitudes and includes three subcategories, which are (a) the Hawthorne effect, (b) John Henry effect (i.e., compensatory rivalry), and (c) compensatory demoralization or resentful demoralization (Ary et al., 2010). The

Hawthorne effect means participants want to perform better than usual when they know they are being observed. For example, participants tended to listen very carefully to announcements and follow the instructions without hesitating in the simulation because they would like to have decent scores. To mitigate the Hawthorne effect, no experimenter appeared in the "cabin" except when directing participants to lower tray tables. Although it did not eliminate the effect as participants knew their behaviors are being recorded, it could reduce the Hawthorne effect.

The John Henry effect refers to the fact that some participants in the control group know they will not receive the treatment and want to defeat the treatment group because they do not want to be the losers (Ary et al., 2010). By contrast, the compensatory demoralization means the control group knows they are expected to be the losers, so they become resentful and do not make an effort in the study (Ary et al., 2010). For example, if one group of participants knew they did not receive intensive auditory cues, they might force themselves to make quick responses to defeat the group that received intensive auditory cues (i.e., John Henry effect), or they became too relaxed. They did not make proper responses to the instructions (i.e., resentful demoralization). Therefore, the design was a repeated-measures, and there was only one group in this study.

Diffusion. This threat appears when one group tells the other group about the treatment so that the other group applied the same treatment resulting in similar performance (Ary et al., 2010). However, this threat did not impair the internal validity of the current study because there was only one group. Also, the treatments were various sounds, so learning about the treatments should not help participants' performance.

Location. It means the sessions are not conducted in the same location, so the difference in the results may be due to the location rather than the treatment. Similar to diffusion, a threat of location was not a problem because the effects were compared within-subjects, and people participated in the same lab space. Therefore, it was not necessary to take this threat into consideration.

Confounding variables. In addition to the threats mentioned above, confounding variables could impair internal validity as well. A confounding variable, which was also called a spurious variable, had effects on IVs and DVs, making it show a relationship between IVs and DVs. A possible confounding variable was the time of the day. For example, if a participant got up early and took part in the study in the morning, the participant might be sleepy. In this case, he or she might have a relatively low arousal level and poor performance, and it seemed there was a direct relationship between the arousal level and the participant's performance. Another confounding variable could be excitement. It was possible that a participant was very excited when participating in the simulation study, so the participant had a high arousal level and good performance. For this reason, resting levels of physiological signals were collected before stimulations for use as a baseline.

Treatment verification and fidelity. Previously, possible threats to internal validity have been discussed, and the control for threats has also been introduced. Another concern is how to ensure the process of the study was conducted as designed. Therefore, it was necessary to attend to the treatment fidelity of the study. Borrelli et al. (2004) defined treatment fidelity as a process to enhance the reliability and the validity of interventions from the perspective of methodology. The goal is to make certain that the variations in the DVs are due to the manipulations of IVs, and it can increase the internal validity, external validity, construct validity, and statistical power (Borrelli et al., 2004; Moncher & Prinz, 1991). Specifically, Moncher and Prinz stated treatment fidelity mainly concerns two issues, which are treatment integrity and treatment differentiation. Treatment integrity means that the treatment is deployed as intended. Treatment differentiation refers to the degree to which different levels in an IV can be distinguished from each other. In other words, the manipulation of IVs is made as planned. Borrelli et al. introduced a treatment fidelity framework, which provides five aspects of the enhancement of treatment fidelity. They are (a) design, (b) training, (c) delivery, (d) receipt, and (e) enactment. As these were developed for public health clinical trials, some conditions are different from the current study, but most of them can be applied.

The perspective of design is about the factors that need to be considered when designing the study and the factors that are necessary for the evaluation and replication of the study (Borrelli et al., 2004). Borrelli (2011) asserted that investigators or experts, if possible, need to check the treatment and measurements before the study to assess the fidelity to study design. The treatment in the current study was the auditory cue. Many related studies and publications have been reviewed when designing the treatment. The levels of auditory cues were established based on the findings of past studies, so the levels were differentiable. Also, committee members supervised the design of the current study because it was a dissertation. Moreover, the auditory cues, which were considered treatments, were pre-recorded for consistency purposes in all the sessions. The design and the treatment were described in detail for the purpose of evaluating and replicating the study.

The training was for the experimenters who provided treatment in the study (Borrelli et al., 2004). The goal is to ensure experimenters implement the treatment with the same standard and equivalent skill levels (Borrelli, 2011). In the current study, the treatment was given automatically without the intervention of humans. There was only one experimenter, who reminded participants to lower tray tables if they did not follow the instructions. The same experimenter conducted all the sessions, so the criterion and the skill were constant. Moreover, the experimenter was familiar with the design and knew the relevant knowledge about the study.

The delivery of treatment means the interventions can be delivered to participants as designed (Borrelli et al., 2004). The treatment, which was the auditory cue, was pre-recorded and was played using a Bluetooth speaker. There was a jet engine sound that was played at the volume of 75 dB. The auditory cue was set at 90 dB, so it was more than 13 dB higher than the background noise. The speaker was placed under the seat, the volume to participants' ears was about 75 dB, whereas the engine sound was approximately 60 dB. Therefore, it could ensure the auditory cues were transmitted to participants, and they were loud and clear enough.

The treatment receipt refers to whether or not participants can receive and utilize the treatment (Borrelli, 2011; Borrelli et al., 2004). As stated above, the auditory cues were 15 dB higher than the background noise. The difference in the volume between the engine sound and the auditory cues were larger than the just-noticeable difference (JND). Participants were able to notice the presence of the auditory cues. Moreover, participants were required to be proficient in English and had normal hearing abilities. They were able to hear the announcements and understand the instructions easily.

The treatment enactment means participants can properly perform the treatment-related skills in their daily lives (Borrelli, 2011). It was closely related to clinical studies but not to simulation studies. There were no follow-up measures after the sessions. Therefore, this aspect was not focused on in the current study.

Data Analysis

Data treatment. The data about the performance were analyzed directly, and two types of data were modified. One was the subjective scales, which included PLU, PLR, and attitude. The other one was physiological data.

Subjective mental measures. Each questionnaire had five parts of measures, and the content was the same. A part included three factors, which were PLU, PLR, and annoyance. Each factor provided two reverse-scored items. The paired items were transformed into the same scale, and the average of two items was calculated then. It was not necessary to include the scores of two items; instead, each factor only had one value, which was the average of two items. The parts of different questionnaires were reordered because the order in which auditory cues were played to participants after the simulation were also counterbalanced.

Physiological data treatment. As for physiological data, it was necessary to identify and remove artifacts, which were the contaminants caused by human behaviors or outside stimuli unrelated to the treatment. Also, the physiological data throughout the 40-minute session were grouped. The grouping was mentioned in the previous section in this chapter (Figure 3.1). Briefly, there were nine segments, and eight of them were meaningful, which were the segments for eight announcements. In each segment, the data for the 4.5-minute duration was separated into three parts (Figure 3.2). The parts lasted for 120, 30, and 120 seconds, respectively.

00:00.0	Segment (4:30)	2 min Before	2:00
00:30.0			
01:00.0			
01:30.0			
02:00.0		Stimulation	0:30
02:30.0		2 min After	2:00
03:00.0			
03:30.0			
04:00.0			

***Figure 3.2.* Segment.**

It was proposed that the average values of 2-minutes-before were used as a baseline, and the average data during the stimulation part indicated stimulated arousal levels (refer to Figure 3.2). However, the long duration offsets the effects of the stimulations because the elevation of physiological signals quickly diminished. Therefore, only 30 seconds before and the 10-second duration with the peak after the stimulations were used to calculate the average values. For the HR analysis, the average HR during the 10 seconds before each auditory cue and the average HR during the 10-second period with the peak after each auditory cue were calculated within each part. The data were processed using MATLAB. Raw ECG data recorded the voltages in millivolts generated by atrionectors on human skin every four milliseconds. They were inputted into MATLAB and transformed into an analyzable format. The coding program automatically extracted the data within the simulation duration and detected heartbeats based on the R-peak in each QRS-complex (Figure 3.3). During rest phases, the voltages were around zero volts, and it increased to approximately .6 mv at R-peaks because the atrionector released currents to contract cardiac muscles. Then, it dropped down to -.4 mv, which is S on the ECG diagram. There were several criteria used to detect heartbeats: (a) the beat should be at least .20 mv higher than 20 ms before and at least .3 mv higher than 20 ms after, (b) the beat should not be more than 1.2 mv (this voltage should be an artifact), (c) the beat should be no lower than the past 20-ms period and higher than the next 20-ms period. If the signals were missing for one or two seconds due to the temporary misworking of sensors, up to three heartbeats were inserted to the duration based on adjacent intervals. However, some heartbeats were unable to be detected because the changes in the voltages were not significant enough, and artifacts could be considered heartbeats. Therefore, the real-time heart rate based on each R-R interval was checked. Usually, the heart rate should be between 50 bpm for athletes and 120 bpm. In this case, if the heart rate is below 60 bpm (i.e., 120 ÷ 2), it meant that one heartbeat could be missed. Also, if the HR is over 100 bpm (i.e., 50 × 2), it could have incorrectly detected heartbeats between two correct heartbeats. The detected artifacts could include valid heartbeats, but this method was able to detect all potential artifacts. An ECG diagram around each detected artifact was created for the experimenter to determine if this was an artifact manually. The detected artifacts were all removed and recorded textually. If several heartbeats were not well detected due to the small changes in voltages, a reasonable number of heartbeats were added. These were based on (a) prominent R peaks if the diagrams were correct or (b) the interval between two correctly detected heartbeats and the adjacent heartbeat intervals if the patterns were confused. The HR elevation was an increase from the baseline (i.e., a HR reading before the simulation) to the aroused level (i.e., the HR reading after the stimulation), and it was calculated in percentage instead of the difference. For example, if the HR for one stimulation was 60 bpm before the stimulation and 72 bpm after hearing the cue, the HR elevation would be 20% (i.e., (72-60) ÷ 60). HRV was the ratio of low frequency (LF) to high frequency (HF) within 30 seconds during the stimulations. A higher LF/HF ratio indicates a higher arousal level. As for the skin conductance, the (a) average value within 10-second before stimulations and (b) average value within 10-second that included the peak after stimulations of GSR were acquired. The GSR that was analyzed was GSR elevation, which was the increase in GSR from the baseline before the stimulation to the stimulated level after the stimulation, which was the same as HR elevation. As participants had different baselines and sensitivity to stimuli, the changes were the increases in percentage instead of exact increasing values. The codes are in Appendices H and I.

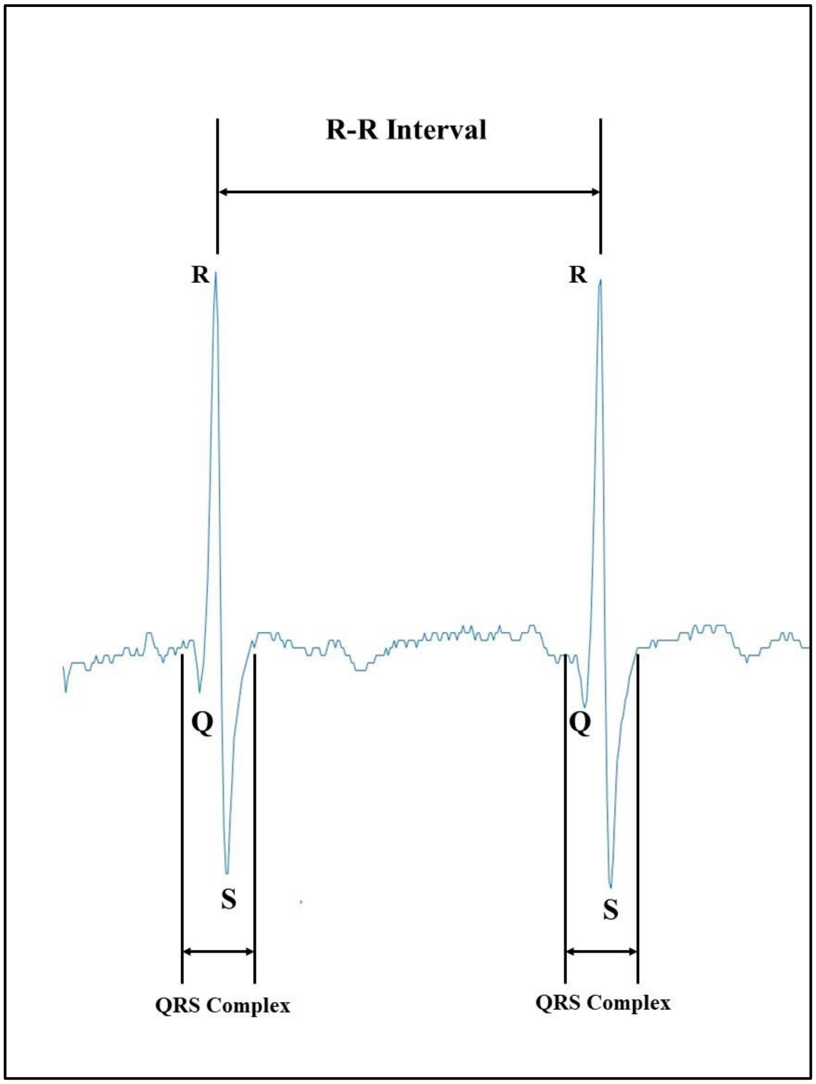

Figure 3.3. **ECG demonstration.**

After the data treatment, the data were analyzed by applying descriptive statistics and inferential statistics. Descriptive statistics were calculated to identify the central tendency and variability. However, inferential statistics were used to identify the significance of the relationships between studied variables.

Descriptive statistics. Initially, the portion of each gender was calculated. The mean, range, and standard deviation (*SD*) regarding the age were also acquired to identify the sample representativeness. In addition to demographics, the measures of arousal levels, perceptions, performance, and attitudes will be analyzed with descriptive statistics. If the variable is continuous, the number of valid data, mean, range and *SD* were identified. If it is a categorical variable, the frequency and percentage were identified.

Inferential statistics. The inferential statistics included three parts: a preliminary analysis, primary analysis, and secondary analysis. In the preliminary analysis, missing data were checked and reported, and outliers were identified. Also, normality as an assumption of used inferential statistical methods were examined. The instrument reliability was also determined.

The primary analysis had the analyses of the following effects and relationships: (a) effect of set A on set B, (b) effect of set A on set C, (c) effect of set A on set D, (d) effect of set A on set E, (e) relationship between set B and set C, (f) relationship between set B and set D, (g) relationship between set C and set D, (h) effect of set B on set E, (i) effect of set C on set E, and (j) effect of set D on set E. Initially, according to the primary purpose of the current study, the effects of auditory cues on performance needed to be assessed. There were two types of data included in the performance. Compliance was categorical, whereas the response time was continuous. To analyze the influence of auditory cues on compliance for a within-subjects effect with multiple levels, a Cochran's Q test was run to test the significance of the omnibus test. The assumptions for the test include (a) one DV with two and mutually exclusive groups, (b) one IV with three or more categorical and related groups, (c) random sampling, and (d) large sample size to interpret the asymptotic *p*-value ("Cochran's Q Test," n.d.). If there was an effect, McNemar tests would be used to make pairwise comparisons between every two auditory conditions individually. The assumptions needed to be checked before running the main analyses, which contained (a) a nominal dichotomous DV and an IV with two connected groups, (b) two groups in the DV are mutually exclusive, and (c) random sampling (Stephanie, 2015). To analyze the effect of auditory cues on response time, a one-way within-subjects ANOVA was applicable. However, when the DV set included a categorical and a continuous DV, it would be difficult to analyze the data. It was determined to combine the DVs and create a new DV, which was named performance scores. Based on the design, if a participant responded to an instruction more than 10 seconds after the announcement, it was considered no compliance. Therefore, the performance score ranged from 0 to 10, which was calculated by subtracting the response times from 10. In this case, a 10 referred to it taking participants 0 seconds to initiate responses, and 0 meant participants did not comply or did not make requested responses within 10 seconds. To analyze the effect of auditory cues on performance scores, a repeated-measures ANOVA was conducted to identify the difference in one parameter at one time among five auditory cues. The assumptions, including (a) normality (i.e., the population where a small sample is from is normally distributed), (b) independence within groups (i.e., participants are observed independently within groups), (c) homogeneity of variance (i.e., the variances in different populations are equal), and (d) homogeneity of covariance (i.e., participant scores in different groups are related), were checked before the analyses (Privitera, 2012). If it showed a significant result, Fisher's least significant difference (LSD) tests were run as post-hoc tests to identify the significant pairs because the variances were not likely to be equal. Similar to the effect of auditory cues on performance scores, a repeated-measures ANOVA was utilized to analyze the effect on (a) physiological parameters, including HR elevation, HRV, and GSR elevation, (b) perceptions, which included PLU and PLR, and (c) annoyance. It was also necessary to test the correlation MV sets by running RMCORR. The assumptions for the RMCORR should be tested. They include assumptions for linear correlation except for independence (Bakdash and Marusich, 2017).

As arousal levels, perceptions, and attitudes were hypothesized to be MVs, and it was mandatory to check influences of MVs on performance scores. RMCORR was run between continuous DV and MVs.

Several potential extraneous variables were also analyzed in the secondary analysis. These included groups regarding the order of counterbalanced auditory cues, gender, age, language, culture, commercial flight experience, and commercial flight habits. Categorical variables were analyzed using between-subjects ANOVA for the variable with more than two group (i.e., group) and independent-samples *t*-test for dichotomous variables (i.e., gender, language, and culture), and continuous variables (i.e., age, commercial flight experience, and commercial flight habits) were examined by conducting Pearson's *r* correlations.

Chapter 4

Results

Introduction

In this chapter, the results of the cabin simulation study are presented. First, the descriptive statistics results are reported. Then, Inferential statistics include three parts, which are preliminary analysis, primary analysis, and secondary analysis. Several preliminary analyses focused on missing data, outliers, normality tests, and reliability of scales. As for the primary statistical analyses, the effects of the independent variables (IV) on mediating variables (MVs) and dependent variables (DVs), effects of MVs on DVs, and relationships among MVs. It also included a secondary analysis, which examined the effects of potential extraneous variables on interesting variables. These were the group (i.e., order of cues), gender, age, language (i.e., native or not), culture (i.e., Western or not), experience in commercial flights, and habits on commercial flights. The puzzle scores were determined not to be analyzed because some participants cheated when solving Sudoku puzzles by using hints. The hypotheses testing results are reported, and a summary of the results.

Descriptive Statistics

This section includes the primary variables and secondary factors. The primary variables were comprised of performance scores, physiological data (i.e., HR elevation, HRV, and GSR elevation), perceptions, and attitudes. The secondary factors were demographics, cultural background, and attitudes toward the experiment.

Demographics. There were 46 participants consisting of 26 males (57%), and 20 females (43%). The age ranged from 18 to 36, with a mean of 20.89 and a standard deviation (*SD*) of 3.62. Thirty-six participants (78%) were native speakers. Participants reported 6.22 commercial flights on average with an *SD* of 8.95. Also, participants were asked about their safety-related behaviors on commercial flights, including (a) complying with the fasten-seatbelt instruction, (b) complying with the airplane mode instruction, and (c) listening to the pre-flight safety instruction. The frequency of participants' compliance with in-flight instructions is displayed in Table 4.1.

Table 4.1

Compliance with Instructions

Instructions	*N*	Mean	*SD*	Minimum	Maximum
Fasten Seatbelts	46	89%	24%	10%	100%
Airplane Mode	46	75%	38%	0%	100%
Pre-flight Briefings	46	51%	34%	0%	100%

Note. The unit is in the percentage of the frequency of each instruction.

Cultural background. Most participants were American. The nationality of 29 (63%) participants was the Unites States, and 32 (70%) participants were born in the United States. Thirty-four (74%) participants attended US high schools, and 43

(94%) had been or were enrolled in an undergraduate degree in the United States. Also, 36 participants grew up in Western countries, and 38 participants were educated during their teen years in Western countries. This included two participants that were born in China but were educated in the United States.

Performance. The performance measure included compliance and response times regarding action-required instructions. There were four action-required instructions out of eight total instructions following auditory cues with four levels of intensity. The data were collected based on the recordings of simulations, so some data were missing due to technical failure of the camera and exclusion of outliers.

For four instructions, the mean response times were between one and two seconds. Means, *SD*s, and ranges are presented in Table 4.2 and Figure 4.1. Compliance is displayed in Figures 4.2.

Table 4.2

Response Time

Parameters	*N*	Mean	*SD*	Minimum	Maximum
Intensive 1	42	1.57	.84	.42	4.63
Intensive 2	43	1.29	.88	.27	4.21
Intensive 3	40	1.03	.43	.19	1.96
Intensive 4	42	1.53	1.34	.54	9.03

Note. The value unit is in second.

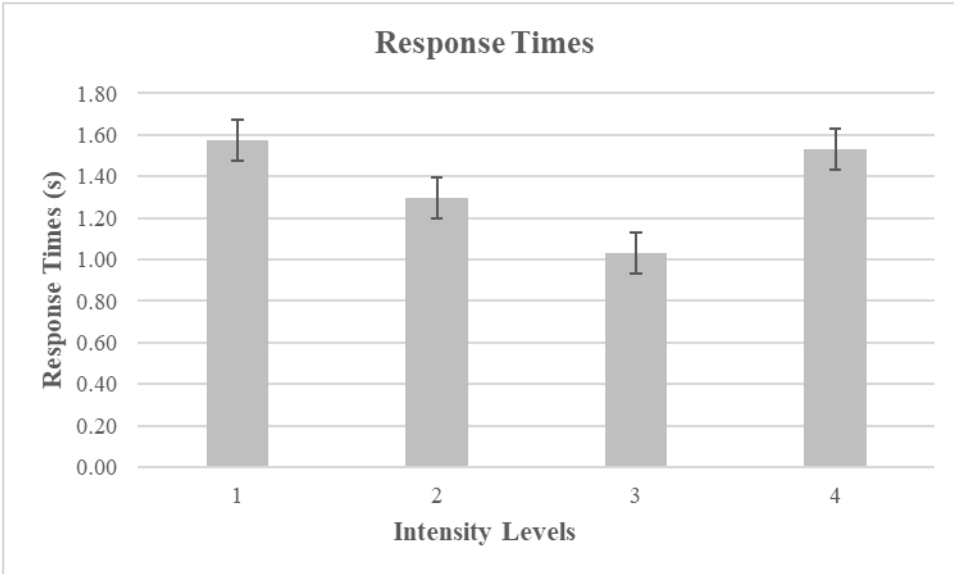

Figure 4.1. **Response times. This figure shows intensity levels 1 to 4.**

Figure 4.2. **Compliance with instructions. This figure shows intensity levels 1 to 4.**

As there were two DVs, and one of them was dichotomous variables, it is challenging to conduct an omnibus test to test the experimentwise significance. For the convenience of analyzing performance, response time is combined with compliance, where a high score indicated good performance. The detail of the data transformation is described in Chapter 3. The descriptive statistics of performance scores are presented in Table 4.3, and the trend is demonstrated in Figure 4.3.

Table 4.3

Performance

Parameters	*N*	Mean	*SD*	Minimum	Maximum
Intensive 1	42	8.43	.84	5.37	9.58
Intensive 2	43	8.71	.88	5.79	9.73
Intensive 3	40	8.97	.43	8.04	9.81
Intensive 4	42	8.47	1.34	.97	9.46

Note. The unit is the performance score, where a 1-point increase indicates a 1-second decrease in response time. The descriptive results included outliers and cases with missing data.

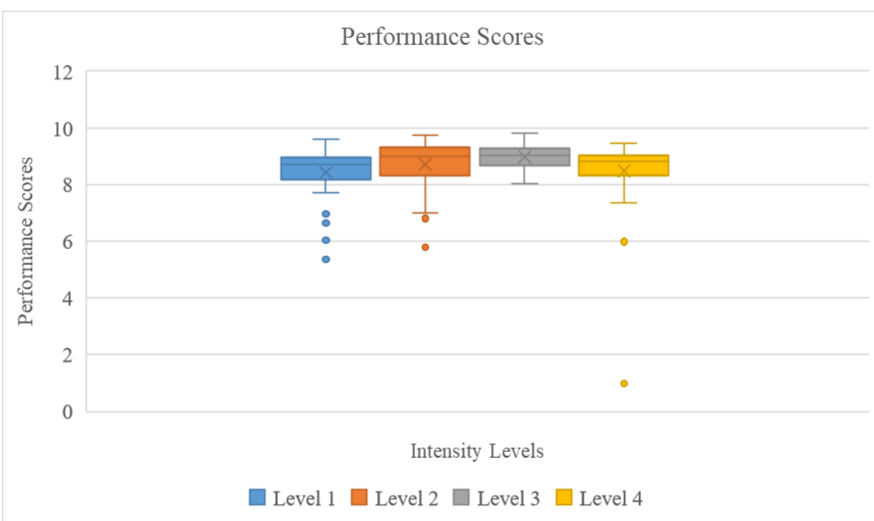

***Figure 4.3*. Performance score. This figure shows intensity levels 1 to 4.**

Physiological data. In addition to four instructions, the baseline is also included. There were four general instructions, so the baseline was the average of the four data points from the baseline levels before general instructions. It needs to be noted that heart rate (HR) and skin conductance (i.e., GSR) were the increases from the baseline to increased levels in percentages. The unit of HR elevation and GSR elevation was in the form of a percentage due to a wide variety of baselines among participants, which was the difference from the physiological signals before the auditory cue. The heart rate variability (HRV) was the ratio of low frequency to high frequency. The descriptive statistics of HR, HRV, and GSR elevation are summarized in Tables 4.4, 4.5, and 4.6. The physiological data were presented in Figures 4.4, 4.5, and 4.6. It should be noted that baselines were much lower partly because they did not need to move their bodies after hearing the general instructions.

Table 4.4

Heart Rate Data

Levels	*N*	Mean	*SD*	Minimum	Maximum
Baseline	45	4.23%	5.49%	-8.04%	20.83%
Intensity 1	45	13.54%	13.31%	-9.27%	55.23%
Intensity 2	45	11.57%	12.32%	-12.66%	45.42%
Intensity 3	44	13.06%	11.92%	-10.65%	43.86%
Intensity 4	45	11.08%	12.41%	-9.54%	34.32%

Note. The values show the increases in HR from the baseline to the simulated level in percentage. The descriptive results included outliers and cases with missing data.

Table 4.5

Heart Rate Variability

Levels	*N*	Mean	*SD*	Minimum	Maximum
Baseline	45	1.44	0.18	1.16	2.28
Intensity 1	45	1.50	0.18	1.19	1.98
Intensity 2	44	1.46	0.23	1.12	2.51
Intensity 3	44	1.48	0.20	1.18	2.20
Intensity 4	45	1.47	0.30	0.00	2.09

Note. The values show the HRV calculated using LF/HF. The descriptive results included outliers and cases with missing data.

Table 4.6

Skin Conductance Data

Levels	*N*	Mean	*SD*	Minimum	Maximum
Baseline	33	5.32%	7.87%	-0.15%	39.94%
Intensity 1	33	13.32%	18.32%	-11.37%	74.87%
Intensity 2	33	10.77%	13.50%	-1.73%	58.03%
Intensity 3	33	11.84%	18.32%	-1.27%	98.07%
Intensity 4	33	9.43%	12.40%	-0.28%	54.03%

Note. The values show the increases in GSR from the baseline to the simulated level in percentage. The descriptive results included outliers and cases with missing data.

Figure 4.4. **Heart rate increase. This figure shows the baseline level and four intensity levels 1 to 4.**

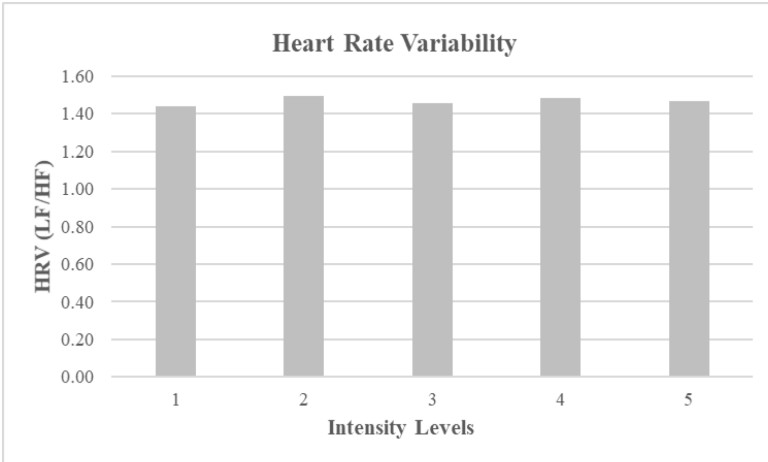

Figure 4.5. **Heart rate variability. This figure shows the baseline level and four intensity levels 1 to 4.**

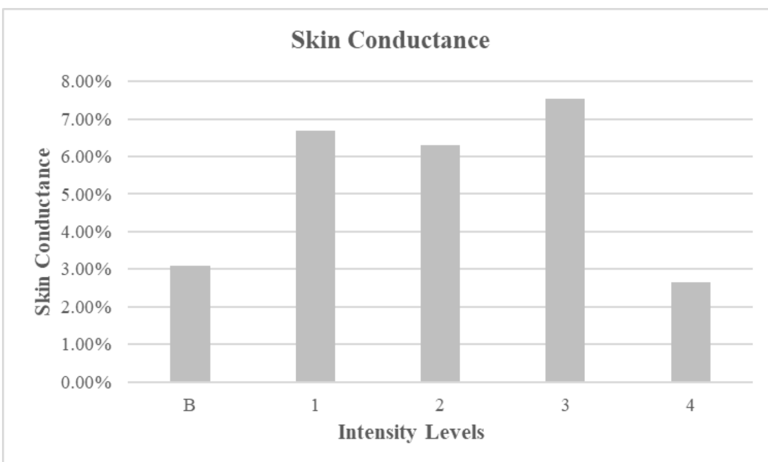

Figure 4.6. **Skin conductance percentage. This figure shows the baseline level and four intensity levels 1 to 4.**

Perceptions. Perceptions include two predictors that indicate participants' perceptions regarding auditory cues, including perceived level of urgency (PLU) and perceived level of risk (PLR). The means and *SD*s for each factor are displayed in Tables 4.7 and 4.8. Trends of measurements are illustrated in Figure 4.7.

Table 4.7

Perceived Urgency Level

Levels	***N***	**Mean**	***SD***	**Minimum**	**Maximum**
Baseline	46	4.05	1.91	1.0	8.0
Intensity 1	46	5.34	1.75	1.0	8.0
Intensity 2	45	5.57	1.74	2.0	9.0
Intensity 3	46	7.09	1.40	3.5	9.0
Intensity 4	46	7.35	1.81	1.0	9.0

Note. The values are based on a scale of 1 to 9. The descriptive results included outliers and cases with missing data.

Table 4.8

Perceived Risk Level

Levels	***N***	**Mean**	***SD***	**Minimum**	**Maximum**
Baseline	46	3.35	1.50	1.0	6.5
Intensity 1	46	4.70	1.78	1.0	8.0
Intensity 2	46	4.77	1.60	1.0	8.0
Intensity 3	46	6.50	1.58	2.0	9.0
Intensity 4	46	6.77	1.76	1.0	9.0

Note. The values are based on a scale of 1 to 9. The descriptive results included outliers and cases with missing data.

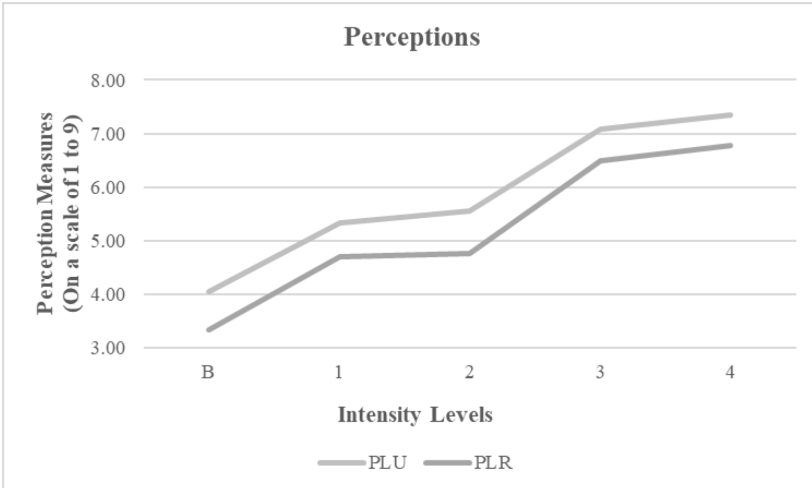

Figure 4.7. **Perception trends. This figure shows the baseline level and four intensity levels 1 to 4.**

Attitudes toward cues. Attitudes refer to participants' annoyance with cues. The descriptive statistics are displayed in Table 4.9, and the trend is presented in Figure 4.8.

Table 4.9

Annoyance

Levels	*N*	Mean	*SD*	Minimum	Maximum
Baseline	46	4.28	1.95	1.0	9.0
Intensity 1	46	5.33	1.70	1.5	9.0
Intensity 2	46	5.76	1.67	2.0	9.0
Intensity 3	46	6.62	1.64	3.0	9.0
Intensity 4	46	6.95	1.73	2.0	9.0

Note. The values are based on a scale of 1 to 9. The descriptive results included outliers and cases with missing data.

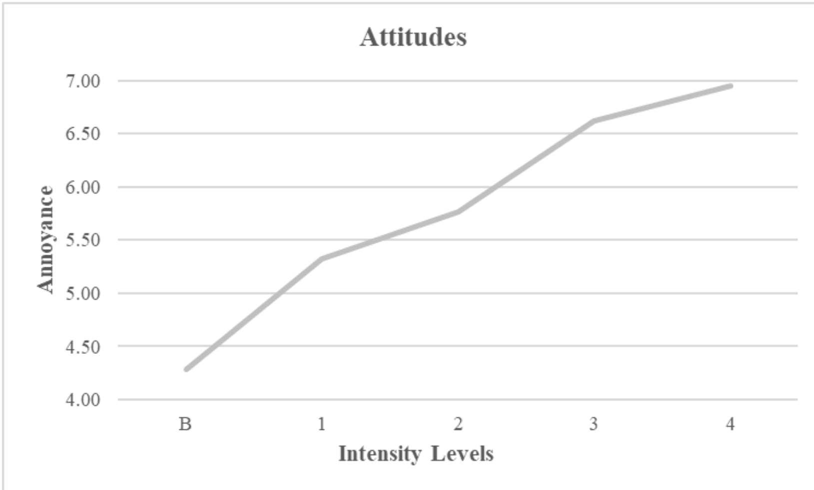

Figure 4.8. **Annoyance. This figure shows the baseline level and four intensity levels 1 to 4.**

Attitudes toward simulation. Participants were asked about their boredom, anger, and interests regarding the study. In the questionnaire, the item that measured anger asked how mad participants were at the simulation, but it collected the feeling of anger instead of insanity. Interest is a reverse score of boredom. The descriptive attitudes toward the study are displayed in Table 4.10.

Table 4.10

Attitude Toward Experiments

Variables	*N*	Mean	*SD*	Minimum	Maximum
Boredom	46	2.91	1.03	1	5
Anger	46	1.37	0.68	1	4
Interest	46	3.46	0.84	1	5

Note. The values are based on a scale of 1 to 5.

Inferential Statistics

The purpose of the current study was to (a) examine the effects of IV on DV, (b) determine the influences of IV on MVs, and (c) test the mediation effect of MVs between IV and DV. The relationships among MVs were also examined. The IV was the auditory cue intensity level. The DV indicated passengers' performance. MVs were comprised of perception, arousal levels (i.e., HR elevation, HRV, GSR elevation), and annoyance (i.e., attitude toward cues). The preliminary analysis checked missing data, potential outliers, and determine normality. The reliability of subjective scales was identified as well. In the primary analysis, the effect of the IV on the DV and MVs, the effects of MVs on the DV, and the relationships between each pair of MVs were analyzed. When analyzing the effects of IV, within-subjects analysis of variance (ANOVA) was conducted because there was one group with more than two levels. As for the effects of MVs on the DV and relationships among MVs, they were all continuous variables that were within-subjects, so repeated-measure correlation (RMCORR) was run to analyze the effects. The secondary analysis analyzed the effects of several extraneous variables on interesting factors.

Preliminary analysis. In this part, the missing data and outliers were discussed. Then, the normality of each variable was checked because normality is an assumption of almost every inferential analysis. The reliability of scales was calculated.

Missing data. There were three cases that included missing data of compliance because of the failure of cameras. As for response times, 14 cases were found to have missing data. If the participant did not follow the instruction, no corresponding response time was recorded. Also, some participants performed other tasks first (e.g., participants fasten seatbelts first, which was not asked, before lowering the tray table), so the response times were not able to be determined. For physiological data, the number of missing data regarding HR and HRV was two, and galvanic skin response (GSR) contained 13 cases with missing data. The reasons were (a) failure of the equipment due to low battery, (b) the devices did not contact with participants' skins well when they moved their bodies, and (c) the sensors suffered from insensitivity due to sweat. As psychological measure items were created in pairs, no participant missed a pair of items, and then no data were missing. The missing data were objectively measured, so they were not plugged with the means.

Outliers. Outliers were detected before descriptive and inferential analyses. Most statistical methods require no significant outlier, and descriptive statistics are more reasonable to be based on the treated data. Jackknife's distance analysis was used to recognize potential outliers statistically.

Performance scores were calculated based on compliance, which is a categorical variable, and response time, which is a continuous variable. It was determined to use performance scores to illustrate two variables because compliance is a categorical DV, leading to difficulty of data analyses. According to the performance scoring method, those who did not comply with the instruction were scored 0. As many data were close to 10, most cases with 0 included were identified as outliers. However, if cases including 0 (i.e., not complied) were removed, performance scores only demonstrate the performance of the subjects who complied with the instruction, and it was the same as the response time only. Whereas, if they were included, large variabilities would sabotage the significance of the auditory cues' effects on participants' performance. Therefore, it is necessary to test the significance of compliance to determine if compliance data should be included. A Cochran's Q was run to analyze the difference in a related-measures categorical variable with four levels. The result appeared non-significant, $\chi^2(3) = 3.00$, $p = .39$. Considering the non-complied cases were each below 12% (3, 1, 5, and 3 out of 45), outliers of the performance scores were not included in the analysis. There were 32 valid cases included as some other response times were missing due to technical difficulties in experiments.

With respect to the physiological data, three cases for HR, five cases for HRV, and five cases for GSR were recognized as outliers, and seven cases were identified as outliers regarding mental measures. Detected outliers were within reasonable ranges, so they were kept for inferential statistics. As participants were sweating, the sensitivity of the GSR sensor was decreasing during sessions, so it needs to be noted that many GSR readings in the second half of the sessions could be inaccurate. Perceptions and attitudes did not include outliers, but one person rated 9 and 9 on two reverse-scored items, and they were removed as outliers.

Normality. There were four levels included in the performance score. Intensity levels 1, 2, and 4 were not compliant with normal distribution based on the statistical methods, where Kolmogorov-Smirnov and Shapiro-Wilk results were below .05. However, based on Q-Q plots, the distribution of these levels can be considered normal regarding a less strict criterion (Figures J.1, J.2, and J.3). The scattered points of level 1 were roughly aligned with reference lines. The data points of levels 1 and 4 were somewhat off the reference line, which indicated that the data were not very normal, but the analysis proceeded anyway.

As for physiological data, all levels of HR were normally distributed, because *p*-values for Kolmogorov-Smirnov and Shapiro-Wilk tests were all above .05. Kolmogorov-Smirnov tests showed non-significant for baseline, 1, and 2 levels of HRV but significant for others. Shapiro-Wilk tests were significant for each level. The Q-Q plots showed that they were normal with a less rigorous standard because they were aligned with the reference line (Figures J.4, J.5, J.6, J.7, and J.8). However, distributions of GSR levels were found not normal as one participant had sensitive GSRs to auditory cues. The Q-Q plots showed they could be treated as normal distributions when removing high scores (Figures J.9, J.10, J.11, J.12, and J.13).

The intensity levels 3 and 4 of PLU and the intensity level 4 of PLR were not normal based on the tests. Intensity levels 3 and 4 of annoyance were also non-normal. However, the points on Q-Q plots were roughly aligned with the normal line, so they can be considered normal (Figures J.14, J.15, J.16, J.17, and J.18).

Besides MVs and DVs, demographics also included some continuous variables, which were (a) age, (b) the annual frequency of commercial flights, and (c) frequency of listening to or compliance with safety instructions. Kolmogorov-Smirnov and Shapiro-Wilk tests showed that the distribution of age was not normal. The Q-Q plot (Figure J.19) presented that several points with high values were off the reference line, so it was not normally distributed, but the analyses proceeded with this distribution.

The flight variable was also not normally distributed due to two extremely high numbers, which were 40 and 50. This may have been attributed to pilots possibly counting the flights on which they were operators. They were also recognized as outliers by conducting Jackknife's distance analysis. As correlation requires two normal-distributed variables, they should be excluded before inferential analyses. Although Kolmogorov-Smirnov and Shapiro-Wilk tests still showed a significant difference from a normal distribution, the Q-Q plot presented an acceptable pattern (see Figure J.20).

As for participant's safety instruction habits, tests of normality showed significant results. The scattered points of the fasten-seatbelt instruction were somewhat off the reference normal-distributed line, and the reason was that most participants reported 100% of compliance, which made it left-skewed, but it was analyzed anyways (Figure J.21). Distributions of the airplane-mode instruction and the safety briefings were nearly normal (Figures J.22 and J.23). It also showed significant Kolmogorov-Smirnov and Shapiro-Wilk results. Whereas, Q-Q plots showed the data approximately normally-distributed patterns (Figures J.24 and J.26) except for anger (Figures J.25). As most participants did not report they were mad, the distribution was skewed to the right, and anger was analyzed with unnormal distribution.

Subjective scales reliability. It is also necessary to test the reliability of mental measure scales. The items within each pair should be correlated with each other negatively to indicate high equivalent-forms reliability. The results showed a significant correlation between items, and p-values were all below .01. Therefore, the scores in paired items were combined. Also, the internal consistency of mental measures was checked. Cronbach's α of the PLU, PLR, and annoyance were .84, .78, and .85, respectively.

Summary. In the preliminary analysis, (a) normality of continuous variables, (b) influences of the factors other than IVs on MVs and DVs, and (c) reliability of scales were checked. Normality was good for most levels of variables, and it was determined that the analyses proceeded with several unnormal variables. Reliability was acceptable after removing extreme data.

Primary analysis. The model included (a) a DV, which was performance, (b) MVs, which had physiological signals, perception, and annoyance, and (c) an IV, which was the intensity level of auditory cues. Variables were within-subjects, and they were in a mediation model. Therefore, an omnibus test was not appropriate. There were five levels of auditory cue intensity levels, including a baseline level and intensity levels 1 to 4. The baseline level was featured with 0.5 kHz, sine wave, one pulse, and 600 ms. Intensity level 1, compared to the baseline level, increased the frequency from 0.5 kHz to 1.0 kHz. Level 2 had all level 1 characteristics except for waveform, which was altered to become a triangle wave. The level-3 auditory cue had two pulses of tones with other acoustic parameters the same as level 2. Level 4 increased the length of each pulse from 600 ms to 1.2 seconds, and other features were the same as level 3.

As mentioned before, within-subjects ANOVA and RMCORR were conducted. The assumptions for ANOVA include (a) independent observations, (b) normality, and (c) sphericity. Independent observations mean that each individual was measured independently, and variables were independent and identically distributed. Normality refers to the requirement that each variable was normally distributed, which was tested. Sphericity means the scores have equal variances, and it can be tested with Mauchly's test. RMCORR has the assumptions of (a) level of measurement, (b) related pairs, (c) no outliers, (d) normality, (e) linearity, and (f) homoscedasticity. These assumptions have been introduced in the previous section.

For the overall post-hoc analyses, Fisher's least significant difference (LSD) was applied to test differences between levels. Bonferroni analyses are popular; however, the analysis simply divides the *p*-value equally. As variances were not equal, it is appropriate to conduct LSD.

Effect of cues on performance scores. A Mauchly's test was run to examine sphericity, and the sphericity assumption was violated. Therefore, the Greenhouse-Geisser test was conducted instead, which showed a significant result, $F(2.38, 73.81) = 5.33$, $p < .01$, $\eta_p^2 = .15$. It formed a significant quadratic relationship, $F(1, 31) = 13.49$, $p < .01$, $\eta_p^2 = .30$. Participants had better performance when hearing cues of Intensity levels 2 and 3 than 1, and performance scores of level 3 were higher than 4. Pairwise comparisons are displayed in Table 4.11. The descriptive information was illustrated in descriptive statistics.

Table 4.11

Performance Pairwise Comparisons

Level 1	Mean	Level 2	Mean	Differences	Significance
I1	8.32	I2	8.87	- .56	< .01
I1	8.32	I3	8.96	- .64	< .01
I3	8.96	I4	8.64	.32	.03

Note. I = Intensity. The unit is the performance score ranging from 0 to 10, where a 1-point increase indicates a 1-second decrease in response time.

Effect of cues on the perception. Perception includes two factors, which are PLU and PLR. Each variable included five levels, but performance scores had only four levels. Therefore, the variables were analyzed twice. One analysis included five levels, and the other one had four levels with baseline excluded. The five-level analysis was not very meaningful as performance only included four levels, which meant the results could not be combined with performance. Therefore, five-level analyses would simply provide more information for readers. The four-level analyses were analyzed and described exhaustively, which identified potential within-subjects contrasts (e.g., linear, quadratic, etc.). When five levels are included, the sphericity was satisfactory. It showed a significant effect, $F(4, 176) = 38.13$, $p < .01$, $\eta_p^2 = .46$. Post-hoc analyses demonstrated that the baseline-level PLU was lower than intensity levels, and intensity levels 3 and 4 were higher than levels 1 and 2. Afterward, another analysis was conducted with four levels included. Assumptions were all satisfactory, and it showed the difference in PLU was significant, $F(3, 132) = 25.64$, $p < .01$, $\eta_p^2 = .39$. It also showed a significant linear effect, $F(1, 44) = 59.76$, $p < .01$, $\eta_p^2 = .58$. At intensity levels 3 and 4, participants perceived higher levels of urgency than those at levels 1 and 2 (Table 4.12). As for PLR, when the baseline level was included, the sphericity was violated. A Greenhouse-Geisser method was used, and it showed a significant result, $F(3.19, 143.45) = 44.07$, $p < .01$, $\eta_p^2 = .50$. The post-hoc pattern was similar to PLU, where the baseline was lower than intensity levels, and levels 3 and 4 were higher than levels 1 and 2. Also, a repeated-measures ANOVA was conducted without the baseline level. Sphericity was violated, so Greenhouse-Geisser was run.

The difference was significant, $F(2.37, 106.83) = 25.96$, $p < .01$, $\eta_p^2 = .37$, and it showed a significant linear effect, $F(1, 45) = 42.58$, $p < .01$, $\eta_p^2 = .49$. Pairwise comparisons showed the same difference pattern as PLU (see Table 4.12).

Table 4.12

Perception Pairwise Comparisons

PLU Level 1	Mean 1	PLU Level 2	Mean 2	Differences	Significance
I1	5.36	I3	7.07	-1.71	< .01
I1	5.36	I4	7.33	-1.98	< .01
I2	5.57	I3	7.07	-1.50	< .01
I2	5.57	I4	7.33	-1.77	< .01
PLR Level 1	**Mean 1**	**PLR Level 2**	**Mean 2**	**Differences**	**Significance**
I1	4.70	I3	6.50	-1.80	< .01
I1	4.70	I4	6.77	-2.08	< .01
I2	4.77	I3	6.50	-1.73	< .01
I2	4.77	I4	6.77	-2.00	< .01

Note. I = Intensity. The values are based on a scale of 1 to 9.

Effect of cues on the attitude. An ANOVA with five levels of attitudes was employed first, and the sphericity was violated, so a Greenhouse-Geisser correction was conducted. It showed a significant effect, $F(3.22, 144.96) = 28.83$, $p < .01$, $\eta_p^2 = .39$. A post-hoc result showed that the baseline-level attitude was lower than intensity levels, and levels 1 and 2 were also lower than levels 3 and 4. When only intensity levels were analyzed, the sphericity was satisfactory. The result was significant, $F(3, 135) = 16.53$, $p < .01$, $\eta_p^2 = .27$, where participants reported intensity levels 1 and 2 were less annoying than 3 and 4. The results of pairwise comparisons are summarized in Table 4.13. Also, it showed a significant linear relationship, $F(1, 45) = 37.71$, $p < .01$, $\eta_p^2 = .46$.

Table 4.13

Attitude Pairwise Comparisons

Level 1	Mean	Level 2	Mean	Differences	Significance
I1	5.33	I3	6.62	-1.29	< .01
I1	5.33	I4	6.95	-1.62	< .01
I2	5.76	I3	6.62	-0.86	< .01
I2	5.76	I4	6.95	-1.18	< .01

Note. I = Intensity. The values are based on a scale of 1 to 9.

Effect of cues on heart rates. A repeated-measures ANOVA was run with the baseline level, and the sphericity was satisfactory. The result showed a significant effect, $F(4, 172) = 8.75$, $p < .01$, $\eta_p^2 = .17$. A post-hoc analysis indicated that the HR of the baseline level was significantly lower than the ones of intensity levels, and there was no difference among intensity levels. Pairwise results are displayed in Table 4.14. There was also a significant linear effect, $F(1, 43) = 13.17$, $p < .01$, $\eta_p^2 = .23$, a quadratic effect, $F(1, 43) = 15.21$, $p < .01$, $\eta_p^2 = .26$, and a cubic effect, $F(1, 43) = 7.55$, $p < .01$, $\eta_p^2 = .15$. When the baseline level was excluded, the sphericity was still not violated, but the effect was no longer significant, $F(3, 129) = 1.10$, $p = .35$, $\eta_p^2 = .03$ (power = .29).

Table 4.14

HR Pairwise Comparisons

Level 1	Mean	Level 2	Mean	Differences	Significance
B	-2.45%	I1	3.77%	-6.23%	< .01
B	-2.45%	I2	3.50%	-5.95%	< .01
B	-2.45%	I3	1.77%	-4.23%	< .01
B	-2.45%	I4	2.50%	-4.95%	< .01

Note. B = Baseline. I = Intensity. The values show the increases in HR from the baseline to simulated level in percentage.

Effect of cues on heart rate variability. When five HRV levels were analyzed, sphericity was violated, so a Greenhouse-Geisser correction was used. There was no significant effect identified, $F(2.35, 101.05) = .82$, $p = .52$, $\eta_p^2 = .02$ (power = .26). If only four levels were included, the sphericity was violated as well, so the Greenhouse-Geisser correction was employed. The result showed no significant effect, $F(2.26, 97.06) = .42$, $p = .69$, $\eta_p^2 = .01$ (power = .13).

Effect of cues on skin conductance. If all levels were included, it showed the sphericity was violated, and a Greenhouse-Geisser correction was conducted. The result was significant, $F(2.82, 87.26) = 3.05$, $p = .04$, $\eta_p^2 = .09$. The baseline level was significantly lower than the other levels, and no significant difference was shown among intensity levels. The results are presented in Table 4.15. It also showed a significant quadratic effect, $F(1, 31) = 6.15$, $p = .02$, $\eta_p^2 = .17$. When only four levels were included without the baseline, the sphericity was violated as well, and a Greenhouse-Geisser correction was used. The result showed no significant effect, $F(3, 93) = .61$, $p = .61$, $\eta_p^2 = .02$ (power = .17).

Table 4.15

GSR Pairwise Comparisons

Level 1	Mean	Level 2	Mean	Differences	Significance
B	4.25%	I1	-7.25%	-7.25%	< .01
B	4.25%	I2	-6.16%	-6.16%	< .01
B	4.25%	I3	-4.94%	-4.94%	< .01
B	4.25%	I4	-4.94%	-4.16%	.03

Note. B = Baseline. I = Intensity. The values show the increases in GSR from the baseline to the simulated level in percentage.

Effect of perception on performance scores. A correlation between two within-subjects variables, a repeated measures correlations (RMCORR) package developed by Bakdash and Marusich (2017) was run in the RStudio. The assumptions were those that applied to between-subjects correlation, expect independence, and were checked. The linear correlation between PLU and performance was not significant, $r = .06$ (95% CI: -.12 to .23), $df = 120$, $p = .53$ (see Figures 4.9), and the power was around 20% (Bakdash & Marusich, 2017). The correlation between PLR and performance was also not significant, $r = .04$ (95% CI: -.14 to .22), $df = 120$, $p = .64$, where the power was below 20% (Figures 4.10). Nevertheless, it should not form a linear relationship. If it was congruent with the Yerkes-Dodson Law, the trends of data should form inverted-U shapes between perception and performance. Figures 4.11 and 4.12 showed inverted-U curves, and quadratic trendlines were added to emphasize the curve. The significance could not be determined because the relationship of the Yerkes-Dodson law was inverted-U but not specified, such as quadratic. It was discovered that the effect of auditory cues intensity on performance formed a significant quadratic relationship, $F(1, 31) = 13.49$, $p < .01$, $\eta_p^2 = .30$, which was an inverted-U shaped. In figures 4.11 and 4.12, points 2 and 3 were significantly larger than 1 vertically, and point 3 was higher than 4. The effect of auditory cue intensity on perception showed a significant linear relationship, $F(1, 44) = 59.76$, $p < .01$, $\eta_p^2 = .58$ for PLU and $F(1, 45) = 42.58$, $p < .01$, $\eta_p^2 = .49$ for PLR. Points 3 and 4 were significantly higher than 1 and 2 longitudinally. The direct nonlinear effect of perception on performance was not determined as it was a repeated-measures design. It can be implicated that the relationship between perceptions and performance followed the Yerkes-Dodson law.

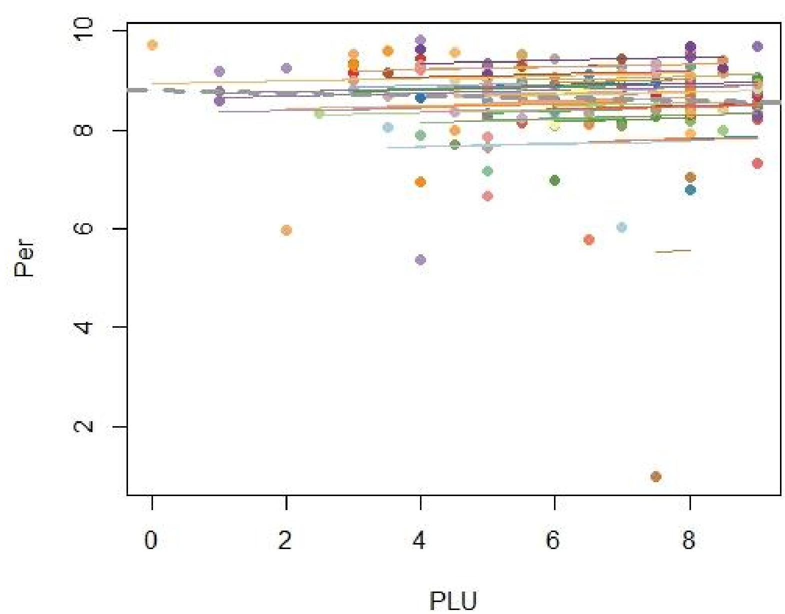

Figure 4.9. **Repeated-measure correlation between PLU and performance.**

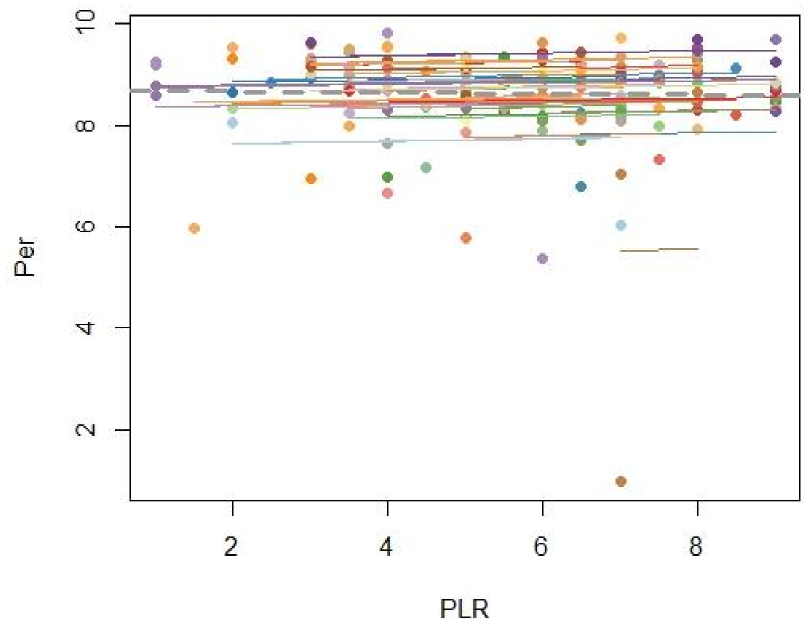

Figure 4.10. **Repeated-measures correlation between PLR and performance.**

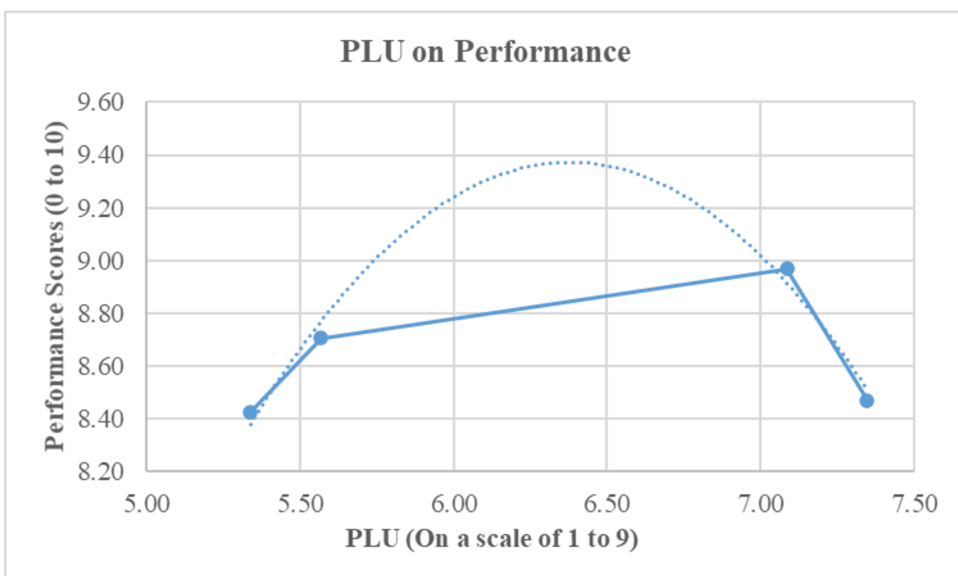

Figure 4.11. **Nonlinear relationship between PLU and performance.**

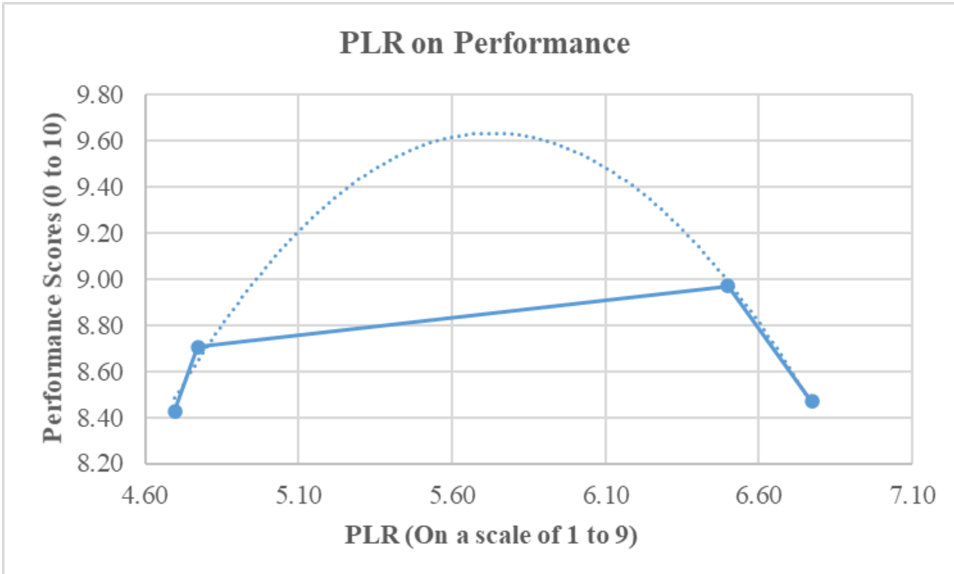

Figure 4.12. **Nonlinear relationship between PLR and performance.**

Effect of attitudes on performance scores. The Bakdash and Marusich's (2017) RMCORR package was conducted again to analyze the correlation between annoyance and performance. The linear correlation was not significant, $r = .03$ (95% CI: -.15 to .20), $df = 120$, $p = .78$, where the power is less than 20% (Figure 4.13). Similar to perceptions, it was discovered that the effect of auditory cues intensity on performance formed a significant quadratic relationship, $F(1, 31) = 13.49$, $p < .01$. The participants' performance when hearing the auditory cues of intensity level 3 was significantly better than levels 1 and 4, and the performance score of level 2 was higher than 1. The effect of auditory cue intensity on attitudes showed a significant linear relationship, $F(1, 45) = 37.71$, $p < .01$. Participants reported they had more annoyance with levels 3 and 4 than 1 and 2. The effect of auditory cue intensity on attitudes was not combined, so the direct nonlinear effect of perception on performance was not determined as it was a repeated-measures design. Nevertheless, the relationship between attitudes and performance showed an inverted-U curve, as displayed in Figure 4.14.

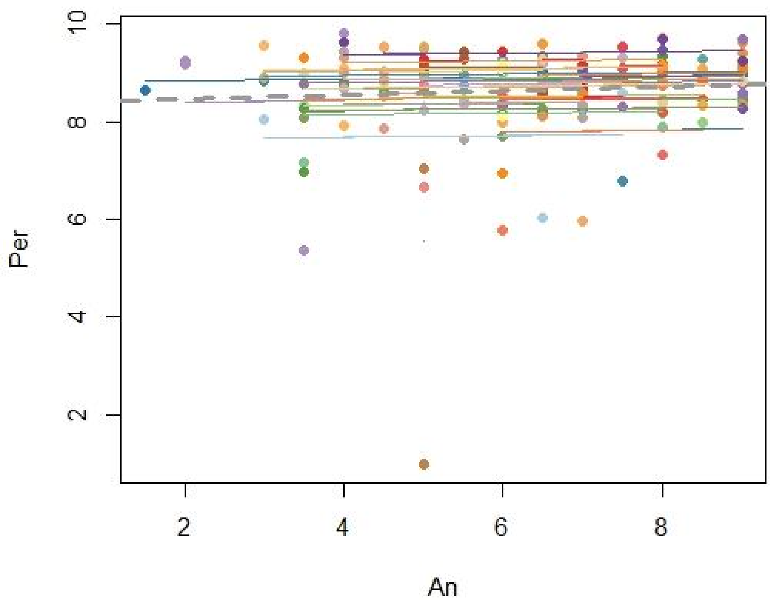

Figure 4.13. **Repeated-measure correlation between annoyance and performance.**

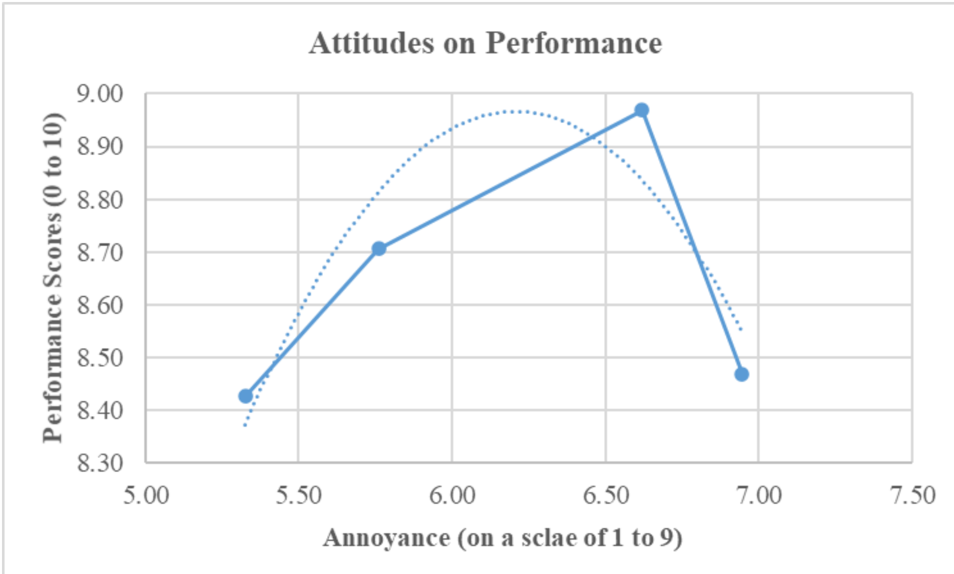

Figure 4.14. **Nonlinear relationship between annoyance and performance.**

Effect of heart rates on performance. An RMCORR test was conducted to test the effect. No significant effect of HR on performance was found, $r = .10$ (95% CI: -.08 to .28), $df = 117$, $p = .28$, and the power is about 20% (Bakdash & Marusich, 2017; Figure 4.15). Similarly, the linear relationship was not significant. The effect of auditory cues intensity on performance formed a significant quadratic relationship, $F(1, 31) = 13.49$, $p < .01$, $\eta_p^2 = .30$, which was an inverted-U shaped. When hearing the auditory cues of intensity level 3, participants performed significantly better than when hearing levels 1 and 4, and the performance score of level 2 was higher than 1. Nevertheless, there was no significant linear effect of auditory cues on HR elevations without the baseline level, and it did not show an inverted-U trend (Figure 4.16).

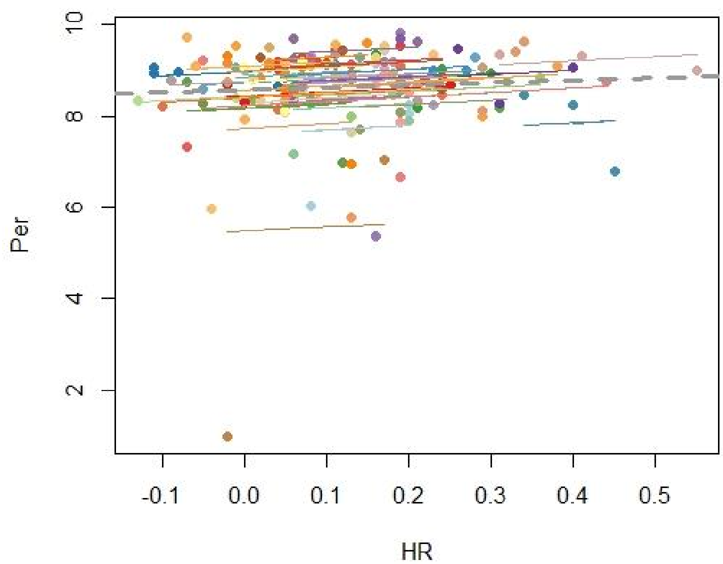

Figure 4.15. **Repeated-measure correlation between HR and performance.**

Figure 4.16. **Nonlinear relationship between HR and performance.**

Effect of heart rate variability on performance. An RMCORR test was conducted to test the effect. A non-significant effect of HR on performance was found, r = -.03 (95% CI: -.21 to .15), df = 115, p = .72, and the power was below 20%. The linear relationship was not significant, and it did not show an inverted-U relationship (Figure 4.17).

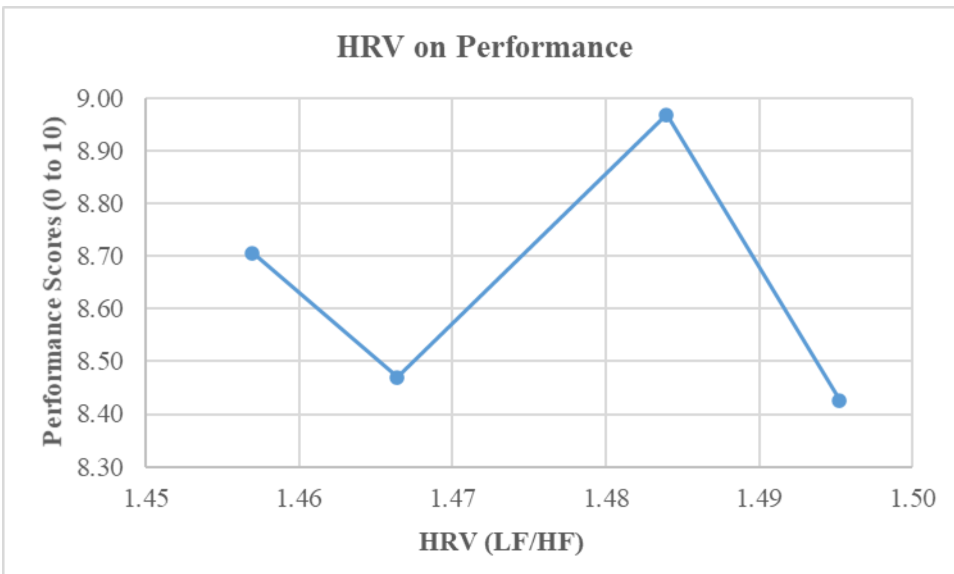

Figure 4.17. **Linear relationship between HRV and performance.**

Effect of skin conductance on performance. An RMCORR test was also run for this analysis. There was no significant linear relationship, r = -.12 (95% CI: -.32 to .10), df = 85, p = .27, where the power is slightly above 20% (Figure 4.18). It showed that the effect of auditory cues intensity on performance formed a significant quadratic relationship, $F(1, 31) = 13.49$, $p < .01$, $\eta_p^2 = .30$, which was an inverted-U shaped. Participants' performance scores of intensity level 3 were significantly higher than levels 1 and 4, and the performance score of level 2 was higher than 1. However, there was no significant effect of intensity levels on GSR when only four intensive levels were included. Also, the scattered plot did not show a possible inverted-U curve (Figure 4.19).

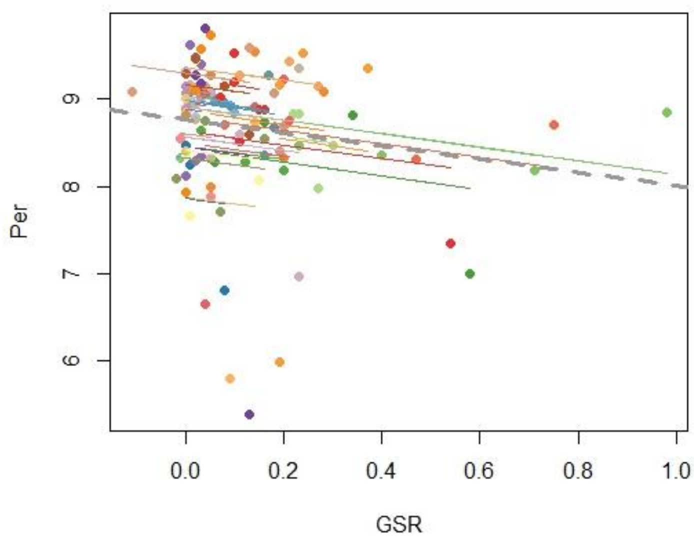

Figure 4.18. **Repeated-measure correlation between GSR and performance.**

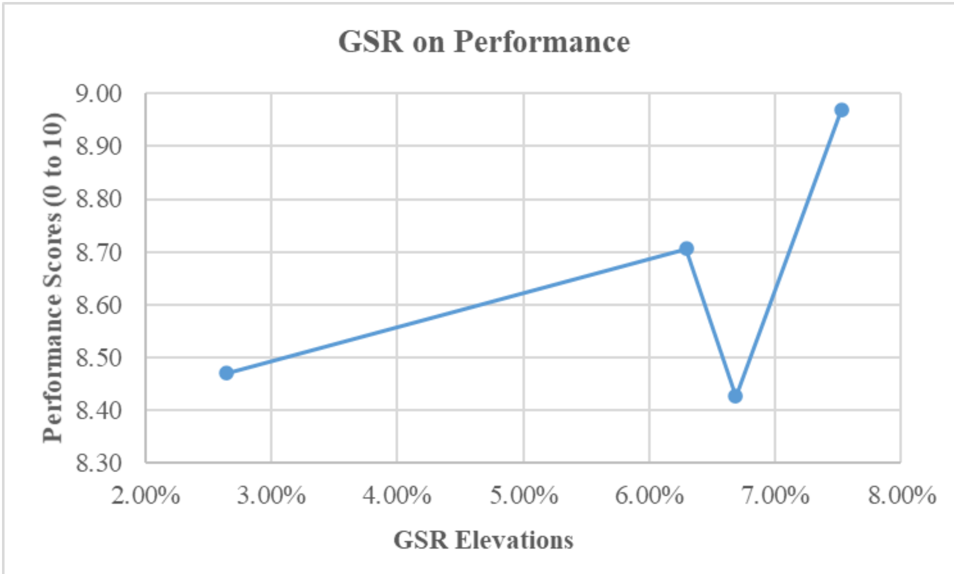

Figure 4.19. **Nonlinear relationship between GSR and performance.**

Relationships among MVs. The PLU and PLR showed a significant correlation, where $r = .78$, $df = 137$, $p < .01$. The PLU and annoyance also showed a significant correlation, $r = .43$, $df = 137$, $p < .01$. In addition, PLR and annoyance was significant correlated, $r = .66$, $df = 137$, $p < .01$. HR was correlated with HRV, $r = .23$, $df = 129$, $p = .01$, and HR was also correlated with GSR, $r = .24$, $df = 95$, $p = .02$. HRV and GSR was not correlated with each other. Also, physiological variables were not correlated with other variables.

Secondary analysis. In this section, several uninteresting variables were analyzed to see if they influenced the interesting variables. The reason was that some could act as extraneous variables. They included group, gender, age, native language, culture, experience, habits, and attitudes toward simulations. The results were summarized in Table 4.16.

As the participants were systematically assigned into different groups to counterbalance the auditory cues, the difference among groups should be tested. Demographics are not included in the current model, so it is necessary to identify if participants' inherent characteristics affected MVs and DVs. These were (a) gender, (b) age, (c) language, (d) culture, (e) experience, and (f) habits. The purpose of the test for inherent characteristics was to eliminate effects. Therefore, it was tested regardless of the experimentwise significance. The omnibus test with all DVs included is redundant. This was conducted to identify potential extraneous variables. If an uninterested variable affected one of the variables in the current study, the extraneous variable was included in the primary analysis for that affected variable.

Group. Outliers were excluded, and normality of continuous variables was examined, so the effects of uninterested factors on MVs and DVs were ready to be tested. As the group has four levels, a one-way between-subjects ANOVA was conducted.

Performance. There were four levels in performance, which included the performance scores of intensity levels 1 to 4. Four between-subjects ANOVAs were run. The homogeneity of variances was satisfactory for levels 1 to 3, so a

Welch ANOVA was run for level 4 instead, and the results revealed no significant effect of group on performance.

Perceptions and attitudes. Perceptions included PLU and PLR. There were five levels in PLU and five levels in PLR, which included a baseline level and intensity levels 1 to 4. A total of 10 ANOVAs were conducted to analyze the effect of the group on PLU and PLR. The homogeneity of variances was not violated for each of the levels; there was no significant effect detected. As for attitudes, there were five levels, and five ANOVAs were run as well. The level 2 of attitudes violated homogeneity of variances, so a Welch ANOVA was conducted. It did not show a significant effect of group on attitudes.

Physiological measures. Physiological signals included HR, HRV, and GSR, and each variable contained five levels. The homogeneity of variance assumption was satisfactory for each level of HR. Results showed that grouping affected intensity levels 3 and 4, $F(3, 40) = 4.27$, $p = .01$ for level 3, and $F(3, 41) = 6.23$, $p < .01$. Sidak post-hoc analyses were conducted for levels 3 and 4 as there were four levels in the group. When analyzing level 3, group 2 (M = - 1.25%) had significantly lower HR elevation than group 3 (M = 6.27%) as group 2 had a decrease in HR when hearing level-3 auditory cue. It also showed group 4 (M = 10.60%) had significantly larger increases in HRs than group 1 (M = - .17%) and 3 (M = 0.00%). For HRV, assumptions were satisfactory, and there was no significant difference between groups. As for GSR, homogeneity of variance assumption for intensity levels baseline, 1, and 3 were violated, so Welch's tests were conducted for these three levels. Two between-subjects ANOVAs were run for levels 2 and 4. No significant results were recognized for each level.

Gender. Afterward, the effect of gender on variables were analyzed. Gender included two levels, which were male and female. Therefore, independent-samples *t*-tests were conducted.

Performance. There were four levels in performance, which included the performance scores of intensity levels 1 to 4. Equal variances were satisfactory for all performance levels. The results showed that male subjects ($M = 9.11$) performed significantly better than females ($M = 8.81$) after hearing level-3 auditory cues, $t(38) = 2.25$, $p = .03$.

Perceptions and attitudes. Perceptions included PLU and PLR. There were five levels in PLU and five levels in PLR, which included a baseline level and intensity levels 1 to 4. Attitudes included five levels as well. Each level in perceptions and attitude had equal variances, and no significant difference was found between genders

Physiological measures. Physiological signals included HR, HRV, and GSR, and each variable contained five levels. The baseline level and intensity levels 1, 2, and 3 of GSR violated the homogeneity of variances, so the Welch-Satterthwaite method was used, whereas the assumption was satisfactory for others. There was no significant effect of gender on HR and HRV levels. For GSR, when hearing the baseline auditory cue, females (*M* = 6.38%) had higher increases in GSR than males (2.13%), $t(18.90) = 2.65$, $p = .02$. Also, females' GSR (*M* = 17.5%) increased higher when hearing intensity level 1 than males (*M* = 5.50%), $t(18.99) = -2.37$, $p = .03$. Other GSR levels did not show a significant difference.

Age. Age was a continuous variable, so Pearson's *r* was run to identify a potential correlation between age and other interesting variables. The scatter plots between the standardized residual and the standardized predicted value showed no obvious pattern for each intensity level of variables. Hence, the homoscedasticity of variance assumption was satisfactory. Results showed no significant correlation between age and performance scores, perceptions, and attitudes. As for physiological data, HR had no significant correlation with age. Nevertheless, it showed a significant reverse correlation between age and the intensity level 2 of HRV, $r = -.32$, $n = 44$, $p = .04$. Intensity levels 2 and 3 of GSR also had negative correlation with age, $r = -.35$, $n = 32$, $p < .05$ for level 2, and $r = -.35$, $n = 32$, $p < .05$ for level 3.

Language. Language included two levels, which were native English speaker and non-native speaker. As there were two between-subjects levels, independent-samples *t*-tests were run. If equal variances were violated, it utilized the Welch-Satterthwaite method.

Performance. There were four levels in performance, which included the performance scores of intensity levels 1 to 4. Intensity level 1 violated the equal variance assumption, and the Welch-Satterthwaite method was employed. The results showed that there was no significant effect of language on performance.

Perceptions and attitudes. Perceptions included PLU and PLR, each of which had five levels, and they were a baseline level and intensity levels 1 to 4. When hearing level-3 auditory cues, there were unequal variances in two groups for PLU and PLR. They were analyzed using the Welch-Satterthwaite method. There were no significant relationship between language and perceptions or attitudes was identified. Attitudes included five levels as well. Each level had equal variances, and no significant difference was found.

Physiological measures. Physiological signals included HR, HRV, and GSR, and each variable contained five levels. The intensity level 3 of GSR violated the homogeneity of variances, so the Welch-Satterthwaite method was used. The results showed that native speakers (M = 11.16%) experienced higher increases in GSR in the condition of level-3 cue than non-native speakers (M = 2.14%), $t(26.68) = 3.58$, $p < .01$.

Culture. Culture is also a dichotomous factor, which included Western and non-Western. With two groups, independent-samples *t*-tests were run. Equal variances were checked as well, and the Welch-Satterthwaite method was used if it was violated.

Performance. Performance has four levels, which included the performance scores of intensity levels 1 to 4. Intensity level 1 violated the equal variance, and the Welch-Satterthwaite method was employed. The results showed that there was no significant effect of culture on performance.

Perceptions and attitudes. Perceptions included PLU and PLR. Each of PLU, PLR, and attitudes had five levels, which were a baseline level and intensity levels 1 to 4. When hearing level-3 auditory cues, there were unequal variances in two groups for PLU, PLR, and attitudes, and level 4 of PLU violated it. These violated levels were analyzed using the Welch-Satterthwaite method. There was no significant relationship between culture and perceptions or attitudes was identified.

Physiological measures. Physiological signals included three variables, which were HR, HRV, and GSR, and each variable contained five levels. The intensity level 3 of GSR violated the homogeneity of variances, so the Welch-Satterthwaite method was used. The results showed that Westerners (M = 10.74%) had larger GSR elevation than others regarding level 3 (M = .80%), $t(28.83) = 4.6$, $p < .01$.

Experience. Experience referred to the frequency of commercial flights annually. The factor was a continuous variable. Pearson's r was run to detect potential effects. Assumptions were checked, and there was no assumption violated. The flight frequency was significantly correlated with intensity level 4 of PLU, $r = .36$, $n = 44$, $p = .15$, other correlations were not significant.

Habits. Habits were usual behaviors for commercial flights regarding how likely they followed safety behavior instructions, which included fastening seatbelts, activating airplane mode, and attending safety briefings. The habits variables were all continuous. The Pearson's *r* correlations were used.

Performance. Performance has four levels, which included the performance scores of intensity levels 1 to 4. Intensity level 4 of performance scores was significantly correlated with habitual compliance with fasten-seatbelt instruction, $r = .40$, $n = 42$, $p < .01$.

Perceptions and attitudes. Perceptions included PLU and PLR. Each of PLU, PLR, and attitudes included five levels, which were a baseline level and intensity levels 1 to 4. There was significant correlation of habitual compliance with airplane-mode instruction with (a) intensity level 3 of PLU, $r = -.34$, $n = 46$, $p = .02$, and (b) intensity level 4 of PLR, $r = -.37$, $n = 46$, $p = .01$.

Physiological measures. Physiological signals included three variables. They were HR, HRV, and GSR, and each variable contained five levels. No significant linear relationship was detected.

Attitudes toward simulation. Attitudes included boredom, madness, and interests. Madness referred to how participants were angry with the simulation. These factors were all continuous variables. The intensity level 4 of performance was reversely correlated with madness, $r = -.35$, $n = 42$, $p = .02$. Boredom had significant correlations with PLU for level 2, $r = .30$, $n = 45$, $p < .05$, level 3, $r = .49$, $n = 46$, $p < .01$, and level 4, $r = .29$, $n = 46$, $p < .05$. Boredom was also correlated with intensity levels 3 and 4 of PLR, $r = .42$, $n = 46$, $p < .01$ for level 3, and r =.60, n = 46, p < .01 for level 4, and level 3 of attitudes, r = .33, n = 46, p = .03.

Table 4.16

Influences of Uninterested Factors on Studied Variables

Factors	Variables	Intensity Levels
Group	HR	3, 4
Gender	Performance Scores	3
	GSR	Baseline, 1
Age	HRV	2
	GSR	2
Language	GSR	3
Culture	GSR	3
Experience	PLU	4
FS Instruction	Performance Scores	4
AM Instruction	PLU	3
	PLR	4
Boredom	PLU	2, 3, 4
	PLR	3, 4
	Annoyance	3
Anger	Performance	4

Note. FS = Fasten-seatbelt. AM = Airplane mode.

Results of Hypotheses Testing

Hypothesis 1 in the null form: that auditory cues will not have a significant effect on passengers' performance. The hypothesis was rejected because the auditory cues were shown to influence performance.

Hypothesis 2 in the null form: auditory cues will have no significant effect on passengers' perceptions and attitudes. The hypothesis was rejected because perceptions and attitudes were affected by intensity levels of cues.

Hypothesis 3 in the null form: if the auditory cues have a zero-order relationship with passengers' performance, the relationship will not be mediated by at least one perception or attitude variable. The hypothesis was retained as MVs did not show significant influences on the DV; nevertheless, it implied a relationship between MVs (except for HRV) and DV that is congruent with the Yerkes-Dodson law.

Summary

The outliers were detected and removed, and the scales were also tested and showed relatively reliable. As for the inferential statistics, the IV had significant effects on the DV and MVs. The effects of the IV on the DV showed an inverted-U trend, where the intensity level 3 was significantly higher than levels 1 and 4, and level 2 was higher than level 1. PLU and PLR, which are included in the perception, showed significant differences among auditory cues. Levels 3 and 4 were perceived higher than levels 1 and 2, and all intensity levels led to higher PLU and PLR than the baseline level, where the result was aligned with how they were designed. Intensity levels 3 and 4 were reported to be more annoying than levels 1 and 2 also, and the baseline level had lower values than intensity levels 1 to 4. HR and GSR showed significant differences; however, the only difference was that the baseline was significantly smaller from other levels. As participants were not instructed to move bodies when hearing baseline auditory cues, the difference did not show the effect of auditory cues merely. Auditory cues did not have a significant effect on HRV. Therefore, the intensity level of auditory cues did not show significant influences on physiological data. Moreover, each MV had no linear correlation with the DV. Nevertheless, the MVs except for HRV each formed a potential inverted-U relationship with the performance, but it is not practical to test if they were significant, especially for the within-measures analysis.

Chapter 5

Conclusions, Implications, and Recommendations

Summary of Study

Purpose. The purpose of the study was to examine the effects of intensity levels of auditory cues that were followed by in-flight announcements on passengers' performance. There were also two secondary purposes, which were to (a) identify the influence of auditory cues intensity levels on perceptions, physiological signals, and attitudes and (b) examine the mediating effect of perceptions, physiological signals, and attitudes on performance.

Variables. The dependent variable (DV) was the participants' performance, indicated by performance scores. The independent variable (IV) was the intensity level of auditory cues, including five levels (i.e., one baseline and four intensive levels). There were three groups of mediating variables (MVs). They were arousal levels (i.e., HR, HRV, and GSR), perception (i.e., PLU and PLR), and attitude (i.e., annoyance).

Research design. The study was conducted on the basis of the risk homeostasis theory (RHT) and arousal levels. It employed a within-subjects design. There was one IV with five levels. Each participant experienced each cue. They were also instructed to take actions according to the announcements following the intensive auditory cues, and the announcements after the baseline cues did not require participants' responses. During the simulation, participants' physiological responses and behaviors were recorded digitally. When the whole experiment was over, each cue was played to participants in different orders to counterbalance. The participants were then asked to report their perceived level of urgency (PLU), perceived level of risk (PLR), and annoyance for each cue. They also rated their feelings about the experiment to see if they were bored or mad during the simulation.

Population and sample. The target population consisted of 1,011,100,232 flight passengers in the United States in 2018, as reported by the Bureau of Transportation Statistics (2019), whereas the accessible population was the students at Florida Tech. A convenience sampling strategy was utilized, consisting of 46 participants. The sample included 25 males and 20 females ranging in age from 18 to 36.

Treatment. The IV was the intensity level of the auditory cues. There were five levels, which included one baseline and four intensive levels. The baseline tone was one 600-ms pulse of a sine wave with a frequency of 0.5 kHz. The intensity level 1 had an increasing frequency, which was 1.0 kHz, and other parameters were the same as the baseline. The level 2 intensity used a triangle wave with the same features as level 1. The difference between level 3 and level 2 was the number of pulses in level 3 was two. The level 4 cue had longer pulses than level 3, where the length was 1,200 ms.

Instrument. The study setting was built with four rows of aircraft seats with three seats in each row. The seats were reclinable and equipped with tray tables and seatbelts. A speaker was placed under the middle seat in the third row, which is where participants sat when the announcements played. The announcements were pre-recorded, and four announcements followed intensive cues and instructed participants to fasten their seatbelts, close tray tables, lower tray tables, and adjust their seatback. As for the other four announcements, they were played after baseline auditory cues and did not include action-required instructions. Another speaker played a jet engine sound to simulate a real cabin. Participants were asked to wear Equivital sensors to measure their heart rates (HRs), heart rate variability (HRV), and galvanic skin responses (GSRs). A camera was placed over participants to record their behaviors, which were used to measure their behaviors and response times to instructions. Participants were provided with scales to measure their feelings about the cues and the experiment. Each cue scale included the items for PLU, PLR, and annoyance. Two reverse-scored bipolar adjective items were employed to measure one factor, which method was perceived easier to report than Likert scales (Ary, Jacobs, Sorensen, & Asghar, 2010). The experiment scale included three 5-point Likert scales for participants to report interests, boredom, and anger. It was employed to see if participants had negative attitudes to the simulation, which could become an extraneous variable and affect their performance.

Statistical strategy. The study tested the (a) the effect of the IV on the DV, (b) the effect of IV on MVs, and (c) the mediating effect of MVs on the DV. The IV was categorical with five within-subjects levels. Whereas, the DV and MVs were all within-subjects continuous variables. The effect of the IV on the DV and MVs were analyzed by running repeated-measures analysis of variance (ANOVA), and a repeated-measure correlation (RMCORR) was conducted to examine the relationships between MVs and the DV. As no significant linear relationship between MVs and the DV was identified, mediation analyses were not run as proposed.

Summary of Findings

Data collection methods. Data were collected in simulated experiments by the experimenter. Participants were asked to follow the instructions stated in in-flight announcements, and their responses were recorded for the experimenter to measure participants' performance using a stopwatch based on the response time after simulations. Also, they reported their PLU, PLR, and annoyance regarding the auditory cues when hearing the cues in a counterbalanced order after the simulation. The ratings were transformed into values manually and reordered by Microsoft Excel formulas. Participants wore several sensors to receive their HR, HRV, and GSR throughout the simulations, which were treated by the experimenter manually with the help of self-coded programs in MATLAB.

Inferential analysis results. The results showed that the intensity levels of auditory cues had significant effects on performance, PLU, PLR, and annoyance but not on HR elevation, HRV, and GSR elevation. Regarding research question 1, there was a significant difference in performance among intensity levels, $F(2.38, 73.81) = 5.33$, $p < .01$, $\eta_p^2 = .15$, and participants performed better when hearing medium levels of auditory cues than extreme intensity levels, including levels 1 and 4 (2 > 1, 3 > 4, 3 > 1). For research question 2, participants reported higher PLU, $F(3, 132) = 25.64$, $p < .01$, $\eta_p^2 = .39$, PLR, $F(2.37, 106.83) = 25.96$, $p < .01$, $\eta_p^2 = .37$, and annoyance, $F(3, 135) = 16.53$, $p < .01$, $\eta_p^2 = .27$, when listening to auditory cues with higher intensity levels than low levels (3, 4 > 1, 2). Hypothesis 3, relative to research question 3, was rejected as MVs did not have mediating effects. No linear correlation between (a) perceptions and performance, (b) annoyance and performance, or (c) arousal levels and performance was identified. Therefore, physiological signals were removed from the model because it was not significantly affected by the IV nor influenced the DV. Nevertheless, the trends between perceptions and performance followed an inverted-U curve and were considered compliant with the Yerkes-Dodson law. However, the non-linear significance could not be determined because the formula for the Yerkes-Dodson law was unknown. The results for the hypotheses are summarized in Table 5.1.

Table 5.1

Hypothesis Results Summary

Hypotheses	Results
The auditory cues have significant effects on performance.	Auditory cues had significant effects on performance.
Intensity levels have significant effects on perceptions and arousal levels.	There was a significant effect of intensity level on perceptions but no effect on arousal levels.
If the auditory cues have a zero-order relationship with passengers' performance, the relationship will be mediated by at least one MVs.	The relationship was not mediated by any MV.

Conclusions and Inferences

Primary Analysis. This section discusses three research questions, whether these questions were rejected, and whether these questions were consistent with the literature. The research questions were (a) the effect of IV on DV, (b) the effect of IV on MVs, and (c) the mediating effect of MVs between the IV and DV. These are introduced and discussed in three sections.

Research question 1. Research question 1 was: what is the effect of auditory cues (set A) on passengers' performance (set E)? Set A, which included the intensity level of played auditory cues, had significant effects on participants' performance. When hearing the level-3 auditory cue, participants performed better than levels 1 and 4 by .64 and .32 units of performance scores. As the remaining performance scores were merely calculated from the response times, it can be interpreted that the increase from intensity level 1 to level 3 made flight passengers initiate responses to safety instructions .64 seconds faster on average, and it is able to reduce the response time by .32 seconds on average if the intensity level dropped from level 4 to 3. Nevertheless, the level-3 auditory cue included 2 pulses, so it was .60 seconds longer than level 1, which offset the decrease in the response time. Compared to level 1, participants had better performance when a level-2 auditory cue was played with .56 shorter in the response time. Also, the level-2 auditory cue had a raised frequency with the same length compared to level 1. Considering there was no significant difference in performance scores between levels 2 and 3, level 2 could be a better option than 3. In summary, the middle levels stimulated participants to perform better. It was compliant with the arousal theory. When the stimulation of auditory cues increased from a low level, participants' performance was improved. After a certain point, which is called the optimal level of arousal (OLA), the performance was impaired as the level of stimulation increases. In the current study, the OLA should be around level 3, which contained two pulses of a 600-ms sine wave with a frequency of 0.5 kHz. It was congruent with Duffy's (1957) findings that the optimal activation degree is intermediate, which indicated that the arousal theory applies to this scenario. In other words, the arousal theory should be an appropriate explanation of the outcome that participants had worse performance when the level was low or high than medium. Nevertheless, it needs to be demonstrated that it did not imply that the designed acoustic parameters (e.g., waveform, the number of pulses, pulse lengths) could have effects on performance. The variable was the intensity level of auditory cues, which accumulated at each higher level by increasing the acoustic parameter intensity instead of parameters. Therefore, although there was a small possibility that the changes in the DV and MVs could be due to the change of an acoustic parameter, it was more likely the accumulative intensity of the cue that affected the parameter.

Research question 2. Research question 2 was: what is the effect of auditory cue intensity levels (set A) on physiological responses, perceptions, and attitudes (sets B, C, and D)? The set A had a significant influence on set B, where intensive levels of HR and GSR were higher than the baseline, and there was no significant difference identified in HRV. It indicated that when hearing intensive auditory cues, participants' HR increases were higher than the baseline by 6.2%, 6.0%, 4.2%, and 5.0%. Similarly, participants' GSR elevations were higher when hearing the intensity levels of auditory cues than the baseline by 7.25%, 6.16%, 4.94%, and 4.16%. Nevertheless, it needs to be stated that participants were instructed to make body movements when hearing intensive auditory cues because they were followed by action-required announcements. By moving bodies, people are expected to have higher HRs and GSRs, so the difference in physiological signals was not entirely contributed by the intensity level of auditory cues. The intensive levels did not show a significant difference from each other.

Set A showed a significant effect on sets C and D. For PLU, PLR, and attitudes, when hearing the baseline level, participants scored significantly higher than the intensity levels, and within intensity levels, levels 3 and 4 were higher than the levels 1 and 2. They were ordinal variables without units, so it was not practical to interpret results.

Research question 3. Research question 3 was: if the auditory cues (set A) have a zero-order relationship with flight passengers' performance (set E), to what extent is the relationship mediated by arousal levels, perceptions, and attitude (sets B, C, and D) variables? It was detected that the IV affected the MVs. Regarding PLU, PLR, and annoyance, intensive levels 3 and 4 were significantly higher than levels 1 and 2, and they were all significantly higher than the baseline. However, the relationship between MVs and the DV could not be determined to be significant. MVs did not have significant linear repeated-measures correlations with the DV. Nevertheless, according to the Yerkes-Dodson law, it should not linearly be correlated with each other. Instead, it should form an inverted-U shape. Perceptions and annoyance were aligned with the designed auditory intensity levels, where it was reported that levels 3 and 4 had significantly higher PLU, PLR, and annoyance than levels 1 and 2, and they all showed linear effects caused by the intensity level. The issue was that there was no formula to express the Yerkes-Dodson law, and the direct nonlinear relationship between them and performance could not be determined. In other words, it could not be determined whether the relationship between the stimulation or the arousal level and performance was quadratic or other inverted-U curves. However, with a linear effect of auditory cues on perceptions and attitudes and a quadratic effect of auditory cues on performance.

It was possible that there were also quadratic effects of perceptions and attitudes on performance. Also, the trends indicated an inverted-U shape though the significance was unknown. Physiological responses did not have linear relationships with performance scores. HR and GSR showed potential U-shaped effects on performance scores, but the significance could not be determined in a proper way.

Based on the results summarized above, the assumed model has been edited. Physiological data, as previously mentioned, were removed from the model. One reason was that these data were designed to be exploratory variables due to the unreliable readings caused by inadequate sensitivity to participants' physical responses. Another reason was that the relationships between MVs and the DV could be implied but not examined. Also, the solid arrows that indicated effects were modified. The updated model is displayed in Figure 5.1. It demonstrated the effects of the IV on MVs and the DV and potential curvilinear correlations between the MVs and the DV. From this section, the set numbering was also changed.

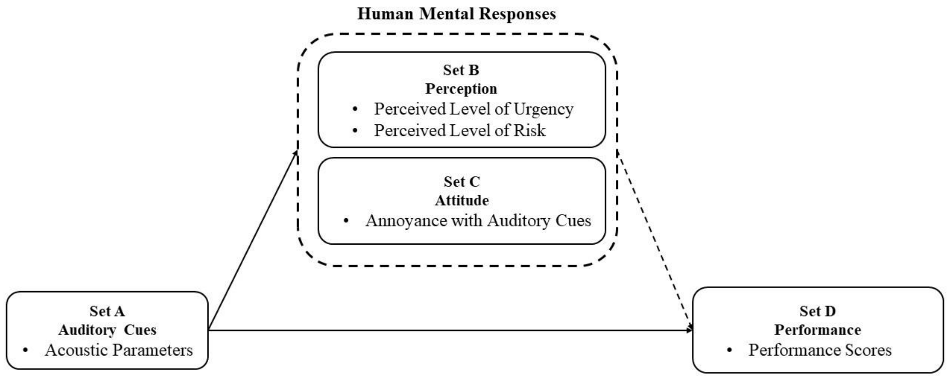

Figure 5.1. **Updated model.**

Secondary analysis. Potential extraneous variables were stated and examined if they affected the interesting variables. The result demonstrated that they all affected some levels of the DV or MVs. They included group, gender, age, language, culture, flight experience as commercial flight passengers, commercial flight habits, and attitudes toward the cabin simulation in current research.

Grouping referred to the order of auditory cue intensity levels that were counterbalanced. There were four groups, and the orders were different from each other, where the level of the first intensive auditory cue was the same as the group number followed by others sequentially. For example, the order in group 1 was 1, 2, 3, and 4, but the order was 2, 3, 4, and 1 in group 2. It showed that when hearing the level-3 cue, group 3 (level-3 cue was the first cue in this group) had higher HR elevation than group 2 (level-3 cue was the second cue in this group), and there was no reasonable explanation for this. For level 4, group 4 (level-4 cue was the first cue in this group) had higher HR elevation than group 1 (level-4 cue was the fourth cue in this group). A plausible explanation was that group 4 participants heard the extremely intensive cue during the first stimulation, whereas the participants had experienced other gradually intensive auditory cues in group 1.

Gender affected performance and physiological measures. It showed when hearing level-3 auditory cue, males performed better than female participants. It could be due to male people were more sensitive to a sound with 1 kHz, triangle wave, and two 500-ms pulses, but there was a 5% chance that the difference was due to the sample selection. Moreover, female people had higher skin conductance after hearing the baseline level and level 1 than male participants. It could be due to (a) different sensitivity or tolerance of gender, (b) different sweat amount, or (c) chance.

Age had influences on HRV and GSR. It showed that when hearing level 2 of auditory cue, there was a negative correlation between age and HRV. It made sense that younger people tend to have more extensive HR ranges, and the HRV was high under the same stimulation. The relationship between age and HRV of other levels were not significant, probably because a smaller effect size, which made the significance hard to be detected with small sample size. It also showed a negative correlation between age and GSR when hearing level-2 cues. A possible reason was that younger people were more sensitive to stimulations.

Language indicated whether or not a participant was a native speaker. Native English speakers had more GSR elevations when hearing the cue of level 3. It was probably because native speakers could fully understand the potential risks that the announcements addressed and had higher arousal levels, or it could be that native speakers were more likely to be Caucasians and African Americans who sweat differently than other races. Also, the difference that was identified was possibly due to chance.

The culture was categorized into Westerners and others, which indicated that Westerners had higher GSR elevations when hearing level-3 auditory cue than others. The reasons could be the same. Westerners were likely to sweat differently, or the difference was identified by chance.

Commercial flight experience meant the frequency of flights as passengers. It was positively correlated with PLU after hearing level-4 auditory cue. It could be due to that experienced flight passengers were more likely to perceive a cue with the acoustic parameters of level 4 was unusual, or it was due to chance.

Flight habits referred to the compliance with fasten-seatbelt instructions, compliance with airplane-mode announcements, and frequency of listening to pre-flight safety-briefings. The compliance with fasten-seatbelts instruction was positively correlated with performance scores when hearing level-4 auditory cue, and this pattern followed their routine behaviors. The compliance with airplane-mode instructions was related to PLU after the level-3 cue and PLR after level-4 cues. The difference could be because the compliance with airplane-mode instructions reflected participants' concerns regarding flight safety, or the difference occurred by chance.

Attitudes toward simulations measured how bored, anger, and interested the participants were concerning the whole simulation study. Boredom was positively correlated with (a) PLU after hearing the auditory cues of levels 2, 3, and 4, (b) PLR after hearing the cues of levels 3 and 4, and (c) annoyance with cues after hearing the tones of levels 3 and 4. That was a popular trend across the levels of all mental measure and was probably because bored people were more sensitive to stimulations. Moreover, madness measuring anger was reversely correlated with performance scores when hearing level-4 auditory cues. A plausible explanation was that participants who were angry with the simulation did not respond appropriately to in-flight instructions when they listened to a largely intensive auditory cue.

Implications

This section introduces the implications of the results. It includes implications relative to theories, prior research, and aviation practice. The theories are RHT, arousal theory, and aviation auditory warnings design. Prior research was discussed from three similar aspects. They are RHT, arousal theory, and auditory parameter effects. Implications for aviation practice demonstrates (a) how aviation researchers reference this current study regarding future research, (b) how aviation manufacturers and airlines use the results to enhance aviation safety, and (c) how aviation authorities apply the findings to practice.

Implications relative to theory. There were two theories and several acoustic parameter design references included in the theory section. The theories were the RHT and the arousal theory. Acoustic parameter design included the design in cockpits and the effects of acoustic parameters on PLUs.

Risk homeostasis theory. The RHT states that when people have higher PLRs, they tend to perform safe behaviors to reduce the risks being experienced to achieve the equilibrium (Wilde, 2014). Nevertheless, it usually reaches homeostasis after a long time. In laboratory settings, the risk compensation theory (RCT) was much easier to be studied, which maintained that people perform safely in risky conditions and make riskier behaviors when PLRs were low (Hoyes, Neville, & Taylor, 1996; Wilde, 1982a; Wilde, Claxton-Oldfield, & Platenius, 1985). In other words, the RCT does not mention PLRs will finally decrease or increase back to target levels of risk (TLRs). According to the RCT, passengers were supposed to always have safe behaviors when their PLR increased (Trimpop, 1996). It appeared that the RCT was not valid in the current study as the performance of participants was impaired when the PLRs were extremely high. One reason was that in prior research, which validated the RCT in laboratory settings (Hoyes et al., 1996; Wilde et al., 1985), there were two levels. It was inadequate to detect an inverted-U-shaped relationship. A non-linear relationship could be discovered in the current study as there were more than two levels of the IV. With two levels, it had to show a linear relationship. Instead, with at least three levels, it becomes possible to identify non-linear relationships. Another reason could be that the factor was the safeness of participants' behaviors in the discussed research, but the DV was the performance in the current study. It is possible that the safeness could not be inferred by their performance. It needs to be advised that the performance was also affected by the intensity levels, which is discussed below with the arousal theory.

Arousal theory. Even though the RHT and RCT could not be researched in a laboratory setting in a scientific manner, the research demonstrated that the relationship between participants' performance and the intensity level of auditory cues was congruent with the arousal theory. According to Duffy (1957), the relationship between the arousal levels and the performance should be compliant with an inverted-U shaped relationship. Specifically, when the arousal level was at a low level, the performance was poor. When the arousal level increases until the optimal level, the performance was better as well. After this point, the performance is impaired as the arousal level further increases. The trends of the effects of auditory cues and perceptions on performance formed an inverted-U shape as what was stated in the Yerkes-Dodson law. Level 3 resulted in better performance than levels 1 and 4, and the level 2 performance score was also higher than level 1.

Audio warning design. Moreover, the design of auditory cues based on the literature was successful. The effective frequency of warnings that Patterson (1982) identified for the cockpit, which was introduced above in chapter 2, was also applicable to the cabin. Participants could recognize the sounds apart from the jet engine background. Also, the perceptions of the auditory cues discovered that the designed intensity levels were aligned with their actual stimulation. It indicated that the frequency and repetition were able to increase the PLU of auditory cues. Still, the increase in length from 0.5 s to 1.0 s and the change from sine waves to triangle waves could not significantly increase the PLU of auditory cues. Nevertheless, they also increased participants' PLUs to a small degree. It showed that the effects of the increase in frequency from 0.5 kHz to 1.0 kHz and the number of pulses increase from one to two on hearers' PLUs, which were recognized as effective acoustic parameters from the literature, were applicable to the current study. Nevertheless, the cues with distinct acoustic parameters were designed to alter the accumulative intensity level rather than examine the effect of each acoustic parameter. The change in a single acoustic parameter with other acoustic parameters changed may not have similar effects. In other words, there can be interactions among acoustic parameters. Consequently, the effects of acoustic parameters on the DV and MVs detected in the current study should be generalized carefully.

In summary, the RCT partly applied to passengers' performance in the cabin. It should be incorporated with the arousal theory to determine the holistic effect of the intensive auditory cues on performance if it examines the effects on performance instead of desire for safeness. The effects of sound frequency and repetition were effective, but other parameters were not utilized in this study, and the effect of each acoustic parameter could be interacted by other parameters.

Implications relative to prior research.

***Risk homeostasis theory*.** Wilde et al. (1985) demonstrated that the RCT was valid in a laboratory setting. Participants made compensation regarding the PLRs at different risk levels. Hoyes et al. (1996) also recognized the participants' risk compensation in the study. However, the current study was not compliant with the RCT. Based on RCT, when participants had a higher PLU, they should perform more safely, however high the PLU was. On the contrary, the performance of participants was better if the PLU increased to a specific extent, but the performance was impaired by overly high PLU. There are two possible reasons for the difference between the previous studies. One reason, as mentioned above, was that participants might have made safer behaviors when the intensity level increased in the current study. However, their performance was also affected by their arousal levels, which impaired performance after the OLA. Namely, the participants wanted to make safer behaviors at high PLU levels, but the performance decreased because it was also affected by the arousal level. Another reason was that Wilde et al. and Hoyes et al. only designed two levels, so they could not identify the changes for riskier conditions. In the current study, if there were only level 1 and level 3, the results should demonstrate similar effects as it could not identify the decrease at level 4.

***Arousal theory*.** With respect to the arousal theory, Ünal, de Waard, Epstude, and Steg (2013) stated that the presence of music and its loudness could increase participants' arousal levels and performance, but it did not show a decrease in the performance. Smolders and de Kort's (2014) employed lighting conditions as stimulation to increase arousal levels, and participants' performance decreased when the lighting levels were higher. In these two studies introduced above, the performance simply increased or decreased when the arousal level rose. However, based on the arousal theory, there should be a non-linear relationship. As the arousal level was increasing, the performance should increase to a certain point and decrease when the arousal level was exceeding. A reason why the previous findings did not indicate an inverted-U shaped curve as what was discovered in the current study could be that those researchers did not select the stimuli within a proper range. All the levels could be below the OLA in the first study and above the OLA when designing lighting conditions.

***Audio warning design*.** Wang, Guo, Ma, & Li (2016) discovered that the volume, frequency, and the IOI of a sound could influence the PLU and reaction. The parameters were different from the auditory cues design in the current study. However, the parameters utilized in the current study were also effective factors, and they also have effects on PLU and performance. Bodendörfer, Kortekaas, Weingarten, and Schlittmeier (2015) did not use the matrix containing every combination of the manipulated parameters. They changed one parameter each time to reduce the number of sounds largely. This method was employed in the current study, and it appeared to be valid to increase PLR and auditory intensity levels. However, by using this method, it was impossible to distinguish the difference between the effect of the accumulated intensity level and the effect of parameter changes.

Implications for aviation practice. The study provides aviation researchers with the support that the arousal theory is applicable to flight passengers' behaviors. In this case, it demonstrated that stimulations, such as intensive auditory cues, can affect passengers' performance compliant with a reversed-U shaped curve. However, the increase in the arousal levels could not be directly acquired in the current study. Therefore, some other stimulations, no matter whether manipulated or unmanipulated, can be initiated to ensure passengers' safety. Moreover, RCT can be partly demonstrated in aviation research. When passengers perceived higher levels of risk, they may perform safer behaviors, which indicated that the PLR should influence their desires to perform safely. However, the outcomes were affected by the arousal levels. When the intensity level increased, even if the passengers wanted to make safer behaviors, their performance was impaired. Another implication was that the auditory parameters that worked for the cabin and other conditions were also effective for passengers in a cabin who were not performing any tasks related to flying. However, the parameters used in the current study may not have the same effects on the passengers' performance in a real situation because passengers may have high arousal levels, so the intensity level should be reduced to some degree.

For aircraft manufacturers and carriers, intensive auditory cues showed significant effects on passengers' performance regarding response times to initiate responses to in-flight announcements. In this case, aircraft manufacturers should consider, or aircraft carriers can require manufacturers to create several auditory cues with different intensity levels for the particular aircraft type. Nevertheless, the intensity level must not be exceedingly high. One reason was that it could impair passengers' performance as their high arousal levels may inhibit their responses; also, intensive auditory cues annoy passengers. The optimal level should be identified for the specific cabin with layout considerations. Based on the current study results, the sound, including two pulses of 600-ms triangle waves with a 1.0-kHz frequency, was an optimal intensive auditory cue in a simulated cabin. However, the level-3 cue was longer than level 2, and considering the cue-playing time, the time from the cue being played to participants made responses for level 2 was shorter than level 3. Therefore, the auditory cue with the acoustic parameters designed for level 2 was optimal. It is necessary to demonstrate that the optimal set of parameters can lead to a decent performance in a laboratory simulation. They may not be appropriate parameter ranges on a real aircraft.

Aviation authorities such as the DOT, FAA, and also NTSB can apply the findings of the current study to the establishment or modification of regulations.

The effects of warning systems on operators' performance were studied in proper passengers' safety is also significant, considering flight passengers usually outnumber pilots and the cabin crew. It may not be practical to train passengers as pilots and flight attendants. An alternative way is to require commercial aircraft to have appropriate intensity levels of auditory cues to help optimize passengers' performance and enhance passengers' safety. They should also consider passengers' annoyance and uncomfortableness when hearing harsh sounds.

Generalizability, Limitations, and Delimitations

Generalizability. Generalizability discussed in this section includes population generalizability and ecological generalizability. The target population was flight passengers in the United States, including domestic and international passengers. There were 10 (22%) non-native speakers recruited in the study, and 233,181,102 (23%) passengers in the United States were international travelers in 2018 (Bureau of Transportation Statistics, 2019). Nevertheless, the accessible population was the students and employees at FL Tech. The sample was comprised of the students at FL Tech who volunteered to participate. As a convenience sampling strategy was employed, it may limit the population generalizability to a small degree. The age ranged from 18 to 36, with a mean of 20.89 and a standard deviation (*SD*) of 3.62. Thirty-six participants (78%) were native speakers. In this sense, the results may not be able to reflect the effects regarding the passengers outside of the age range of the participants in the current study, which may slightly weaken the generalizability of the study especially considering that elder adults tend to have retrogressive hearing abilities and slower movements. The portion of native speakers and the percentage of domestic passengers matched. Therefore, the results can be generalized to the highly educated flight passengers whose ages are between 18 and 36.

Due to safety concerns, the experiment was conducted in a laboratory room instead of a flying aircraft. Efforts were made to ensure participants had similar experiences to that of a real aircraft. Several rows of reclinable aircraft seats with seatbelts and tray tables were used. The seat pitch was set by referencing the seat pitch of multiple US airlines, where the mean was 32 inches, which was also the distance between rows in the laboratory. Also, jet engine noise was played to simulate the experience in a cabin. The only difference was that participants could not feel the motion changes in the experiments. Therefore, the results should be generalizable to real conditions. A factor that impacts generalizability is that the propeller sounds were different and probably louder than jet engines.

Study limitations and delimitations.

Limitations 1: Sample size. A limitation was the sample size, which was 46. The IV showed significant effects on the MVs and the DV. However, the intensity level did not affect HR and GSR. Also, no linear correlation between MVs and the DV was recognized. If more participants were recruited, it could show significant results. Nevertheless, physiological signals were not sensitive. For example, HR is usually measured throughout the sections, which means participants should be receiving stimulation or engaged in the tasks for several minutes. The changes can be detected in this case because it should minimize the effects of the variability of the HR. On the contrary, two HRs were measure within less than 30 seconds in the current study, and participants' physiological sensitivity was high enough to show detectable differences. As for correlation, as stated above, the relationship should be curvilinear but not linear. Therefore, it was not supposed to detect significant correlations though the powers were not high either.

Limitations 2: Demographics. The sample was not randomly selected, so this may have restricted the population generalizability. The participants were students at FL Tech. The age and behaviors of participants could not be representative of US passengers. The results could not be generalizable to passengers that are older than the population tested. Some people might be less sensitive to intensive auditory cues. However, it should still have high generalizability considering other demographic variables.

Limitations 3: Authenticity of responses. The simulations were accomplished in a lab setting, which could not simulate a real-world dangerous situation. On a flying and shaking aircraft, passengers may have higher arousal levels, which impair their performance with optimal intensity levels identified in the study. Their behaviors and compliance might also deviate from the simulated results. Moreover, participants were volunteers, and it was possible that they participated in the study to receive extra credit or raffle prizes but did not care about the data they provided. Therefore, subjective responses were designed to have reverse-scored items, and the scale reliability was good. The pairs, where reverse-scored items rated with the same or close extreme values, were considered outliers.

Delimitations 1: Auditory cue design. The intensity levels of auditory cues were distinguished from each other by modifying acoustic parameters. The levels from baseline to intensity level 4 changed only one parameter at each time. It increased the designed intensity level. However, as only one parameter was changed for an increase of one level, the difference in the DV and MVs among intensity levels might have been because of a specific parameter instead of the intensity level. For example, the difference in performance between the intensity levels 1 and 2 could increase the frequency from 0.5 kHz to 1.0 kHz instead of the increase in intensity level.

Delimitations 2: Laboratory experiment. Another delimitation was the laboratory setting. The study was conducted in a lab, which did not provide shaking and gravity changes for the aircraft seats. Also, participants knew they were safe and did not care about potential dangers, which could alter their responses to instructions. Nevertheless, the room was set to be realistic to an aircraft cabin. Jet engine sounds were also played. The interference of the noise outside the room was measured too. The volume of the people who were speaking outside the room was at 40 dB in the morning and 45 dB in the afternoon. These dB levels were completely covered by the engine noise.

Delimitations 3: Data collection method. The simulations were recorded with a camera, and response times were acquired by observing participants' movements in video recordings. The data should be accurate and objective. The perceptions were measured with a bipolar adjective scale, so the responses may be different if they are measured with another scale. Psychological signals were collected by asking participants to wear an Equivital vest and GSR sensors, and the signals can be different when being collected with different devices.

Recommendations for Research and Practice

Aviation researchers may perform future research studies by testing more cues with more levels and more parameters. Also, it can be more convenient to analyze the relationship between the PLU or PLR and performance with a between-subjects design. The reason is that the effect can then be analyzed using the MRA, whereas RMCORR is not appropriate in this study. Moreover, the fluctuation in physiological signals caused by short stimulations is hard to be detected due to the insensitivity of human physiological responses.

For aircraft manufacturers and airlines, it is necessary to apply intensive auditory cues in the cabin to draw passengers' attention and stimulate passengers to follow instructions. It needs to be emphasized that an exceedingly intensive auditory cue may inhibit passengers' responses. A moderate intensity level is ideal.

The FAA may consider requiring aircraft manufacturers to test the optimal levels for particular cabins and provide different levels of auditory cues for the crew to play in different situations. For example, if the announcement is not important, a baseline or no auditory cue is played to reduce passengers' annoyance. When the announcement is significant and related to safety, an auditory cue with a proper intensity level should be played.

Recommendations for research relative to study limitations. One limitation stated was the sample size. There were 46 participants recruited because it was time-consuming and hard to include too many participants in an experiment compared to a survey. The consequence of a relatively small sample was that several analyses, including the effects of the IV on HR, HRV, and GSR and effects of MVs on the DV, did not show significant results or high powers. The power analyses were conducted using medium effect sizes, but the actual effects sizes were small. It is recommended to replicate the study with a larger sample size to determine if physiological data are affected by the auditory cues within a short duration and test if there is a significant correlation between MVs and the DV. Also, a between-subjects design may be deployed, which can be more appropriate to detect the effects of MVs on the DV by running Pearson's *r* correlation instead of RMCORR. Moreover, it is recommended that researchers should ask for the demographic information that is available for the target population to examine the representativeness of the study. Researchers can also employ a quota sampling strategy to recruit participants.

Another limitation was that the sample was not randomly selected, which could impair the population generalizability and representativeness. Therefore, it is also recommended to randomly select participants from US passengers in future studies if practical. If it is not practical, future studies may include participants with a diversity of demographic characteristics.

A third limitation was identified because participants may have only participated in the study for prizes and did not have realistic responses as on a real aircraft. Therefore, they might not provide valid data. The recommendation regarding this can be that multiple data collection methods should be employed to verify the authenticity of the data. Also, additional monetary incentives can be provided for participants with better performance. A simulation in a more realistic cabin with simulated lighting and motion is recommended. It would be more generalizable to conduct the study in a flying aircraft if safety could be ensured.

Recommendations for research relative to study delimitations. One of the identified delimitations was that the intensity level of auditory cues was designed by changing each acoustic parameter for each level increase. Although the accumulative intensity level increased, the acoustic parameters were also different. The difference between the two separate levels could be the change in acoustic parameters rather than the intensity level. It is recommended in future research, to change only one acoustic parameter or increase the intensity level differently by counterbalancing the acoustic parameters when increasing the intensity levels.

The experiments were conducted in a laboratory setting, where the cabin and events were simulated, to control uninterested variables and strengthen the validity of the study. The design may affect generalizability. In this case, a design with extraneous variables controlled on a real cabin setting is recommended.

In the current study, data were collected with newly-developed scales. They were validated by conducting reliability analyses, but the study could show different findings if it is replicated with different scales. Therefore, other instruments can be used to collect data in future studies and compare the results in order to determine ecological generalizability.

Recommendations for future research relative to implications. With respect to one of the implications that RCT was not fully demonstrated in the current study because (a) it was performance but not safe behaviors that were measured, but the performance was also affected by the arousal levels and (b) the two reviewed studies used only two levels and could not detect the U-shaped relationship, two recommendations were provided. One is that to apply RCT in the studies where stimulations were deployed; the desire to make safer behaviors should be measured instead of the performance. Another recommendation can be that at least three levels should be designed when studying RCT, and a long-term outcome should be tested regarding RHT.

In the current study, the arousal theory and audio warning design were applicable to flight passengers during cruising. Therefore, the arousal theory can be further studied regarding passengers' safety to determine the combination of stimulations on passengers' performance, which can include visual warnings and manipulated unusual motions. Another recommendation is that frequency and number of cue pulses can be valid parameters to increase hearers' PLR and PLU, but they can also increase their annoyance with the auditory cues.

In addition to implications, there are several miscellaneous recommendations about the study design. One recommendation is that physiological signals should not be measured within a short duration because they can be insensitive, but a large sample size may show significant patterns. If the study examines multiple variables, especially when mediating effects are tested, repeated-measures design may not be applicable. Therefore, if the hypothesized relationship between variables is not linear, a recommendation can be that it is more practical to utilize between-subjects design.

Recommendations for practice relative to implications. Combining RCT and arousal theory, it appeared that the optimal auditory cue intensity level was moderate, where passengers have better performance than other levels. Therefore, several moderate levels of auditory cues can be used in aircraft cabins to lead emergency in-flight announcements, which can draw passengers' attention and help them administer good performance regarding the instruction. The optimal parameters may need to be designed for specific cabins. However, the intensity level should not be extremely high. Otherwise, the performance could be impaired after a certain level, and passengers would be considerably annoyed with sounds.

To design cues, the increase in frequency from 0.5 kHz to 1.0 kHz, and the change from one pulse to two pulses were demonstrated to be effective in the cabin. Aviation manufacturers and airlines should consider adding several levels of auditory cues to draw passengers' attention and improve performance. The factors of frequency and number of pulses can be included in the design of auditory cues.

References

1 dead, scores hurt after United jet hits turbulence. (1997, December 28). *World News*. Retrieved from http://www.cnn.com/WORLD/9712/28/japan.turbulence.update/

Aarons, R. N. (2014). Lessons learned-Asiana flight 214. *Business & Commercial Aviation*, *110*(8), 68-71. Retrieved from http://search.proquest.com.ezproxy.libproxy.db.erau.edu/docview/1647367817?accountid=27203

Arousal Level. (2012). In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning*. doi:10.1007/978-1-4419-1428-6

Ary, D., Jacobs, L. C., Sorensen, C., & Asghar, R. (2010). *Introduction to research in education* (8th ed.). Retrieved from http://www.modares.ac.ir/uploads/Agr.Oth.Lib.12.pdf

Attitude. (2013). In G. T. Kurian (Ed.), *The AMA dictionary of business and management*. New York, NY: AMACOM, Publishing Division of the American Management Association.

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measure correlation. *Front Psychology*, *8*, 456. doi:10.3389/fpsyg.2017.00456

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191-215. doi:10.1037/0033-295X.84.2.191

Belz, S. M., Robinson, G. S., & Casali, J. G. (1999). A new class of auditory warning signals for complex systems: Auditory Icons. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *41*(4), 608-618. doi:10.1518/001872099779656734

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. Retrieved from https://babel.hathitrust.org/cgi/pt?id=mdp.39015003862425&view=1up&seq=1

Berlyne, D. E. (1972). Humor and its kin. In J. H. Goldstein (Ed.), *The psychology of humor: Theoretical perspectives and empirical issues* (pp. 43-60). doi:10.1016/B978-0-12-288950-9.50008-0

Bodendörfer, X., Kortekaas, R., Weingarten, M., & Schlittmeier, S. (2015). The effects of spectral and temporal parameters on perceived confirmation of an auditory non-speech signal. *The Journal of the Acoustical Society of America*, *138*(2), EL127-EL132. doi:10.1121/1.4927636

Borrelli, B. (2011). The assessment, monitoring, and enhancement of treatment fidelity in public clinical trials. *Journal of Public Health Dentistry*, *71*(s1), S52-S63. doi:10.1111/j.1752-7325.2011.00233.x

Borrelli, B., Sepinwall, D., Ernst, D., Bellg, A. J., Czajkowski, S., Breger, R., … Orwig, D. (2004). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology*, *75*(5), 852-860. doi:10.1037/0022-006X.73.5.852

Bureau of Transportation Statistics. (2019). *Passengers: All carriers – all airports*. Retrieved from https://www.transtats.bts.gov/Data_Elements.aspx

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Retrieved from https://www.sfu.ca/~palys/Campbell&Stanley-1959-Exptl&QuasiExptlDesignsForResearch.pdf

Cochran's Q Test Using SPSS Statistics. (n.d.). *Lærd Statistics*. Retrieved from https://statistics.laerd.com/spss-tutorials/cochrans-q-test-in-spss-statistics.php

Cohen, J. (1992a). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98-101. doi:10.1111/1467-8721.ep10768783

Cohen, J. (1992b). A power primer. *Psychological Bulletin*, 112, 155-159. doi:10.1111/1467-8721.ep10768783

Cole, C. R., Blackstone, E. H., Pashkow, F. J., Snader, C. E., & Lauer, M. S. (1999). Heart-rate recovery immediately after exercise as a predictor of mortality. *The New England Journal of Medicine*, *341*(18), 1351-1357. doi:10.1056/NEJM199910283411804

Davies, A. (2013, July 11). Yes, you should buckle your seatbelt on an airplane. *Business Insider*. Retrieved from http://www.businessinsider.com/why-you-keep-your-seatbelt-on-in-a-plane-2013-7

Duffy, E. (1957). The psychological significance of the concept of arousal of activation. *Psychological review*, *64*(5), 265-275. Retrieved from https://search-proquest-com.portal.lib.fit.edu/docview/82048637?pq-origsite=summon&accountid=27313

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design. *Human Factors*: *The Journal of the Human Factors and Ergonomics Society*, *33*(2), 205-231. doi:10.1177/001872089103300206

Eysenck, M. W. (1982). *Attention and arousal: Cognition and performance*. doi:10.1007/978-3-642-68390-9

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2013). G*Power Version 3.1.7 [Computer software]. Uiversität Kiel, Germany. Retrieved from http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register

Galvanic Skin Response. (2009). In M. D. Binder, N. Hirokawa, & U. Windhorst (Eds.), *Encyclopedia of neuroscience* (p. 115). Springer, Berlin, Heidelberg.

Galvanic Skin Response. (2012). In R. Sell, M. A. Rothenberg, & C. F. Chapman (Eds.), *Dictionary of medical terms* (6th ed.). Hauppauge, NY: Barron's Educational Series.

Galvanic Skin Response. (2015). In the Editors of the American Heritage Dictionaries (Ed.), *The American heritage dictionary of medicine* (2nd ed.). Boston, MA: Houghton Mifflin.

Haight, F. A. (1986). Risk, especially risk of traffic accident. *Accident Analysis and Prevention*, *18*(5), 359-366. doi:10.1016/0001-4575(86)90009-6

Hellier, E., & Edworthy, J. (1989). Quantifying the perceived urgency of auditory warnings. *Canadian acoustics*, *17*(4), 3-11. Retrieved from https://jcaa.caa-aca.ca/index.php/jcaa/article/view/611

Hellier, E., & Edworthy, J. (1999). On using psychophysical techniques to achieve urgency mapping in auditory warnings. *Applied Ergonomics*, *30*(2), 167-171. doi:10.1016/S0003-6870(97)00013-6

Hellier, E. J., Edworthy, J., & Dennis, I. (1993). Improving auditory warning design: Quantifying the predicting the effects of different warning parameters on perceived urgency. *Human Factors*: *The Journal of the Human Factors and Ergonomics Society*, *35*(4), 693-706. doi:10.1177/001872089303500408

How to use GPower. (n.d.). *Mormonsandscience.* Retrieved from http://www.mormonsandscience.com/gpower-guide.html

Hoyes, T., Neville, A., & Taylor, R. G. (1996). Risk homeostasis theory: A study of intrinsic compensation. *Safety Science*, *22*, 77-86. doi:10.1016/0925-7535(96)00007-0

In-flight Seat Belt Requirements. (n.d.) *Skybrary*. Retrieved from https://www.skybrary.aero/index.php/In-flight_Seat_Belt_Requirements

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*(3), 247-266. doi:10.1016/0272-7358(91)90103-2

Moran, L. (2013, June 4). Passengers can't keep breakfast down when plane hits turbulence mid-meal. *Daily News*. Retrieved from http://www.nydailynews.com/news/world/breakfast-flying-plan-hits-turbulence-article-1.1362435

Mead, N. L. (2007). Compliance. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of Social Psychology* (pp. 161-162). Retrieved from http://web.b.ebscohost.com.portal.lib.fit.edu/ehost/ebookviewer/ebook/bmxlYmtfXzQ3NDMwNV9fQU41?sid=68791524-cb94-4c60-a7bb-280153ac6030@pdc-v-sessmgr03&vid=0&format=EB&rid=1

National Institute for Occupational Safety and Health. (1998). *Criteria for a Recommended Standard*: *Occupational noise exposure* (Publication No. NIOSH-98-126). Retrieved from https://www.cdc.gov/niosh/docs/98-126/pdfs/98-126.pdf

O'Neill, B., & Williams, A. (1998). Risk homeostasis hypothesis: A rebuttal. *Injury Prevention*, *4*(2), 92-93. doi:10.1136/ip.4.2.92

Patterson, R. D. (1982). *Guidelines for auditory warning systems on civil aircraft*. Retrieved from https://pdfs.semanticscholar.org/cd93/db7b798ce1b6fcdb0fc9762b5f9a6f154c65.pdf?_ga=2.107024706.1797249288.1562510308-1149360979.1562510308

Patterson, R. D., & Mayfield, T. F. (1990). Auditory warning sounds in the work environment [and discussion]. *Philosophical Transactions of the Royal Society of London*, *327*(1241), 485-492. doi:10.1098/rstb.1990.0091

Perala, C. H., & Sterling, B. S. (2007). Galvanic skin response as a measure of soldier stress (Performing Organization Report No. ARL-TR-4114). Retrieved from https://permanent.access.gpo.gov/LPS116970/army/www.arl.army.mil/arlreports/2007/ARL-TR-4114.pdf

Privitera, G. J. (2012). Statistics for the behavioral sciences. Thousand Oaks, CA: SAGE Publications, Inc.

Ribble, R. G. (2010). Optimal arousal theory. In *Salem Health: Psychology & mental health* (pp. 1328-1332). Ipswich, MA: Salem Press.

Scales, J. (2018, April 20). Aircraft seat belts: Friend or foe? *The Irish Times*. Retrieved from https://www.irishtimes.com/life-and-style/travel/aircraft-seat-belts-friend-or-foe-1.3468380

Sensory Cue. (2018, October 26). In *Wikipedia*. Retrieved November 2018, from https://en.wikipedia.org/wiki/Sensory_cue#Auditory_Cues

Sheridan, T. B. (2007). Attention and its allocation: Fragments of a model. In A. F. Kramer, D. A. Wiegmann, & A. Kirlik (Eds.), *Human technology interaction series: Attention: From theory to practice* (pp. 16-26). Oxford, England: Oxford University Press.

Sirkka, A., Fagerlönn, J., Lindberg, S., & Frimalm, R. (2014). An auditory display to convey urgency information in industrial control rooms. *Proceedings of the Engineering Psychology and Cognitive Ergonomics*, *Greece*, *LNAI 8532*, 533-544. Retrieved from https://link-springer-com.portal.lib.fit.edu/content/pdf/10.1007%2F978-3-319-07515-0.pdf

Smolders, K. C. H. J., & de Kort, Y. A. W. (2014). Bright light and mental fatigue: Effects on alertness, vitality, performance and physiological arousal. *Journal of Environmental Psychology*, *39*, 77-91. doi:10.1016/j.jenvp.2013.12.010

Stephanie. (2015). McNemar test definition, examples, calculation. *Statistics How to*. Retrieved from https://www.statisticshowto.datasciencecentral.com/mcnemar-test/

Strohminger, N. (2014). Arousal theory (Berlyne). In S. Attardo (Ed.), *Encyclopedia of humor studies* (pp. 62-63). Thousand Oaks, CA: Sage Publications.

Suied, C., Susini, P., & McAdams, S. (2008). Evaluating warning sound urgency with reaction times. *Journal of Experimental Psychology*: *Applied*, *14*(3), 201-212. doi:10.1037/1076-898X.14.3.20

Toohill, K. (2015, May 10). The real reason you should wear your seat belt on airplanes. *Attn:*. Retrieved from http://www.attn.com/stories/1635/why-its-important-wear-your-seatbelt-airplane

Turbulence accidents that killed airline passengers. (2012, October 18). *Airsafe.com*. Retrieved from http://www.airsafe.com/events/turb.htm

Trimpop, R. M. (1996). Risk homeostasis theory: Problems of the past and promises for the future. *Safety Science*, *22*, 119-130. doi:10.1016/0925-7535(96)00010-0

Ünal, A. B., de Waard, D., Epstude, K., & Steg, L. (2013). Driving with music: Effects on arousal and performance. *Transportation Research Part F*: Psychology and Behaviour, *21*, 52-65. doi:10.1016/j.trf.2013.09.004

Walther, A., Breidenstein, J., Bösch, M., Selfidan, S., Ehlert, U., Annen, H., … La Marca, R. (2018). Association between digit ratio (2D4D), mood, and autonomic stress response in healthy men. *Psychophysiology*, *56*(5), 1-13. doi:10.1111/psyp.13328

Wang, L., Guo, W., Ma, X., & Li, B. (2016). Analysis of influencing factors of auditory warning signals' perceived urgency and reaction time. *Proceedings of Engineering Psychology and Cognitive Ergonomics*, *Canada*, *9736*, 452-463. doi:10.1007/978-3-319-40030-3_44

Wilde, G. J. S. (1982a). The theory of risk homeostasis: Implications for safety and health. *Risk Analysis*, *2*, 209-225. doi:10.1111/j.1539-6924.1982.tb01384.x

Wilde, G. J. S. (1982b). Critical issues in risk homeostasis theory. *Risk Analysis*, *2*, 249-258. doi:10.1111/j.1539-6924.1982.tb01389.x

Wilde, G. J. S. (1988). Risk Homeostasis theory and traffic accidents: Propositions, deductions and discussion of dissension in recent reactions. *Ergonomics*, *31*(4), 441-468. doi:10.1080/00140138808966691

Wilde, G. J. S. (1989). Accident countermeasures and behavioral compensation: The position of risk homeostasis theory. *Journal of Occupational Accidents*, *10*(4), 267-292. doi:10.1016/0376-6349(89)90021-7

Wilde, G. J. S. (2014). *Target Risk 3: Risk Homeostasis in Everyday Life*. Retrieved from https://is.muni.cz/el/1423/podzim2016/PSY540/um/64998189/64998284/targetrisk3_1.pdf

Wilde, G. J. S., Claxton-Oldfield, S. P., & Platenius, P. H. (1985). Risk homeostasis in an experimental context. In: L. Evans, & R. C. Schwing (Eds), *Human behavior and traffic safety* (pp. 119-149). doi:10.1007/978-1-4613-2173-6_7

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459-482. doi:10.1002/cne.920180503

Appendix A

Bipolar Adjective Scales

Auditory Cue 1

How do you describe your sense of urgency in the situation after hearing this auditory cue?

Voluntary										Urgent
Critical										Needless

To what extent did you perceive the risk in the situation after hearing the auditory cue?

Risky										Harmless
Safe										Dangerous

How annoying did the auditory cue seem to you?

Agreeable										Annoying
Aggravating										Pleasant

Auditory Cue 2

How do you describe your sense of urgency in the situation after hearing this auditory cue?

Voluntary										Urgent
Critical										Needless

To what extent did you perceive the risk in the situation after hearing the auditory cue?

Risky										Harmless
Safe										Dangerous

How annoying did the auditory cue seem to you?

Agreeable										Annoying
Aggravating										Pleasant

Auditory Cue 3

How do you describe your sense of urgency in the situation after hearing this auditory cue?

Voluntary										Urgent
Critical										Needless

To what extent did you perceive the risk in the situation after hearing the auditory cue?

Risky										Harmless
Safe										Dangerous

How annoying did the auditory cue seem to you?

Agreeable										Annoying
Aggravating										Pleasant

Auditory Cue 4

How do you describe your sense of urgency in the situation after hearing this auditory cue?

Voluntary										Urgent
Critical										Needless

To what extent did you perceive the risk in the situation after hearing the auditory cue?

Risky										Harmless
Safe										Dangerous

How annoying did the auditory cue seem to you?

Agreeable										Annoying
Aggravating										Pleasant

Auditory Cue 5

How do you describe your sense of urgency in the situation after hearing this auditory cue?

Voluntary										Urgent
Critical										Needless

To what extent did you perceive the risk in the situation after hearing the auditory cue?

Risky										Harmless
Safe										Dangerous

How annoying did the auditory cue seem to you?

Agreeable										Annoying
Aggravating										Pleasant

Appendix B

Instruction

Instruction

This is an in-flight simulation experiment. Assume you are flying on a commercial airliner from Orlando to New York. During the experiment, you are required to behave as if you are on a commercial flight and obey all current in-flight regulations. You are not allowed to talk, listen to music, or leave your seat. There will be several announcements, which are similar to real in-flight announcements. Please listen to and adhere to the information provided in the announcements (**make sure your responses are observable**). Also, please solve the Sudoku on a tablet. Your performance will affect the possibility to win the raffle prize. <u>If you feel uncomfortable during the experiment, you can quit at any time.</u>

Appendix C

IRB Approval Letter

Florida Institute of Technology
Institutional Review Board

Notice of Expedited Review Status
Certificate of Clearance for Human Participants Research

Principal Investigator:	Tianhua Li
Date:	October 8, 2019
IRB Number:	19-152
Study Title:	The Effect of Intensive Auditory Cues on Flight Passengers' Safe Behaviors and Attitudes

Your research protocol was reviewed and **approved** by the IRB Chairperson. Per federal regulations, 45 CFR 46.110, your study has been determined to involve no more than minimal risk for human subjects. Federal regulations define minimal risk to mean that the probability and magnitude of harm are no more than would be expected in the daily life of a normal, healthy person.

Unless you have requested a waiver of consent, participants must sign a consent form, and the IRB requires you give each participant a copy of the consent form for their records. For online surveys, please advise participants to print out the consent screen for their files.

All data, which may include signed consent form documents, must be retained in a locked file cabinet for a minimum of three years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained on a password-protected computer if electronic information is used. Access to data is limited to authorized individuals listed as key study personnel.

Prompt reporting to the IRB is required in the following conditions:

- Procedural changes increasing the risk to participants or significantly affecting the conduct of the study
- All adverse or unanticipated experiences or events that may have real or potential unfavorable implications for participants
- New information that may adversely affect the safety of participants or the conduct of the study.

This study is approved for one year from the above date. If data collection continues past this date, a Protocol Renewal Form must be submitted.

High Tech with a Human Touch™

150 West University Boulevard, Melbourne, FL 32901-6975 • (321) 674-7316

Appendix D

Informed Consent

Informed Consent

Please read this consent document carefully before you decide to participate in this study. The researcher will answer any questions before you sign this form.

Study Title: Cabin Simulation Experiment

Purpose of the Study: The purpose of the study is to test passengers' behaviors under different situations.

Procedures: This is an in-flight simulation experiment. You will be assigned a seat. During the entire experiment, you are required to obey all current in-flight regulations. You are not allowed to leave the seat or talk. There will be several announcements, which are similar to real in-flight announcements. Your involvement in this simulation will be approximately 40 minutes. After the simulation, you will be asked to fill out a questionnaire.

Potential Risks of Participating: Risks of participating in this experiment are no more than everyday life.

Potential Benefits of Participating: Your participation will help us understand passenger behavior.

Confidentiality: Your responses in this study will be anonymous. The simulation will be recorded to collect the information about your behaviors, and the camera will capture your partial facial features. Only researchers directly involved in this study will have access to the data and video recordings. Recordings will not be used by anyone other than directly-involved researchers consisting of my doctoral committee and myself. In order to protect the confidentiality of your responses, I will provide each participant with a random ID for the study. Collected data and recordings will be entered and stored on a password-protected computer. They will be stored for three years after any publication in my dissertation chair's locked office, and then will be shredded. Your name will not be used in any report.

Voluntary participation: To be in this study, you must (a) be fluent in English, (b) be at least 18 years old, (c) have normal hearing ability, and (d) have commercial flight experience as a passenger on a commercial flight. Your participation in this study is completely voluntary. There is no penalty for not participating. You may also refuse to answer any questions asked of you.

Right to withdraw from the study:
You have the right to withdraw from the study at any time without consequence.

Whom to contact if you have questions about the study:

Tianhua Li, Ph.D. student, Aviation Sciences
Email: tli2017@my.fit.edu Phone: 386.212.3718

Dr. Carstens, Dissertation Chair, College of Aeronautics
Email: Carstens@fit.edu Phone: 321.674.8820

Whom to contact about your rights as a research participant in the study:

Dr. Jignya Patel, IRB Chairperson
150 West University Blvd.
Melbourne, FL 32901
Email: FIT_IRB@fit.edu Phone: 321-674-7347

Agreement:
I have read the procedure described above. I voluntarily agree to participate in the procedure, and I have received a copy of this description.

Participant: ____ Date: _________________

Email: ____ (announce the winner)

Principal Investigator: __________________________________ Date: _________________

Appendix E

Announcements

Announcements

General Announcement (Emergency Exit)
Ladies and gentlemen, if you are seated next to an emergency exit and do not wish to perform the functions in the event of an emergency, please ask a flight attendant to reseat you.

General Announcement (No Smoking)
Ladies and gentlemen, we remind you that this is a non-smoking flight. Smoking is prohibited on the entire aircraft. Tampering with, disabling, or destroying the lavatory smoke detectors is prohibited by law.

Action-Required Announcement (Fasten Seatbelts)
Ladies and gentlemen, the Captain has turned on the fasten seat belt sign. We are now crossing a zone of turbulence. Please return to your seats and keep your seat belts fastened. Thank you.

Action-Required Announcement (Close Tray Tables)
Ladies and gentlemen, we have detected a possible mechanical issue. The crew is now sorting out the problem. Please be seated and close your tray tables. Thank you.

Action-Required Announcement (Lower Tray Tables)

Ladies and gentlemen, in a few minutes, the flight attendants will be passing around the cabin to serve dinner. To help flight attendants serve quickly, please lower your tray tables. Thank you.

General Announcement (Food)

Today, we will be serving beef, turkey, and chicken. And we also offer you hot or cold drinks. Alcoholic drinks are available at a nominal charge.

General Announcement (Destination Weather)

We will arrive in New York soon. The weather in New York is clear and sunny, with a high of 25 degrees for this afternoon. If the weather cooperates, we should get a great view of the city as we descend.

Action-Required Announcement (Adjust Seat Backs)

Ladies and gentlemen, as we start our descent, please make sure all carry-on luggage is stowed in the overhead compartment. And please make sure your seat backs are in their full upright position. Thank you.

Appendix F

Demographic Questionnaire

Gender:
Male________________ Female_______________

Age: _________________

In your best judgment, do you have a hearing impairment?

Yes ___________ No ___________

If yes, explain: ________________________

Are you fluent in English?

Yes ___________ No ___________

What is your nationality: ___________________

What is the nation of your birth? ___________________

What is your first language: ____________________

In which country did you attend high school: ________________

In which country did you do/are you doing your bachelor's degree: ____________

How frequently do you fly (how many times per year on average)? _______

How frequently do you comply with the fasten-seatbelt instructions (1-100%)? _______

How frequently do you comply with the instructions to place your cell-phone on airplane-mode (1-100%)? _______

How frequently do you pay attention to the pre-flight safety briefing (1-100%)? _______

Appendix G

Attitude Questionnaire

1. I was bored during the simulation.

○	○	○	○	○
Strongly disagree	Disagree	Neutral	Agree	Strongly Agree

2. I was mad at the simulation.

○	○	○	○	○
Strongly disagree	Disagree	Neutral	Agree	Strongly Agree

3. I found the simulation interesting.

○	○	○	○	○
Strongly disagree	Disagree	Neutral	Agree	Strongly Agree

Appendix H

MATLAB Programming Codes for ECG Analyses

```
[Initialize]
time=inputdlg({'HH','MM','SS'},'Start Time');
if exist('format1','var')==0
   format1=0;
end
if exist('format2','var')==0
   format2=0;
end
if isempty(time)==0
if isempty(str2num(time{1,1}))==0
   if isempty(str2num(time{2,1}))==0
      if isempty(str2num(time{3,1}))==0
thh=str2num(time{1,1});
tmm=str2num(time{2,1});
tss=str2num(time{3,1});
      else
         thh=str2num(time{1,1});
         tmm=str2num(time{2,1});
         tss=00;
      end
t0=T2421(thh,tmm,tss);
t2h=hour(New{height(New),1});
t2m=minute(New{height(New),1});
t2s=second(New{height(New),1});
tl=T1224(T2421(t2h,t2m,t2s)-t0);
tlh=tl{1,1};
tlm=tl{1,2};
tls=tl{1,3};
questdlg({'Do you want to modify the length of the session?',sprintf('The total
length is %d hour %d minutes %.2f second',tlh,tlm,tls)},'Session
Length','Yes','No','No');
      answer=ans;
      if sum(char(answer))==sum('Yes')
         length=inputdlg({'HH','MM','SS'},'Length');
         if isempty(str2num(length{1,1}))==0
            if isempty(str2num(length{2,1}))==0
               if isempty(str2num(length{3,1}))==0
lhh=str2num(length{1,1});
lmm=str2num(length{2,1});
lss=str2num(length{3,1});
               else
                  lhh=str2num(length{1,1});
```

```
                lmm=str2num(length{2,1});
                lss=1;
            end
        else
            lhh=str2num(length{1,1});
            lmm=38;
            lss=1;
        end
    else
        lhh=0;
        lmm=38;
        lss=01;
    end
elseif sum(char(answer))==sum('No')
    lhh=0;
    lmm=38;
    lss=01;
elseif sum(char(answer))==sum('')
    error
end
t2=t0+T2421(lhh,lmm,lss);
if t2>T2421(t2h,t2m,t2s)
    error('The selected length exceeds the length of the data.')
end
if format1==0
for i=1:width(New)
    if strcmp(New.Properties.VariableNames{1,i},'TimeHHmmss000')==1
        it=i;
        break
    end
end
for i=1:width(New)
    if strcmp(New.Properties.VariableNames{1,i},'ECGLead1')==1
        ie=i;
        break
    end
end
if ie<width(New)
    New(:,ie+1:width(New))=[];
end
if ie>it+1
    New(:,it+1:ie-1)=[];
```

```
end
if it>1
   New(:,1:it-1)=[];
end
New{:,3}=hour(New{:,1});
New{:,4}=minute(New{:,1});
New{:,5}=second(New{:,1});
New{:,1}=New{:,1}-fix(datenum(New{1,1}));
New=addvars(New,datenum(New{:,1}),'Before',1);
New.Properties.VariableNames{1}='Time';
New(:,2)=[];
format1=1;
end
if format2==0
h0=1;
h2=height(New);
n=1;
while 2^n<h2
   n=n+1;
end
n2=n+1;
h1a=h0;
h1b=h2;
for n=1:n2
   h1=h1a+fix((h1b-h1a)/2);
   if New{h1,1}>=t0
      if New{h1-1,1}<t0
         h11=h1-1;
         break;
         disp(1);
      else
         h1b=h1;
      end
   else
      h1a=h1;
   end
end
n=1;
while 2^n<h2-h1
   n=n+1;
end
n2=n+1;
```

```
h1a=h1;
h1b=h2;
for n=1:n2
   h1=h1a+fix((h1b-h1a)/2);
   if New{h1,1}>=t2
      if New{h1-1,1}<t2
         h12=h1;
         break;
      else
         h1b=h1;
      end
   else
      h1a=h1;
   end
end
New(h12:h2,:)=[];
New(h0:h11,:)=[];
format2=1;
end
New.Properties.VariableNames{2}='ECG_1';
New.Properties.VariableNames{3}='Hour';
New.Properties.VariableNames{4}='Minute';
New.Properties.VariableNames{5}='Second';
ecgtime=1;
open New;
disp('Initialization is finished. Detecting heart beats.');
DetectHBs%VoltageDif(default)
disp('Removing zeroes.');
RemoveZero%
disp('Detecting artefacts.');
DetectArtefacts%Lower&upper
   end
end
end

[DetectHBs]
mv1=0.2;%0.2
mv2=0.3;%0.3
skip=0;
if width(New)>5
   questdlg({'Heart beats have already been detected.','Do you want to clear current
heart beats?'},'Warning','Yes','No','No');
```

```
    answer=ans;
    if sum(char(answer))==sum('Yes')
        New(:,6)=[];
    elseif sum(char(answer))==sum('No')
        error('Heart beats cannot be detected without removing current data.');
    elseif sum(char(answer))==sum('')
        skip=1;
    end
end
if skip==0
h0=7;
h1=height(New)-6;
for h=h0:h1
    if New{h,2}>=New{h-5,2}+mv1
        if New{h,2}>=New{h+5,2}+mv2
            if New{h,2}<1.2
            for i=h-5:h-1
                if New{h,2}<=New{i,2}
                    ir1=0;
                    break
                else
                    ir1=1;
                end
            end
            if ir1==1
                for i=h+1:h+5
                    if New{h,2}<New{i,2}
                        ir2=0;
                        break
                    else
                        ir2=1;
                    end
                end
                if ir2==1
                    New{h,6}=1;
                end
            end
            end
        end
    end
end
disp('DetectHBs is finished');
```

```
end

[RemoveZero]
h=1;%3m
h2=height(New);
New(:,7)=New(:,6);
h12=h;
it=0;
while h<h2;
   if h==h2-1
      break
   end
   if New{h,2}<-5
      %Find ones around 0
      h1=h;
      h11=h1-1;
      h12=h1+1;
      while 0<1
         h11=h11-1;
         if h11==1
            it=1;
            break
         end
         if find(New{h11,6},1)==1
            break
         end
      end
      if it==1
      else
      while 0<1
         h12=h12+1;
         if h12==h2
            break
         end
         if find(New{h12,6},1)==1
            break
         end
      end
      td=New{h12,3}*3600-New{h11,3}*3600+New{h12,4}*60-
New{h11,4}*60+New{h12,5}-New{h11,5};
      ti1=td/1;
      ti2=td/2;
```

```
        ti3=td/3;
        %Find more 1
        h1=h11;
        h10=h11-50;
        while 0<1
          h10=h10-1;
          if find(New{h10,6},1)==1
             break
          end
        end
        td0=New{h11,3}*3600-New{h10,3}*3600+New{h11,4}*60-
New{h10,4}*60+New{h11,5}-New{h10,5};
        td1=abs(ti1-td0);
        td2=abs(ti2-td0);
        td3=abs(ti3-td0);
        if td1<td2
          if td1<td3
             n1=1;
          else
             n1=3;
          end
        else
          if td2>td3
             n1=3;
          else
             n1=2;
          end
        end
        if n1==1
        else
          h111=h11;
          tit=td/n1;
          for ht=h11+1:h12-1
             h112=ht;
             tdt=New{h112-1,3}*3600-New{h111,3}*3600+New{h112-1,4}*60-
New{h111,4}*60+New{h112,5}-New{h111,5};
             if tdt>tit
                New{ht,7}=1;
                n=n+1;
                h111=ht;
             else
                New{ht,7}=0;
```

```
            end
         end
      end
      h=h12;
      end
   end
   h=h+1;
end
Artefacts{1,1}=0;
disp('RemoveZero is finished.');
detect=1;

[DetectArtefacts]
close all
count=0;
if detect==1
   ll=60;
   ul=100;
elseif detect>1
   answer=questdlg({'Manually define the heart rate range?','Default Limits: 60 and
100'},'Define Range','Yes','No','No');
   if sum(char(answer))==sum(char('Yes'))
      range=inputdlg({'Lower Limit (60)','Upper Limit (100)'},'Range');
      if isempty(str2num(range{1,1}))==0
         ll=str2num(range{1,1});
         if isempty(str2num(range{2,1}))==0
            ul=str2num(range{2,1});
         else
            ul=100;
         end
      else
         ll=60;
         if isempty(str2num(range{2,1}))==0
            ul=str2num(range{2,1});
         else
            ul=100;
         end
      end
   elseif sum(char(answer))==sum(char('No'))
      ll=60;
      ul=100;
   elseif sum(char(answer'))==sum(char(''))
```

```
        return
    end
end
Artefacts(:)=[];
l=width(New);
ht=1;
while 0<1
    if New{ht,l}==1
        break
    end
    ht=ht+1;
end
t22=t0+T2421(lhh,lmm,lss-1);
h0=ht+1;
h2=height(New);
n=1;
while 2^n<h2
    n=n+1;
end
n2=n+1;
h1a=h0;
h1b=h2;
for n=1:n2
    h1=h1a+fix((h1b-h1a)/2);
    if New{h1,1}>=t22
        if New{h1-1,1}<t22
            h12=h1;
            break;
        else
            h1b=h1;
        end
    else
        h1a=h1;
    end
end
j=h0;
i=1;
for h=h0:h12
    if New{h,l}==1
        td=New{h,3}*3600-New{j,3}*3600+New{h,4}*60-New{j,4}*60+New{h,5}-
New{j,5};
        if td<60/(ul+1)
```

```
            Artefacts{i,1}=h;
            Artefacts{i,2}=sprintf('High');
            Artefacts{i,3}=sprintf('%d',fix(60/td));
            Artefacts{i,4}=New{h,3};
            Artefacts{i,5}=New{h,4};
            Artefacts{i,6}=New{h,5};
            i=i+1;
        else
            if td>=60/ll
                Artefacts{i,1}=h;
                Artefacts{i,2}=sprintf('Low');
                Artefacts{i,3}=sprintf('%d',fix(60/td));
                Artefacts{i,4}=New{h,3};
                Artefacts{i,5}=New{h,4};
                Artefacts{i,6}=New{h,5};
                i=i+1;
            end
        end
        j=h;
    end
end
open Artefacts;
if size(Artefacts,1)>0
    figure=1;
    artefact=1;
    open ECG;
    if detect==1
        if exist('Correction','var')==0
            Correction{1,1}=1;
        end
    end
    open Correction;
    open Artefacts;
    msgbox(sprintf('%d artefacts were identified',size(Artefacts,1)));
    Chirp
else
    if width(New)==7
        msgbox('No artefact is found. Process proceeds.');
        artefact=0;
        EditArtefacts;
    elseif width(New)==8
        if artefact==1
```

```
        Train
      questdlg('No Artefact is found. Do you want to modify the treatment
number?','No New Artefact','Yes','No','No');
      answer=ans;
      if sum(char(answer))==sum('Yes')
        custreat=1;
        treatnum=inputdlg('Number of Treatment','Treatment');
        treatment=str2num(treatnum{1,1});
        CalculateHRs
        ListTimes%Interval&Segments
        MeanHRs
        CalculateDif
        ListResults
        Chirp
        return
      elseif sum(char(answer))==sum('No')
        custreat=0;
        treatment=8;
        CalculateHRs
        ListTimes%Interval&Segments
        MeanHRs
        CalculateDif
        ListResults
        Chirp
        return
      elseif sum(char(answer))==sum(char(''))
        custreat=0;
        return
      end
      end
   end
end
detect=detect+1;
pre0=500;
aft0=600;
disp('DetectArtefacts');

[EditArtefacts]
if size(Correction,2)>1
if isempty(Correction{1,2})==1
      Correction(1,:)=[];
end
```

```
else
   if size(Correction,1)>0
   Correction(1,:)=[];
   end
end
i0=1;%1m/20
is=size(Correction);
i2=is(1);
hs=size(New);
if is(2)==1
   New(:,8)=New(:,7);
elseif hs(2)<8
   New(:,8)=New(:,7);
end
for i=i0:i2
h0=Correction{i,1};
h2=Correction{i,2};
n1=Correction{i,3};
td=New{h2-1,3}*3600-New{h0,3}*3600+New{h2-1,4}*60-
New{h0,4}*60+New{h2-1,5}-New{h0,5};
ti=td/n1;
if ti<=60/120
   message=sprintf('High in %d: %.2f (HR). Do you still want to proceed?',i, 60/ti);
   questdlg(message,'Yes','Yes','No','No');
   answer=ans;
   if sum(char(answer))==sum('Yes')
   elseif sum(char(answer))==sum('No')
      error('Please correct errors.');
   end
elseif ti>=60/50
   message=sprintf('Low in %d: %.2f (HR).Do you still want to proceed?',i, 60/ti);
   questdlg(message','Yes','Yes','No','No');
   answer=ans;
   if sum(char(answer))==sum('Yes')
   elseif sum(char(answer))==sum('No')
      error('Please correct errors.');
   end
end
h11=h0;
n=1;
for h=h0+1:h2-1
   New{h,8}=0;
```

```
end
for h=h0+1:h2
   if n==n1
      break
   end
   h12=h;
   td=New{h12-1,3}*3600-New{h11,3}*3600+New{h12-1,4}*60-
New{h11,4}*60+New{h12,5}-New{h11,5};
   if td>ti
      New{h,8}=1;
      n=n+1;
      h11=h;
   end
end
end
disp('EditArtefacts is finished');
DetectArtefacts

[ECG]
if exist('count','var')==0
   count=0;
   close all
end
skip=1;
while skip==1
pre=500;%500
aft=600;%600
if pre~=pre0
   figure=1;
   disp('Remember to reset the "pre" variable');
elseif aft~=aft0
   figure=1;
   disp('Remember to reset the "aft" variable');
end
if figure>size(Artefacts,1)
   questdlg('All artefacts have been treated. Edit artefacts?','All has
finished','Yes','No','No');
   answer=ans;
   if sum(char(answer))==sum('Yes')
      if isempty(Artefacts)==0
         Artefacts(1,:)=[];
         EditArtefacts
```

```
      return
    end
  elseif sum(char(answer))==sum('No')
    return
  end
else
rn=Artefacts{figure,1};
h=figure;
while 0<1
  if h==size(Artefacts,1)
    break
  end
  if Artefacts{h+1,1}>Artefacts{h,1}+600
    break
  end
  h=h+1;
end
int=Artefacts{h,1}+aft-rn;
fx=fix(rn/10000)*10000;
if rn~=Artefacts{figure,1}
  plot(rn-fx-intbef:rn-fx+intaft,New.ECG_1(rn-intbef:rn+intaft));
  figure=1;
else
  if rn-int>=1
    if rn>=pre+1
      if rn+int<=height(New)
        temp=0;
      else
        temp=2;
      end
    else
      temp=1;
    end
  else
    temp=1;
  end
open Artefacts;
i1=rn-1;
while 0<1
  if i1==1
    i1='beginning';
    break
```

```
    end
    if New{i1,width(New)}==1
        break
    else
        i1=i1-1;
    end
end
i2=rn+1;
while 0<1
    if i2==height(New);
        i2=='end';
        break
    end
    if New{i2,width(New)}==1
        break
    else
        i2=i2+1;
    end
end
if temp==0
    plot(i1-pre-fx:rn+int-fx,New.ECG_1(i1-pre:rn+int),'-
p','MarkerIndices',[pre+1,rn-i1+pre+1,i2-i1+pre+1,Artefacts{h,1}-
i1+pre+1],'MarkerFaceColor','red','MarkerSize',15);
    disp(sprintf('%d(%d) %d(%d) %d(%d)',i1,i1-fx,rn,rn-fx,i2,i2-fx));
elseif temp==1
    plot(1:rn+int-fx,New.ECG_1(1:rn+int),'-
p','MarkerIndices',[i1,rn,i2,Artefacts{h,1}-
fx],'MarkerFaceColor','red','MarkerSize',15);
    disp(sprintf('Begin %d(%d) %d(%d) %d(%d)',i1,i1-fx,rn,rn-fx,i2,i2-fx));
elseif temp==2
    plot(i1-pre-fx:height(New)-fx,New.ECG_1(i1-pre:height(New)),'-
p','MarkerIndices',[pre+1,rn-i1+pre+1,i2-i1+pre+1,Artefacts{h,1}-
i1+pre+1],'MarkerFaceColor','red','MarkerSize',15);
    disp(sprintf('%d(%d) %d(%d) %d(%d) End',i1,i1-fx,rn,rn-fx,i2,i2-fx));
end
if count==0
    pause(6)
    count=1;
end
if str2num(Artefacts{figure,3})>100
```

```
   questdlg({'Correct or
Skip?',sprintf('High: %d',str2num(Artefacts{figure,3})),sprintf('Showing %d/%d',fi
gure,size(Artefacts,1))},'Result','Correct','Determine','Skip','Skip');
   answer=ans;
   if sum(char(answer))==sum('Skip')
      skip=1;

   elseif sum(char(answer))==sum('Correct')
      pause(6);
      IfOnes
      skip=0;
      break
   elseif sum(char(answer))==sum('Determine')
      pause(6);
      DetermineIntervals
      skip=0;
      break
   elseif sum(char(answer))==sum('')
      skip=0;
      break
   end
elseif str2num(Artefacts{figure,3})<60
   questdlg({'Correct or
Add?',sprintf('Low: %d',str2num(Artefacts{figure,3})),sprintf('Showing %d/%d',fig
ure,size(Artefacts,1))},'Edit Option','Correct','Add','Skip','Skip');
   answer=ans;
   if sum(char(answer))==sum('Correct')
      questdlg('Correct or Determine?','Correct','Correct','Determine','Determine');
      if sum(char(answer))==sum('Correct')
         pause(6);
         IfOnes
         skip=0;
         break
      elseif sum(char(answer))==sum('Determine')
         pause(6);
         DetermineIntervals
         skip=0;
         break
      elseif sum(char(answer))==sum('')
         skip=0;
         break
      end
```

```
    elseif sum(char(answer))==sum('Add')
        if str2num(Artefacts{figure,3})<40
            pause(6)
        else
            pause(3);
        end
        GiveOne;
        skip=0;
        break
    elseif sum(char(answer))==sum('Skip')
        skip=1;
    elseif sum(char(answer))==sum('')
        skip=0;
        break
    end
else
    open ECG;
    skip=0;
    break
    disp(sprintf('Now at %d',figure));
end
ha=Artefacts{h,1};
pre0=pre;
aft0=aft;
bottom=0;
figure=figure+1;
end
end
end

[IfOnes]
correct=inputdlg({'Point 1','Point 2','Interval'},'Correction');
if size(correct,1)>0
if isempty(str2num(correct{1,1}))==0
    if isempty(str2num(correct{2,1}))==0
        if isempty(str2num(correct{3,1}))==0
i1=str2num(correct{1,1});
i2=str2num(correct{2,1});
it=str2num(correct{3,1});
if i1+fx==Correction{size(Correction,1),1}
    if it==Correction{size(Correction,1),3}
        if i2+fx==Correction{size(Correction,1),2}
```

```
      open ECG
      error('This correction has already been added');
    end
  end
end
fx=fix(rn/10000)*10000;
fe1=0;
fe2=0;
if New{fx+i1,width(New)}==1
  if New{fx+i2,width(New)}==1
    message=sprintf('Data are checked');
    ih=size(Correction,1)+1;
    Correction{ih,1}=fx+i1;
    Correction{ih,2}=fx+i2;
    Correction{ih,3}=it;
  else
    f1=i2;
    f2=i2+1;
    while 0<1
      if New{f1+fx,width(New)}==1
        fo=f1;
        break
      end
      if New{f2+fx,width(New)}==1
        fo=f2;
        break
      end
      f1=f1-1;
      f2=f2+1;
    end
    fi2=fo;
    if (fo-i2)^2<=100
      fe2=1;
    else
      error(sprintf('%d (%d) is not 1 (%d is one)',i2,i2+fx,fi2));
    end
  end
else
  f1=i1;
  f2=i1+1;
  while 0<1
    if New{f1+fx,width(New)}==1
```

```
      fo=f1;
      break
    end
    if New{f2+fx,width(New)}==1
      fo=f2;
      break
    end
    f1=f1-1;
    f2=f2+1;
  end
  fi1=fo;
  if (fo-i1)^2<=100
    fe1=1;
  else
    error(sprintf('%d (%d) is not 1 (%d is one)',i1,i1+fx,fi2));
  end
end
if fe1==1
  if fe2==1
    ih=size(Correction,1)+1;
    Correction{ih,1}=fx+fi1;
    Correction{ih,2}=fx+fi2;
    Correction{ih,3}=it;
    message={sprintf('%d (%d) is not 1 (%d is one)',i1,i1+fx,fi1),sprintf('%d (%d)
is not 1 (%d is one)',i2,i2+fx,fi2)};
  else
    ih=size(Correction,1)+1;
    Correction{ih,1}=fx+fi1;
    Correction{ih,2}=fx+i2;
    Correction{ih,3}=it;
    message=sprintf('%d (%d) is not 1 (%d is one)',i1,i1+fx,fi1);
  end
else
  if fe2==1
    ih=size(Correction,1)+1;
    Correction{ih,1}=fx+i1;
    Correction{ih,2}=fx+fi2;
    Correction{ih,3}=it;
    message=sprintf('%d (%d) is not 1 (%d is one)',i2,i2+fx,fi2);
  end
end
h=1;
```

```
if size(Artefacts,1)==1
   questdlg('All artefacts have been treated. Edit artefacts?','Yes','Yes','No','No');
   answer=ans;
   if sum(char(answer))==sum('Yes')
      Artefacts(1,:)=[];
      EditArtefacts
   elseif sum(char(answer))==sum('No')
   end
else
if Artefacts{h,1}>fx+i2
else
bottom=0;
while 0<1
   if Artefacts{h,1}<fx+i2
      h=h+1;
   else
      if h==size(Artefacts,1)
      bottom=1;
      break
      else
         Artefacts(1:h,:)=[];
      break
   end
end
end
end
end
if bottom==1
   questdlg('All artefacts have been treated. Edit artefacts?','Yes','Yes','No','No');
   answer=ans;
   if sum(char(answer))==sum('Yes')
      Artefacts(1,:)=[];
      EditArtefacts
   elseif sum(char(answer))==sum('No')
   end
   EditArtefacts
else
figure=1;
end
      else
         error('Enter all values');
      end
```

```
   else
      error('Enter all values');
   end
else
   error('Enter all values');
end
   questdlg(message,'Next Artefact','Next Artefact','Keep Correcting','Keep
Correctin');
   answer=ans;
   if sum(char(answer))==sum('Next Artefact')
      ECG
   elseif sum(char(answer))==sum('Keep Correcting')
   end
end

[DetermineIntervals]
determine=inputdlg({'Reference Point','Point 1','Point 2'},'Interval');
if size(determine,1)>0
if isempty(str2num(determine{1,1}))==0
   if isempty(str2num(determine{2,1}))==0
      if isempty(str2num(determine{3,1}))==0
i1=str2num(determine{1,1});
i2=str2num(determine{2,1});
i3=str2num(determine{3,1});
m=i1-6+fx;
for h=i1-5+fx:i1+5+fx
   if New{m,2}<New{h,2}
      m=h;
   end
end
if m==i1+fx
else
   if abs((m-i1-fx))>=10
      error(sprintf('%d is higher than %d by more than 10',m-fx,i1));
   else
      i1=m-fx;
   end
end
m=i2-6+fx;
for h=i2-5+fx:i2+5+fx
   if New{m,2}<New{h,2}
      m=h;
```

```
    end
end
if m==i2+fx
else
    if abs((m-i2-fx))>=10
        error(sprintf('%d is higher than %d by more than 10',m-fx,i2));
    else
        i2=m-fx;
    end
end
m=i3-6+fx;
for h=i3-5+fx:i3+5+fx
    if New{m,2}<New{h,2}
        m=h;
    end
end
if m==i3+fx
else
    if abs((m-i3-fx))>=10
        error(sprintf('%d is higher than %d by more than 10',m-fx,i3));
    else
        i3=m-fx;
    end
end
dif0=(New{i2+fx,1}-New{i1+fx,1})*24*60*60;
dif=(New{i3+fx,1}-New{i2+fx,1})*24*60*60;
n=1;
temp1=dif;
while 0<1
    dif1=dif/n;
    temp2=abs(dif1-dif0);
    if temp2>temp1
        in=n-1;
        break
    else
        temp1=temp2;
    end
    n=n+1;
end
msgbox({sprintf('The interval is suggested to be} %d',in),sprintf('Referenced HR
is %.2f',60/dif0),sprintf('Difference is %.2f',abs(60/dif0-60*in/dif))});
pause(2);
```

```
IfOnes;
        else
            error('Enter all values');
        end
    else
        error('Enter all values');
    end
else
    error('Enter all values');
end
end

[GiveOne]
I=inputdlg({'Point 1','Point 2 (Optional)','Point 3 (Optional)'},'One');
if size(I,1)>0
if isempty(str2num(I{1,1}))==0
for ni=1:size(I,1)
    if isempty(str2num(I{ni,1}))==1
        break
    end
end
for n=1:ni-1
i=str2num(I{n,1});
m=i-6+fx;
for h=i-5+fx:i+5+fx
    if New{m,2}<New{h,2}
        m=h;
    end
end
if m==i+fx
    New{i+fx,width(New)}=1;
    message=sprintf('%d is the peak',i);
    mis=0;
else
    New{m,width(New)}=1;
    if abs(m-i-fx)>=10
        mis=1;
    else
    message=sprintf('%d is higher than %d',m-fx,i);
    mis=0;
    end
end
```

```
h=m+1;
while New{h,width(New)}==0
   h=h+1;
end
bottom=0;
if Artefacts{1,1}<=h
   if h==size(Artefacts,1)
      bottom=1;
      break
   else
      for ha=1:size(Artefacts,1)
         if fx+i>=Artefacts{ha,1}
         else
            break
         end
      end
      Artefacts(1:ha-1,:)=[];
   end
end
open ECG;
if mis==1
   error(sprintf('%d is higher than %d',m-fx,i));
end
figure=1;
end
if bottom==1
   questdlg('All artefacts have been treated. Edit artefacts?','All has
finished','Yes','Yes','No','No');
   answer=ans;
   if sum(char(answer))==sum('Yes')
      Artefacts(1,:)=[];
      EditArtefacts
   elseif sum(char(answer))==sum('No')
   end
else
questdlg(message,'Next or Add?','Next Artefact','Keep Correcting','Keep
Correcting');
answer=ans;
if sum(char(answer))==sum('Next Artefact')
   ECG
elseif sum(char(answer))==sum('Keep Correcting')
   open GiveOne
```

```
end
end
else
    error('Enter all values');
end
end

[CalculateHRs]
ht=1;%1m
nt=0;
while 0<1
    if New{ht,6}==1
        nt=nt+1;
        if nt==2
            break
        end
    end
    ht=ht+1;
end
h0=ht+1;
h1=height(New);
hi=ht;
hn=1;
for h=h0:h1
    if New{h,8}==1,
        %Duration
        d=New{h,1}-New{1,1};
        dh=fix(d*24);
        dm=fix((d*24-dh)*60);
        ds=round(((d*24-dh)*60-dm)*60,5);
        HR{hn,1}=dm;
        HR{hn,2}=ds;
        %Interval
        i=New{h,3}*3600-New{hi,3}*3600+New{h-1,4}*60-
New{hi,4}*60+New{h,5}-New{hi,5};
        HR{hn,3}=60/i;
        hi=h;
        hn=hn+1;
    end
end
open HR;
if custreat==0
```

```
Time={'B1',6,0;'B2',10,30;'T1',15,0;'T2',19,30;'T3',24,0;'B3',28,30;'B4',33,0;'T4',37,
30};
elseif custreat==1
   Train
   treatname=listdlg('ListString',{'B1','B2','T1','T2','T3','B3','B4','T4'});
Time0={'B1',6,0;'B2',10,30;'T1',15,0;'T2',19,30;'T3',24,0;'B3',28,30;'B4',33,0;'T4',3
7,30};
   for tr=1:size(treatname,2)
      trh=1;
      while 0<1
         if sum(char(Time0{treatname(tr),1}))==sum(char(Time0{trh,1}))
            Time(tr,1)=Time0(trh,1);
            Time(tr,2:3)=Time0(tr,2:3);
            break
         end
      trh=trh+1;
      end
      Time(size(treatname,2)+1:size(Time,1),:)=[];
   end
end
open Time;
AveHR{1,1}=Time{1,1};
disp('CalculateHRs');

[ListTimes]
hs=size(AveHR);
if hs(1)>0
   AveHR(:)=[];
end
ti=5;%Interval
kn=5;%Segments
if ti*kn>=30
   error('Cannot exceed 30 seconds');
end
h0=1;
h2=(kn+1)*treatment;
i1=1;
k=1;
for h=h0:h2
   if i1==1
      i='B';
```

```
        AveHR{h,1}=Time{k,1};
        if Time{k,3}==0
            AveHR{h,3}=Time{k,2}-1;
        else
            AveHR{h,3}=Time{k,2};
        end
        if Time{k,3}==0
            AveHR{h,4}=60-ti;
            AveHR{h,6}=0;
        else
            AveHR{h,4}=Time{k,3}-ti;
            AveHR{h,6}=Time{k,3};
        end
    else
        i=(i1-1)*ti;
        AveHR{h,3}=Time{k,2};
        AveHR{h,4}=Time{k,3}+ti*(i1-2);
        AveHR{h,6}=Time{k,3}+ti*(i1-1);
    end
    AveHR{h,2}=i;
    AveHR{h,5}=Time{k,2};
    if i1==kn+1
        i1=1;
        k=k+1;
    else
        i1=i1+1;
    end
end
open AveHR;
disp('ListTimes');

[MeanHRs]
ai=0;%Adjust
h0=1;
hs=size(AveHR);
h2=hs(1);
n=1;
for h=h0:h2
    j11=AveHR{h,3};
    j12=AveHR{h,4};
    j21=AveHR{h,5};
    j22=AveHR{h,6};
```

```
   while 0<1
      if HR{n,1}==j11
         if HR{n,2}>j12
            t0=n;
            break
         end
      end
      n=n+1;
   end
   while 0<1
      if HR{n,1}==j21
         if HR{n,2}>j22
            t2=n;
            break
         end
      end
      n=n+1;
   end
   H=0;
   for j=t0+ai:t2+ai-1
      H=H+HR{j,3};
   end
   AveHR{h,7}=H/(t2-t0);
end
disp('MeanHRs');

[CalculateDif]
h0=1;
hs=size(AveHR);
h2=hs(1);
i=h2/treatment;
i1=1;
for h=h0:h2
   if i1==1
   else
      AveHR{h,8}=AveHR{h,7}-AveHR{h-i1+1,7};
   end
   if i1==i
      i1=1;
   else
      i1=i1+1;
   end
```

```
end
ns=size(AveHR);
h=1;
hs=size(AveHR);
h2=hs(1);
i=kn+1;
i1=1;
while h<h2
   if i1==1
      ht=h;
      i1=2;
   else
      nt=h+1;
      for n=h+1:h+i-2
         if AveHR{nt,8}>AveHR{n+1,8}
         else
            nt=n+1;
         end
      end
      AveHR{ht,8}=nt-ht;
      h=h+i;
      i1=1;
   end
end
disp('CalculateDif');

[ListResults]
hs=size(AveHR);
if hs(2)==9
   AveHR(:,9)=[];
end
h2=hs(1);
i=h2/treatment;
for n=1:treatment
   AveHR{2*n-1,9}=round(AveHR{(n-1)*i+1,7},2);
   AveHR{2*n,9}=round(AveHR{(n-1)*i+AveHR{(n-1)*i+1,8}+1,7},4);
end
disp('ListResults');
close all;

[ListHRV]
if size(AveHR,1)>0
```

```
   AveHR(:)=[];
end
treatment=8;
h0=1;
h2=treatment;
k=1;
for h=h0:h2
      if Time{k,3}==0
         AveHR{h,2}='V';
         AveHR{h,3}=Time{k,2}-1;
         AveHR{h,4}=30;
         AveHR{h,5}=Time{k,2};
         AveHR{h,6}=30;
      else
         AveHR{h,2}='V';
         AveHR{h,3}=Time{k,2};
         AveHR{h,4}=0;
         AveHR{h,5}=Time{k,2}+1;
         AveHR{h,6}=0;
      end
      k=k+1;
end
h0=1;
hs=size(AveHR);
h2=hs(1);
n=1;
for h=h0:h2
   j11=AveHR{h,3};
   j12=AveHR{h,4};
   j21=AveHR{h,5};
   j22=AveHR{h,6};
   while 0<1
      if HR{n,1}==j11
         if HR{n,2}>j12
            t0=n;
            break
         end
      end
      n=n+1;
   end
   while 0<1
      if HR{n,1}==j21
```

```
            if HR{n,2}>j22
                t2=n;
                break
            end
        end
        n=n+1;
    end
    L=HR{t0,3};
    H=HR{t0,3};
    for j=t0:t2-1
        if HR{j,3}<L
            L=HR{j,3};
        elseif HR{j,3}>H
            H=HR{j,3};
        end
    end
    AveHR{h,7}=H/L;
end
e=0;
for h=1:2*treatment
    if e==0
        HRV2{h,participant}=0;
        e=1;
    elseif e==1
        HRV2{h,participant}=AveHR{h/2,7};
        e=0;
    end
end
participant=participant+1;
clearvars -except HRV2 participant

[T1124]
%T1224(T1)
%out{1,1:3}
function out=T1224(in)
th=fix(in*24);
tm=fix((in*24-th)*60);
ts=round(((in*24-th)*60-tm)*60,5);
if ts==60
    tm=tm+1;
    ts=00;
end
```

```
if tm>=60
   th=th+1;
   tm=tm-60;
end
out={th tm ts};
```

```
[T2421]
%T2421(h,m,s)
function out=T2421(th,tm,ts)
out=round(th/24,10)+round(tm/(24*60),10)+round(ts/(24*3600),10);
```

```
[Train]
load train
sound(y,Fs)
```

```
[Chirp]
load chirp
sound(y,Fs)
```

Appendix I

MATLAB Programming Codes for GSR Analyses

```
[Initialize]
time=inputdlg({'HH','MM','SS'},'Start Time');
if exist('format1','var')==0
   format1=0;
end
if exist('format2','var')==0
   format2=0;
end
if isempty(time)==0
if isempty(str2num(time{1,1}))==0
   if isempty(str2num(time{2,1}))==0
      if isempty(str2num(time{3,1}))==0
thh=str2num(time{1,1});
tmm=str2num(time{2,1});
tss=str2num(time{3,1});
      else
         thh=str2num(time{1,1});
         tmm=str2num(time{2,1});
         tss=00;
      end
t0=T2421(thh,tmm,tss);
if format1==0
t2h=hour(New{height(New),1});
t2m=minute(New{height(New),1});
t2s=second(New{height(New),1});
end
tl=T1224(T2421(t2h,t2m,t2s)-t0);
tlh=tl{1,1};
tlm=tl{1,2};
tls=tl{1,3};
if tlm<38
questdlg({'Do you want to modify the length of the session?',sprintf('The total
length is %d hour %d minutes %.2f second',tlh,tlm,tls)},'Session
Length','Yes','No','No');
      answer=ans;
      if sum(char(answer))==sum('Yes')
         length=inputdlg({'HH','MM','SS'},'Length');
         if isempty(str2num(length{1,1}))==0
            if isempty(str2num(length{2,1}))==0
               if isempty(str2num(length{3,1}))==0
lhh=str2num(length{1,1});
lmm=str2num(length{2,1});
```

```
lss=str2num(length{3,1});
                else
                    lhh=str2num(length{1,1});
                    lmm=str2num(length{2,1});
                    lss=1;
                end
            else
                lhh=str2num(length{1,1});
                lmm=38;
                lss=1;
            end
        else
            lhh=0;
            lmm=38;
            lss=01;
        end
    elseif sum(char(answer))==sum('No')
        lhh=0;
        lmm=38;
        lss=01;
    elseif sum(char(answer))==sum('')
        error
    end
else
    lhh=0;
    lmm=38;
    lss=01;
    custreat=0;
end
t2=t0+T2421(lhh,lmm,lss);
if t2>T2421(t2h,t2m,t2s)
    error('The selected length exceeds the length of the data.')
end
if format1==0
for i=1:width(New)
    if strcmp(New.Properties.VariableNames{1,i},'TimeHHmmss000')==1
        it=i;
        break
    end
end
for i=1:width(New)
    if strcmp(New.Properties.VariableNames{1,i},'microSiemens')==1
```

```
      ie=i;
      break
   end
end
if ie<width(New)
   New(:,ie+1:width(New))=[];
end
if ie>it+1
   New(:,it+1:ie-1)=[];
end
if it>1
   New(:,1:it-1)=[];
end
New{:,3}=hour(New{:,1});
New{:,4}=minute(New{:,1});
New{:,5}=second(New{:,1});
New{:,1}=New{:,1}-fix(datenum(New{1,1}));
New=addvars(New,datenum(New{:,1}),'Before',1);
New.Properties.VariableNames{1}='Time';
New(:,2)=[];
format1=1;
end
if format2==0
h0=1;
h2=height(New);
n=1;
while 2^n<h2
   n=n+1;
end
n2=n+1;
h1a=h0;
h1b=h2;
for n=1:n2
   h1=h1a+fix((h1b-h1a)/2);
   if New{h1,1}>=t0
      if New{h1-1,1}<t0
         h11=h1-1;
         break;
         disp(1);
      else
         h1b=h1;
      end
```

```
    else
        h1a=h1;
    end
end
n=1;
while 2^n<h2-h1
    n=n+1;
end
n2=n+1;
h1a=h1;
h1b=h2;
for n=1:n2
    h1=h1a+fix((h1b-h1a)/2);
    if New{h1,1}>=t2
        if New{h1-1,1}<t2
            h12=h1;
            break;
        else
            h1b=h1;
        end
    else
        h1a=h1;
    end
end
New(h12:h2,:)=[];
New(h0:h11,:)=[];
format2=1;
end
New.Properties.VariableNames{2}='GSR';
New.Properties.VariableNames{3}='Hour';
New.Properties.VariableNames{4}='Minute';
New.Properties.VariableNames{5}='Second';
open New;
ListTimes
MeanGSR
CalculateDif
ListResults
    end
end
end
```

[ListTimes]

```
if tlm<38
treatnum=questdlg('Do you want to change the number of treatments?','Treatment
Number','Yes','No','No');
if sum(char(treatnum))==sum(char('No'))
   custreat=0;
elseif sum(char(treatnum))==sum(char('Yes'))
   custreat=1;
end
end
if custreat==0

Time={'B1',6,0;'B2',10,30;'T1',15,0;'T2',19,30;'T3',24,0;'B3',28,30;'B4',33,0;'T4',37,
30};
elseif custreat==1
   Train
   treatname=listdlg('ListString',{'B1','B2','T1','T2','T3','B3','B4','T4'});

Time0={'B1',6,0;'B2',10,30;'T1',15,0;'T2',19,30;'T3',24,0;'B3',28,30;'B4',33,0;'T4',3
7,30};
   for tr=1:size(treatname,2)
      trh=1;
      while 0<1
         if sum(char(Time0{treatname(tr),1}))==sum(char(Time0{trh,1}))
            Time(tr,1)=Time0(trh,1);
            Time(tr,2:3)=Time0(tr,2:3);
            break
         end
      trh=trh+1;
      end
      Time(size(treatname,2)+1:size(Time,1),:)=[];
   end
end
treatment=size(Time,1);
if exist('AveGSR','var')==1
   removeave=questdlg('Do you want to remove AveGSR?','Remove
AveGSR','Yes','No','No');
if sum(char(treatnum))==sum(char('No'))
elseif sum(char(treatnum))==sum(char('Yes'))
   AveGSR=[];;
end
end
open Time;
```

```
ti=5;%Interval
kn=5;%Segments
if ti*kn>=30
   error('Cannot exceed 30 seconds');
end
h0=1;
h2=(kn+1)*treatment;
i1=1;
k=1;
for h=h0:h2
   if i1==1
      i='B';
      AveGSR{h,1}=Time{k,1};
      if Time{k,3}==0
         AveGSR{h,3}=Time{k,2}-1;
      else
         AveGSR{h,3}=Time{k,2};
      end
      if Time{k,3}==0
         AveGSR{h,4}=60-ti;
         AveGSR{h,6}=0;
      else
         AveGSR{h,4}=Time{k,3}-ti;
         AveGSR{h,6}=Time{k,3};
      end
   else
      i=(i1-1)*ti;
      AveGSR{h,3}=Time{k,2};
      AveGSR{h,4}=Time{k,3}+ti*(i1-2);
      AveGSR{h,6}=Time{k,3}+ti*(i1-1);
   end
   AveGSR{h,2}=i;
   AveGSR{h,5}=Time{k,2};
   if i1==kn+1
      i1=1;
      k=k+1;
   else
      i1=i1+1;
   end
end
open AveGSR;
```

```
[MeanGSR]
ai=0;%Adjust
h0=1;
h2=size(AveGSR,1);
n=1;
for i=1:size(AveGSR,1)
h0=1;
h2=height(New);
n=1;
while 2^n<h2
   n=n+1;
end
n2=n+1;
t0=T2421(0,AveGSR{i,3},AveGSR{i,4})+T2421(thh,tmm,tss);
h1a=h0;
h1b=h2;
for n=1:n2
   h1=h1a+fix((h1b-h1a)/2);
   if New{h1,1}>=t0
      if New{h1-1,1}<t0
         h11=h1;
         break;
      else
         h1b=h1;
      end
   else
      h1a=h1;
   end
end
t2=T2421(0,AveGSR{i,5},AveGSR{i,6})+T2421(thh,tmm,tss);
h1a=h0;
h1b=h2;
for n=1:n2
   h1=h1a+fix((h1b-h1a)/2);
   if New{h1,1}>=t2
      if New{h1-1,1}<t2
         h12=h1-1;
         break;
      else
         h1b=h1;
      end
   else
```

```
        h1a=h1;
    end
end
    H=0;
    for j=h11:h12
        H=H+New{j,2};
    end
    AveGSR{i,7}=H/((t2-t0)*24*3600);
end

[CalculateDif]
h0=1;
h2=size(AveGSR,1);
i=h2/treatment;
i1=1;
for h=h0:h2
    if i1==1
    else
        AveGSR{h,8}=AveGSR{h,7}-AveGSR{h-i1+1,7};
    end
    if i1==i
        i1=1;
    else
        i1=i1+1;
    end
end
h=1;
i=kn+1;
i1=1;
while h<h2
    if i1==1
        ht=h;
        i1=2;
    else
        nt=h+1;
        for n=h+1:h+i-2
            if AveGSR{nt,8}>AveGSR{n+1,8}
            else
                nt=n+1;
            end
        end
        AveGSR{ht,8}=nt-ht;
```

```
        h=h+i;
        i1=1;
    end
end

[ListResults]
if size(AveGSR,2)==9
    AveGSR(:,9)=[];
end
i=size(AveGSR,1)/treatment;
for n=1:treatment
    AveGSR{2*n-1,9}=round(AveGSR{(n-1)*i+1,7}*1000,2);
    AveGSR{2*n,9}=round(AveGSR{(n-1)*i+AveGSR{(n-
1)*i+1,8}+1,7}*1000,2);
end
close all;

[Train]
load train
sound(y,Fs)

[Chirp]
load chirp
sound(y,Fs)

[T1224]
%T1224(T1)
%out{1,1:3}
function out=T1224(in)
th=fix(in*24);
tm=fix((in*24-th)*60);
ts=round(((in*24-th)*60-tm)*60,5);
if ts==60
    tm=tm+1;
    ts=00;
end
if tm>=60
    th=th+1;
    tm=tm-60;
end
out={th tm ts};
```

[T2421]

```
%T2421(h,m,s)
function out=T2421(th,tm,ts)
out=round(th/24,10)+round(tm/(24*60),10)+round(ts/(24*3600),10);
```

Appendix J

Normality Test Q-Q Plots

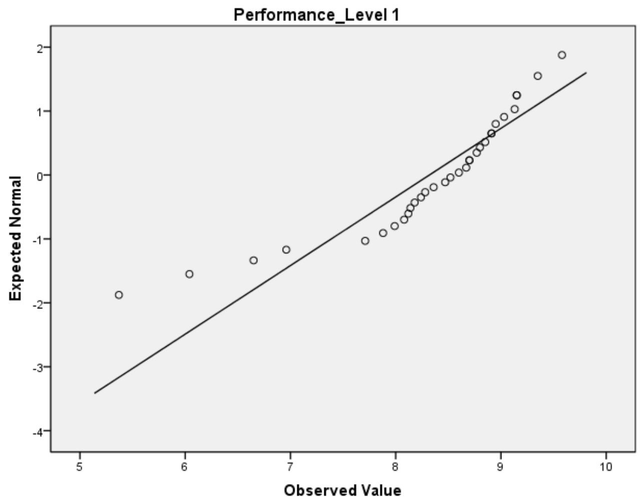

Figure J.1. **Performance score for intensity level 1 normality.**

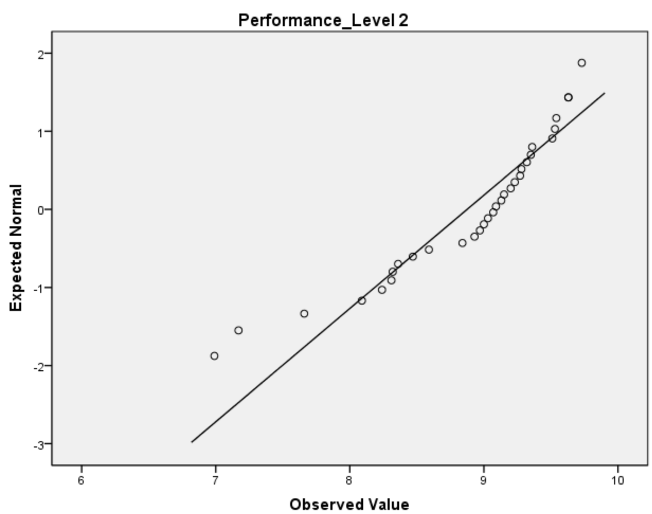

Figure J.2. **Performance scores for intensity level 2 normality.**

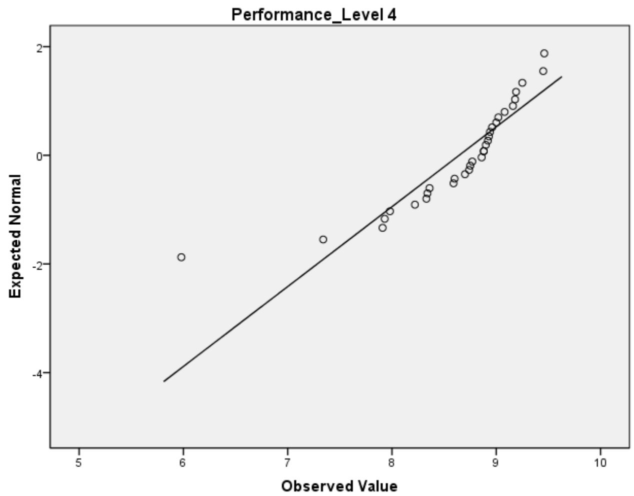

Figure J.3. **Performance scores for intensity level 4 normality.**

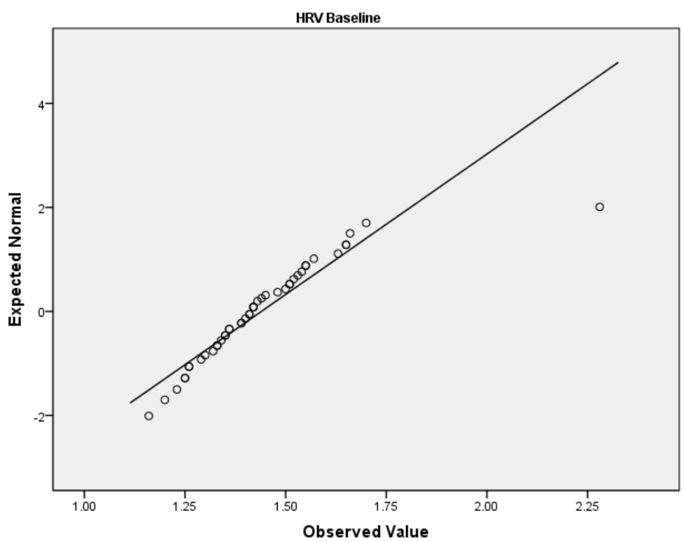

Figure J.4. **HRV baseline normality.**

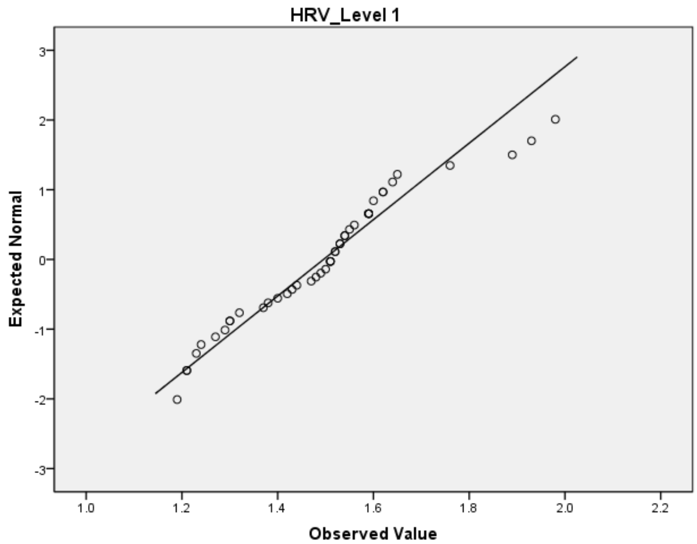

Figure J.5. **HRV baseline normality.**

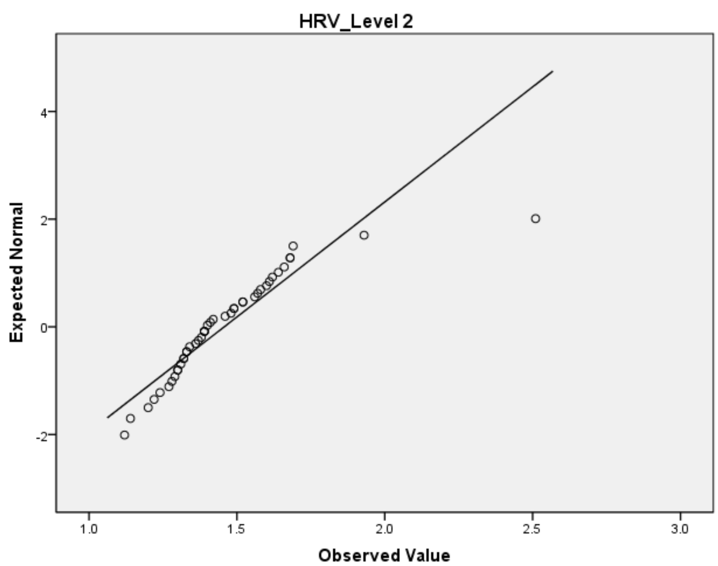

Figure J.6. **HRV baseline normality.**

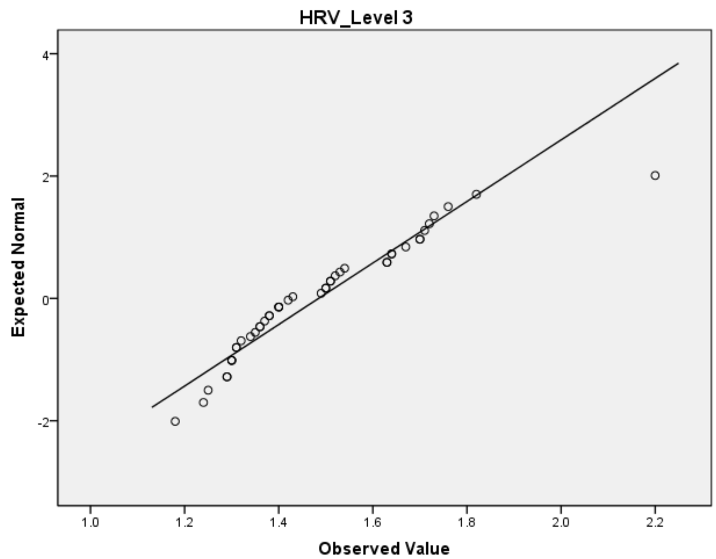

Figure J.7. **HRV baseline normality.**

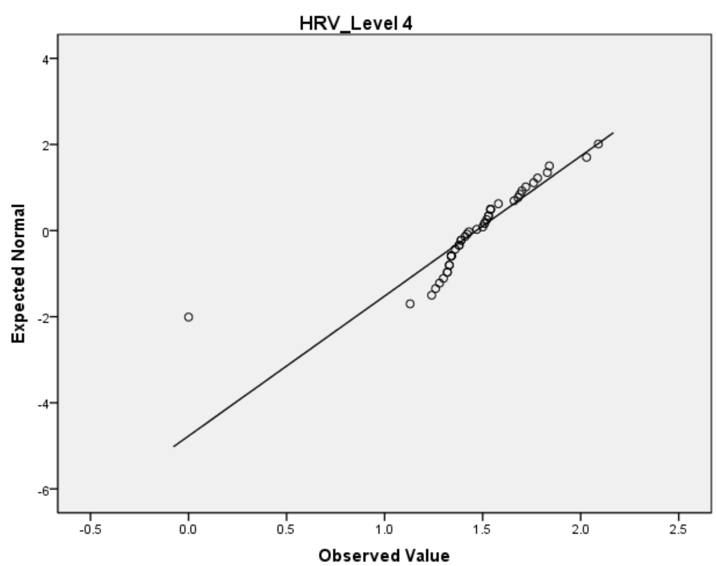

Figure J.8. **HRV baseline normality.**

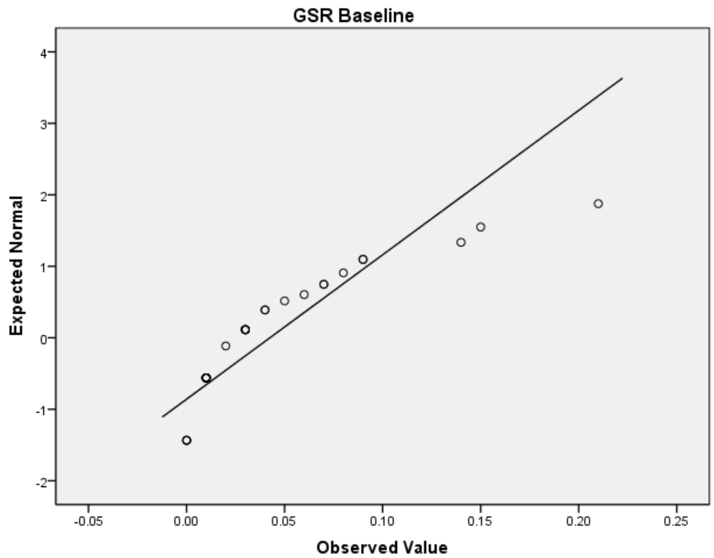

Figure 4.9. **GSR baseline normality.**

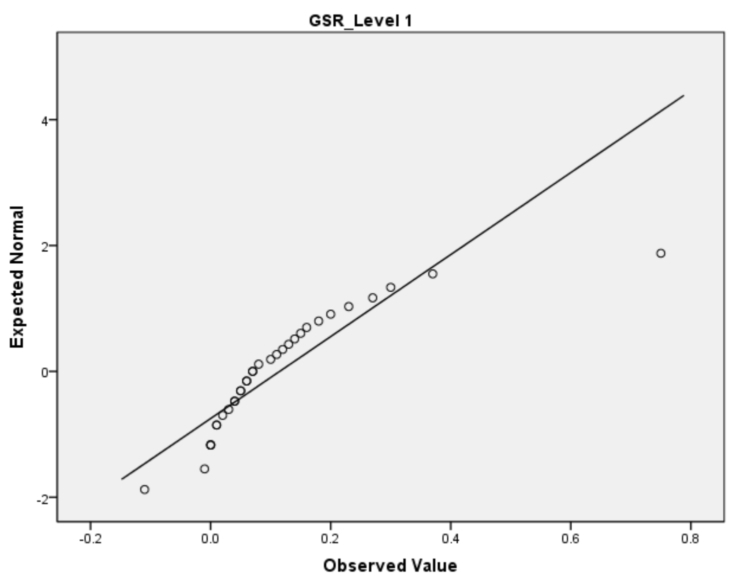

Figure J.10. **GSR level 1 normality.**

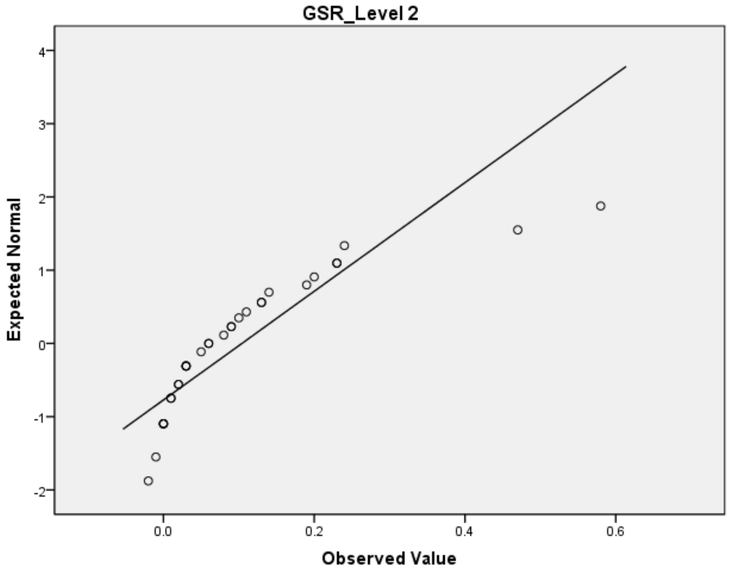

Figure J.11. **GSR level 2 normality.**

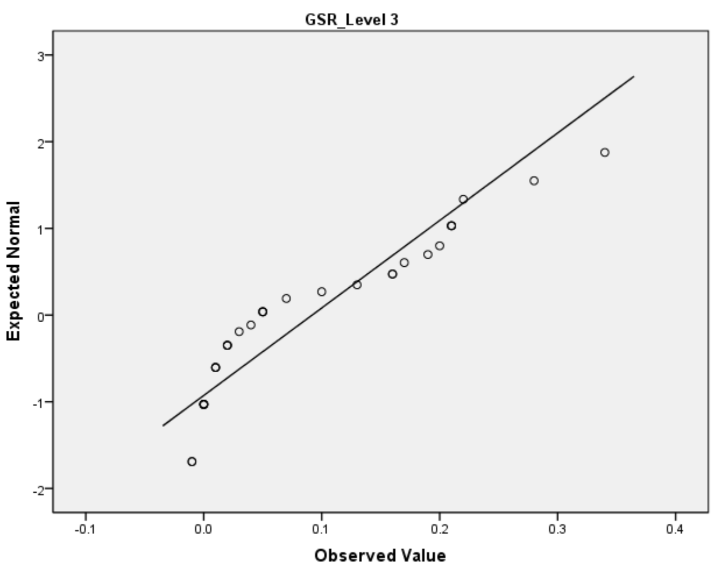

Figure J.12. **GSR level 3 normality.**

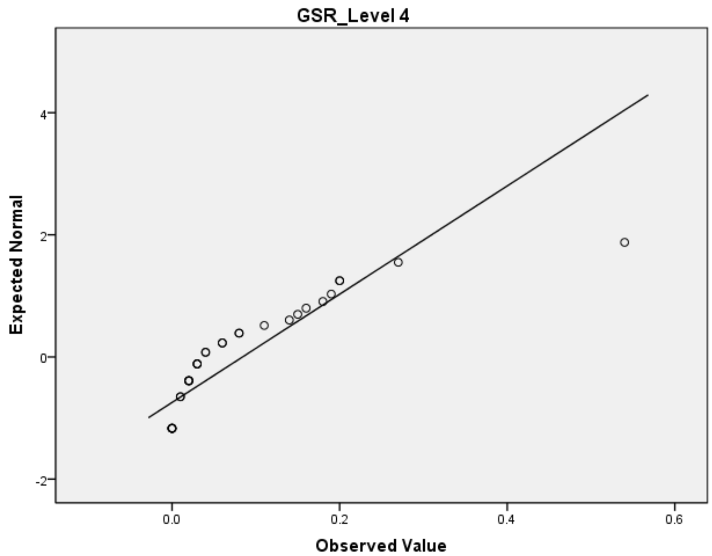

Figure J.13. **GSR level 4 normality.**

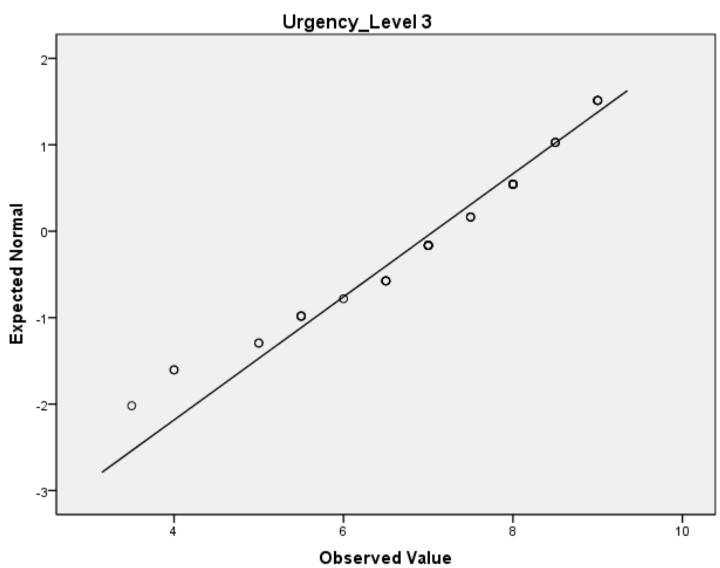

Figure J.14. **Perceived level of urgency intensity level 3.**

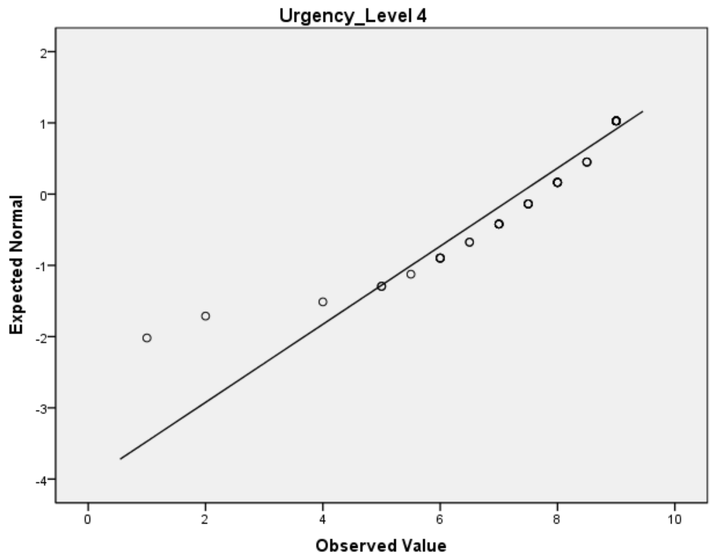

Figure J.15. **Perceived level of urgency intensity level 4.**

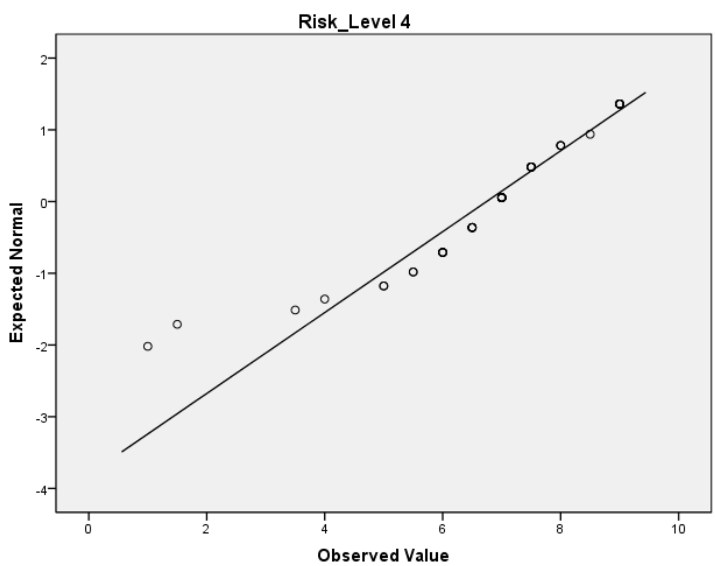

Figure J.16. **Perceived level of annoyance intensity level 4.**

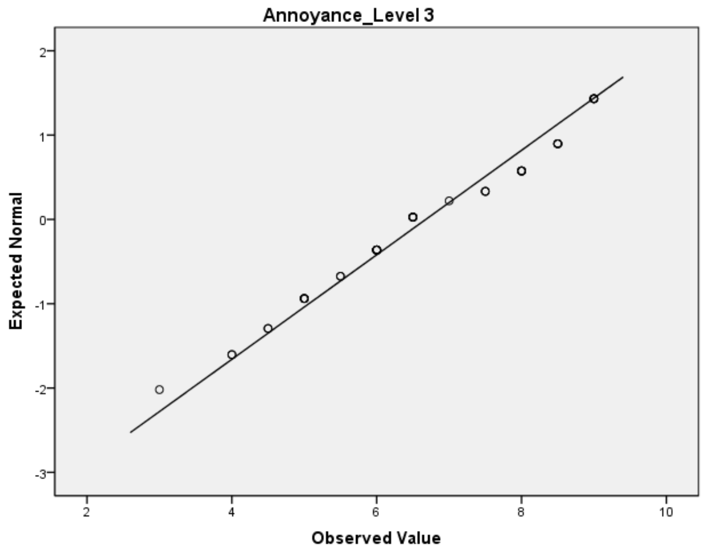

Figure J.17. **Annoyance intensity level 3.**

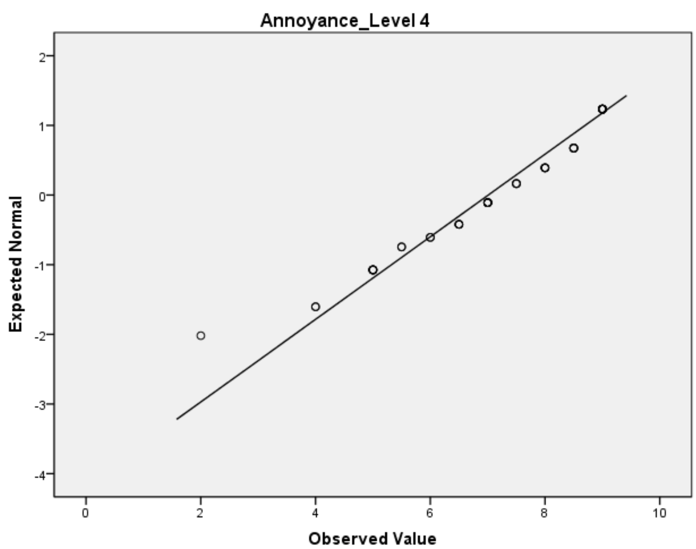

Figure J.18. **Annoyance intensity level 4.**

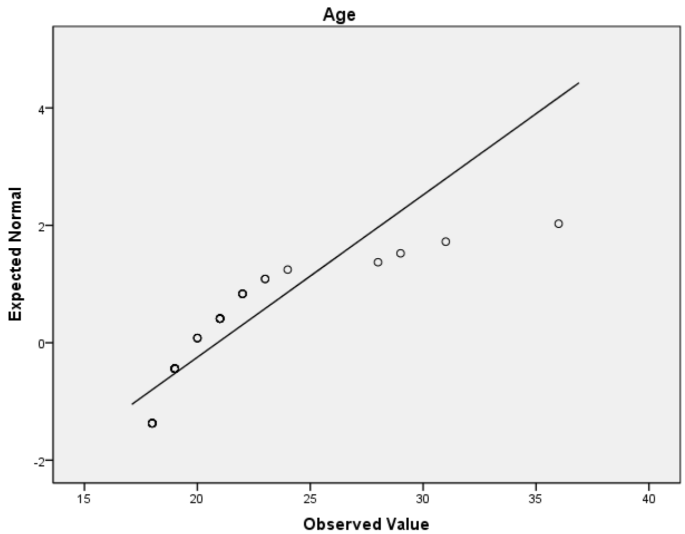

Figure J.19. **Age normality.**

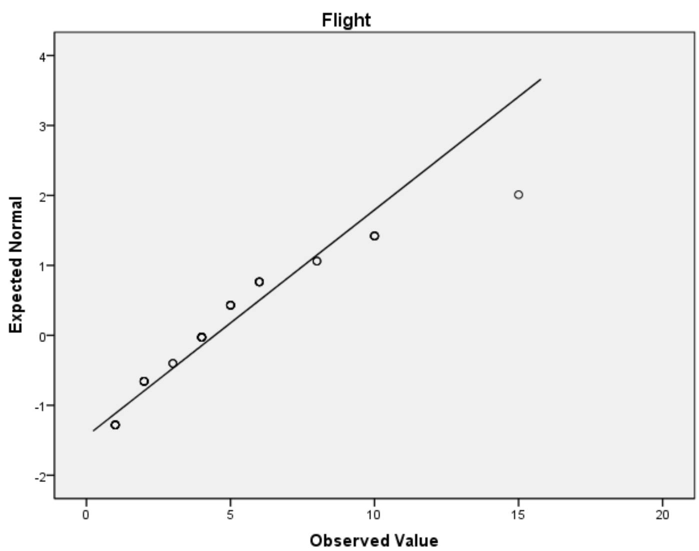

Figure J.20. **Flight normality.**

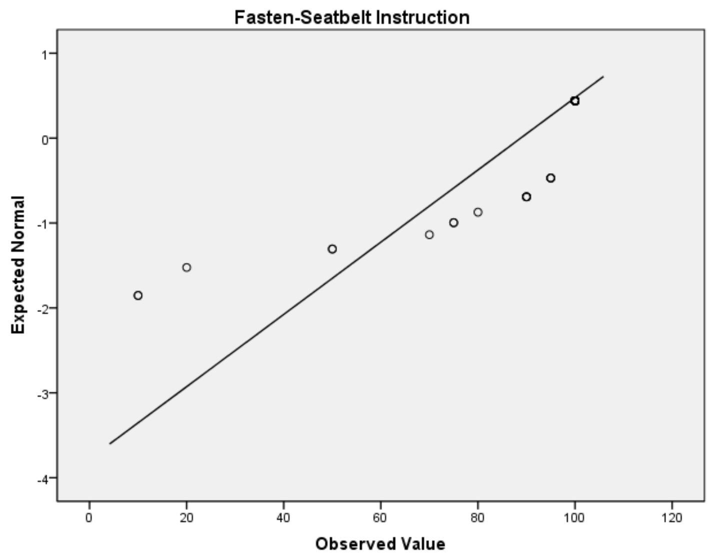

Figure J.21. **Fasten-seatbelt instruction normality.**

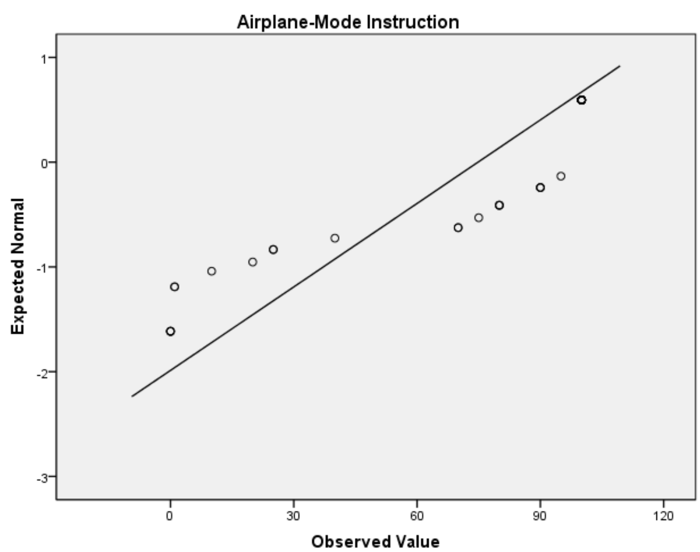

Figure J.22. **Airplane-mode instruction normality.**

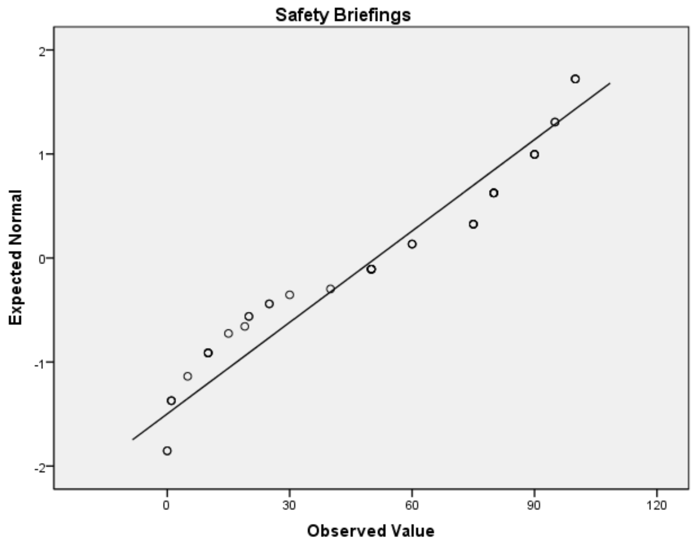

Figure J.23. **Safety briefing normality.**

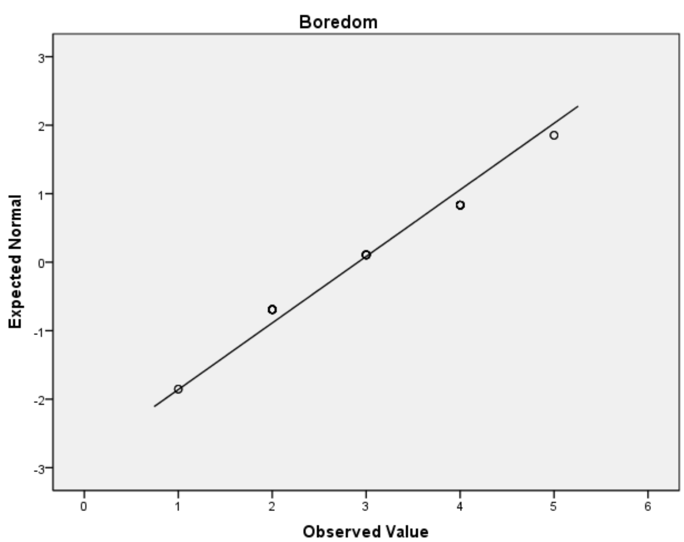

Figure J.24. **Boredom normality.**

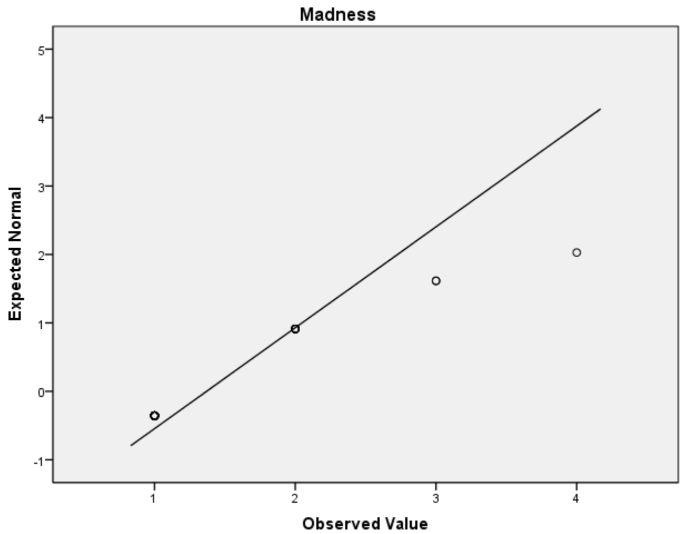

Figure J.25. **Anger normality.**

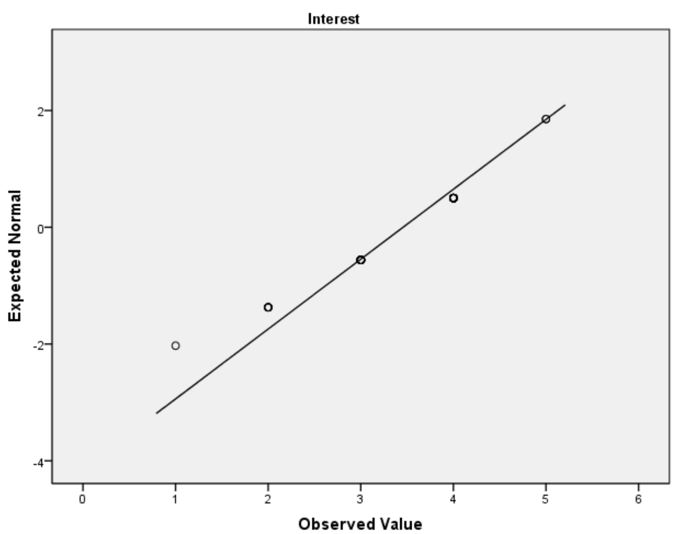

Figure J.26. **Interest normality.**