Project Report
ATC-243

The Effects of Compression-Induced Distortion of Graphical Weather Images on Pilot Perception, Acceptance, and Performance

A. T. Lind
A. Dershowitz
D. Chandra
S. R. Bussolari

21 November 1997

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS

Prepared for the Federal Aviation Administration,
Washington, D.C. 20591

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
ATC-243		

4. Title and Subtitle	5. Report Date
The Effects of Compression-Induced Distortion of Graphical Weather Images on Pilot Perception, Acceptance, and Performance	21 November 1997
	6. Performing Organization Code
7. Author(s) A.T. Lind, A. Dershowitz, D. Chandra, and S. Bussolari	8. Performing Organization Report No. ATC-243
9. Performing Organization Name and Address MIT Lincoln Laboratory 244 Wood Street Lexington, MA 02173-9108	10. Work Unit No. (TRAIS)
	11. Contract or Grant No. DTFA01-93-Z-02012
12. Sponsoring Agency Name and Address Department of Transportation Federal Aviation Administration Systems Research and Development Service Washington, DC 20591	13. Type of Report and Period Covered Project Report
	14. Sponsoring Agency Code

15. Supplementary Notes

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, under Air Force Contract F19628-95-C-0002.

16. Abstract

The Graphical Weather Service (GWS) is a data link application that will provide near-real-time graphical weather information to pilots in flight. To assess the effect of GWS, as well as to aid in the proper design, implementation and certification of the use of GWS in aircraft, two human factors studies have been conducted. The second study conducted (Phase Two) is the topic of this report. Phase Two was conducted to determine the maximum level of compression-induced distortion that would be acceptable for transmission of weather images to the cockpit. To make this determination the following data were collected and analyzed: pilot subjective ratings of the perceived amount of distortion of a compressed image, pilot subjective ratings of the acceptability of a compressed image for use in the flight task, and pilot route selections as a function of the amount of compression presented in an image. Results indicated that images of low to moderate compression levels were generally acceptable for transmission to the cockpit, while images that were highly compressed were generally unacceptable. In addition, computed measures of image quality have been identified to enable the establishment of a criteria for transmitting images to aircraft.

17. Key Words		18. Distribution Statement
data link	human factors	This document is available to the public through the National Technical Information Service, Springfield, VA 22161.
general aviation	weather graphics	
data compression	distortion	
pilot performance	cockpit display	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	106	

FORM DOT F 1700.7 (8-72) Reproduction of completed page authorized

EXECUTIVE SUMMARY

BACKGROUND

MIT Lincoln Laboratory, with the sponsorship of the Federal Aviation Administration, is developing a data link application, Graphical Weather Service (GWS), that will provide graphical weather information to the general aviation (GA) pilot in the cockpit. The initial GWS product is a composite RADAR precipitation graphic. The transmission of these complex images is made possible through application of image compression algorithms developed at MIT Lincoln Laboratory. These algorithms introduce image distortion.

To assess the effects of this distortion, as well as to aid in the proper design, implementation, and certification of use of the Graphical Weather Service (GWS) in aircraft, two human factors studies were conducted. The first study, Phase One, was documented in ATC Report-215: "The Influence of Data Link-Provided Graphical Weather on Pilot Decision Making" [1]. The results of that study demonstrated that GWS had a significant positive effect on pilot weather-related decision making. Given the fact that images have to be compressed to enable the timely transmission of images to the cockpit, Phase Two was conducted to determine the maximum level of compression-induced distortion that would be acceptable for the transmission of weather images to the cockpit. The images were compressed using a polygon-ellipse compression method, for precipitation data, developed at MIT Lincoln Laboratory. In this method each region of weather is approximated using a polygon or an ellipse. Parameters to describe these regular shapes require less data than the original images, hence the image is compressed [2]. The second study, Phase Two, is the topic of this report.

STUDY METHOD

Twenty volunteer instrument-rated pilots participated in the study. Subjects had a range of total flight time from 525 to more than 28,000 hours and a range of actual instrument time from 55 to more than 2,600 hours. The experimenters conducted the study in an office setting using custom software running on a Macintosh personal computer. All weather information and images displayed on the computer were constructed from actual recorded data provided by WSI Corporation.

The study tested the effect of various levels of compression of GWS images on pilot perception of distortion, opinion of acceptability, and route selection. The objectives of this phase were to determine: 1) what amount of compression is acceptable for transmission of images to an aircraft, and 2) whether there is a computational measure of image quality that can be used to predict the acceptability of images.

Subjective Ratings of Distortion and Acceptability

In the Distortion Rating and Acceptability Rating Tasks, the subject saw pairs of GWS weather images. Each pair contained an uncompressed image (original) and compressed image (altered version). In rating distortion, the subject judged the degree to which the compressed image had been distorted relative to the uncompressed image using a magnitude estimation technique [4].

In the Acceptability Rating Task, the subject judged the operational acceptability of the compressed image as a replacement for the uncompressed image in the context of its use in flight.

Route Selection Task

The subject saw a series of single GWS weather images presented in a random order. Each image was either an uncompressed image or a compressed image at high, moderate, or low compression. For each image the subject was asked to draw the best route of flight from one designated point, indicated as "A", to another designated point, "B". In addition to drawing the best route, the subject reported whether or not, in the context of a flight, the route would be attempted (Go, No Go), and rated the degree of hazard perceived in the depicted weather.

RESULTS

Several analyses were performed. First, ratings of distortion and acceptability were analyzed independently and then compared against one another. Next, computational measures of distortion were correlated with the subjective ratings. The effects of compression were then evaluated in the context of route selection. A statistic, called Normalized Route Difference, was devised to quantify the difference between two routes. This statistic enables a quantitative comparison of routes drawn for uncompressed weather images with those drawn for compressed weather images. Routes were also compared in terms of their proximity to different precipitation intensities.

Distortion and Acceptability Ratings

Results of the Distortion Rating Task indicated that subjects were in general agreement in their perception of the amount of distortion in the images. However, there was a large amount of between-subjects variance in the acceptability of compressed images. That is, subjects differed on how many of the most distorted images they were willing to call acceptable. While subjects found images of moderate compression to be acceptable, subject comments indicated the main objection to the highly compressed images was the lack of detail and the altered shape of the weather elements. When using the Polygon-Ellipse Algorithm, higher compression increases the use of ellipses to approximate weather regions. Most of the images with large ellipses or many ellipses were unacceptable. Overall, the subjects found all but two of the highly-compressed images to be unacceptable.

Several computed measures of image distortion were studied to determine the measure that best predicted pilot ratings (both distortion and acceptability). The best quantitative predictor of these ratings was a compression ratio defined by the number of bits in the undistorted image (when coded by a lossless run-length encoding) divided by the number of bits in the distorted images (when coded by the polygon-ellipse technique).

Route Selection Task

To determine whether there were changes in pilot performance as a function of distortion, the routes drawn were analyzed using the following measures: Normalize Route Difference, route length, and proximity to each level of precipitation. The area enclosed by two routes with the same end points is a function of how different the routes are. This area is then normalized by the average of the two route lengths between the departure and destination points; this is called the Normalized Route Difference. A Normalized Route Difference of zero means that the two routes are identical, while a large value indicates that the two routes are very different from each other.

An analysis of variance (ANOVA) was performed to determine the effect of image and compression level on Normalized Route Difference. Results of the ANOVA indicated that while there was some small significant variations in route difference, they were not found to be operationally significant.

Route length and proximity to each level of precipitation were assessed. Route length did not vary as a function of compression level. The nearest approach to each weather level was calculated for each route that was drawn. There was again a minor significant difference in only the proximity to Level 1 weather, however again this was not operationally significant.

CONCLUSIONS

Graphical weather images of low and moderate compression, as used in this experiment, were found to be generally acceptable by pilots. A computed measure of image quality has been identified that will enable the establishment of selection criteria for transmitting images to an aircraft. Pilot performance, as measured by the route selection task, was not significantly affected by low and moderate compression. High compression resulted in statistically significant, but operationally insignificant, differences in route selection and proximity to weak precipitation intensity. At very high image compression ratios, the Polygon-Ellipse algorithm represents areas of precipitation as ellipses, which the subjects generally found to be unacceptable. While the Polygon-Ellipse Algorithm preserves the fidelity of representation of precipitation intensity levels, the configuration of these levels were considered by subjects to be "too distorted", and to appear to be "unnatural", when a high degree of compression was applied. However, the subjects generally accepted the weather images compressed to a low or moderate degree. These findings have lead to the exploration of use of another compression algorithm that will provide a more faithful representation of the weather image under conditions of high compression.

ACKNOWLEDGMENTS

The authors are grateful to the twenty pilots who volunteered their time and effort to the study. Their interest in and dedication to the advancement of General Aviation, through the assessment of newly developed technology, made this study possible.

We are also grateful to the pilots of the MIT Lincoln Laboratory Flight Facility who helped test our procedures and measures.

We thank the Federal Aviation Administration for funding this research. Thanks are given to Rhonda Thomas (AND-720), Data Link Applications Program Manager, and Hugh McLaurin (AND-310), Data Link Integrated Product Team Leader, for their continued support.

We acknowledge the contributions of the following members of the staff of the Air Traffic Surveillance Group at MIT Lincoln Laboratory, including:

D. Jonathan Bernays, Data Link Project Leader, for his support,

Robert D. Grappel for technical advice,

Kenneth E. Stringham for software support, and

Katharine R. Krozel for graphics and technical editing.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

1.1 BACKGROUND

Among the most important information that affects the situational awareness of pilots of both transport category and general aviation (GA) aircraft is the location and severity of hazardous weather. The flight crews of commercial transport aircraft have a variety of on-board systems to assist them with maintaining awareness of potentially dangerous weather. Many of these aircraft are equipped with airborne weather radar, which detects hazardous weather ahead of the aircraft. Weather information and advisories are provided via VHF radio (voice and text datalink) by company airline dispatchers and staff meteorologists on the ground. In contrast with the airline crew, the GA pilot has much less information available, and has no second crew member to share the workload, nor any of the available supporting technology.

MIT Lincoln Laboratory, with the sponsorship of the Federal Aviation Administration (FAA), is developing a data link application that will provide graphical, as well as text, weather information to the GA pilot in the cockpit. The goal is to provide relevant and timely information at an affordable cost to the GA community.

To assess the effects, as well as to aid in the proper design, implementation, and certification of use of the Graphical Weather Service (GWS) in aircraft, two human factors studies were conducted. This report documents the findings of the second human factors study, Phase Two. The first study, Phase One, was documented in ATC Report-215: "The Influence of Data Link-Provided Graphical Weather on Pilot Decision Making" [1] and is summarized in Section 1.4.

1.2 GRAPHICAL WEATHER SERVICE: A DATA LINK APPLICATION

The first graphical weather product to be developed for GWS is a composite precipitation image derived from an array of ground-based weather radars. The radar composite is a commercial product provided by WSI Corporation and is a nationwide image of the six National Weather Service precipitation levels with a resolution of 2 kilometer x 2 kilometer (km). For this study, WSI provided images every 15 minutes that covered the New England region. The weather levels represent the intensity of the radar echoes from the precipitation, and are a function of the precipitation intensity.

The data link transmission of the raw precipitation image would require more bandwidth than is available with any practical data link implementation. However, the transmission of these complex images is accomplished through application of a compression algorithm developed at MIT Lincoln Laboratory. Figure 1-1 shows an uncompressed and compressed weather image. The Polygon-Ellipse algorithm [2] is based upon the underlying geometric structure of weather phenomena. Each weather region of a given weather level is approximated as either a polygon or an ellipse. The parameters to describe these regular shapes require less data than the original images, hence the image is compressed. The algorithm attempts to keep the correct area for each region. If it is necessary to distort the higher weather levels it will increase rather than decrease the size of any region, making the weather look more severe rather than less.

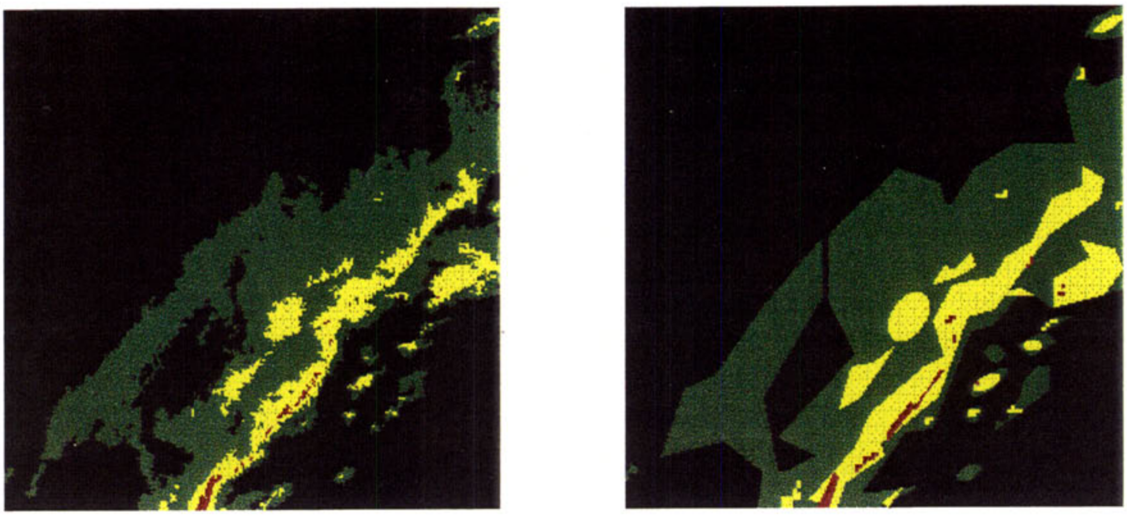


Figure 1-1. Uncompressed and compressed weather images. Without data compression, the 256x256 km image on the left would require 131,000 bits to transmit. The image on the right has been compressed to 2413 bits using the Polygon-Ellipse algorithms.

Figure 1-2 illustrates the components of GWS. To use GWS, the aircraft must be equipped with a data link "modem" such as a Mode S transponder or a VHF data radio that transmits and receives the data link messages. Polygon-Ellipse (Poly-Ell) was optimized to operate over Mode S, however compression is required for transmission of graphical weather over any type datalink, and Poly-Ell could be used with any other datalink system. An onboard Control and Display Unit is used by the pilot to request services and for the system to display information. It is estimated that the required avionics will cost approximately $5000 to $8000 [3]. To receive a GWS image, a data link request for a specific image is received from an aircraft; it is passed to a ground-based image compression processor; the processor selects the appropriate image area from a weather data base (based on location, time, and scale specified in the request); then, the processor compresses the image and encodes it for transmission to the requesting aircraft. The image is decoded on-board the airplane and displayed to the pilot.

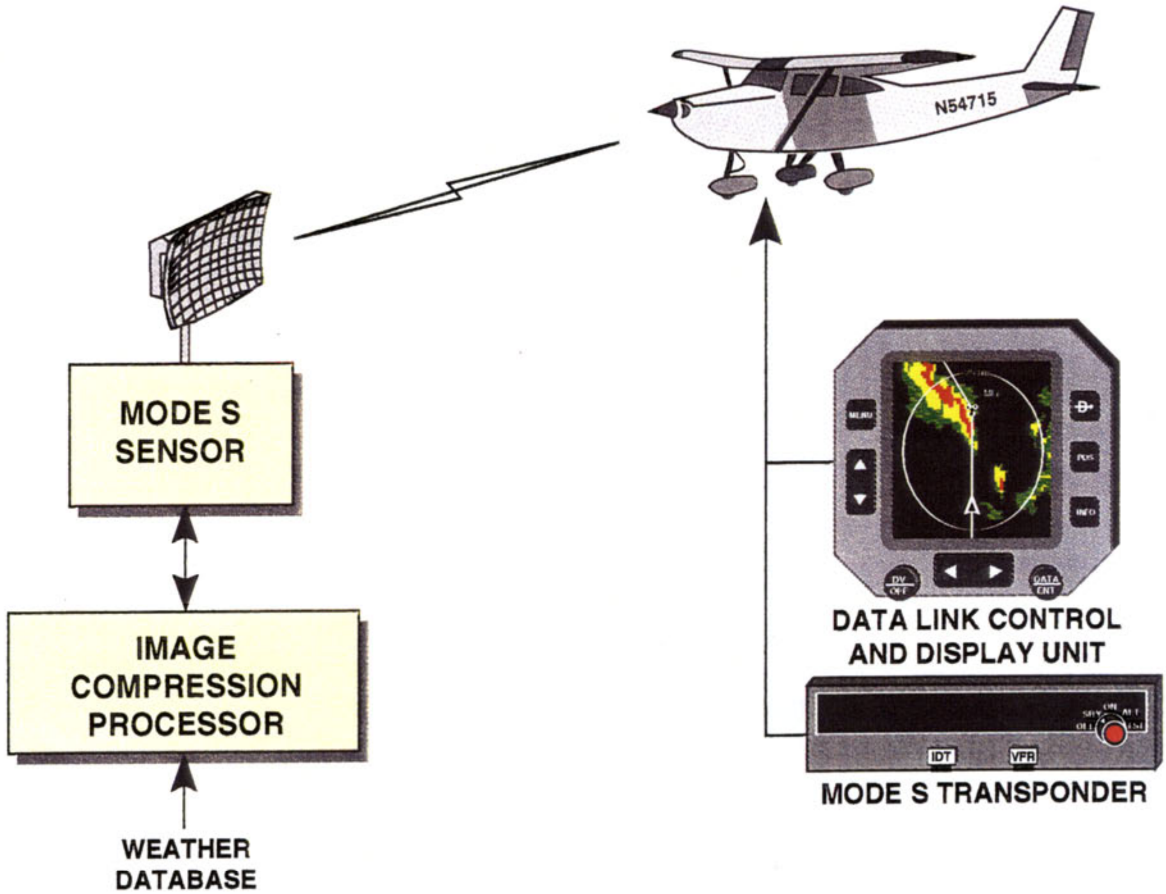


Figure 1-2. The components that make up the Graphical Weather Service.

1.3 PURPOSE OF THE PHASE ONE AND PHASE TWO STUDIES

The availability of near-real-time graphical weather information via data link will significantly affect pilot situational awareness and decision making. Phase One was conducted to assess the overall effect of GWS on pilot decision making. It was seen as a first step in validating the need for GWS and as a proof of concept. An overview of Phase One is provided below. Once Phase One findings validated that GWS is useful and effective, it was necessary to determine pilot response to the compressed images and to determine what amount of compression would be acceptable for transmission of images to an aircraft.

Since these complex images need to be compressed due to limited bandwidth, the resulting image may be somewhat altered from the original image. Therefore, the key issue in Phase Two was to determine how much distortion, associated with the compression, is considered acceptable for transmission of images to the aircraft and at what point is the level of compression no longer acceptable, both in terms of subjective and performance measures. Phase Two also addressed the issue of determining whether there is a computational measure of image quality that could be used to predict subjective acceptability of images.

1.4 PHASE ONE — OVERVIEW

Phase One tested the effect of GWS on decision making during hypothetical flights in challenging weather conditions. It was documented in ATC Report-215: "The Influence of Data Link-Provided Graphical Weather on Pilot Decision Making" [1]. Twenty volunteer instrument rated pilots participated in the study. Subjects had a minimum of 555 hours and a maximum of 28,000 hours of flight time, with a mean of 5,318 hours and a median of 2,925 hours. These subjects had a range of actual instrument hours from 35 to 2,700 hours, with a mean of 427 hours and a median of 170 hours.

Each subject participated in four hypothetical flights in an office setting. For each flight half of the subjects had access to GWS and half of the subjects did not. This design enabled the testing of the GWS versus No GWS Condition. Prior to each flight, the subject received a prepared flight plan, relevant navigational charts and weather briefing materials.

The subject was questioned at each of three decision points within the flight. The first decision point was at departure, prior to starting the aircraft engine. The second decision point was in the cruise portion of the flight, and the third was near the destination. Since the subject did not have the benefit of the sensory experience of flight, the experimenter told the subject what the pilot would be experiencing, e.g., ride quality, visibility, and precipitation. The subject was then asked what action he would take. The subject could respond immediately or could seek additional information using GWS (in the GWS Condition) or via queries to Air Traffic Control or Flight Watch (FW) (in the GWS and No GWS Condition). An experimenter, sitting in the room with the subject, played the role of ATC and FW personnel, using scripted responses.

For each decision point in which the subject had GWS, experimental images could be accessed for four locations (present position, departure, destination, alternate), at four different ranges (25, 50, 100, 200-nautical mile (nmi) radius). The route of flight was in the 200-nmi range.

The subject was asked to "think aloud" throughout the experimental session. Verbal request for information from ATC and FW, choices of GWS images, comments and action taken at each decision point were recorded. Actions taken included Go and No Go decisions and, decisions to deviate or to proceed on course. After selecting the action to be taken, the subject gave two ratings: a rating of confidence in his ability to assess the weather situation, given the information available, and a rating of the level of hazard presented.

Results indicated that GWS had a substantial positive effect on weather-related decision making. This was found for pilots with varying levels of instrument experience. Subject confidence in their own ability to assess the weather situational was markedly increased when GWS was used. Subjects with GWS made fewer requests for weather information to weather dissemination ground personnel, thus indicating a potential decrease in ground personnel workload. Subject comments indicated that GWS was found to be very useful and subjects were enthusiastic about receiving data link services in the GA cockpit in the future.

1.5 PHASE TWO — PURPOSE AND APPROACH

Phase Two was conducted to determine the effects of compression and to identify the level of compression at which an image is no longer useful for the flight task, and, therefore, should not be transmitted to an aircraft. The research questions asked were:

- As compression level is increased, is the pilot's perception of distortion affected, and if so, to what extent?
- How does distortion affect the operational acceptability of images as reflected by pilot acceptance and decision making supported by images?
- Is pilot performance affected by image distortion, and if so how much?
- Is there a computed measure of the quality of the compressed image that can be used as a good predictor of pilots' subjective ratings of distortion and acceptability? Does the measure identify a threshold value that reliably discriminates images that are acceptable to pilots from images that are unacceptable to pilots?

To obtain the data needed to answer the above questions, Phase Two was composed of three experimental tasks performed on a Macintosh computer: the Distortion Rating Task, the Acceptability Rating Task, and the Route Selection Task. In this section, the function of each task and a brief description of the task is given.

Task One, the Distortion Rating Task, was designed to measure the pilot's perception of distortion of compressed images, i.e., a perceptual judgment of the amount of distortion in an image. Data from this task are applied in answering the question: As compression level is increased, is the pilot's perception of distortion affected, and if so, to what extent? The subjects were presented with images in pairs, one uncompressed (original), and one compressed (altered version). They were then asked to determine the quantitative amount of distortion in the compressed image, as compared to the raw image. A numerical value was reported based on this distortion value.

Task Two, the Acceptability Rating Task, was designed to determine the usefulness of a compressed image for the flight task. Data from this task are applied in answering the question: How does compression affect the subjective acceptability of images? The subject again saw a series of pairs of GWS weather images. As in the Distortion Rating Task, each pair contained an uncompressed image and a compressed image. The subject was asked to judge the acceptability of the compressed image as a replacement for the uncompressed image in terms of the compressed image's functionality for the flight task, without explicit regard for image distortion.

The data from the Distortion Rating Task and the Acceptability Rating Task are applied in answering the questions: Is there a computed measure of the quality of the compressed image that can be used as a good predictor of pilots' subjective ratings of distortion and acceptability? Does the measure identify a threshold value that reliably discriminates images that are acceptable to pilots from images that are unacceptable to pilots?

Task Three, the Route Drawing Task provides behavioral data on changes in pilot performance as a function of the amount of compression. Data from this task are applied in answering the question: Is pilot performance affected by the level of compression and accompanying distortion in a weather image?

In the Route Drawing Task, the subject saw a series of single GWS weather images, presented with a North up orientation. Each image was either an uncompressed image or a compressed image (no designation of compression level was made to the subject). The subject was asked to draw a route of flight from one designated point to another designated point, indicated on the Macintosh screen as Points "A" and "B". The route was drawn by using the mouse and clicking. In addition to drawing the route, the subject answered two questions regarding willingness to go on the flight and the level of hazard in the depicted weather.

Use of these subjective and performance measures enabled the determination of the amount of distortion perceived, the usefulness of images compressed to various levels, and the effects of compression on pilot performance. In addition, data can be correlated to determine if what pilots said was unacceptable actually resulted in a measurable change in performance.

1.6 REPORT ORGANIZATION

Section 2 provides an overview description of the experimental design of Phase Two. Section 3 includes a detailed account of the methodology of the experiment. Section 4 describes the analyses performed and the results obtained. Section 5 contains conclusions.

2. EXPERIMENTAL DESIGN

2.1 INDEPENDENT VARIABLE

The independent variable studied was Level of Distortion. To assess the affect of the distortion introduced by the Polygon-Ellipse Algorithm, the subjects were shown both uncompressed or "raw" images, and compressed images. The images used in the experimental tasks were compressed to three different levels, considered to be "High, Moderate, and Low" Distortion. Figure 2-1 is an example of an image in the following states: uncompressed, and then High, Moderate, and Low distortion.

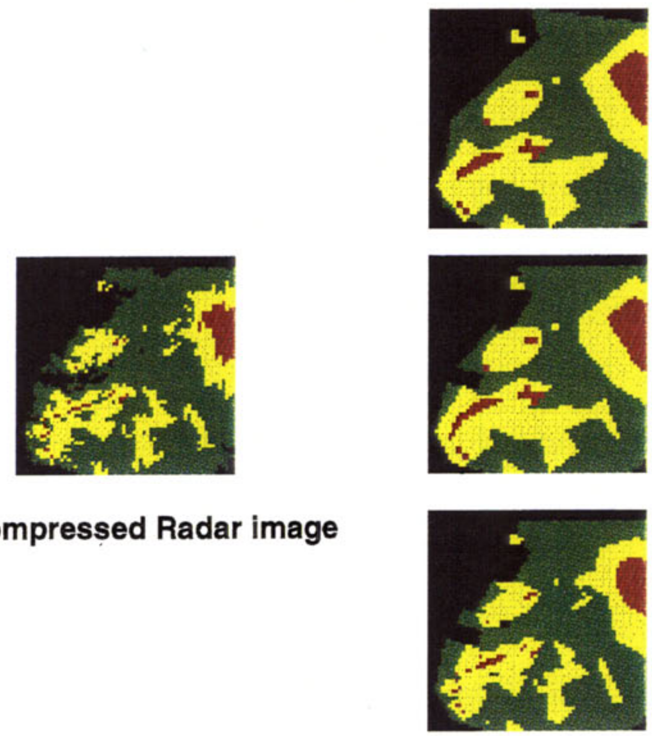


Figure 2-1. Sample image at different compression levels.

The Polygon-Ellipse algorithm replaces each region of a given level of precipitation intensity with a region defined as either a polygon or an ellipse. In order to compress the image, the regular shapes do not perfectly match with the original image. Before compressing an image, Poly-Ell is given a target number of bits that it attempts to compress the image into. The smaller the number of bits, the more the image is compressed, and the more it is distorted. Thus, the compression of Polygon-Ellipse was used to introduce distortion into the images, and the effect of this distortion is what was studied.

By having a range of distortion levels a range of reported distortion and acceptability ratings and a range of pilot behaviors in response to the route drawing task were expected. For a description of how the original (uncompressed) test images were selected and the process for compressing them to three levels, refer to Section 3.2.2—Stimuli. In that section is Figure 3-2 which shows the number of bits in each image compressed to each level of compression.

2.2 DEPENDENT VARIABLES

2.2.1 Subjective Ratings

Two types of subjective ratings were taken: Distortion Ratings and Acceptability Ratings. The Distortion Rating represents the subject's subjective assessment of the amount of distortion of an image. The Acceptability Rating represents the subject's subjective assessment of the usefulness/acceptability of the image for use in actual general aviation flight. These ratings are described below. The ratings were chosen for several reasons. They have face validity, and it is difficult to objectively measure something that is subjective, such as perceived distortion. One of the intents of this experiment was to come up with a measure of computed distortion that can be substituted for this subjective distortion measure.

2.2.1.1 Distortion Rating

The subject was shown an uncompressed image and a compressed image, side by side, and was asked to judge the degree to which the compressed image had been distorted relative to the uncompressed image. The rating was based on the quantitative amount of distortion of the compressed image. Images were presented in random order.

2.2.1.2 Acceptability Rating

As in the Distortion Rating Task, the subject was shown an uncompressed image and a compressed image, side by side. However, rather than rating distortion, the subject was asked to rate the acceptability of images for operational use in flight, regardless of the degree of image distortion. Images were presented in random order. A four point scale was used in rating acceptability, Very Poor (unacceptable), Poor (unacceptable), Good (acceptable) and Excellent (acceptable).

2.2.2 The Route Selection Task

To enable the collection of data on pilot performance as it is affected by level of compression, a Route Selection Task was devised. In this task, the subject was presented with a departure point, and a destination point shown on a single precipitation image, at a time. Each image was either uncompressed or distorted to high, moderate, or low level. No indication was given to the subject to denote the distortion level. Images were presented in random order. For each weather image with the same source undistorted image, the same departure point and destination point were presented. The subject was asked to select the best route from the departure point to the destination point.

Along with each image that was presented, two questions were asked of the subject. The subject was asked to make a Go/No Go decision. In other words, he was asked, "Will you go on this flight?" Yes or No? Even if the subject answered "No" they were still asked to choose a route that they would fly if forced. He was also asked, "How hazardous is the weather between A and B?" He selected a response from 1 (Not at all) to 5 (Very). The responses to these questions provide subjective comparisons of the effects of compression.

2.2.3 Survey Data

In addition to the measurements of the dependent variables defined earlier in this section, data were collected by use of surveys. The "Pilot Background Questionnaire" (see Appendix A) was completed by each subject prior to participation in the experiment. The responses provided information on the weather interpretation and flying experience of the pilot.

The "Post-Route Drawing Task Questionnaire" (see Appendix B) was completed by each subject following completion of the Route Drawing Task. the responses provided information on the pilot weather-related decisions and the routes selected.

The "Exit Interview" (see Appendix C) was completed by each subject after participation in the entire experiment, i.e., after all route selections and subjective ratings of distortion and acceptability were made. The responses provided pilot reaction to the procedures and measurements used in the experiment, as well as, additional information on the weather-related decisions of the pilot.

3. METHOD

3.1 SUBJECTS

Twenty instrument-rated pilots from the New England area participated in the study. The subjects had single engine and/or light twin engine experience. Sections 3.1.1 and 3.1.2 describe the recruitment process and provide background information on the flying experience of the subjects.

Subjects were required to be instrument-rated to participate. This criteria was set to enable the assessment of the effect of compression on a group of pilots who were quite knowledgeable and experienced in dealing with weather conditions. These pilots had a significant amount of actual instrument experience and, therefore, experience in judging actual weather situations, and using weather radar images.

3.1.1 Recruitment

Pilots were recruited as subjects for participation in the Phase One Study through use of an advertisement in the Atlantic Flyer, an aviation newspaper. The advertisement specified that instrument-rated pilots with single engine and/or light twin engine experience were needed as volunteer-subjects in the evaluation of a new air/ground data link service being developed by the Federal Aviation Administration (FAA). The GWS Program and plans for a series of studies were briefly described in the advertisement. The pilots who participated in Phase One were notified of Phase Two and asked to participate. The majority of the twenty pilots who participated in Phase One participated in Phase Two. However, several were not available and pilots were added from the potential-subject pool. This pool comprised pilots who originally responded to the advertisement, but were not selected for Phase One since the goal of twenty test participants was reached.

3.1.2 Subject Background

Each subject was sent a Pilot Background Questionnaire (see Appendix A) that he could complete at his leisure with time to refer to his log books in answering questions regarding flight hours. Each subject returned the questionnaire to the experimenters on the day of participation in the study.

Responses to the questionnaire provided information on the pilot and related flying experience, including pilot age, license held, ratings held, flight hours, level of familiarity with the New England Region, types of navigational or weather detection equipment that are in the aircraft usually flown by the pilot, any training the pilot may have had in weather interpretation, and how the pilot usually obtains the pre-flight weather briefing.

Subject age ranged from 27 to 63 years, with a mean of 42 years. All subjects were male. Several female pilots responded to the advertisement, but did not meet the criteria of 40 actual instrument hours set for subject selection.

The subjects had a wide range of licenses and ratings. There were four private pilots, seven commercial pilots, and nine Airline Transport Pilots (ATP). Sixteen of the twenty subjects had multi-engine ratings. Thirteen subjects were flight instructors. All of the subjects were rated in single engine airplanes; additionally, seven had helicopter ratings and four had glider ratings.

The subjects had a wide range of flying experience. For the purpose of calculating pilot experience, hours in single-engine aircraft and multi-engine aircraft were combined. The mean for single-engine and multi-engine combined was 3,773 hours, with a range of 270 to 28,000 hours and a median of 2,085 hours. For "total flight hours", combining single-engine aircraft and multi-engine aircraft experience does not adequately describe the experience of all subjects. For example, the subject with the least amount of hours shown above (270 hours) actually was a much more experienced pilot. Most of his experience was in rotorwing aircraft, reporting a total of 4,750 hours in that category. There were six other subjects who also had extensive experience in the rotorwing category. However, we do not have a count of their experience in this area since it was not an item on the background questionnaire.

The subjects had a range of total actual instrument hours from 55 to 2,600 hours, with a mean of 326 hours and a median of 123 hours. There was also a wide range of recent Instrument Flight Rules (IFR) experience.

The subject was asked what percentage of his IFR time had been single pilot IFR. Fifty percent of the subjects reported that 0% to 19% of their IFR time as being "single pilot IFR." Ten percent of subjects reported 21% to 39% of their IFR time as being "single pilot IFR". Forty percent of subjects reported 80% to 100% of their IFR time as "single pilot IFR".

To determine recent flying experience, the subject was asked about his flying experience in the past year, including the number of instrument approaches flown, the number of actual instrument hours, and the distance of their average IFR flight. Table 3-1 presents these results.

TABLE 3-1

Instrument Experience in the Past Year

Instrument Experience	Mean	Range
Number of Instrument Approaches	41	3 to 200
Actual Instrument Hours	18	0 to 60
Distance (nmi) of Average IFR Flight	160	50 to 500

The subject was asked to indicate the percentage of intended IFR flights that he canceled due to weather in the past year. The mean response was 5%, with a range of 0 to 20%. Among the weather conditions listed by subjects, as the cause of the cancellation, were thunderstorms (embedded or widespread), freezing rain, and icing.

The subject was asked how often he flies for each of several different reasons. In each case, the subject was asked to give an answer from 1 (Never) to 5 (Always). The primary reason for flying (indicated as an "Always" or "Usually" response) by 60% of the subjects was business. The primary reason for flying (indicated as an "Always" or "Usually" response) for 30% of the subjects was recreation. The remaining 20% of the subjects reported flying for a combination of reasons including both recreation and business. The total adds up to more than 100% because subjects were able to select "Always" or "Usually" for more than one reason.

The weather images used in the study were actual weather images of New England weather. All subjects were residents of New England and routinely fly in the area. Subjects rated how familiar they were with flying in the New England region on a scale of 1 (Not at All Familiar) to 5 (Very Familiar). All subjects indicated a rating of 4 or 5 (More Than Moderately Familiar or Very Familiar).

To determine how familiar the subject was with in-flight weather detection equipment, he was asked what weather equipment is on board the aircraft that they usually fly. Two subjects listed Stormscope, and five subjects listed weather radar.

The subject was asked if he had any weather training beyond basic pilot training. Eleven of the subjects said that they had additional weather training, including: college meteorology classes, military and airline training, and radar training courses.

To determine the way in which the subject gets his pre-flight weather briefings, he was asked to indicate a rating from 1 (Never) to 5 (Always) for each option. The number and percentage of people who answered "Usually" or "Always" were then calculated for each option. Please note that this can result in a total percentage of greater than 100. Table 3-2 lists the source of pilot pre-flight briefing and the number and percentage of pilots who usually or always use that source.

TABLE 3-2

Sources of Pilot Pre-Flight Briefings
"Usually" and "Always" Responses Combined

Source of Briefing	Number	Percentage
Over the Phone from FSS Personnel	14	70
In Person from FSS Personnel	0	0
DUAT	8	40
Other Computer Service	1	05
Facsimile Service (Weather Fax / Jepp Fax)	1	05
Other Service	4*	20

* These four subjects usually use the Weather Channel in addition to either DUAT or phoning an FSS.

3.2 FACILITIES, STIMULI, AND APPARATUS

3.2.1 Facilities

The study was conducted in an office at MIT Lincoln Laboratory in Lexington, MA. The equipment necessary to display the weather images was a Macintosh personal computer. The office contained chairs for the subject and the experimenters and a desk large enough to accommodate the Macintosh and study questionnaires.

3.2.2 Stimuli

3.2.2.1 Image Selection

All of the weather images used in the study were constructed from actual recorded data (NOWrad™ images), made available to us from WSI, a commercial vendor. Whenever "interesting" weather was forecast to move through the region, WSI was contacted to record the data. Periods of interest included times when IFR or thunderstorm activity was predicted in the New England region. A three-hour block of radar data was then recorded; one image every 15 minutes. These 15-minute images were quality controlled by a human at WSI before they were recorded. The images recorded were the standard radar product issued by WSI. The raw images had a 2-km resolution, were centered on Worcester, MA, and covered a region 760 by 480 km in extent. Each image was a composite of decluttered data from all of the National Weather Service (NWS) Weather Service Radar (WSR)-57/74 radars that were operating within the image region shown in Figure 3-1. The data are substantially less subject to attenuation and terrain blocking effects that degrade single site radar data.

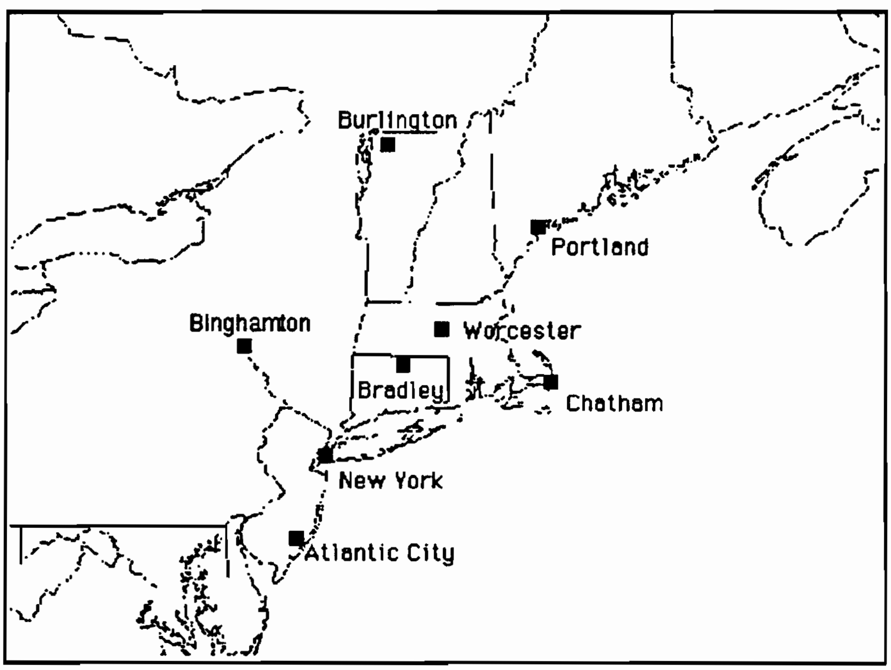


Figure 3-1. Region where weather radar images were stored and used for the experiment.

To describe the meaning of the three levels of distortion selected (low, moderate, and high), we begin with a brief description of the goal of the image-selection process and how the original (uncompressed) images for use in the study were selected. The goal of the image-selection process was to choose a series of images that would represent a range of distortion and provide a potential for a range of pilot decisions concerning route planning. It was intended that the pilots would be presented with images that would represent some easy and some difficult instrument flight tasks.

The weather images, selected from the recorded data, were chosen so that the pilots would have to make some decisions in the placement of a route from Point A (departure point) to Point B (destination). Each selected weather image represented enough weather so that the pilots would not simply draw a straight line from Point A to Point B. In each weather image, there were areas of Level 1 (weak precipitation intensity, depicted by green) and Level 2 (moderate precipitation intensity, depicted by yellow) in the region. In some of the weather images, Level 3 and above (strong to extreme precipitation intensity, depicted by red) were also present.

The area between Point A and Point B contained regions of precipitation, thus encouraging the subjects to alter this route around the weather. Additionally, in the more difficult cases, the weather images were selected so that, in the opinion of the experimenter (an instrument-rated flight instructor), a flight was possible, but problematic. Thus, the range of weather images and routes selected were from straightforward IFR flying to difficult and potentially dangerous flying that pilots might want to avoid in a light single-engine aircraft.

Once the weather images had been selected by the experimenter, they were compressed to different levels using the Polygon-Ellipse Algorithm. These levels of compression include: High (maximum compression), Low (minimum compression), and Moderate (approximately midway between maximum and minimum compression). The following process enabled the creation of the experimental stimuli, i.e., weather images at three levels of compression:

To create the high-compression images, the experimenter specified a target number of bits to which the algorithm then attempted to compress the image. There is a minimum number of bits required to depict each object in an image, therefore, the algorithm cannot compress the image below a certain value. If the algorithm was unable to compress the image to the specified level, the number of bits was automatically increased until compression was accomplished.

To create the low-compression images, each image was run through the algorithm with a much larger number of bits allowed. The values were selected somewhat arbitrarily by the experimenter in the range where allowing the algorithm extra bits had little to no effect on the compression process. These values are essentially the least distortion that the algorithm can introduce while performing compression.

Finally, to create the moderate-compression images, the experimenter selected the number of bits resulting in an image that, in the opinion of the experimenter, represented the midpoint between the high and low compression of the image in question.

Figure 3-2 shows the number of the image and the number of bits per image. By consulting the key, the reader can see which images were categorized as representing High, Moderate, or Low Compression. Generally, the high compression group included images that were compressed to less than approximately 2,000 bits. There is some overlap of these categories (as seen in Table 3-3), since the available range of bit values for each image depends on the image complexity.

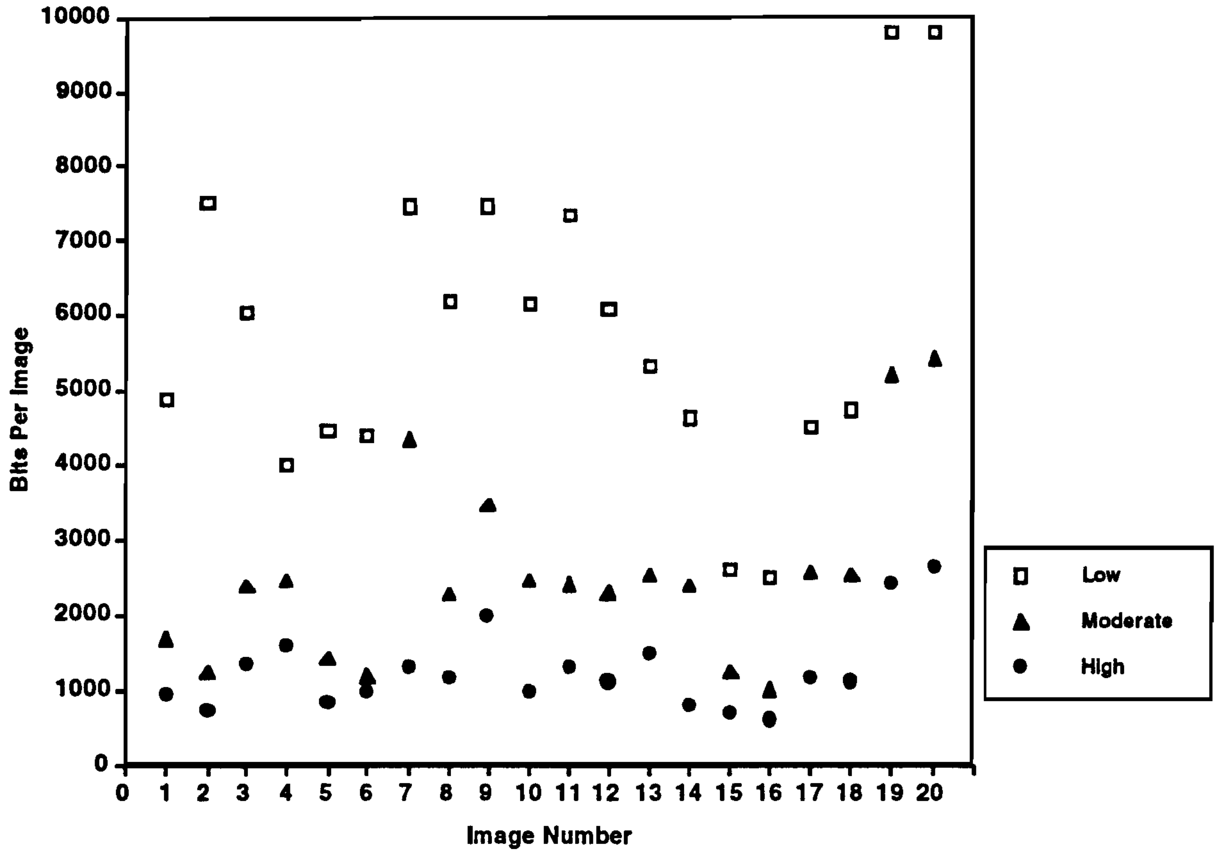


Figure 3-2. Compression Level.

TABLE 3-3

Range of Bits in the Various Compression Groups

Compression Group	Bit–Range
High	606 to 2628
Moderate	999 to 5429
Low	2599 to 9758

3.2.2.2 Training Images and Experimental Images

Table 3-4 shows the list of training and experimental blocks presented for each training session and experimental task. Prior to each experimental task, subjects received training as indicated by the "Practice" blocks. The practice images were chosen to be representative of a wide range of conditions typical of what was to be seen in the test images.

The Route Selection Task practice trials consisted of six images. In the experiment itself, there were 112 image-presentations. Fourteen images were selected, and each one was shown at four different distortion levels (compressed to the high, moderate, and low distortion level and undistorted). Each image was presented twice, and the order of the images was random, without any representation to the subject of distortion level.

The Route Selection Task was performed before the subjective ratings so that the subjects had not compared pairs of images yet. It was felt that this might bias their judgment of the routes selected.

Practice trials for the Distortion Rating Task consisted of eight pairs of images showing an uncompressed image next to an image compressed to one of the three levels of compression (High, Moderate, and Low), for a total of twenty four practice trials. In the Distortion Rating Task, for data collection purposes there were twenty pairs of images showing an uncompressed image next to an image compressed to one of the three levels of compression. These pairs were presented three times in random order, for a total of 180 trials.

Practice trials for the Acceptability Rating Task consisted of five pairs of images showing an uncompressed image next to a compressed image. In the Acceptability Rating Task, for data collection purposes there were twenty pairs of images showing an uncompressed image next to an image compressed to one of the three levels of compression, for a total of 60 trials.

TABLE 3-4

Schedule of Trials

Block	Task	Number of Trials
Practice	Route Selection	6
1	Route Selection	28
2	Route Selection	28
3	Route Selection	28
4	Route Selection	28
Practice	Distortion Rating	24
1	Distortion Rating	60
2	Distortion Rating	60
3	Distortion Rating	60
Practice	Acceptability Rating	5
4	Acceptability Rating	60

The subjects were allowed to take as much time for the completion of each task as they desired. They were also allowed to take breaks between any of the blocks. The subjects typically took approximately four hours to complete the total experiment.

3.2.3 Apparatus

Weather images were presented on a Macintosh personal computer with a color monitor. Custom software was written for both the Route Drawing and the Distortion and Acceptability parts of this experiment. Each of the programs was written so that no experimenter input was necessary after the initial startup information was entered. After that point, the subject was able to handle all of the input to the computer.

The Distortion and Acceptability program displayed the images in a quasi-random order (not truly random, as the identical images were forced not to be displayed consecutively). The software was able to first display the training images, then to go on to run the experiment. The software recorded all of the subject inputs to a data file for subsequent data analysis.

The Route Drawing software was able to select images in a quasi-random order for display. The image was displayed with a starting point and destination point. The Hazard and Go/No Go questions were displayed on the bottom of the screen. The subject was required to draw a route and answer both questions before moving on to the next image. The route drawing itself was designed to follow a typical Macintosh interface. The subject was able to click with the mouse to define a waypoint or click on the route itself to define a new waypoint. Finally, the subject was able to select a waypoint, and click "Delete Point" to clear a waypoint. This interface was learned very quickly and was generally found to be easy to use. The software recorded each waypoint that was defined, as well as the answers to the questions, in a data file that was analyzed after the testing was completed. The software also displayed the training images before the actual data images were used.

3.3 PROCEDURE

Sections 3.3.1 and 3.3.2 describe the training that the subject received at the beginning of the study and prior to each task of the study and the procedures followed during the experiment. Each subject participated individually, i.e., no other subjects were present.

3.3.1 Training

Before beginning the experiment the subject was given training material to read (see Appendix D). The material generally took about 15 minutes for the subject to review. The training material explained the purpose of the study and gave a brief description of GWS, including its purpose, information content, how information would be provided to the pilot via data link, and a brief description of compression. This introductory material was then followed by a written overview of the two parts of the study: Part One – Route Drawing and Part Two – Subjective Ratings. In the training materials, a table listed each experimental block, task, and number of trials so that the subject would be aware of the flow and extent of the study.

Prior to beginning each experimental task, practice trials were provided. During practice for the Route Drawing Task, the subject had an opportunity to become familiar with the route drawing process, i.e., practicing how to draw routes by performing any or all of the following: selecting, adding, deleting, and moving waypoints. He also had an opportunity to become familiar with clicking with the mouse to designate their answers to the two questions related to the routes. During the practice for the subjective rating tasks (distortion rating and acceptability rating), the subject had an opportunity to become familiar with the specific rating systems used and the response-entry process.

3.3.2 Experiment Procedures

3.3.2.1 Part One / Route Drawing Task

Written instructions were provided for this task and are included in Appendix D, which summarizes those instructions. The subject was instructed that he would see a series of GWS weather images on a Macintosh computer. He was asked to draw a route of flight from one designated point to another designated point, indicated on the screen as point "A" and "B". He was given detailed instructions on how to draw a route by using the mouse and clicking. When he clicked the mouse button, a waypoint was defined at the location selected. The subject could move that waypoint as desired. The subject also could delete the entire route and start over again, if desired. The subject was instructed on how to select, add, delete, and move waypoints. An example is shown in Figure 3-3. In this example, a route has already been drawn in by a subject.

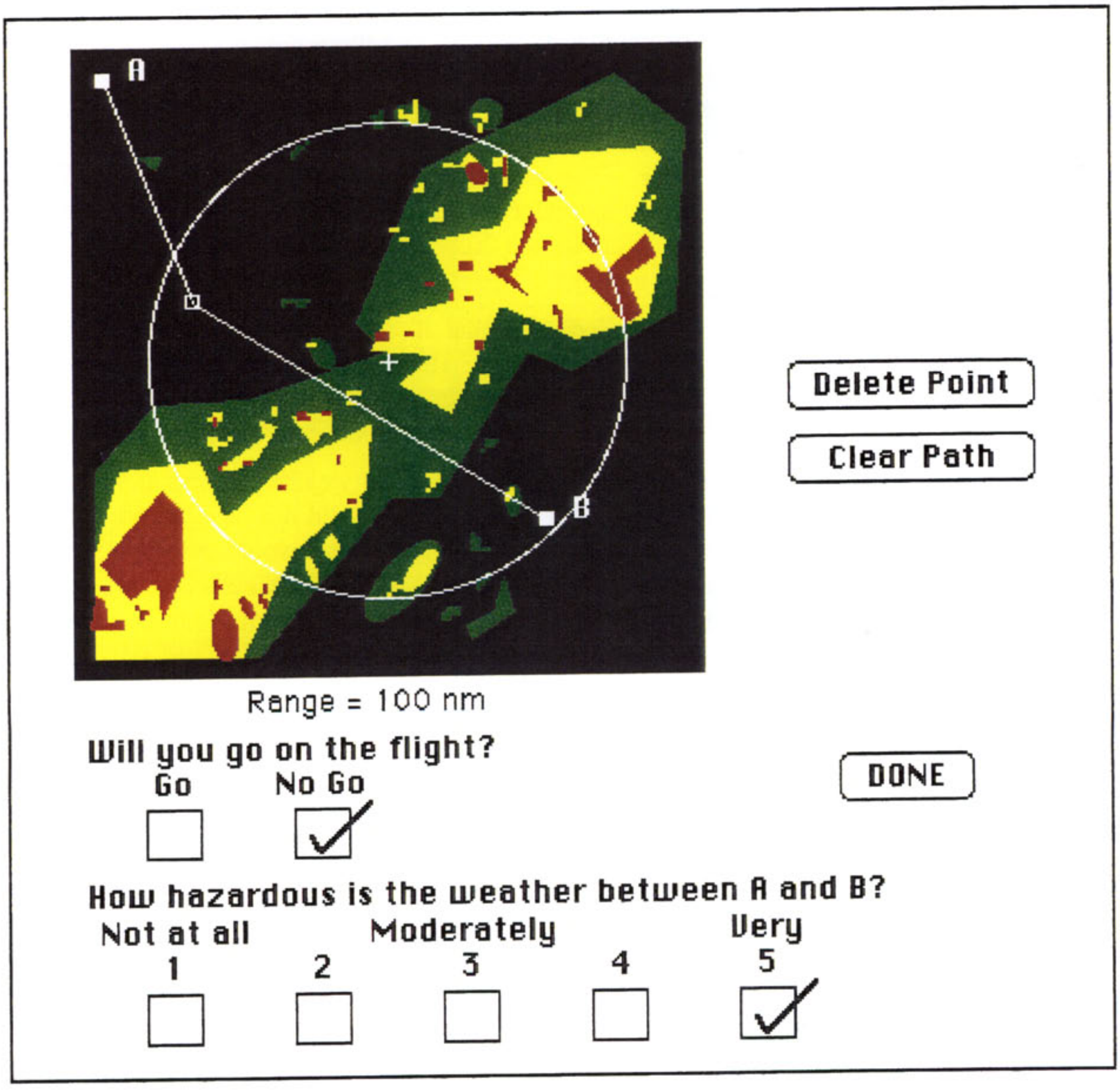


Figure 3-3. Route Drawing Task.

In addition to drawing the route, the subject was asked to answer two questions that appeared on the screen with the image, as shown in Figure 3-3. The subject was asked to make a Go/No Go decision. In other words, he was asked, "Will you go on this flight?" Yes or No? He was also asked, "How hazardous is the weather between A and B?" He selected a response from 1 (Not at all) to 5 (Very). The responses to these questions provide subjective comparisons of the effects of compression. The subject was told that he could complete these steps in whatever order he desired, i.e., answer question(s) first or draw a route first. However, he could not exit that screen until all steps had been completed. The subject responded to the two questions by clicking with the mouse and a check mark appeared in the box which corresponded with his selected response. All responses were made by clicking with the mouse. The data were saved to a file in a format that could be read by Microsoft Excel.

In completing this task, the subject was asked to make certain assumptions regarding the type of aircraft he would be flying, his intention in taking this flight, and the extent of weather information available. He was instructed that the aircraft is a light, single engine piston aircraft, such as a Cessna 172. The instructions indicated that the aircraft has conventional IFR avionics, including dual navigation/communication radios, Distance Measuring Equipment (DME) and Automatic Direction Finder (ADF.) The instructions indicated that the aircraft does not have LORAN, Stormscope, or weather radar. The subject was also told that it is equipped for ILS (Instrument Landing System) and has no autopilot or HSI (Horizontal Situation Indicator). He was instructed to assume that he had full fuel for the flight and could assume that he was planning to travel with one passenger who is not a pilot.

Regarding intent, the instructions indicated that it was important for him to reach the destination, but that it was not a matter of life or death. He was told to be concerned with getting to the given destination in a timely fashion, while maintaining flight safety. The emphasis was on planning a route that reflects usual consideration of the balance between safety and convenience.

Regarding weather, the subject was instructed that the weather information available is limited to what appears on the GWS weather image, i.e., he would not have access to any other information sources. The subject was told that all the weather shown would be actual weather that was recorded during the summer months in New England and that the weather would be depicted North-up. In addition, he was told that the time of the weather image should not be considered in his decision and to assume that each image is current. Since the images would be snapshots in time of weather situations, the subject was told that although in actual flight the weather is changing over time and moving and that he would be thinking of where the weather will be when he reaches a certain point; in this task, he must assume that the weather depicted is stationary.

The instructions indicated that there are no right or wrong answers, and that our purpose in conducting the study was to understand how pilots select routes in relation to weather. The subject was also told that he would see images that may look familiar from previous trials. However, instead of trying to remember an earlier route and accompanying responses, he was instructed to consider the image on the screen and respond. The subject was encouraged to ask questions at any time during the trials. Following the practice trials, the subject began the first block of test trials. After each block of 28 images, we asked the subject if he would like a break. The subject could take a break between blocks as he desired. All subjects were told to take a break after completing the four test blocks of the Route Drawing Task.

3.3.2.2 *Part Two / Subjective Ratings*

3.3.2.2.1 Distortion Rating. Written instructions were provided for this task and are included in Appendix D, which summarizes those instructions. The subject was instructed that he would, as in the Route Drawing Task, see a series of GWS weather images on a Macintosh computer. However, this time, he would see a pair of images on the screen. He was reminded that the compressed image is an altered version of the uncompressed image. He was asked to judge the degree to which the compressed image had been distorted relative to the uncompressed image. The subject's task was to assign a numerical value to the level of distortion perceived. He was asked to base his rating on the quantitative amount of distortion of the compressed image; not on the usefulness and functionality of the compressed image. He was informed that he would have a chance to rate functionality later in the Acceptability Task.

Subjects rated distortion, through use of a magnitude-estimation method [4]. In using magnitude-estimation, the experimenter asks the observer to assign a number to the perceived magnitude of each stimulus. In this case, the stimulus was the compressed image which appeared next to the uncompressed image. The rating was based on the quantitative amount of distortion of the compressed image.

A stimulus-measuring technique frequently used in research is direct scaling (for example, a 1 to 10 rating scale). The basic assumption of direct scaling is that the observer is able to match experimenter-prescribed numbers to his perceptions. Direct scaling is a closed scale, i.e., having upper and lower limits prescribed by the experimenter. Direct scaling methods restrict the observer to equal intervals, ratios, and pair comparisons.

As an alternative to direct scaling, magnitude-estimation was selected as the stimulus-measuring technique for the Distortion Rating Task. It was selected in an attempt to avoid restrictions and to encourage the observer to assign the numbers he feels are appropriate without any of the biases which may be associated with a response system devised by the experimenter. In the case of rating distortion, this open-ended scale allows the subject to choose a higher value for an image that he feels is more distorted than any image previously viewed. Magnitude-estimation was also selected since it provides a workable means for obtaining the subject's rank ordering of a large number of stimuli without actually displaying all of the stimuli at once, an obviously difficult task when the subject is asked to view a large number of stimuli. Once the ratings are obtained through use of magnitude estimation, the experimenter can assign ranks regardless of the subject's own rating scale.

Using magnitude-estimation, a defined attribute of any set of stimuli can be scaled; for example, visual brightness, intensity of odors, the saltiness of solutions, or the beauty of works of art. Usually a fixed set of stimuli covering a wide range of a certain attribute is presented to the observer.

In using this method, first, the experimenter presents the observer with a standard stimulus and defines the subjective value of that as the observer's modulus. In this case the modulus, or anchor, was the raw image, and the subjects were told that it had a distortion magnitude of 10. Next, the subject was asked to report a distortion value for the test image, and that if he felt that the compressed/altered image does not distort the weather picture at all (in terms of being a substitute for the uncompressed image), he should enter a response of "10." He was instructed to assign higher numbers to more distorted images and that he could respond with any numerical value (greater than or equal to "10", the value of the uncompressed image). He was asked to try to make the numbers proportional to the distortion of the compressed image as they perceived it. Verbal examples as well as examples in the written instructions were given. It was emphasized that he could assign any number, and there was no upper-limit on the number assigned, however, try to keep rating proportional.

The subject began with a practice block of 24 image, as seen in Table 3-4. The practice block was provided so that subject could become comfortable with this type of rating and could develop his own internal scale of distortion in a consistent manner before beginning the test trials. The subject was told that there were no right or wrong answers and that the purpose of the study was to understand how pilots judge image distortion.

3.3.2.2.2 Acceptability Rating. Written instructions were provided for this task and are included in Appendix D, which summarizes those instructions. The subject was instructed that he would, as in the Distortion Rating Task, see a series of pairs of GWS weather images on a Macintosh computer. The subject was asked to answer the question: How acceptable is the compressed/altered image as a replacement for the uncompressed image? He was told that this question should be answered in the context of typical general aviation flight in a single or light twin-engine aircraft. He was asked to judge acceptability in terms of the compressed image's functionality for the flight task as compared with the functionality of the uncompressed image for the flight task. The subject was asked to rate acceptability regardless of the degree of image distortion. He was asked not to judge the acceptability of the compressed image in comparison to a situation where no graphical weather image is available to the pilot, but to rate acceptability in comparing the two images. An example image is shown in Figure 3-4.

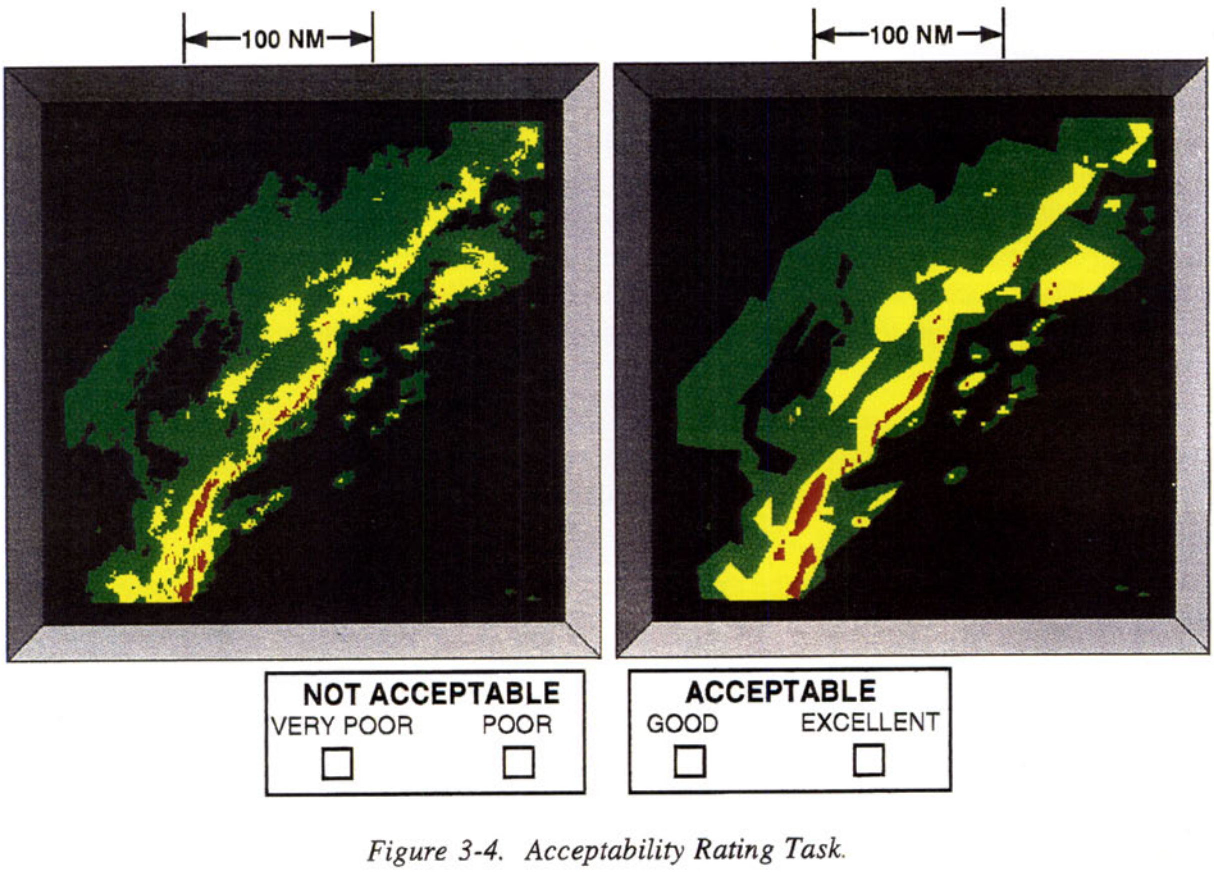


Figure 3-4. Acceptability Rating Task.

A four point scale was used in rating acceptability. The scale was broken into two main categories: acceptable and unacceptable. Then within those two main categories, there were two subheadings: Very Poor and Poor (unacceptable) and Good and Excellent (acceptable). The operational definitions of each of the ratings that were given to the subject are as follows:

Not Acceptable/Very Poor. There are major functional differences between the two images. The deficiencies in the compressed/altered image make its utility for GA operations very low.

Not Acceptable/Poor. There are functional differences between the two images. The deficiencies in the compressed/altered image limits its utility for GA operations.

Acceptable/Good. There are no major functional differences between the two images. The compressed/altered image has no serious deficiencies and is useful for GA operations.

Acceptable/Excellent. There are no functional differences between the two images. The compressed/altered image has no deficiencies and is as useful for GA operations as the uncompressed image.

4. RESULTS

Section 4.1 provides the results of the subjective ratings of distortion and acceptability. The results of the analyses of subjective ratings will aid us in determining what images to send to the aircraft, based on pilot opinion of distortion and acceptability. In Section 4.1, acceptability ratings are considered in relation to a number of computed measures of image quality. By doing so, we begin the process of identifying threshold values or cutoff points that can be used to differentiate images that are acceptable to pilots from images that are not acceptable.

Section 4.2 provides the results of the Route Drawing Task. The results of the analyses of the Route Drawing Task will aid us in determining what images to send to the aircraft, based on pilot behavior

Section 4.3 provides the results of the analysis of the relationship between the subjective ratings and pilot performance (as measured in the Route Drawing Task). Section 4.4 provides the results of the Exit Interview.

4.1 RESULTS OF THE SUBJECTIVE RATINGS

4.1.1 Distortion Rating

4.1.1.1 Within-Subject Consistency in Distortion Ratings

Each subject was exposed to each image three times. The first step in examining the distortion ratings was to determine whether subjects were internally consistent across their three distortion ratings of each image. To determine the level of internal consistent, a series of correlation coefficients were calculated. The formula used was Pearson r, which is a standard statistical procedure for determining the linear relationship between variables. The relationship between each time a subject rated a particular image was analyzed, to determine if that subject was consistent in rating distortion. The following correlations analyses were calculated:

the correlation between repetition 1 and 2

the correlation between repetition 1 and 3

the correlation between repetition 2 and 3

Appendix E lists the results and also gives the minimum and maximum values used by each subject in rating distortion. Results indicated that subjects were generally consistent in their ratings. The correlation coefficients for the majority of subjects were at or above 0.75[1]. For many subjects, correlation coefficients ranged from 0.80 to 0.91, indicating that 64% to 82% of the variance in one repetition is associated with the variance in the other repetition.

The correlation coefficient of 0.91 (highest correlation coefficient reached) or 0.49 (lowest correlation coefficient reached) represent single numbers that conveniently describe the linear relationship between repetitions. It is also useful to know whether the distortion ratings in the one repetition would be associated with distortion ratings in another repetition in the general

---

[1] A correlation coefficient of 0.75 means that 56% (calculated by squaring the correlation coefficient) of the variance in one repetition is associated with the variance in the other repetition.

population, i.e., not just in our sample of twenty pilots. A test of significance was performed to determine if the correlation in the sample of twenty pilots was due to sampling error, or if we can conclude that there is some non zero positive correlation between repetitions (distortion ratings one time versus next time) in the population. For all the cases, a statistical significance was found at the .01 level. It can be concluded that there was good consistency for each pilot, and little learning effect here, and that the pilots adapted well to the open ended scale used for this task.

4.1.1.2 Distortion Ranking

The next step in examining distortion ratings is to consider the results of all subjects combined. Since with the magnitude estimate scaling the subject develops his own internal scale, a means must be identified for minimizing the resulting between subject variability. If the raw distortion ratings were averaged across subjects, ratings from subjects who used large maxima will be weighted more heavily than the ratings from subjects who used smaller maxima (Appendix E lists the minimum and maximum ratings given by each subject). In addition, data from subjects who used a wide range of ratings would be weighted more heavily than data from subjects who used a narrow range of ratings. To reduce these inequities in rating a log transform of each response was used. The analyses were then performed on the log-transformed ratings. A property of the log transform is that the antilog of the mean of log-transformed data is in fact the median value of the raw data. Thus, the log transform has a meaningful interpretation. The log-transform reduces both the within-subject and between-subjects variability, but it does not eliminate these sources of variability.

One way to eliminate the between-subject variability in distortion ratings is to convert each subject's distortion rating into a distortion rank. The rank is generated by putting the images for a given subject into an order based on the image's rating. For each subject's ratings, a rank of 1 was given to the image with the lowest distortion rating, and a rank of 60 was given to the image with the highest distortion rating.

Unlike the log transform, the distortion rank does not preserve the spacing between responses. For example, the fifth image could have a distortion rating of 25, the sixth image a rating of 50, and the seventh a rating of 500. These images would be ranked as 5, 6, and 7, respectively. Ranks were assigned by first averaging the three raw responses that each subject gave to an image. Then the sixty within-subject means were sorted and assigned ranks in order from lowest to highest. If two or more images were given equal distortion ratings, they were all given the mean rank for the set. For example, if the fifth through tenth images all had the same distortion rating, they were each given a rank of 7.5. If the seventeenth through nineteenth images tied, they were given a rank of 18.

Figure 4-1 shows the mean distortion ranks relative to the number of bits in the compressed image. Appendix F lists the mean distortion rankings for each compressed image and the corresponding number of bits in the image. The standard deviation of each rank is also listed. In consulting Appendix F, we see which images were considered to be the most distorted. In general, images in the high compression group were rated to have the highest distortion. The middle compression group were considered the next distorted and the low compression group were considered the least distorted. It was also found that the least distorted images and the most highly distorted images tend to have lower standard deviations, meaning that subjects agreed on which are the "best" and which are the "worst" images.

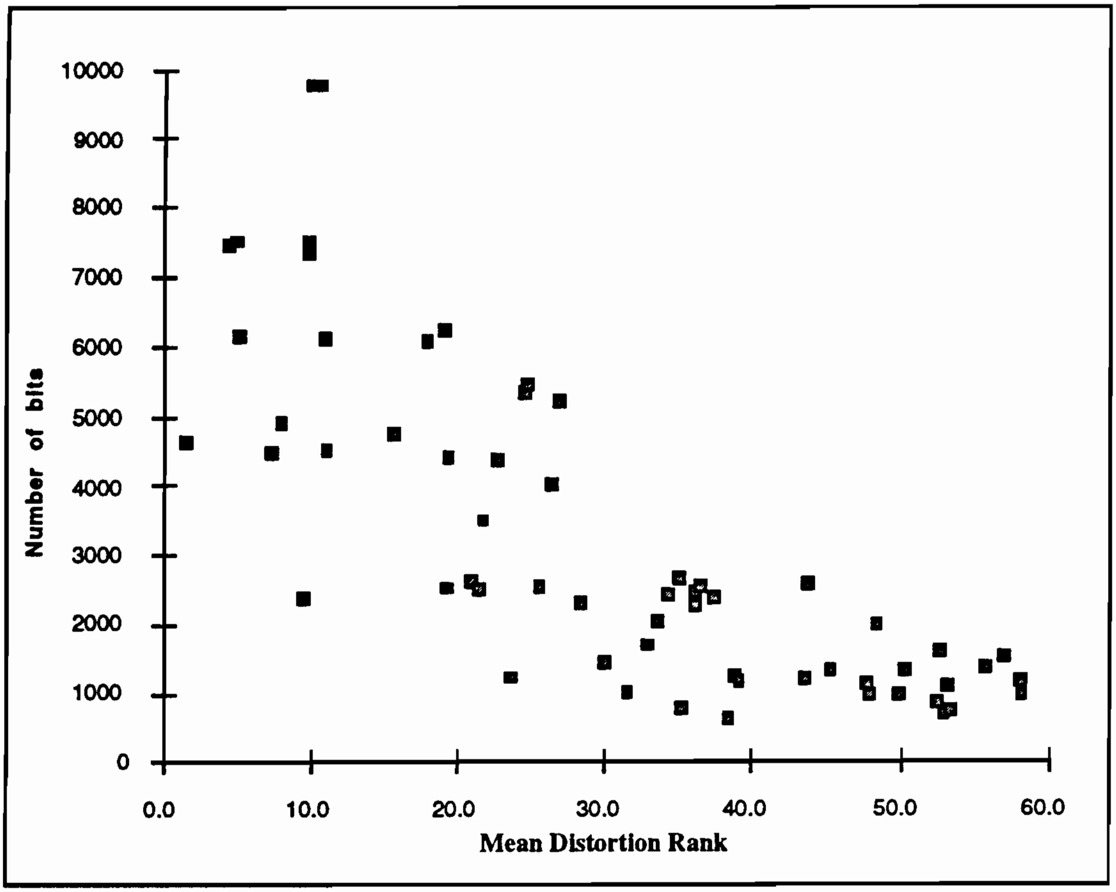


Figure 4-1. Mean distortion rank plotted by Number of Bits.

4.1.2 Acceptability Rating

Figure 4-2 shows the mean acceptability ratings of all subjects combined. They are plotted by the number of bits in each of the sixty compressed image. The supporting data for this figure are found in Appendix G. In Figure 4-2, it is seen that generally as the number of bits per image decreases the acceptability rating decreases. In these cases with significant amounts of weather complexity, images with over 2,000 bits were found to be acceptable.

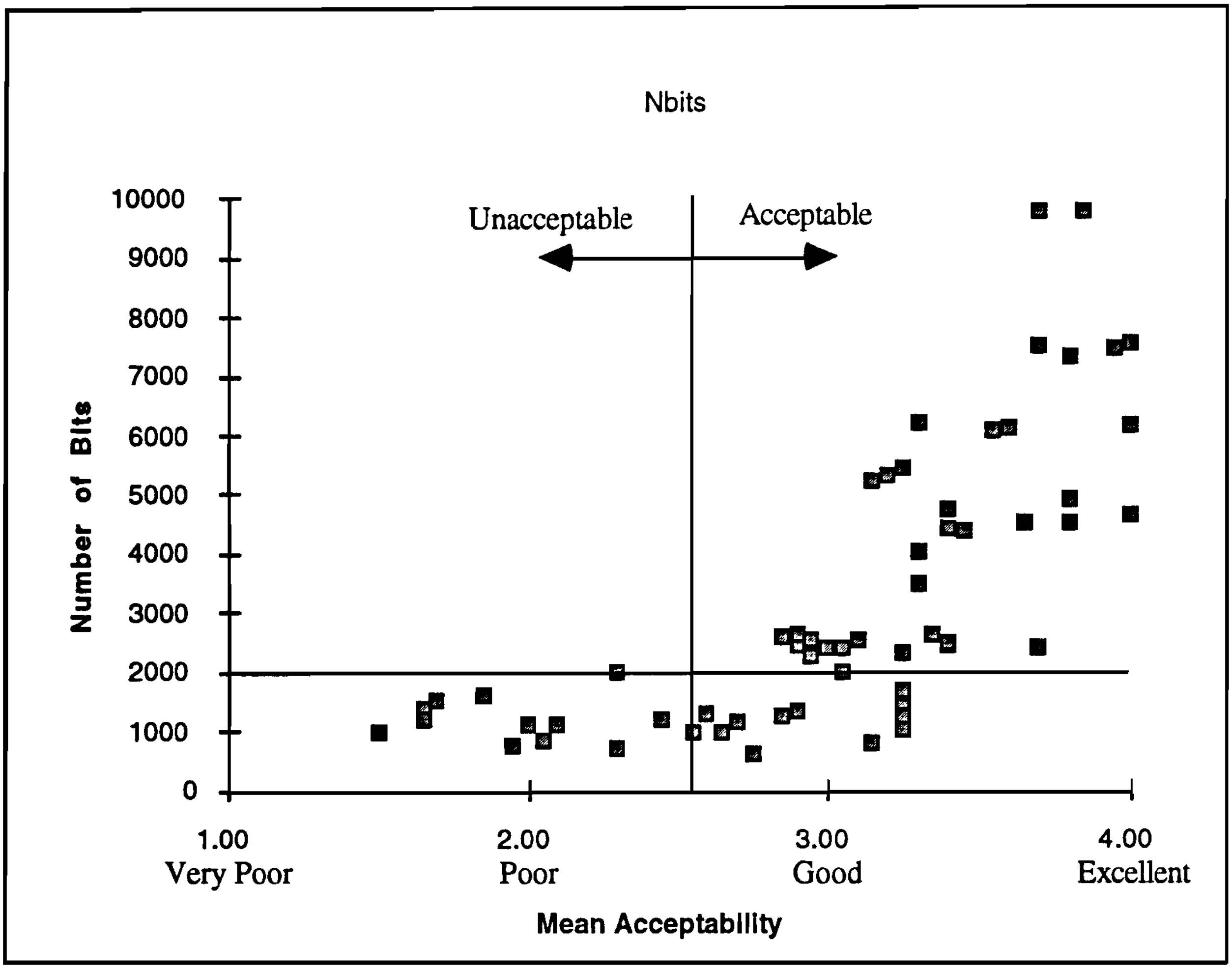


Figure 4-2. Mean acceptability ratings of all subjects combined.

Another way to look at acceptability ratings, rather than using mean acceptability scores, is to determine the percentage of pilots who have determined a particular image to be acceptable or unacceptable. Figure 4-3a shows the group acceptability for each of the images in the high compression group. Acceptable includes ratings of "Good" and "Excellent" and unacceptable includes ratings of "Poor" and "Very Poor". The dotted line on the figure indicates the cut-off point if one declared that images would be acceptable only if 80% of the pilots said they were acceptable.

Figure 4-3b shows the percentage of subjects who found each of the images acceptable in the moderate compression group. The dotted line on the figure indicates the cut-off point if one declared that images would be acceptable only if 80% of the pilots said they were acceptable.

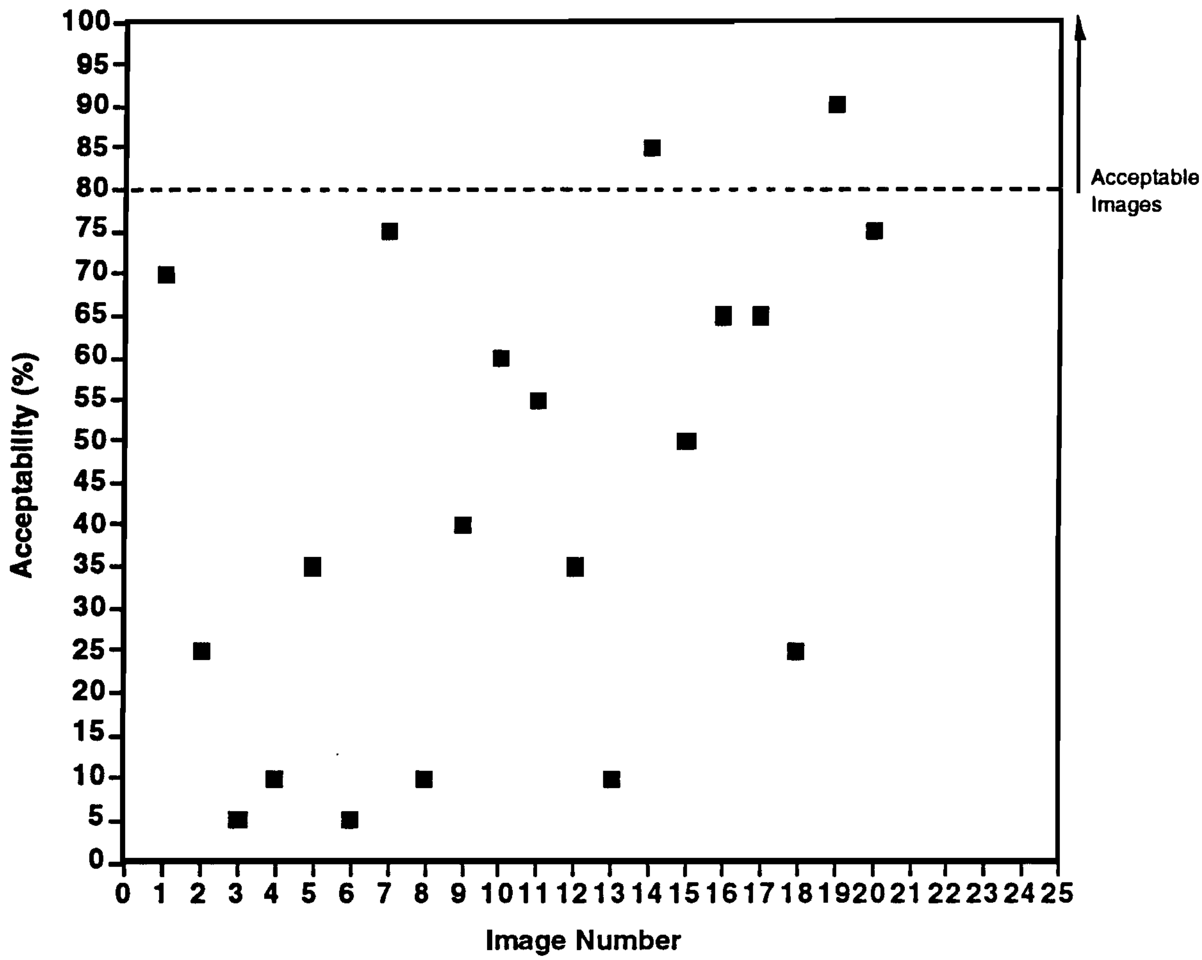


Figure 4-3a. Percentage of subjects who rated each image to be acceptable for the highly-distorted images.

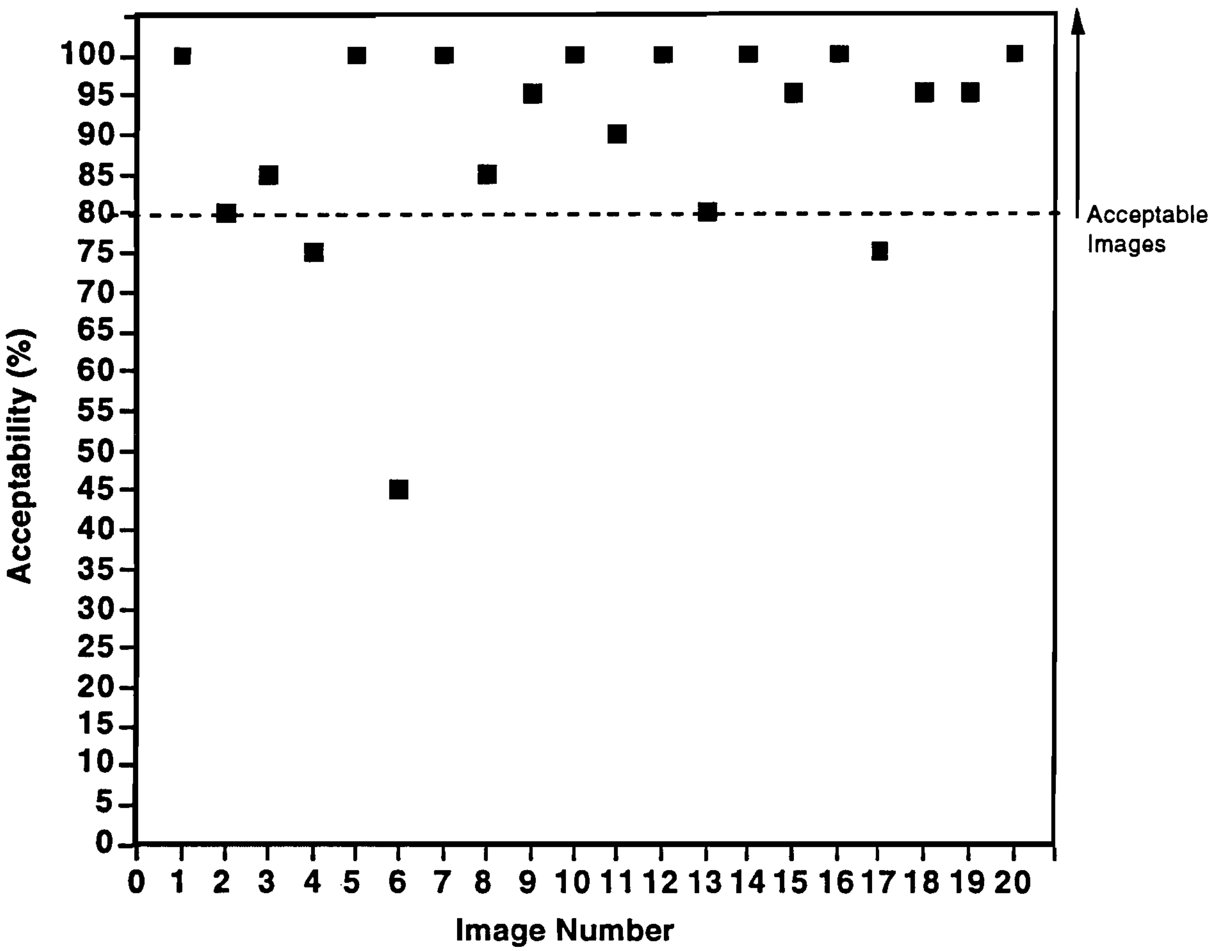


Figure 4-3b. Percentage of subjects who rated each image to be acceptable for the moderately-distorted images.

Figure 4-4 is an example of an image-set with an uncompressed and highly-compressed image. The highly -compressed image was judged to be not acceptable by 80% of the subjects. Subjects were asked to make verbal comments (after each rating of an image) on why they decided a compressed images was acceptable or unacceptable. A review of these comments provides some insight into why subjects may have judged most highly-compressed images to be unacceptable. Subject comments regarding why an image was judged to be unacceptable centered around: a loss of detail in the compressed image, elliptical shape was not trustworthy, and lack of confidence in the image to represent truth (the actual weather).

At the highest distortion level, all but two of the images, which are shown in Figure 4-5, were judged by 80% of the subjects to be unacceptable. These two images, numbers 14 and 19, were acceptable, even at the highest compression level. These images where examined to determine if there was some salient characteristic which rendered the highly compressed version of these images as still being acceptable. For Image 14, 10% of the pixels had non-zero values, and the highly compressed image had 783 bits in it. While Image 19 had 60% non-zero pixels and was compressed to 2406 bits. These numbers were typical, so did not seem to relate to the acceptability of the two images.

The subjects' verbal comments (after each rating of an image set) on why they decided a highly compressed image was acceptable were examined. Comments on the two images in question indicated that the compressed images maintained the basic shape of the uncompressed image the compressed image, thus they were acceptable. Images 14 and 19 are seen in Figure 4-5. For Image 14, comments indicated that the weather in the uncompressed image was somewhat elliptical to begin with. Therefore, the subjects did not have an unfavorable response to the use of ellipses in the compressed image, i.e., the basic shape of the weather was maintained. In Image 19, the weather in the uncompressed image was very non-elliptical to begin with. However, the "high" compression did not result in ellipses, but instead in polygons, i.e., the basic shape of the weather was maintained. Poly-Ell was not able to force many of the regions to be ellipses, and instead required a large number of bits for compression. It thus kept a significant amount of detail.

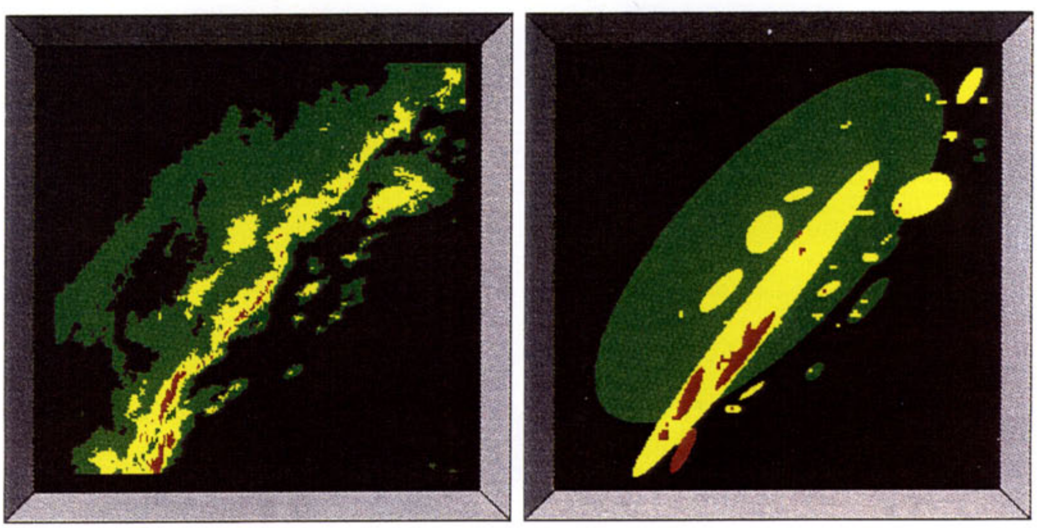


Figure 4-4. A highly-compressed image that was deemed to be unacceptable.

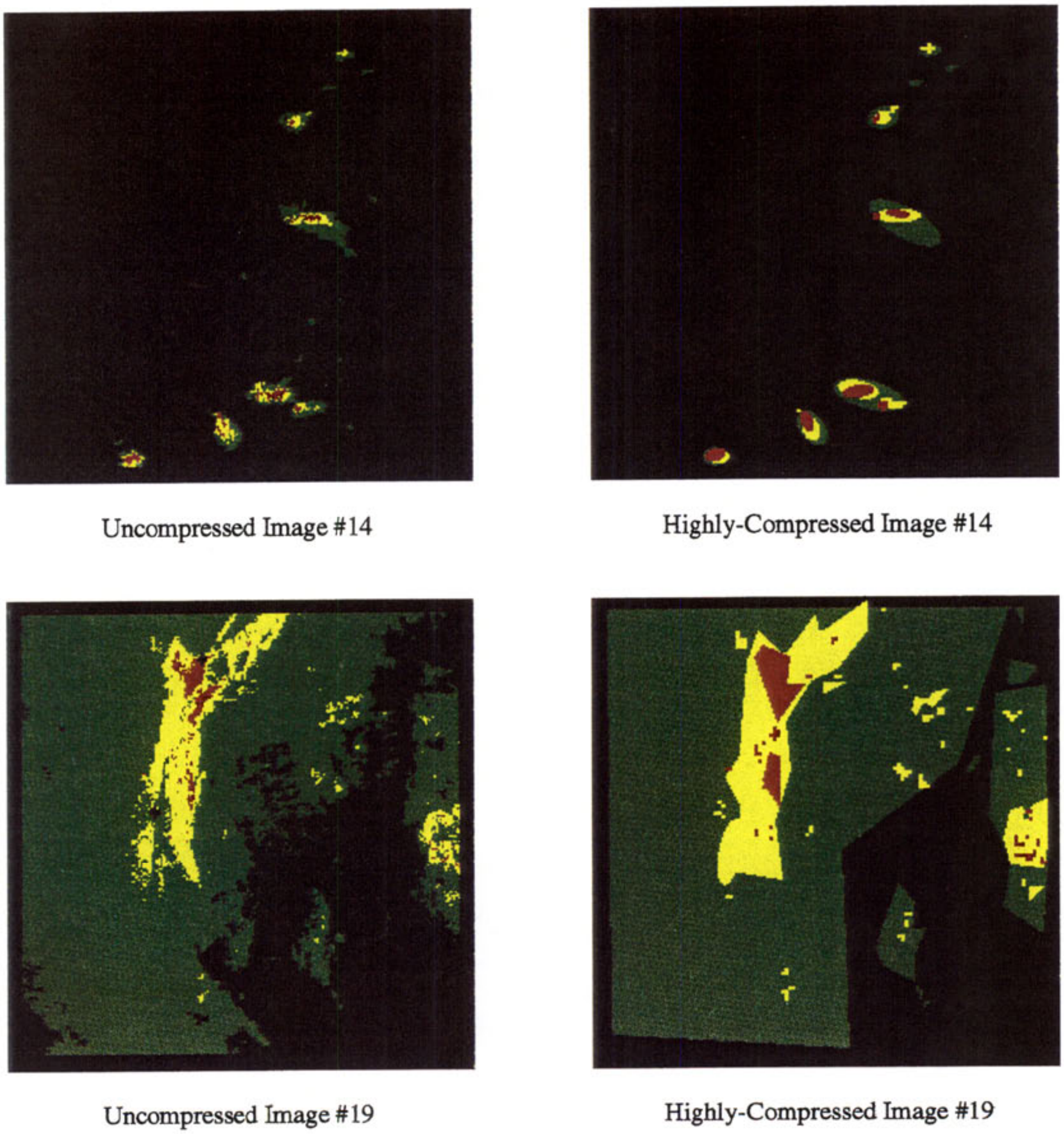


Figure 4-5. Highly-compressed images that were found to be acceptable.

4.1.3 Comparison of Acceptability and Distortion Ratings

4.1.3.1 Between-Subject Analysis

While in the distortion task, it was found that subjects generally agreed on which are the "best" and "worst" images, in the acceptability task it was found that there was a wide range in variability in the types of images that subjects were willing to accept or reject. Table 4-2 shows the percentage of the total number of images that subjects were willing to accept. For example one subject accepted all images, while another subject accepted half of the images. The conclusion that can be made from the standard deviations obtained from the distortion rankings and the percentage of images that subjects accepted is that subjects agree on which are the best and worst images as far as level of distortion, but disagree on cutoff level for acceptability.

TABLE 4-2

Percentage of the Total Number of Images That Each Subject Accepted

Subject Number	% Acceptable
1	80%
2	85%
3	73%
4	90%
5	63%
6	85%
7	87%
8	85%
9	63%
10	78%
11	53%
12	50%
13	68%
14	90%
15	78%
16	78%
17	85%
18	87%
19	100%
20	77%

4.1.3.2 Within-Subject Analysis

The next question we asked in our analysis was: What is the relationship between distortion and acceptability ratings within subjects? That is, within each subject were the distortion and acceptability ratings correlated? Results indicated that the two measures were negatively correlated and the data are shown in Appendix H. This was expected, as images that are more distorted are expected to be less acceptable to the subjects. The size of this correlation can be taken as a reasonable upper bound on the magnitude of the correlation between any physical correlate and the subjective ratings. In other words, the two subjective ratings are expected to be related to each other more strongly than any computable measure is expected to be related to either acceptability or distortion. Distortion and acceptability were highly correlated within subjects, with correlations ranging from -0.503 to -0.879.

4.1.4 The Relationship between Acceptability Ratings and the Computed Measures

One of the goals of this study was to identify an objective computed measure of image distortion that could be used to decide which images to uplink to an airplane. To do that several different computed measures were calculated, the number of bits in the compressed images, the mean square error, and two different compression ratios. Acceptability was then plotted against each of these measures to determine the measure with the most clear-cut threshold value (for separating acceptable and unacceptable images). Ideally that measure would be a good predictor of image acceptability and could be used to determine which images to uplink. In plotting acceptability the following method was used: all images that 80 to 100% of the pilots rated as being acceptable were labeled "acceptable", all images that 70 to 75% were labeled "borderline", and all images that less than 70% of the pilots rated as being acceptable were labeled "unacceptable".

The first of the measures that was considered was the number of bits in each compressed image (NBits). As seen in Figure 4-6, images with less than 2,000 bits were generally considered to be unacceptable. However, a few were borderline acceptable, and conversely a few of the images of less than 2,000 bits were acceptable. Furthermore, this cannot be generalized, since raw images may have very few bits and is strongly dependent on the compression algorithm used.

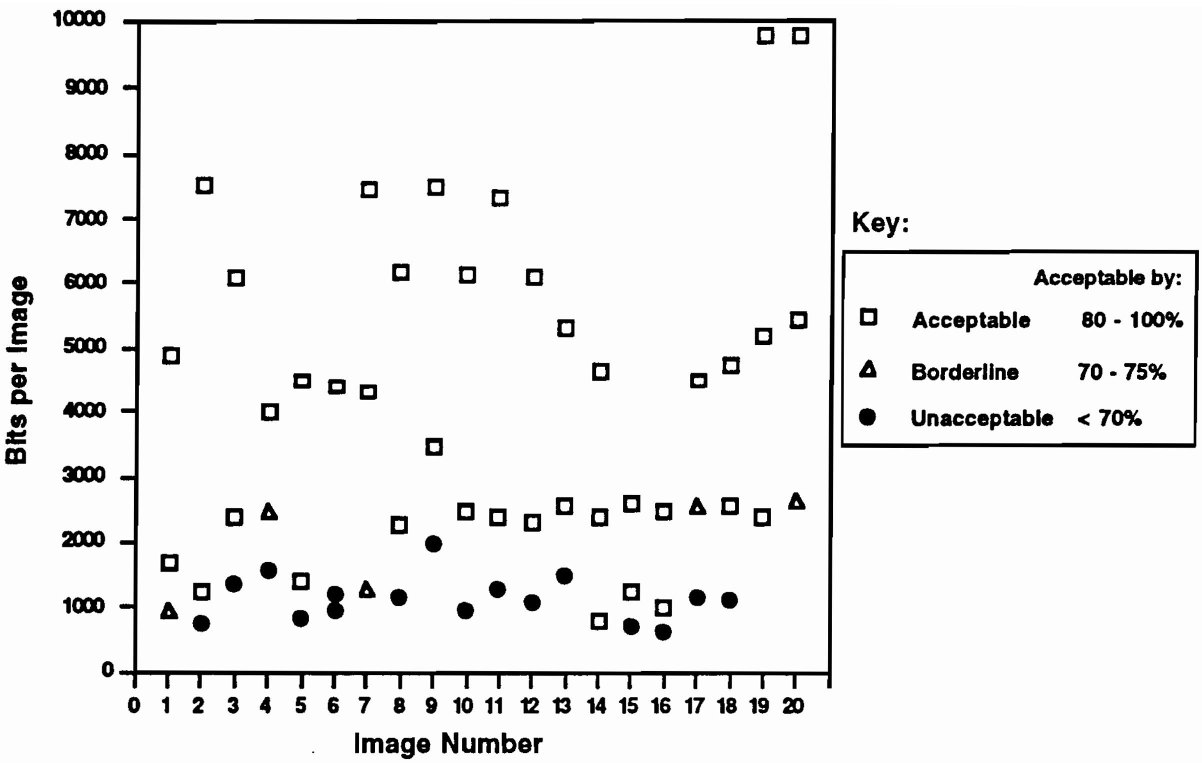


Figure 4-6. Bits per Image as a computed measure of acceptability.

The second measure that was considered was the MSE (mean square error), which is a commonly used measure of image distortion. It is calculated by subtracting the value (where weather level was used for the value) of each pixel in the compressed image, from that in the raw image, squaring that number, and then summing over all pixels. This number was then normalized by the total number of non-zero pixels in the raw image. This number represents the difference between the two images. As seen in Figure 4-7, images with an MSE of greater than 0.25 were considered unacceptable.

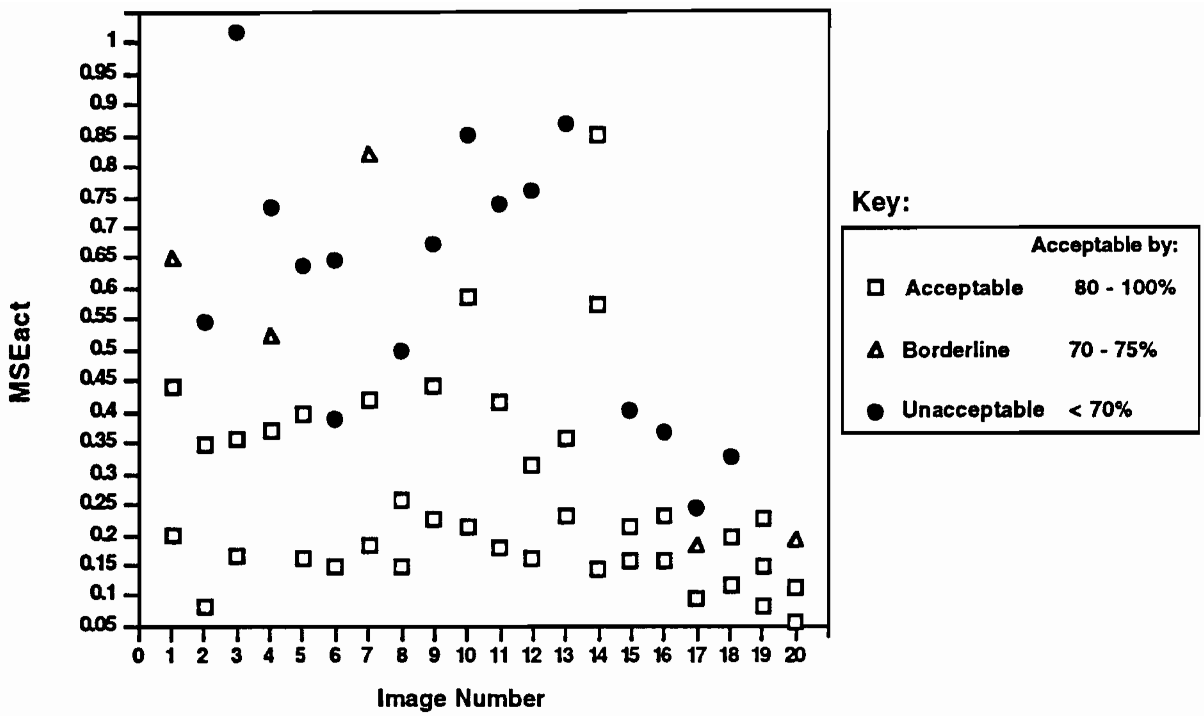


Figure 4-7. Mean Square Error as a computed measure of acceptability.

The next measure that was considered was to compare the Polygon-Ellipse compression method to a standard lossless compression technique. In this case, the Polygon-Ellipse method was normalized by using a standard lossless compression scheme. Each raw image was compressed using a standard run length encoding (RLE) scheme (PackBits on a Macintosh computer). This method was chosen because it is lossless, i.e. there is no distortion introduced, and the number of bits that it generates is a function of the number of weather regions and complexity of the image. The Polygon-Ellipse method, on the other hand, is a lossy algorithm, which means that there is some information lost from the image after compression. This number of bits from RLE, was then divided by the number of bits that Polygon-Ellipse used to compress the same image. This number represents a lossless to lossy compression ratio. Each of these numbers is a function of the amount of information in the image, so this ratio represents the amount of information that is lost from the image, due to the compression induced distortion. As it is possible for the Polygon-Ellipse method to require more bits than RLE, it is possible for this ration to be less than 1. Figure 4-8 shows acceptability versus RLE/NBits. As can be seen in the figure, images with RLE/NBits of greater than 3 were considered unacceptable.

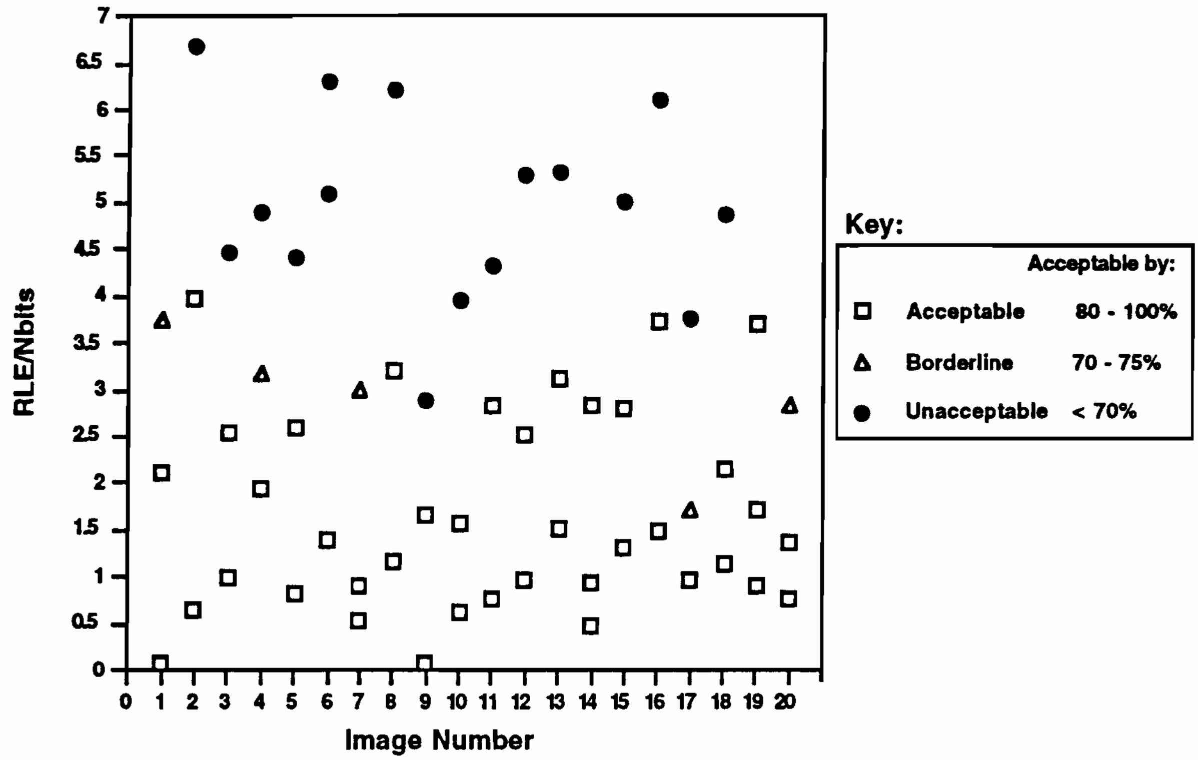


Figure 4-8. Run Length Encoding / Number of Bits as a computed measure of acceptability.

A fourth measure of information was also tested. As each pixel had four possible values, it can be represented with two bits. An approximation for information in an image was S' which was defined as twice the number of non-zero pixels in the image. This measure is again a lossy measure of information, as the image could not be reconstructed from that number of bits. This value was divided by the number of bits in the compressed image (NBits) to again represent lost information due to compression.

Because there were four different computed measures that were of interest, a stepwise regression was performed. The stepwise regression process calculates the partial correlation between each single predictor and the dependent variable, then adds the best predictor of the set to the model, and tries to fit the remaining predictors. The result is that the predictors are ordered in terms of the strength of their relationship to the dependent variable.

Four independent variables were tested in the stepwise regressions: MSEact, NBits, Log(S'/NBits), and Log(RLE/NBits). Log(RLE/NBits) yielded the single highest partial correlation with the raw distortion rating for 19 out of 20 subjects, with a mean value of 0.783 ($p < .001$ for each of the 19 subjects). For the log-transformed distortion ratings, Log(RLE/NBits) yielded the highest partial correlation for all 20 subjects with a mean value of 0.806 ($p < .001$ for all subjects). For acceptability, this measure was the best single predictor for 18 of the 20 subjects with a mean correlation of –0.752 ($p < .001$ for each of the 18 subjects).

In the stepwise regression analyses presented above, the best computational correlate of pilot subjective ratings was found to be Log(RLE/NBits). The high correlation, however, does not guarantee in and of itself that the same measure will yield a clear threshold value to distinguish acceptable images from unacceptable ones.

Another way to determine the relative strengths of the different computational measures is to plot receiver operating characteristic (ROC) curves [5]. The ROC curve arises from a signal detection paradigm where the signal is considered to be an "unacceptable" image. In this case, the "truth" is determined by consensus among the pilots' judgments. A "hit" occurs when the computed measure declares that an image is unacceptable and pilots also rate the image as being unacceptable. A "false alarm" occurs when the computed measure declares the image to be unacceptable when pilots rate the image as being acceptable. The ROC curve is a plot of correct judgments by the computed measure that an image was unacceptable to pilots (hits) versus false judgments by the computed measure that an image was unacceptable to pilots (false alarms).

Before plotting an ROC curve, a question must be answered: How should pilot consensus be computed? One way to do this is to average across each pilot's rating of acceptability for each image. The mean ratings will range from 1 (very poor) to 4 (very good). A cutoff between acceptable and unacceptable could be defined as a mean rating of 2.5. Images with mean ratings above 2.5 would be "acceptable", and those with mean ratings below 2.5 would be "unacceptable". Another way to determine pilot consensus is to look at the proportion of pilots who felt an image was acceptable, or "Group Approval". For example, if 75% or more of the pilots rated the image as good (3) or very good (4), the image is considered to be acceptable. These two measures, mean rating, and proportion of acceptable ratings were considered, and it was found that they are highly correlated with one another ($r = 0.945$). Group Approval was selected due to its face validity. An ROC curve for MSEact, NBits, Log(S'/NBits), and RLE/NBits was drawn (refer to Figure 4-9). In this case, the best performance of the calculated measures was again RLE/NBits.

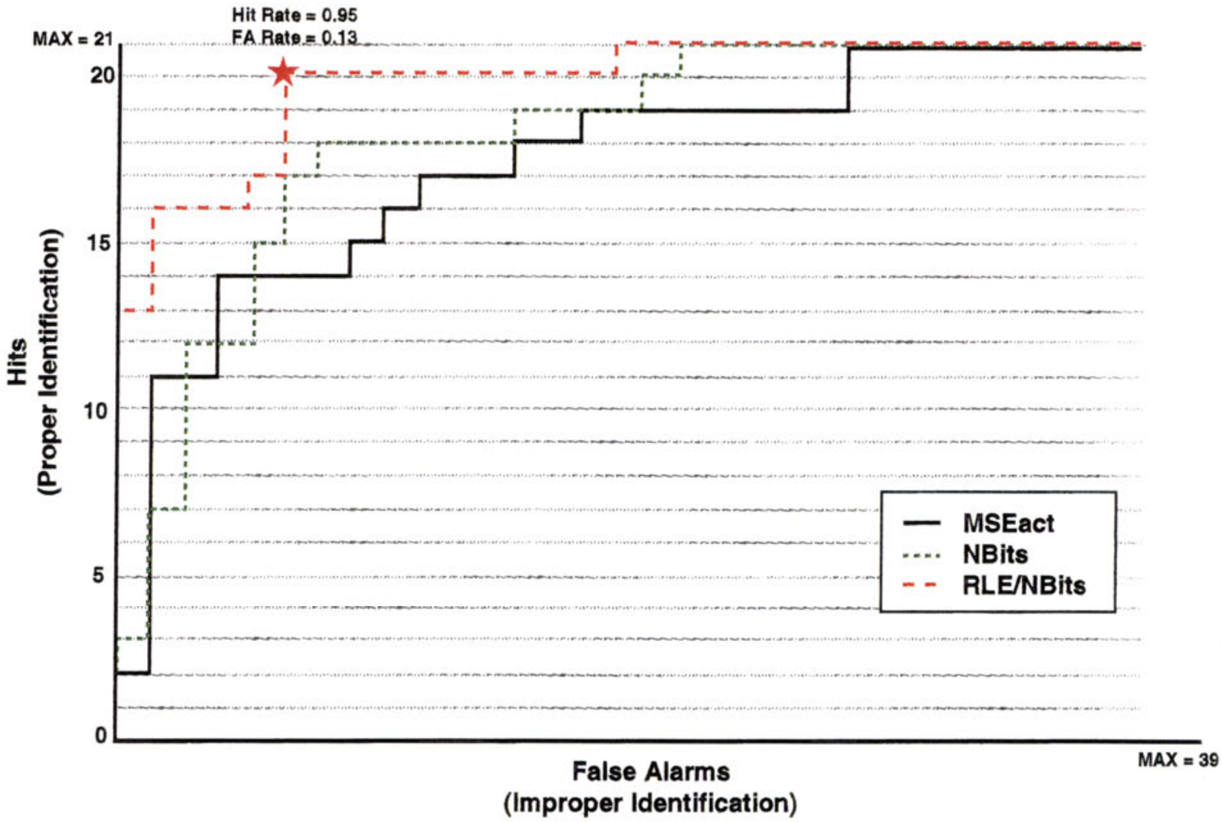


Figure 4-9. ROC curve comparing different computed measures of compression. A hit was a successful detection of an Unacceptable Image.

This ROC curve allows a cutoff value to be chosen based on the desired hit rate of false alarm rate. In other words the number of acceptable images that are labeled acceptable by the algorithm, or the number of acceptable images that are falsely labeled unacceptable. Some example values are shown in Table 4-3 for each of the computed measures. For example if a RLE/Nbits is used, and a cutoff of 22.566 is selected, then one can expect 95% hits and 13% false alarms.

TABLE 4-3

Examples of Prediction of Acceptability by the Various Computed Measures of Compression

Hit and False Alarm Rates for RLE (Given the Selected Cutoff Values)				
Cutoff Value	Number of Hits	% of Hits	Number of False Alarms	% of False Alarms
29.868 and above	16/21	76	1/39	03`
25.356 and above	17/21	81	4/39	10
22.566 and above	20/21	95	5/39	13

Hit and False Alarm Rates for MSE (Given the Selected Cutoff Values)				
Cutoff Value	Number of Hits	% of Hits	Number of False Alarms	% of False Alarms
.498 and above	14/21	67	3/39	08
.364 and above	17/21	81	9/39	23
.326 and above	18/21	86	12/39	31
.245 and above	19/21	90	14/39	36

Hit and False Alarm Rates for S'/NBits (Given the Selected Cutoff Values)				
Cutoff Value	Number of Hits	% of Hits	Number of False Alarms	% of False Alarms
1.364 and above	12/21	57	2/39	05
1.128 and above	14/21	67	5/39	13
1.028 and above	17/21	81	11/39	28
.936 and above	19/21	90	14/39	36

Hit and False Alarm Rates for NBits (Given the Selected Cutoff Values)				
Cutoff Value	Number of Hits	% of Hits	Number of False Alarms	% of False Alarms
1191 and above	12/21	57	2/39	05
1349 and below	15/21	71	4/39	10
1967 and below	18/21	86	6/39	15
2449 and below	19/21	90	12/39	31

4.2 RESULTS OF THE ROUTE DRAWING TASK

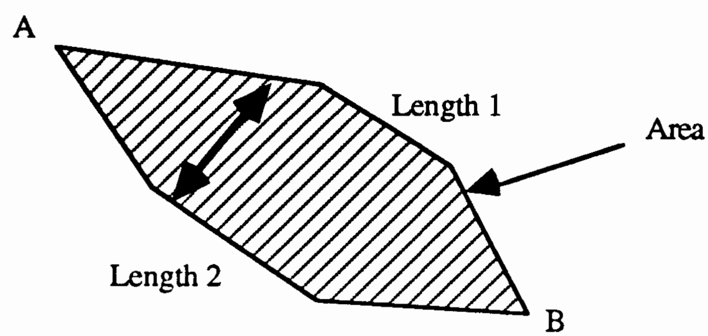
The subjects drew routes on both the raw images, and the images compressed to different levels. This allowed the effect of the compression on the route to be measured. The route selected is an objective measure of the effect of distortion on the subjects.

There is no "correct route" to select, thus it was not possible to look at how "good" the routes were in any objective way. As these subjects were all experienced pilots, any route that they selected is one that by definition, an expert might select. Instead of trying to compare the routes to some arbitrary "good route", the routes were selected without any distortion were used as the controls, and were compared to routes selected at different distortion levels. Several different measures (Normalized Route Difference, route length, and proximity to levels of precipitation intensity) were calculated to look for any differences in route selection. Each of these is discussed in Section 4.2.1.

4.2.1 Route Selection – Analysis

Several different performance measures were used to compare the routes that the subjects selected. These included Normalized Route Difference, Route Length, and Proximity to Levels of Precipitation Intensity.

Normalized Route Difference allows us to compare any two routes. The area enclosed by two routes with the same end points, is a function of both how different the routes are and the distance between the start and end point. This area is then normalized by the average of the two route lengths, to remove the effect of this distance from the calculation and is called the Normalized Route Difference. A Normalized Route Difference of zero means that the two routes are identical, while a large Normalized Route Difference indicates that the two routes are very different from each other. Normalized Route Difference is the average distance between two routes and is shown in Figure 4-10.



$$\text{Normalize Route Difference} = \frac{\text{Area}}{(\text{Length 1} + \text{Length 2}) / 2}$$

Figure 4-10. Normalized route difference.

Normalized Route Difference allows any two routes to be compared. Thus, it is possible to compare the route that was drawn at a given compression level to the route drawn on the same weather, but uncompressed. Even if there was no effect of compression, this measurement would generally not be zero, due to human inconsistency in choice of route and use of a mouse, so it is necessary to have a control. Since each uncompressed image was viewed twice, it is possible to compare these two routes that were selected and to use this value as a control for that weather image.

Route Length is another measure used to compare the routes selected. This is simply the total length (in nautical miles) of the route selected from Point A to Point B. Route length, as it varies as a function compression level, can then be assessed.

Proximity to Levels of Precipitation Intensity is the nearest that a given route passed to a given level of precipitation intensity. For example, a given route might come within 25 nmi of Level 3 precipitation intensity, 20 nmi of Level 2, and 10 nmi of Level 1. This measurement was used in two different ways. First, the proximity measurement was taken for each route as it was drawn on the weather. This represents how close the subject thought that they had gotten to each weather level. As the compression introduced some distortion to the images, this did not represent how close the selected route actually got to the weather. To get this second measurement, the route that was drawn on the compressed images was "pasted" onto the raw image, and the proximity calculation was done again. In this second case, it represents how close the route would have brought the subject to the actual weather, although they might not have realized this due to distortion introduced by the compression.

4.2.2 Results of Performance Measures of Route Selection

4.2.2.1 Results of Normalized Route Difference

As Normalized Route Difference is a method to compare pairs of routes, the proper pairs must be selected. The following analysis was performed for each image and each subject. First, the two uncompressed routes were compared. This provided a control value and showed how much the subjects routes will change from one replicate to the next without compression being an effect. Next, each of the routes that was drawn was compared to each of the raw routes that was drawn. The analysis involved the following thirteen comparisons, illustrated in Figure 4-11.

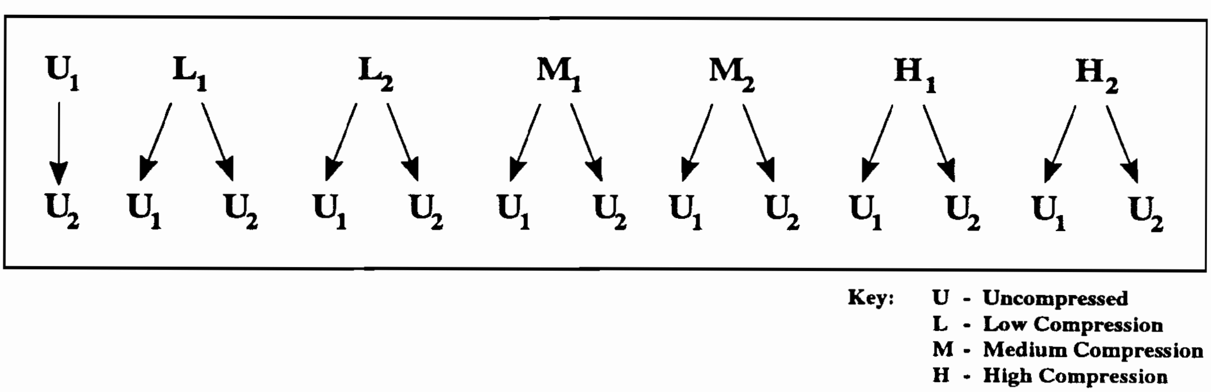


Figure 4-11. Pairwise matching of images for analysis.

This analysis was repeated for each of the images (14) and for each of the subjects (20). This led to a total of 13 x 14 x 20 = 3640 cases.

To assess the main effect of image number, compression level, and replicate on the above paired Normalized Route Differences, an ANOVA was performed. Image number (1 to 14) is a variable since each of the fourteen images was unique, i.e., showing different pre-recorded weather data. There were 4 values for Compression Group referring to the three compression groups studied: Low, Moderate, and High, with the addition of Uncompressed as the fourth group (each of these was compared to a raw image as discussed in the prior paragraph). Performance on the Uncompressed is used as a control / point-of-comparison with performance in response to the compressed images. Replicate (2) is a variable since each image in each compression group was shown two times. The ANOVA was a 14 (number of images) x 4 (number of compression groups, plus uncompressed group) x 2 (number of replicates) analysis. This was a within-subject analysis, i.e., each subject's performance under each condition was compared to his own performance under the various other conditions.

Results of the ANOVA indicated that there was a significant difference in Normalized Route Difference as a function of image number, $F(13, 3622) = 11.624$, $p < .001$. There was also a significant difference as a function of compression group, $F(3, 3622) = p < .003$. However, replicate was not significant, $F(1, 3622) = 0.154$, $p = .695$.

To identify which compression group was causing the significant difference in Normalized Route Difference, a separate ANOVA for Compression Group was performed, followed by a Scheffé Procedure. These results are shown in Table 4-4. To simplify, the ANOVA indicates that there is a significant difference attributable to compression group but it does not identify which compression group accounts for this difference. The Scheffé Procedure is used to identify the group wherein the significant difference lies. One could just perform a series of significance tests to determine whether each of the compression groups is significantly different from the other. When multiple comparisons are made, such as performing a succession of within-subject t-tests, each result on its own is well founded. However, inherent in performing a succession of analyses is the fact that as one increases the number of comparisons, the likelihood of finding "significance" increases. In an attempt to be conservative in interpreting results, an adjustment in the test levels (the level at which the probability of significant difference is accepted) can be made. This adjustment reflects the fact that, just on chance, a proportion of the tests will be labeled as significant. However, one should be aware that in doing so we will err on the conservative side, i.e., we will tend to screen out comparisons that really are significant. One way to make this adjustment is to use the Scheffé Multiple Testing Procedure. There are various multiple testing procedure, however, the Scheffé was selected since it is conservative for pairwise comparisons of means. It requires larger differences between means for significance than most of the other methods.

TABLE 4-4

A Comparison of Normalized Route Difference by Compression Level

Compression Group	Mean	Standard Deviation	Cases
Uncompressed	12.6	16.9	280
Low	12.3	16.7	1120
Moderate	13.7	16.9	1120
High	14.9	18.7	1120

In Table 4-4, the means for each compression group are listed. It was found that the mean difference between the first and second trial of the pair of uncompressed images was 12.6 pixels, which is equivalent to 12.6 nmi. This is the baseline variation between replicates. The results of the Scheffé Procedure indicated that the significance lies in the high compression group. The table shows that the mean normalized area between the high compression routes, and the corresponding raw routes is 14.9 nmi. That group was significantly different from both the low and moderate compression group. The Scheffé Procedure did not find the high compression group to be significantly different from the uncompressed group. It may be that a difference between the high compression and uncompressed group may not have proven to be statistically significant partly due to the rigorous criteria set by the Scheffé Procedure and partly due to the smaller number of samples in the uncompressed group (as indicated in Table 4-4, 280 cases) when compared to the high compressed group. It is also not clear whether the selected routes are operationally different from each other.

4.2.2.2 Route Length

The next measure that was used to compare routes as a function of compression was route length. To assess the main effect of compression level an analysis of variance (ANOVA) was performed. This was done with 20 subjects x 14 images x 4 compression levels x 2 replicates = 2240 cases. Route length was not expected to either increase or decrease with distortion. The results indicated that there was no significant difference in route length as a function of compression level, $F(3, 2236) = .459$, $p = .704$.

4.2.2.3 Proximity to Levels of Precipitation Intensity

One possible effect of compression was to change how close the pilots were willing to get to the depiction of precipitation intensities in each weather image. If the subjects were found to get much closer to high weather levels then compression could become a safety issue, on the other hand if subjects were found to stay much further away from some weather levels, then there might be increased costs such as fuel use, or time enroute. To determine if there was any change in the nearest approach to each weather level, some custom software was written to analyze each route that was drawn and to determine how close that route would have brought the airplane to each weather level. This software found the shortest distance from each point along the selected route to each weather level, in that image.

The analysis of the nearest approach to each weather level was performed in two different ways. First, the calculation was performed with the route selected, paired with the weather that the subject had seen. This represents how close the subject thought that he came to each precipitation level. Next, the same route was superimposed on to the corresponding uncompressed weather, and the same calculation was performed. In this case, the raw weather had a full six levels, even though the subjects saw all the higher levels displayed as red. As the compression introduced some distortion, it was possible for the nearest approach to be different than the subject might have thought. This calculation represented how close the subject would have actually come to each weather level, had they flown the selected route.

All of the fourteen images contained Level 1, 2, and 3. The images that the subjects were shown had the higher levels, 3-6 all shown as red, so there was no way for them to differentiate between levels. Table 4-5 lists the number of cases (images presented) with each of the levels. For the lower weather levels (1-3) there were 14 comparisons done, for each image and subject. The 14 were available, since there were 8 images shown, and one route for each image. The 6 routes that were drawn on the compressed images were then also analyzed on the raw images, leading to the total of 14. For the higher weather levels (4-6), since the subjects did not see them, it was only possible to do the analysis of the 8 routes that were drawn on the raw images, leading to 8 comparisons.

TABLE 4-5

Frequency of Levels of Precipitation Intensity in the Images

Level	Number of Cases	Number of Image Sets
1,2,3	3920	14 (all)
4	1280	8
5	800	5
6	160	1

To assess the effect of compression on nearness to precipitation intensity levels a multiple analysis of variance (MANOVA) was performed. This analysis provides how close pilots drew a route to each precipitation intensity depicted in an image and whether the proximity of that route to precipitation intensity varied as a function of the amount of compression seen by the pilots. We only looked at proximity to the three color-coded levels: Level 1, Level 2, and Level 3 through 6. We did not display Levels 4 through 6 separately, but they were incorporated with Level 3. There were 2240 cases included in the analysis, i.e., the total number of images seen (both compressed and uncompressed are included). The calculation for this figure is: 4 compression levels x 2 replicates x 14 images x 20 subjects. The proximity to weather Level 1 was found to be significant, $F(3, 2236) = 20.84$, $p < .001$.

After it was determined that there was an effect on the proximity to Level 1 precipitation, the next stage of the analysis was to determine which of the compression levels had caused this. A one-way ANOVA and Scheffé Procedure were run, and it was found that the significant difference lies in the high compression group. The relevant descriptive statistics are shown in Table 4-6. They show that in the high compression case, the subjects stayed 2.29 nmi away from the Level 1 precipitation, while in the other cases they stayed from 1.00 to 1.16 nmi away. The highly-compressed case was found to be significantly different from the low and moderate cases. However, the high was not significantly different from the uncompressed case. This was probably due to the small sample size of the uncompressed case, and the strong selectivity of the Scheffé procedure. Although this was statistically significant, it is felt that this is not operationally significant, but was probably due to either the ellipse not looking natural, so they were just avoided, or perhaps due to the removal of single Level 1 pixels along the route that were ignored in any case by the subjects.

TABLE 4-6

Average Distance From Each Precipitation Intensity Level by Compression Group

PRECIPITATION INTENSITY			COMPRESSION GROUPS					
Levels	Uncompressed		Low		Moderate		High	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
1	1.00	2.30	1.05	2.69	1.16	2.67	2.29*	4.54
2	6.62	6.60	6.87	6.64	6.53	6.27	6.56	6.26
3 and above	26.20	18.54	25.50	17.74	25.94	17.68	25.19	17.80

This analysis tells us how close pilots drew a route to each actual (data from uncompressed image) precipitation intensity level. This was done by superimposing the routes selected in response to the compressed image onto the uncompressed image, and then doing the same analysis (MANOVA) as was done above. There were 2240 cases included in the analysis, i.e., the total number of images seen (both compressed and uncompressed are included). There was no significant difference found in how near the pilots got to the actual (uncompressed image) precipitation intensities. Therefore, compression did not significantly affect how close pilots would have gotten to the actual weather.

This is an important finding. As compression distorts the images, one could hypothesize that subjects would be apprehensive of some of the images, if the distortion made them look worse, so they would be more conservative. Or that the distortion might make the weather appear to be less hazardous, so they would get closer to where there was actually severe weather. The above shows that neither of these were true. The subjects behaved in the same way, as measured by nearness to each level, whether the images were compressed or not.

4.2.2.4 Results of analysis of pilots' own rules versus actual behavior

In the "Pilot Exit Interview Questionnaire" the subjects were asked to give their own personal rules for how far away from each level of precipitation they stay when flying. These results are summarized in Figure 4-12. As there was general consensus on these results, conservative values based on these were used to generate an overall safe distance from each weather level rule. Subjects were generally willing to fly through Level 1, and many were also willing to fly through Level 2, and said they would stay further away from each of the higher levels. Next, the number of images that contained each of the levels was counted to find the number of opportunities that were available to violate these rules of thumb. The number of actual violations of these rules was counted. The ratio of number of violations to violation opportunities was also calculated and is shown in Table 4-7. These violations were not limited to just a few pilots, but instead were distributed over the whole range of subjects. (Note that for Level 5, there were not five, but really only four cases where it was possible to violate the rules of thumb, since in one case the Level 5 precipitation intensity was not near the route of flight. Also note that in the one image with Level 6, the weather was also not near any reasonable route of flight.)

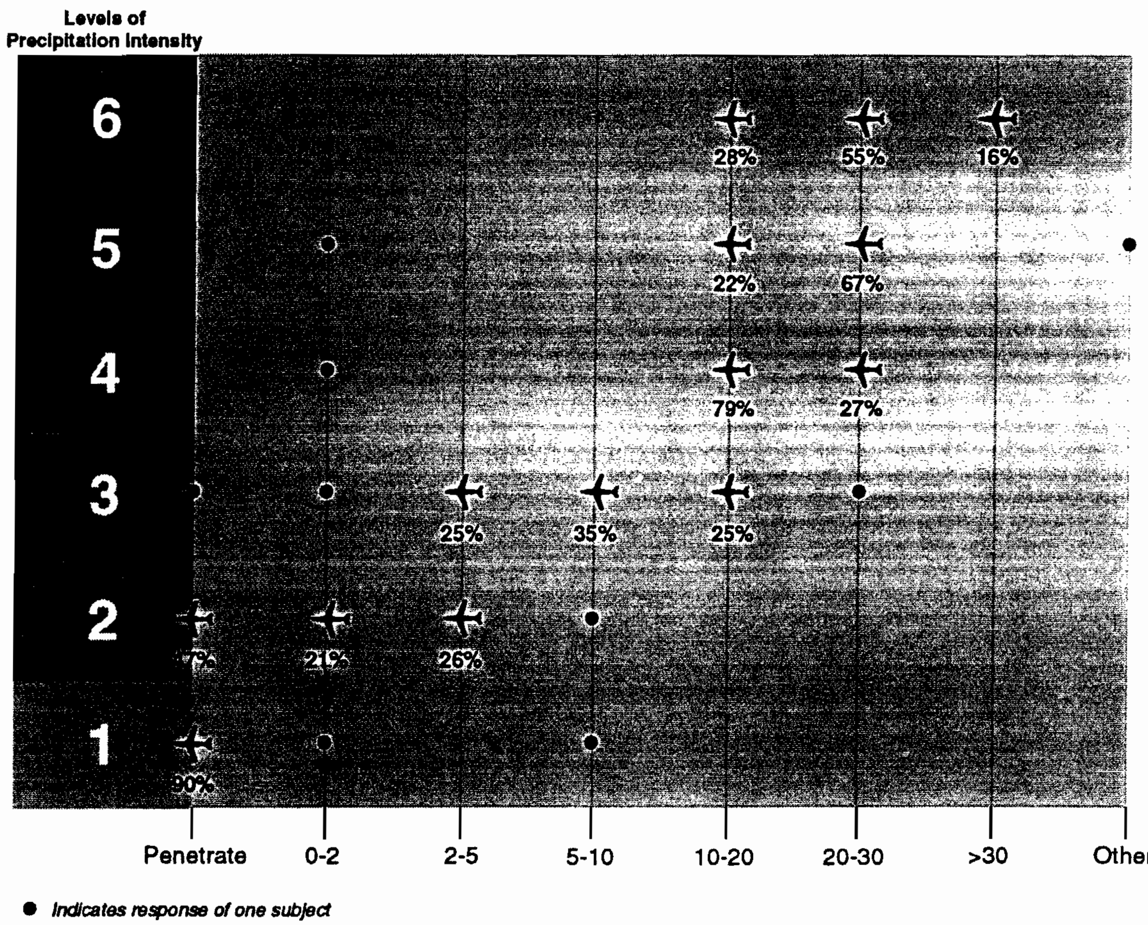


Figure 4-12. Reported rules of thumb for distance kept from each level of precipitation.

TABLE 4-7

Violations of Stated Rules of Thumb

Level of Precipitation	Safe Distance (nmi)	Number of Images	Number of Violation-Opportunities	Number of Violations	Percentage of Violations
3	> 5	14	2240	46	2%
4	> 10	8	1280	149	11.6%
5	> 20	5*	800	350	**43.8%**
6	> 20	1**	160	0	0

* 5 possible, but only 4 had level 5 near enough to make a violation reasonable.

** In this single case level 6 was far from any likely route of flight, so a violation was again not reasonable.

The subjects could not tell the difference between any of the high levels as they were all displayed as red. However the subjects were warned of this in the initial briefing. Additionally, many of the subjects had experience with weather radar which also is generally only 3 colors. This suggests that the same effect may happen when pilots fly with weather radar. The rules that they reported are essentially the same as those that are suggested to pilots. However, it seems that either pilots do not follow them, or that pilots assume that red is always only Level 3 (and nothing higher).

4.2.2 Results of Subjective Measures of Route Selection

Each time that the subject was asked to select a route, he also was asked to make a Go or No Go decision, and asked to rate the Hazard of the weather that was presented. For each of the four level of compression a total of 560 (14 images x 20 subjects x 2 replicates) Go/No Go decisions were made. The percentage of No Go decisions in each compression group was calculated and is included in Table 4-8.

TABLE 4-8

No Go Decisions Sorted by Compression Group

Compression Group	No Go Responses
Uncompressed	17%
Low	16%
Moderate	17%
High	15%

Go and No Go decisions were analyzed by assigning a value of 1 to Go Decisions and a value of 2 to No Go Decisions. An ANOVA was performed to determine whether Go/No Go decisions were significantly affected by the compression group and by the particular image to which the subject was responding. Therefore, the ANOVA tested for significance for the variables of compression and image. Results indicated a significant main effect for image, $F(13, 2184) = 81.686$, $p < .001$, but not for compression, $F(3, 2184) = .658$, $p = .578$. The significant main effect for image was expected, as each of the different images had very different types of weather from each other.

Hazard Ratings ranged from 1 (not at all hazardous) to 5 (very hazardous). Table 4-9 below lists the descriptive statistics for Hazard Ratings. In each compression group, a total of 560 responses were included, i.e., 14 images x 20 subjects x 2 replicates. The mean hazard rating and standard deviation sorted by compression group is listed in Table 4-9.

TABLE 4-9

Hazard Ratings Sorted by Compression Group

Compression Group	Mean	Standard Deviation
Uncompressed	2.88	1.03
Low	2.89	0.99
Moderate	3.00	1.04
High	2.95	1.01

An ANOVA was performed to determine whether hazard ratings were significantly affected by the compression group and by the particular image to which the subject was responding. Results indicated a significant main effect for both image and compression. For image, a significant difference was found, $F(13, 2184) = 113.422$, $p < .001$. For compression, a significant difference was found, $F(3, 2184) = 2.636$, $p < .05$. From considering the means above in Table 4-9, it is seen that all four means are relatively close to each other. The difference in means attributable to compression level appears when we consider uncompressed and low compression groups versus moderate and high compression groups. The difference between these two sub-sets is slight, but the statistical significance derived for compression group indicates that as compression increases, the hazard rating increases. Perhaps with more ellipses and generalized shapes, pilots lose some confidence in the compressed images, and may feel that what they are seeing does not represent the truth, ellipses may not look like real weather most times. Uncertainty of accuracy of depiction of weather situation would logically lead to an increase in hazard rating. Additionally, more weather that is organized into clear "cells" will tend to be more hazardous, and ellipses will tend to appear to be more like these severe "cells". However, even with an increase in hazard rating the mean is still no more than 3 which indicates moderate and not extreme hazard. The weather that was chosen for the experiment was all "Moderately Hazardous" which is what the subjects also reported. Although there is a statistically significant difference, it seems that there will be no operational difference.

4.3 RELATIONSHIP BETWEEN DIFFERENT VARIABLES

4.3.1 Relationship Between Subjective Ratings and Route Drawing Measures (Performance/Behavior)

The next phase of the analysis was to look for a relationship between what the pilots thought of the images, and how they performed on the same images. This was to determine if the subjective ratings were a good indicator of the subsequent performance changes. The subjective acceptability ratings, and distortion ratings were compared with the corresponding route length, and normalized Normalized Route Difference, by calculating correlation coefficients. The results of the correlation testing indicated that neither of the two types of subjective ratings (indicators of pilot opinion) were correlated with the performance measures. This is a linear correlation so what this indicates is that when subjective ratings increase or decrease there is no corresponding increase or decrease in pilot performance. The interpretation is that pilots may judge an image to be more or less distorted and acceptability but this judgment was not linearly related to what their actual performance in the route drawing task. We know that when images are highly compressed there is a significant difference in Normalized Route Difference. Perhaps we do not see a linear relation between this variable and the subjective ratings since once there is a change in behavior we are at a point where distortion and acceptability ratings have plateaued, i.e., distortion ratings are at an even high and acceptability ratings are at an even low.

4.3.2 Relationship Between Calculated Distortion and Subjective Measures

The correlation between RLE/NBits (independent variable) to a number of dependent variables was assessed, including the subjective rating measures of distortion and acceptability and the performance measures of route length and Normalized Route Difference. This analysis was done to determine if the calculated measure of image distortion was consistent with the ratings given by the subjects, and with the performance of the subjects. Table 4-10 lists the correlations and their significance.

TABLE 4-10

Correlations Between RLE/NBits and Subjective and Performance Measures

	Dependent Variable			
Independent Variable	Acceptability	Distortion	Route Length	Normalized Route Difference
RLE/NBits	-.7385*	.8945*	.0292	.0407**

* significant at less than or equal to .01

** significant at less than or equal to .05

To assess the relationship between the independent variable (descriptor of characteristics in images shown, how to categorize images) RLE/NBits and a number of dependent variables a series of two-tail correlations were performed. In analyzing Acceptability, mean ratings were used and in analyzing Distortion, the ranking explained in Section 4.1.1.2 was used.

As indicated in Table 4-10 above, the results of the correlation testing indicated that RLE/NBits was: 1) negatively correlated with acceptability rating, i.e., as RLE/NBits increases the acceptability rating increases, and 2) positively correlated with distortion rating, i.e., as RLE/NBits increases distortion ranking increases. No correlation was found between RLE/NBits and the performance measures: route length and Normalized Route Difference. This means that changes in acceptability and distortion ratings are related to the value of RLE/NBits. However, as indicated above, there was very little change in behavior so it might not be sensitive to changes in RLE/NBits presented. RLE/NBits is a predictor of subjective ratings, but not of changes in pilot performance (as used in study). As there was almost no indication of changes in how close pilots came to different weather levels, it was not expected that there would be any correlation with RLE/NBits so this was not tested.

4.4 POST-ROUTE DRAWING TASK QUESTIONNAIRE

The "Post-Route Drawing Task Questionnaire" (see Appendix B) was completed by each subject following completion of the Route Drawing Task. The responses provided information on pilot flying and weather experience, as well as, information on pilot weather-related decisions and the routes selected. In results of the questionnaire are discussed in this section.

Question 1. Have you piloted in the types of weather conditions represented in these scenarios?

YES
20 pilots/100%

Question 2. Have you piloted in light aircraft in the types of weather conditions represented in these scenarios?

YES
20 pilots/100%

Question 3. Were the weather images representative of weather conditions that might be encountered during IFR flight?

YES
20 pilots/100%

These first three questions show that all subjects had experience in the type of weather that was presented in the experiment.

Question 4. On the flights that you chose to fly (selected "Go") what is your estimate of the likelihood of encountering at least moderate turbulence?

This question was a multiple choice, with the number of subjects who selected each answer shown in Table 4-11. One subject commented that he stayed away from yellow (moderate precipitation), and would only fly in green (light precipitation). Another commented that he would need cloud top information to better be able to guess at the likelihood of turbulence. There was also a comment that there can be turbulence even in clear, and that in the weather that was presented he would expect turbulence on most flights. Another subject commented that he was planning his routes in order to avoid any turbulence. From these data, it seems that most of the subjects were perfectly willing to fly when they were expecting to encounter moderate turbulence. This is an unexpected finding since moderate turbulence can result in an extremely uncomfortable ride.

TABLE 4-11

Probability of Encountering Turbulence on Each Flight

Probability	Number Who Selected
0% - 5%	0
5% - 25%	7
25% - 50%,	9
50% - 75%	3
75% - 100%	1

<u>Question 5.</u> If you had been told that the majority of the area between Point A and B in the images was IMC, would your number of No Go decisions decrease, remain the same, or increase?

The intent of this question was to help understand how pilots avoid regions of precipitation. Instrument Meteorological Conditions (IMC) is a term that indicates that the aircraft is in clouds or fog, so the pilot has no visibility. Twelve subjects reported that the decisions would have stayed the same. Six subjects reported that the number of No Go decisions would have increased. This suggested that they had assumed that the weather along the routes was Visual Meteorological Conditions (VMC), and that they would have been able to use the visibility to make it a safer flight. Two subjects reported that the number of No Go decisions would have decreased. One of these subjects commented that "In some cases visual separation from weather was part of the "Go" decision." This suggests that the subject, and probably both subjects who selected "decrease", may have misunderstood the question since this comment indicates that the subject expected VMC conditions in some cases, and that since he did not have them he would have made more No Go decisions. Overall, the answer to this question suggests that just under half of the subjects either assumed that all of the weather in the study was IMC, or that it would not have mattered to them if it was.

<u>Question 6.</u> If you had been told that the majority of the area between A and B in the images was VMC, would your number of No Go decisions: decrease, remain the same, or increase?

One subject did not answer this question because he had not made any No Go decisions, although he did answer the prior question. Eight subjects reported that the decisions would have stayed the same. Six subjects reported the number of No Go decisions would have decreased, while five subjects reported the number of No Go decisions would have increase. Again it seems that some of the subjects may have misinterpreted this question because it seems that increasing the number of No Go decisions given VMC, seems counterintuitive. There were several comments such as "The pilot's eyes add a lot to the overall picture" that suggest that the pilots had assume IMC for the flight, but that pilots feel that VMC would have made the flights safer, and therefore allowed more Go decisions. This is the expected result.

4.5 EXIT INTERVIEW RESULTS

The Exit Interview is contained in Appendix C. Page One of the Exit Interview contained questions regarding the distortion rating task and Page Two contained questions regarding the acceptability rating task. Page Three contained a few general questions related to both tasks. Page Four contained one question regarding the pilot's rules of thumb for the nearest acceptable distance he would fly to each weather level. Responses to this question have already been reported in Section 4.2.2.3 "Proximity to Levels of Precipitation Intensity". Additional questions were included in the Exit Interview regarding piloting, but were not pertinent to the results of this study and instead used as an opportunity to survey pilots for data to be used in future work.

Responses to the Exit Interview are reported below, as well as a summary of subject comments, and experimenter interpretation.

4.5.1 Exit Interview Results for Distortion Task

Question 1a. Did you have enough practice trials?

YES	NO
20	0

Question 1b. Were you able to develop an internal scale of distortion during the practice block?

YES	NO
18	2

Question 1c. If "Yes", do you think that you used this scale consistently throughout the remaining trials?

YES	NO
18	2

Question 2. After the practice block, did you have any difficulty in assigning the distortion ratings?

YES	NO
20	0

From the above responses, it is seen that overall subjects had enough practice and were able to perform the task. To better understand why some subjects may have had lower correlation coefficients between distortion ratings in each repetition, as discussed in Section 4.1.1.1, the comments of those subjects were explored. The responses of three of the six subjects provide some insight into the difficulties they experienced in the distortion rating task. One subject said that he had enough practice, but then countered this response by responding that he was not able to develop an internal scale of distortion during the practice block, commenting: "the scale took time to develop fully", He also reported that he did not think that he used his scale consistently throughout the trials. He also said that after the practice block, he did not have any difficulty in assigning the distortion ratings but commented: "during the exercise, my rating scale became finer".

Another subject said that he had enough practice, was able to develop an internal scale during the practice, that he felt that he was consistent, and that he had no difficulty in assigning the distortion rating. However, he commented that he was uneasy about the fluid nature of the scale (referring to the fact that he could pick whatever rating scale he liked) and that he would have been more comfortable with a predetermined scale provided by the experimenter. Other subjects made similar comments. For these subjects, this was the first time they had encountered a task of this type.

A third subject said that practice time was adequate and that he was able to develop an internal scale, but that "my definition of scale may have become more consistent toward the end". It may be that for a minority of the subjects more practice time would have helped them to better define their scale and would have increased their consistency of response for repetitions of the same image.

Question 3. In rating distortion, did you find yourself using any particular features as rules or guidelines for giving a substitute image a high rating for distortion versus a lower rating?

Responses indicated that generally subjects based their rating on how closely the compressed image represented the uncompressed image. They tended to look for keeping the same shape of the weather. They were unhappy with what they called, "ovals, blobs, circles, ellipses", i.e., elliptical shapes caused by the compression algorithm. They also had problems with round circles and undefined borders and gave them a high distortion rating. Additionally, images with missing detail were given a higher rating as would be expected. Therefore, in general subjects did not like a loss of detail and in particular they did not like elliptical shapes.

4.5.2 Exit Interview Results for Acceptability Task

Question 1. Did you have any difficulty in assigning the acceptability rating?

YES	NO
7	13

Subjects referred to the fact that "the system could do better", having seen better representation of images in the study. It appears that subjects became very particular and said some images were unacceptable, but in reality would have been somewhat useful. In the instructions the subjects had been asked to report how useful the images were, compared to the raw image, and not as compared to having no image at all. Some subjects said that they would have preferred to have a "Fair" rating, between acceptable and unacceptable, since some images were on the borderline of acceptability. However, the decision to not include a "Fair" rating was made intentionally by the experimenters to force the subjects to make a decision. Some subjects mentioned that after the distortion task they were predisposed to disliking ellipses and that this made it difficult to rate the compressed image as useful based on information conveyed rather than just saying that images that contained ellipse are not good.

Question 2. Did you have any difficulty in switching from assigning distortion ratings (using your own internal scale) to acceptability ratings (selecting one of four ratings)?

YES	NO
4	16

The comments associated with the above numbers suggest that those who said "Yes" were not actually saying that they had difficulty switching from one task to the other. Instead the reason for saying "Yes" was that one would have liked more practice, two said they would have preferred a "fair" rating, and one said that he was predisposed to disliking circular/elliptical images and had to concentrate on fairly rating those types of images during the test, i.e., it took some effort.

Question 3. In rating acceptability, did you find yourself using any particular features as rules or guidelines for giving a substitute image an unacceptable versus an acceptable rating (other than the definitions given for each of the rating)?

Again as in distortion criteria, pilots did not like ellipses. Comments indicated that some subjects were able to move from just not liking them to considering whether or not they preserved the content of the information in the original image. As they progressed through the experiment, they began to consider whether the basic shape of the weather was maintained. Subjects reported that truthfulness and faithfulness of the reproduction of Level 2 and above were important for an acceptable rating. Comments indicated that some subjects felt that the exaggeration (caused by compression) of red (Level 3 and above) was an asset, making it easier to see these potentially hazardous areas. One subject also reported that he judged an image to be acceptable if both images had a similar "optimal" path through Level 1 precipitation.

Question 4. In rating acceptability, what were the key differences (if any) between images that rendered an image unacceptable?

Comments indicated that weather needed to be represented accurately or else it would affect the pilot's confidence in the weather depicted or cause a change in the route of flight. Generally, the subjects did not like ellipses – especially large ellipses. Many subjects referred to misrepresentation of weather by elliptical shapes and the loss of detail. Subjects generally reported that ellipses were tolerable if confined to small areas. Three subjects reported that the key difference between the uncompressed and the compressed image that would render the compressed image as unacceptable was that using the compressed image would result in a change in flight path.

Question 5. What specific flight tasks (if any) were you thinking of when considering the functionality of the compressed image?

Subject responses indicated that it would be used basically to avoid weather. Some subjects said that it would help to determine if they should penetrate weather. One subject mentioned terminal operation and enroute planning. Several subjects mentioned using GWS to determine safety of flight. One subject said for "picking my way through the weather" but did not specify any particular precipitation intensities level. Most comments showed that the most common use was for weather avoidance and flight planning.

4.5.3 Exit Interview Results from General Questions

These questions did not refer to a particular part of the experiment, but were of general interest.

Question 2. What information was most important in these images and how did compression affect this information?

As in responses to previous questions, this question brought comments about the importance of the faithful representation of weather. The subjects liked that there was no compromise in showing the most intense weather levels. The subjects also reported that showing details of the breaks between different weather regions was important, as was the position of severe weather. One subject commented that the most important part was being able to plan a flight without going through any Level 3 precipitation.

Question 3. Any other comments about the ratings, procedures, or the images themselves?

The subjects used this question to mention improvements in the system that they would like to see. Some subjects suggested including information on the direction of movement of the weather and cloud top information.

5. CONCLUSIONS

The study tested the effect of various levels of compression of GWS weather images on pilot perception of distortion, opinion of acceptability, and performance on a route selection task. The main objective of the study was to determine what amount of compression would be acceptable for transmission of images to an aircraft. It was found that, based on subjective reporting that low and moderate levels of compression, using the Polygon-Ellipse compression algorithm, were generally acceptable to pilots.

Several measures of image quality were identified as means for setting criteria to be used in determining if images are acceptable, and therefore should be transmitted up to aircraft. Some of these measures were output from the Polygon-Ellipse compression algorithm, so are not applicable to other compression methods. RLE/NBits (Run Length Encoding/the number of bits that the Polygon-Ellipse method used to encode the same image) was found to be the most promising predictor of subjective ratings of pilot acceptability.

Pilot performance, as measured by the route drawing task, was not significantly affected by low and moderate compression. High compression resulted in statistically significant differences in Normalized Route Difference and proximity to weak precipitation intensity. Pilot comments indicated that the subjects generally found the presence of ellipses to be unacceptable. While the algorithm preserves the fidelity of representation of precipitation intensity levels, the configuration of these levels were considered by subjects to be "too distorted", and not to appear to be "natural", when a high degree of compression was applied. The subjects were however generally accepting of the images that were compressed to a low or moderate degree in the compressed weather images. As a result of this study, a new compression algorithm was developed that does not introduce ellipses, and thus will hopefully be more acceptable to the pilot community.

The new algorithm is the Improved Weather-Huffman method of compression, which is a type of run length encoding. As weather tends to form in regions, the algorithm uses a Hilbert scan rather than a standard row-by-row raster scan. In this method, the scan pattern tends to follow weather regions, leading to longer runs. If this initial scan does not meet the bit limit, then several different steps are taken in different combinations. The algorithm can reduce the resolution of the image, using a pixel averaging technique. It may then throw short runs of lower level weather away. Finally, if it is able to reach the bit limit, but has extra bits available, then these are used to increase the resolution of specific small areas that will most benefit. Finally when the image is decompressed, the neighboring pixels are used to help expand each pixel appropriately. The effect of this is that the images appear to be much more natural, as shown in Figure 5.1.

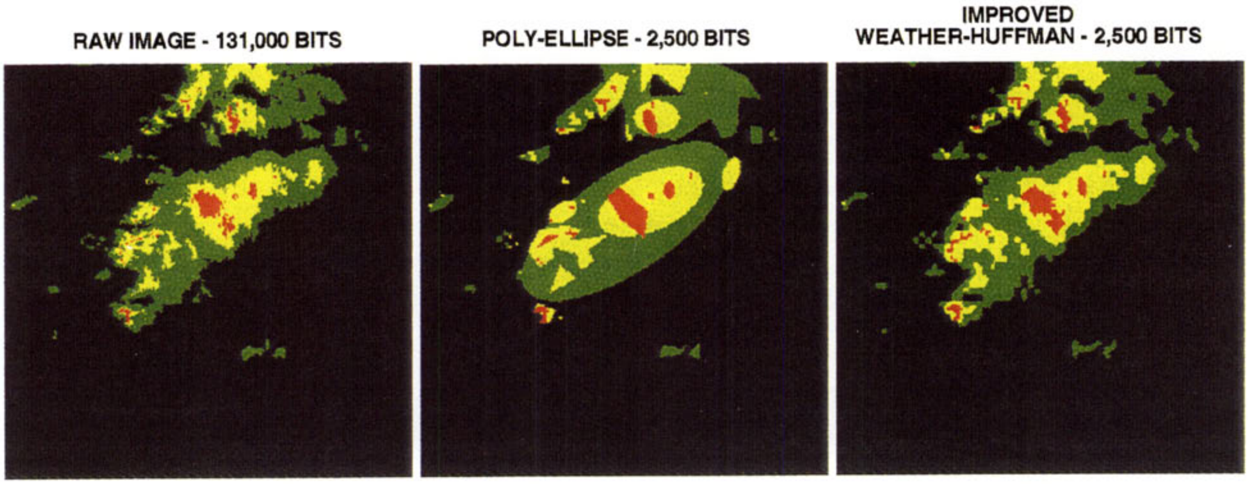


Figure 5-1. Comparison of uncompressed image to image compressed by Polygon-Ellipse algorithm and Improved Weather-Huffman algorithm.

The Improved Weather-Huffman method gives results that look different from the Polygon-Ellipse method. Generally, the Improved Weather-Huffman looks more natural. It does not force objects to be ellipses (as previously mentioned, subjects often found the presence of ellipses to be unacceptable). However, it is not able to compress the images as much as Polygon-Ellipse was able. Both algorithms generally give about the same number of pixels that are different from the original image.

As the best measure of image acceptability was found to depend strongly on the compression algorithm, it would be useful to repeat this study using the same images, but with the new compression method. That would allow for a predictor of image acceptability for the new algorithm. Additionally, it will allow for a better understanding of the effects of the new algorithm on pilot opinion and performance.

GLOSSARY

ADF	Automatic Direction Finder
ANOVA	Analysis of Variance
ATC	Air Traffic Control
ATP	Airline Transport Pilots
DF	Degree of Freedom
DME	Distance Measuring Equipment
FA	False Alarms
FAA	Federal Aviation Administration
FSS	Flight Service Station
FW	Flight Watch
GA	General Aviation
GWS	Graphical Weather Service
HSI	Horizontal Situation Indicator
IFR	Instrument Flight Rules
ILS	Instrument Landing System
IMC	Instrument Meteorological Conditions
MANOVA	Multiple Analysis of Variance
MSE	Mean Square Error
nmi	nautical mile(s)
NWS	National Weather Service
RLE/NBits	Run Length Encoding/Number of Bits
sd	Standard Deviation
VMC	Visual Meteorological Conditions
WSR	Weather Service Radar

APPENDIX A

THE GRAPHICAL WEATHER SERVICE
PILOT BACKGROUND QUESTIONNAIRE

(Subject ID #: ______________) D.O.B: ___________ Date: _________

1. Years as an active pilot _______

2. What type of aircraft do you usually fly? ____

3. License held (circle one): Private Commercial ATP

4. Ratings held (circle those that apply): Multi-Engine Instrument Sea Plane CFI CFII Helicopter Glider

5. Aircraft Experience (approx. hours): Single-Engine ___ Multi-Engine ___ Complex ___

   Actual Instrument hours ____ Simulated Instrument hours ____

6. Please estimate for the past year: # of Instrument approaches flown _____

   Actual Instrument hours ____ Simulated Instrument hours ____

7. During the past year, what percentage of your IFR time has been single pilot IFR? _______________

8. a. During the past year, what percentage of your intended IFR flights did you cancel due to weather? ____

   b. Please describe the weather conditions that would cause you to cancel your IFR flight.

9. Please circle the number that indicates how often you pilot for the following reasons:

	never	occasionally	sometimes	usually	always
recreation	1	2	3	4	5
business	1	2	3	4	5
commuter	1	2	3	4	5
airline	1	2	3	4	5

10. a. During the past year, what has been your most frequent point of origin _____________ and destination ________________.

    b. Please list some of your other destinations in the past year:

(Subject ID #:_________)

11. During the past year, what has been the approximate distance of your average IFR flight (in nmi)?_______

12. How familiar are you with flying in the New England Region?

1	2	3	4	5
Not at all Familiar	Somewhat Familiar	Moderately Familiar	More Than Moderately Familiar	Very Familiar

13. Navigational Equipment -- please circle those that are in the aircraft you usually fly:

VOR	NDB	Loran (IFR certified)	LORAN (non-IFR certified)
GPS	RNAV	DME	Inertial Navigation

other (specify:___________)

14. Please list any weather detection equipment on board the aircraft you usually fly (for example, weather radar, Stormscope):

15. Have you had any training in weather interpretation other than basic pilot training (for example, courses in meteorology)? If yes, please explain.

16. Please circle the number that indicates how often you get your pre-flight weather briefing in the following ways:

	never	occasionally	sometimes	usually	always
telephone FSS personnel	1	2	3	4	5
in person from FSS personnel	1	2	3	4	5
DUAT	1	2	3	4	5
other computerized service (please name: ___________)	1	2	3	4	5
Weather FAX/Jepp FAX	1	2	3	4	5
other (please name: ___________)	1	2	3	4	5

THANK YOU

APPENDIX B

Subject ID: ____________
Date: ____________

POST ROUTE DRAWING TASK QUESTIONNAIRE

1. Have you piloted in the types of weather conditions represented in these scenarios?

   Yes ☐ No ☐

2. Have you piloted in light aircraft in the types of weather conditions represented in these scenarios?

   Yes ☐ No ☐

3. Were the weather images representative of weather conditions that might be encountered during IFR flight?

   Yes ☐ No ☐

   Comment:

4. On the flights that you chose to fly (selected "Go") what is your estimate of the likelihood of encountering at least moderate turbulence? (Circle one)

   0-5% **5%-25%** **25%-50%** **50%-75%** **75%-100%**

   Comment:

5. If you had been told that the majority of the area between Point A and B in the images was IMC, would your number of No Go decisions (Circle one):

   Decrease **Remain the Same** **Increase**

   Comment:

6. If you had been told that the majority of the area between Point A and B in the images was VMC, would your number of No Go decisions (Circle one):

   Decrease **Remain the Same** **Increase**

   Comment:

APPENDIX C

Subject ID: ____________

Date: ____________

EXPERIMENT TO ASSESS SUBJECTIVE EVALUATIONS OF ALTERED WEATHER IMAGES

PILOT EXIT INTERVIEW

<u>Distortion Task</u>: All of the following questions refer to the Distortion Task.

1a. Did you have enough practice trials? Yes ☐ No ☐
Comment:

1b. Were you able to develop an internal scale of distortion during the practice block? Yes ☐ No ☐
Comment:

1c. If "Yes," do you think that you used this scale consistently throughout the remaining trials? Yes ☐ No ☐
Comment:

2. After the practice block, did you have any difficulty in assignment the distortion ratings? Yes ☐ No ☐
Comment:

3. In rating distortion, did you find yourself using any particular features as rules or guidelines for giving a substitute image a high rating for distortion versus a lower rating for distortion? Yes ☐ No ☐
Comment:

4. In rating distortion, did you take into account the scale of the images? Yes ☐ No ☐
Comment:

Acceptability Task: All of the following questions refer to the Acceptability Task.

1. Did you have any difficulty in assignment the acceptability ratings?

   Yes ☐ No ☐

   Comment:

2. Did you have any difficulty in switching from assigning distortion ratings (using your own internal scale) to acceptability ratings (selecting one of four ratings)?

   Yes ☐ No ☐

   Comment:

3. In rating acceptability, did you find yourself using any particular features as rules or guidelines for giving a substitute image an unacceptable versus an acceptable rating (other than the definitions given for each of the ratings)?

   Yes ☐ No ☐

   Comment:

4. In rating acceptability, what were the key differences (if any) between images that rendered an image unacceptable?

5. What specific flight tasks (if any) were you thinking of when considering the functionality of the compressed image?

6. In rating acceptability, did you take into account the scale of the images?

   Yes ☐ No ☐

   Comment:

General:

1. Based on your internal scale of distortion (developed during the Distortion Task), could you give a rough cut-off value, above which the images were generally unacceptable and below which the images were generally acceptable? ________

2. What information was most important in these images and how did the compression affect this information?

3. Any other comments about the ratings, procedures, or the images themselves?

Route Drawing Task: All of the following questions refer to the Route Drawing Task.

1. If you have any rules of thumb for the nearest acceptable distance you will fly to each weather level please check the appropriate box for each level. (For "other," please list the number of nm.)

Level	Penetrate	0-2 nm	2-5 nm	5-10 nm	10-20 nm	20-30 nm	other (nm)
1							
2							
3							
4							
5							
6							

2. If you must fly close to a region of precipitation on which side would you choose to fly?

   Please circle one: **North** **South** **No Preferance**

   Plesae circle one: **East** **West** **No Preferance**

   If you have any other rules of thumb, please explain:

3. Please indicate how large a region of precipitation, in the vicinity of your route, would have to be to effect your route planning? For each level of precipitation, check the diameter in nautical miles that applies. If the presence of a precipitation level would have no effect on your route planning, check "No Effect."

Level	Diameter of Region of Precipitation					
	0-2 nm	2-4 nm	4-10 nm	10-20 nm	20-30 nm	No Effect
1						
2						
3						
4						
5						
6						

4a. You are planning a flight. Consider each of the following conditions as being forecast along your planned route of flight. Please rate the relative significance of each in deciding to deviate from a straight line flight path: (Circle one)

	Irrelevant				**Very Significant**
light rain showers	1	2	3	4	5
moderate rain showers	1	2	3	4	5
heavy rain showers	1	2	3	4	5
thunderstorm activity	1	2	3	4	5
chance of light turbulence	1	2	3	4	5
chance of moderate turbulence	1	2	3	4	5
chance of severe turbulence	1	2	3	4	5
icing	1	2	3	4	5
lightning	1	2	3	4	5
hail	1	2	3	4	5
rapidly changing weather	1	2	3	4	5

4b. When making a decision to deviate from your planned route, how significant is the proximity of airports other than your destination, i.e., availability of a way out? (Circle one)

Irrelevant				**Very Significant**
1	2	3	4	5

5a. You are planning a flight. Consider each of the following conditions as being forecast along your planned route of flight. Please rate the relative significance of each in deciding to cancel a flight (making a No Go decision): (Circle one)

	Irrelevant				**Very Significant**
light rain showers	1	2	3	4	5
moderate rain showers	1	2	3	4	5
heavy rain showers	1	2	3	4	5
thunderstorm activity	1	2	3	4	5
chance of light turbulence	1	2	3	4	5
chance of moderate turbulence	1	2	3	4	5
chance of severe turbulence	1	2	3	4	5
icing	1	2	3	4	5
lightning	1	2	3	4	5
hail	1	2	3	4	5
rapidly changing weather	1	2	3	4	5

5b. When making a decision to cancel a flight, how significant is the proximity of airports other than your destination, i.e., availability of a way out? (Circle one)

Irrelevant				**Very Significant**
1	2	3	4	5

6. Please rate the relative importance of each type of information in pre-flight route planning (Circle one number per line):

	Not Important				**Very Important**
Radar Summary Charts	1	2	3	4	5
Pilot Reports	1	2	3	4	5
Surface Observations	1	2	3	4	5
Terminal Forecasts	1	2	3	4	5
Convective SIGMETs	1	2	3	4	5

7. Please rate the relative importance of each type of information available during flight for route planning (Circle one number per line):

	Not Important				**Very Important**
FSS Verbal Descriptions of Radar	1	2	3	4	5
Pilot Reports	1	2	3	4	5
Surface Observations	1	2	3	4	5
Terminal Forecasts	1	2	3	4	5
Convective SIGMETs	1	2	3	4	5
Stormscope	1	2	3	4	5
Airborne Weather Radar	1	2	3	4	5
Your Eyes (view out the window)	1	2	3	4	5

APPENDIX D

EVALUATIONS OF ALTERED WEATHER IMAGES
INSTRUCTIONS FOR PILOTS

BACKGROUND

MIT Lincoln Laboratory, through the sponsorship of the Federal Aviation Administration, is developing the Graphical Weather Service (GWS) that will provide graphical weather information to the pilot in the cockpit. Your participation in this study will provide valuable information that will be used in the development of this service.

You will be viewing weather radar images that contain real weather radar data acquired from the national array of weather radars. The data are collected from ground stations and is used by WSI Corporation (a commercial weather-information vendor) to build a mosaic national image. This resulting image is similar to what is seen on the TV news.

In this experiment, the complete image depicts weather over a 276-nm square region and no landmarks are shown. The weather image depicts color-coded <u>precipitation</u> information in a graphical format. Table 1 lists the three colors used to convey the precipitation intensities and presents some common features associated with each precipitation level (definitions are from the Airman's Information Manual.)

TABLE D-1

Color-Coded Precipitation Information

Color	Precipitation Intensity	Description
Green	Weak (Level 1)	Light to moderate turbulence is possible with lightning.
Yellow	Moderate (Level 2)	Light to moderate turbulence is possible with lightning.
Red	Strong (Level 3)	Severe turbulence possible, lightning.
	Very Strong (Level 4)	Severe turbulence likely, lightning.
	Intense (Level 5)	Severe turbulence, lightning, organized wind gusts. Hail likely.
	Extreme (Level 6)	Severe turbulence, large hail, lightning, and extensive wind gusts.

When this system is implemented, the weather images will be sent up to aircraft using some form of data link. Data link is a method by which digital information can be transmitted between ground stations and aircraft. Figure 1 illustrates a typical data link system.

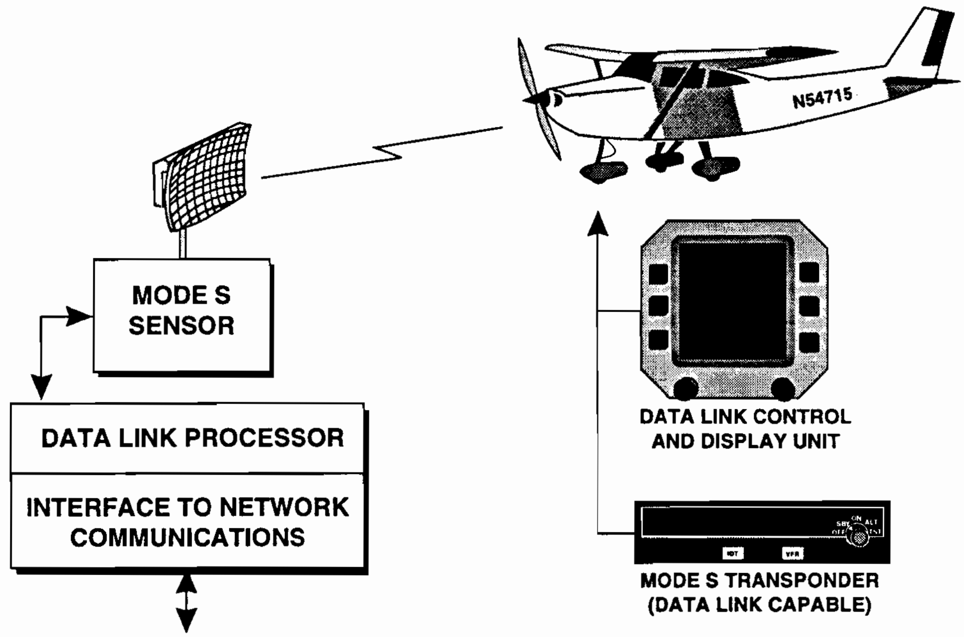


Figure 1. Mode S Data Link Components. The Mode S surveillance sensor at left provides a connection to ground-based data link services. The aircraft is equipped with a data link Mode S transponder and a Control / Display Unit (CDU).

Because weather images contain a large amount of information, they cannot be uplinked to airborne avionics in a timely manner unless the images are compressed. This means reducing the amount of information that is sent. This is done through a process that approximates precipitation regions as polygons or ellipses, resulting in a somewhat modified weather image. Images that are compressed to a variety of levels are shown in Figure 2. As can be seen, when the images are altered (compressed) there is some loss of image fidelity in the compressed image, since the original (uncompressed) image is not made up of exact polygons and ellipses. In this experiment, you will be working with these two types of images (compressed and uncompressed.)

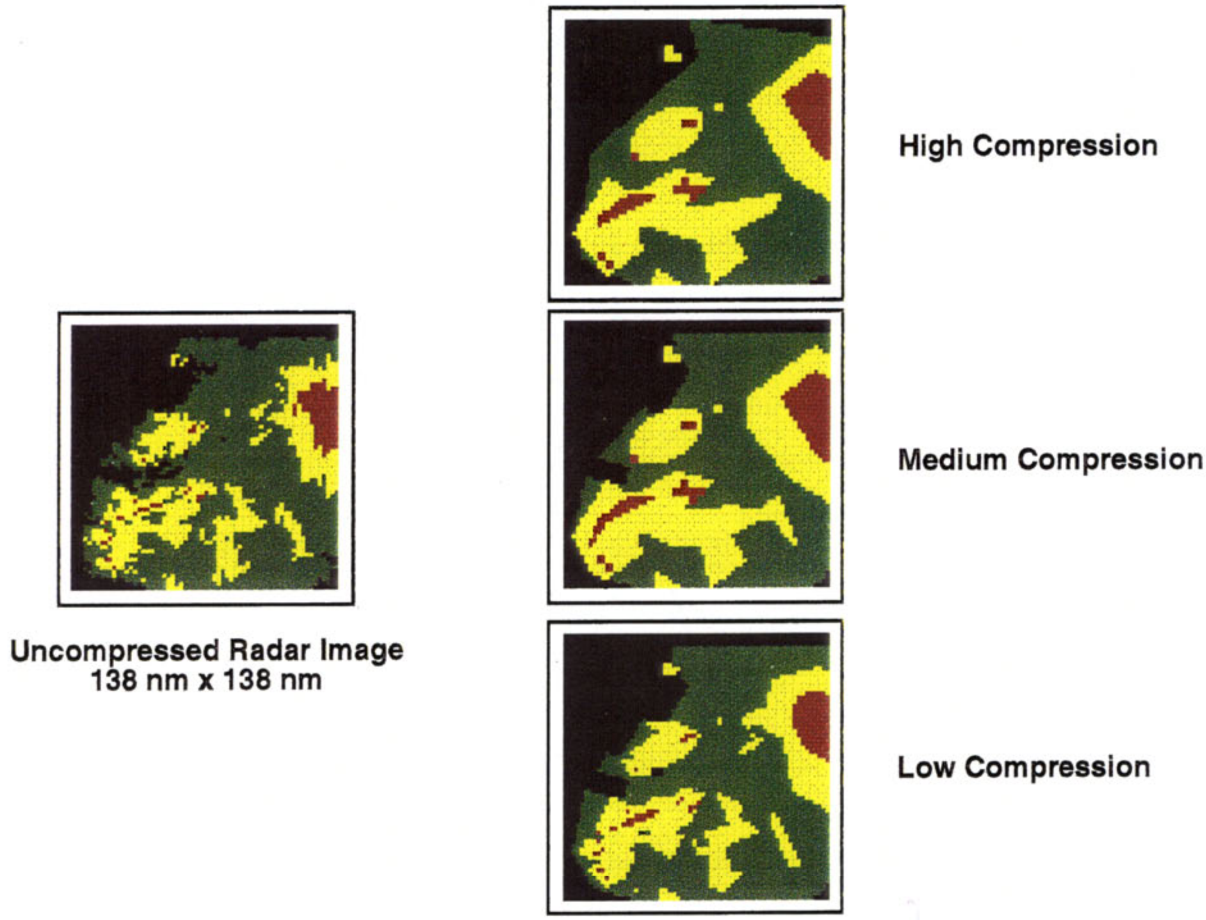


Figure 2. Image compressed to High, Medium, and Low compression levels.

OVERVIEW

There are two parts to the study, each containing "blocks" of trials. Each block contains a number of "trials." Each part involves different tasks and so each task begins with a practice block so that you may become familiar with the task. Below, each task is briefly described and then you will be receiving detailed instructions before beginning each task.

<u>Part One — Route Drawing</u>

You will see a series of weather images on a Macintosh Computer and for each image you will be asked to draw a route of flight from one designated point to another designated point. You will be asked to answer two questions about the flight.

<u>Part Two — Distortion Rating and Acceptability Rating</u>

In the Distortion Task, you will judge the quantitative amount of distortion in the compressed image. In the Acceptability Task, instead of rating distortion, you will judge the acceptability of the compressed image as a substitute for the uncompressed image. As previously mentioned, you will be given detailed instructions before beginning each task.

Table 2 provides an overview of Part One and Table 3 provides an overview of Part Two.

Table 2.

Overview of Part One

Block	Task	Number of Trials
Practice	Route Drawing	6
1	Route Drawing	28
2	Route Drawing	28
3	Route Drawing	28
4	Route Drawing	28

Table 3.

Overview of Part Two

Block	Task	Number of Trials
Practice	Distortion Rating	24
1	Distortion Rating	60
2	Distortion Rating	60
3	Distortion Rating	60
Practice	Acceptability Rating	5
4	Acceptability Rating	60

The total time for the experiment (which includes short breaks between blocks) will be approximately three and a half hours. Feel free to ask questions now or at any time during the experiment.

INSTRUCTIONS FOR ROUTE DRAWING TASK

You will see a series of GWS weather images on a Macintosh Computer. For each image, you are asked to draw a route of flight from one designated point to another designated point. You are also asked to answer two questions about the flight. You may complete the route drawing and questions in whatever order you wish. But you must complete both the route drawing and questions for the image on the screen before proceeding to the next image.

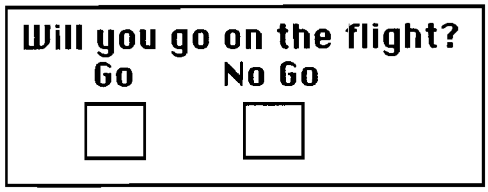
During this task, we ask that you make the following assumptions regarding your aircraft, intentions, and weather:

<u>Your aircraft</u>	Your aircraft is a light, single-engine piston aircraft, such as a Cessna 172. Assume that you have full fuel for this flight. Assume that you are planning to travel with one passenger who is not a pilot. The aircraft has two VOR receivers, one with RNAV. It has an ADF and does not have LORAN, Stormscope, or weather radar. It is equipped for ILS and has no autopilot or HSI.
<u>Your intention</u>	We ask that you plan a route that reflects your usual consideration of the balance between safety and convenience. It is important for you to reach the destination, but it is not a matter of life or death. You should be concerned with getting to the given destination in a timely fashion, while maintaining flight safety.
<u>The weather</u>	The weather information you have is limited to what appears on the GWS Image. You will not have access to any other information sources. All the weather that is shown is actual weather that was recorded during the summer months in New England. The weather is depicted north-up. The time of the weather image should not be a consideration in your decision, so you may assume that each image is current. Although in actual flight the weather is changing over time and moving, and you would be thinking of where the weather will be when you reach a certain point, in this task, assume that the weather depicted is stationary.

Imagine that you are flying from point A to point B. Given the weather depicted on the screen, draw your route of flight on the Macintosh screen. The following paragraph describes how to draw a route on the Macintosh. We will also demonstrate drawing a route and provide some practice trials so that you will be familiar with drawing a route on the Macintosh.

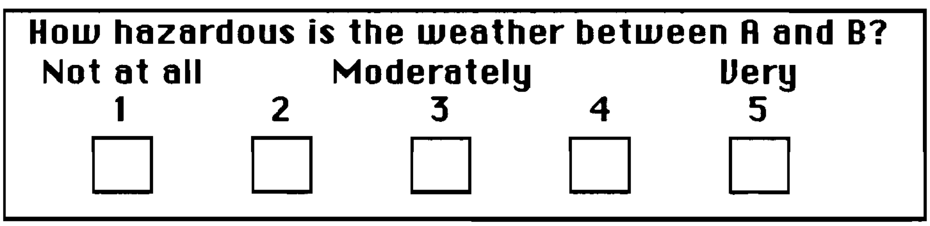
When you click the mouse button, a waypoint is defined at the location you selected. You can still move that waypoint until you release the button. You can delete a waypoint by putting the arrow cursor on it and clicking down on the mouse button, and then moving the arrow cursor to the DELETE POINT button and clicking on it. You can move a waypoint by placing the arrow cursor on it and clicking the mouse button, and then "dragging" the waypoint to the desired location. Finish the route by selecting the destination as the final waypoint. You may add a new waypoint by clicking on any part of your route line, then dragging the new waypoint to a desired location.

At any time, decide whether you will go on the flight, in these weather conditions, in a light single engine aircraft, and then select either Go or No Go.



Make your response by placing the arrow cursor in or near the appropriate box and click the mouse button. A check mark will appear in the box that you have selected. To change your selection, simply click on your new choice and the check mark will re-draw automatically.

At any time, assess the amount of hazard of the weather depicted between A and B and select a rating from one of the five following responses:



Make your response by placing the arrow cursor in or near the appropriate box and click the mouse button. A check mark will appear in the box that you have selected. To change your selection, simply click on your new choice and the check mark will re-draw automatically.

To proceed to the next trial, use the mouse to click on the DONE button or press either RETURN or ENTER key.

There are no right or wrong answers. We would like to understand how pilots select routes in relation to weather. Try to select a route and answer the accompanying questions for each image individually. On some trials, you will see images that are familiar to you from previous trials. Instead of trying to remember an earlier route and accompanying responses, consider the image on the screen and respond. You may ask questions now or at any time during the practice trials.

INSTRUCTIONS FOR DISTORTION TASK

As noted earlier, the compressed image is an altered version of the uncompressed image. We are interested in your judgment of the degree to which the compressed image has been distorted relative to the uncompressed image. Your task is to assign a numerical value to the level of distortion that you perceive, keeping in mind that an image depicts weather information. Remember that you are basing your rating on the quantitative amount of distortion of the compressed image and not on the usefulness and functionality of the compressed image. You will be rating functionality later in the Acceptability Task.

You should judge the distortion of the compressed/altered image in relation to the uncompressed image. For this purpose, the uncompressed image has been assigned a distortion rating of "10" arbitrarily. Thus, if you feel that the compressed/altered image does not distort the weather picture *at all* (in terms of being a substitute for the uncompressed image), you should enter a response of "10". You should assign higher numbers to more distorted images. You may respond with any numerical value (greater than or equal to "10", the value of the uncompressed image). Try to make the numbers proportional to the distortion of the compressed image as you perceive it. For example, if you rated one compressed image with a "20", and you feel that the next compressed image is twice as distorted as the previous one (each relative to its own uncompressed image), you should give the new image a rating of "40".

NOTE: You may assign ANY number and there is no upper-limit on the number that you assign. Thus, on a given trial, if you feel that the compressed image is distorted more heavily than any you have seen up to that point, you should assign it a higher rating than any you have assigned previously.

Try to judge each pair of images independently. On some trials, you will see images that are familiar to you from previous trials. Instead of trying to remember your earlier rating, consider the pair of images on the screen and select a rating.

The first block of 24 trials you will see is for practice. You should use the practice trials to set up an internal scale of distortion for yourself. On the first few trials your ratings will be fairly arbitrary, but you will be getting a sense of the range of distortions that you will see. After the practice trials, you should try to use your internal scale of distortion in a consistent manner for the remaining trials.

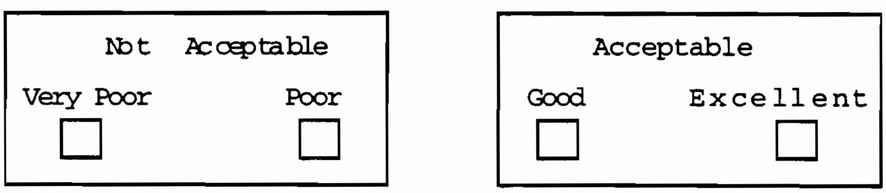
There are no right or wrong answers. We would like to understand how you judge the image distortions. You may ask questions now or at any time during the practice trials.

INSTRUCTIONS FOR ACCEPTABILITY TASK

In the Acceptability Task, you will again see pairs of images. In the Distortion Task, you were asked to assign a numerical value to the amount of distortion present in the compressed/altered image. For the Acceptability Task, you are asked to answer the question: How acceptable is the compressed/altered image as a replacement for the uncompressed image? This question should be answered in the context of typical general aviation flight in a single or light twin-engine aircraft. Remember that you should judge "acceptability" in terms of the compressed image's functionality for the flight task as compared with the functionality of the uncompressed image for the flight task. Remember that you should rate acceptability, regardless of the degree of image distortion.

You should not judge the acceptability of the compressed image in comparison to a situation where no graphical weather image is available to the pilot. Also note that "acceptability" does not refer to the advisability of safety of flight in the depicted weather.

Your judgment of acceptability should be chosen from one of the four following responses:



To make your response, place the arrow cursor in or near the appropriate box and click the mouse button. A check mark will appear in the box that you have selected. To change your selection, simply click on your new choice and the check mark will re-draw automatically. After you have made your selection, please tell the experimenter why you selected that rating. Then to proceed to the next trial, use the mouse to click on the DONE button or press either RETURN or ENTER key.

Definitions of the choices for the Acceptability Task are given below. All of these refer to the GA flight environment (i.e., in a single or light twin-engine aircraft.) You may refer to these definitions at any time during the block of acceptability-task trials.

Not Acceptable/Very Poor. There are major functional differences between the two images. The deficiencies in the compressed/altered image make its utility for GA operations very low.

Not Acceptable/Poor. There are functional differences between the two images. The deficiencies in the compressed/altered image limits its utility for GA operations.

Acceptable/Good. There are no major functional differences between the two images. The compressed/altered image has no serious deficiencies and is useful for GA operations.

Acceptable/Excellent. There are no functional differences between the two images. The compressed/altered image has no deficiencies and is as useful for GA operations as the uncompressed image.

APPENDIX E

INTERNAL CONSISTENCIES IN DISTORTION RATINGS

Subject Number	Correlations Between Repetitions 1 & 2	1 & 3	2 & 3	Minimum Rating	Maximum Rating
1	0.84	0.85	0.89	10	80
2	0.82	0.76	0.90	20	440
3	0.67	0.73	0.73	10	40
4	0.91	0.90	0.90	11	100
5	0.55	0.49	0.62	15	60
6	0.66	0.71	0.71	15	45
7	0.90	0.85	0.90	15	50
8	0.81	0.82	0.87	10	50
9	0.88	0.87	0.93	15	70
10	0.75	0.73	0.81	15	50
11	0.85	0.89	0.88	11	35
12	0.84	0.84	0.91	11	70
13	0.85	0.87	0.90	12	80
14	0.80	0.82	0.88	15	200
15	0.82	0.87	0.81	20	90
16	0.71	0.76	0.86	10	45
17	0.88	0.78	0.83	15	75
18	0.74	0.80	0.79	10	25
19	0.64	0.72	0.78	15	70
20	0.84	0.83	0.87	13	80

APPENDIX F

MEAN DISTORTION RANKINGS FOR COMPRESSED IMAGES
(For All Subjects Combined)

Image Number	Compression Level	Mean Rank	Standard Deviation	Number of Bits
14	Low	1.6	0.9	4631
7	Low	4.5	3.2	7443
2	Low	5.1	2.7	7500
10	Low	5.3	3.6	6124
5	Low	7.5	6.3	4472
1	Low	8.1	4.0	4880
14	Moderate	9.4	4.7	2378
11	Low	9.9	4.1	7317
9	Low	9.9	4.7	7464
20	Low	10.1	6.0	9758
19	Low	10.7	3.8	9748
12	Low	11.0	4.3	6086
17	Low	11.1	6.6	4495
18	Low	15.9	5.3	4726
3	Low	18.0	7.9	6061
8	Low	19.2	5.4	6180
16	Low	19.3	9.1	2486
6	Low	19.5	4.8	4385
15	Low	20.9	5.5	2599
10	Moderate	21.6	6.4	2456
9	Moderate	21.8	7.0	3462
7	Moderate	22.8	10.1	4332
15	Moderate	23.7	6.8	1224
13	Low	24.6	9.1	5303
20	Moderate	24.7	6.6	5429
18	Moderate	25.6	6.8	2534
4	Low	26.5	7.7	3987
19	Moderate	27.0	4.8	5196
12	Moderate	28.4	7.5	2294
5	Moderate	29.9	7.7	1422
16	Moderate	31.6	6.4	999
1	Moderate	32.9	8.6	1685

MEAN DISTORTION RANKINGS FOR COMPRESSED IMAGES
(continued)

Image Number	Compression Level	Mean Rank	Standard Deviation	Number of Bits
11	Moderate	33.6	6.5	1983
19	High	34.4	5.5	2406
20	High	35.1	6.7	2628
14	High	35.3	9.5	783
8	Moderate	36.1	6.4	2257
4	Moderate	36.2	5.4	2449
13	Moderate	36.5	5.0	2537
3	Moderate	37.5	5.8	2378
16	High	38.5	5.8	606
2	Moderate	38.9	5.4	1244
17	High	39.2	6.3	1163
6	Moderate	43.6	8.3	1191
17	Moderate	43.7	4.9	2560
7	High	45.4	6.2	1300
18	High	47.8	5.5	1113
10	High	47.9	5.1	963
9	High	48.4	3.9	1967
1	High	50.0	4.0	952
11	High	50.3	3.7	1297
5	High	52.5	4.1	840
4	High	52.7	3.5	1582
15	High	52.9	3.7	687
12	High	53.3	3.1	1095
2	High	53.4	3.5	742
3	High	55.7	4.4	1349
13	High	57.0	2.4	1486
8	High	58.1	1.6	1165
6	High	58.2	2.4	960

APPENDIX G

MEAN ACCEPTABILITY RATINGS
(For All Subjects Combined)

Image Number	Compression Level	Mean Rating	Standard Deviation	Number of Bits
6	High	1.50	0.612	960
3	High	1.65	0.582	1349
8	High	1.65	0.671	1165
3	High	1.70	0.653	1486
4	High	1.85	0.602	1582
2	High	1.95	0.780	742
2	High	2.00	0.848	1095
5	High	2.05	0.970	840
8	High	2.10	0.765	1113
9	High	2.30	0.820	1967
5	High	2.30	0.820	687
6	Moderate	2.45	0.697	1191
0	High	2.55	0.905	963
1	High	2.60	0.902	1297
1	High	2.65	0.895	952
7	High	2.70	0.562	1163
6	High	2.75	0.653	606
2	Moderate	2.85	0.688	1244
7	Moderate	2.85	0.765	2560
7	High	2.90	0.937	1300
4	Moderate	2.90	0.658	2449
0	High	2.90	0.658	2628
3	Moderate	2.95	0.621	2537
8	Moderate	2.95	0.524	2257
3	Moderate	3.00	0.577	2378
1	Moderate	3.05	0.524	1983
9	High	3.05	0.524	2406
8	Moderate	3.10	0.459	2534
4	High	3.15	0.688	783
9	Moderate	3.15	0.419	5196
3	Low	3.20	0.535	5303
5	Moderate	3.25	0.452	1422
1	Moderate	3.25	0.452	1685
2	Moderate	3.25	0.452	2294
5	Moderate	3.25	0.562	1224
6	Moderate	3.25	0.452	999

MEAN ACCEPTABILITY RATINGS
(continued)

Image Number	Compression Level	Mean Rating	Standard Deviation	Number of Bits
0	Moderate	3.25	0.452	5429
9	Moderate	3.30	0.562	3462
4	Low	3.30	0.478	3987
8	Low	3.30	0.478	6180
5	Low	3.35	0.478	2599
8	Low	3.40	0.496	4726
0	Moderate	3.40	0.496	2456
6	Low	3.40	0.507	4385
6	Low	3.40	0.496	2486
7	Moderate	3.45	0.513	4332
3	Low	3.55	0.507	6061
2	Low	3.60	0.507	6086
7	Low	3.65	0.478	4495
4	Moderate	3.70	0.478	2378
9	Low	3.70	0.478	7464
9	Low	3.70	0.478	9748
1	Low	3.80	0.419	4880
5	Low	3.80	0.419	4472
1	Low	3.80	0.419	7317
0	Low	3.85	0.375	9758
7	Low	3.95	0.229	7443
0	Low	4.00	0.000	6124
2	Low	4.00	0.000	7500
4	Low	4.00	0.000	4631

APPENDIX H

CORRELATION COEFFICIENT OF RAW DISTORTION RATING WITH ACCEPTABILITY RATING

Subject Number	Correlation Coefficient
1	-0.857
2	-0.721
3	-0.737
4	-0.814
5	-0.656
6	-0.695
7	-0.601
8	-0.699
9	-0.813
10	-0.616
11	-0.856
12	-0.879
13	-0.826
14	-0.675
15	-0.724
16	-0.727
17	-0.661
18	-0.682
19	-0.503
20	-0.763

REFERENCES

[1] Lind, A.T., A. Dershowitz, and S.R. Bussolari, "The Influence of Data Link-Provided Graphical Weather on Pilot Decision Making," MIT Lincoln Laboratory: Lexington MA, ATC (Air Traffic Control) Project Report-215, 1994.

[2] Gertz, J.L., "Weather Map Compression for Ground to Air Data Links," *Proceedings of the Aeronautical Telecommunications Symposium on Data Link Integration* (May 15-17, 1990): 131-40.

[3] Bussolari, S.R., "Data Link Applications for General Aviation," in MIT Lincoln Laboratory Journal, Vol. 7, No. 2, Lexington, MA. 1994.

[4] Engen, T., Psychophysics: II. Scaling Methods. in J.W. Kling & L.W. Riggs (Eds.) Woodworth and Schlosberg's Experimental Psychology (3rd ed.), pp. 47-86. New York: Holt Rinehart and Winston.

[5] Sheridan, T.B. and W.R. Ferrell, Man-machine Systems: Information, Control, and Decision Models of Human Performance, MIT Press, Cambridge, MA, 1974.