NASA/TM—2009–215368

On Organization of Information: Approach and Early Work

Asaf Degani
Charles Jorgensen
David Iverson
Michael Shafto
Leonard Olson

NASA Ames Research Center

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California, 94035-1000

May 2009

The NASA STI Program Office . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the Lead Center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA's counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or cosponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results . . . even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at *http://www.sti.nasa.gov*

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA Access Help Desk at (301) 621-0134

- Telephone the NASA Access Help Desk at (301) 621-0390

- Write to:
  NASA Access Help Desk
  NASA Center for AeroSpace Information
  7115 Standard Drive
  Hanover, MD 21076-1320

NASA/TM—2009–215368

On Organization of Information: Approach and Early Work

Asaf Degani
Charles Jorgensen
David Iverson
Michael Shafto
Leonard Olson

NASA Ames Research Center

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California, 94035-1000

May 2009

Composed in 12 pt Adobe Caslon Pro Regular
Printing and Binding by Reproduction Services, Ames Research Center
Designed by the good people of TopSpin Design Works.

Available from:

NASA Center for AeroSpace Information
7115 Standard Drive
Hanover, MD 21076-1320
(301) 621-0390

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
(703) 487-4650

Table of Contents

Abstract vii

1. Introduction 1
- The Party Scene 2
- A Conceptual Framework 3

2. The Framework in Context 7
- Analysis 13
- Summary 18

3. A Tale of Two Maps 19
- Abstraction 21
- Integration 22
- Organization 24
- Summary 25

4. Analysis of Pilot-Autopilot Interaction 27
- Field Study Data 27
- Linear Regression Analysis 29
- Canonical Correlation Analysis 32
- Abstraction and Representation of Data 33
- Integration and Presentation of Information 36
- Organization of Multiple Patterns into a Whole 37
- Summary 44

5. An Integrated Engine Display 46
- Data 46
- Data Abstraction 48
- Application to Aerospace Systems 50
- Abstraction of Engine Parameters 52
- An Old Tradition with a New Application 53
- Analysis of a 15th Century Tilework 57
- From Tilework to Cockpit Displays 61
- Arrangement of Engine Information: Composites 62
- Arrangement of Engine Information: Linking Related Subsystems 64
- Below the Lines 67
- Summary 70

6. Conclusions 71
- Acknowledgements 72
- Observations and Inferences 73

Appendix A. Canonical Correlation Analysis: A Literature Review 74

Appendix B: Signal Transformation to Support Abstraction and Integration 76
- Signal Analysis and Alphabets 76
- Approach 78
- Signal-specific Coding 80
- Representation of Signal Alphabets 82
- Creating Geometrical Structures from Signal Alphabets 83
- A General Approach for Creating 2-D Compositions 85
- Summary 88

References 89

Abstract

In this report we describe an approach for organizing information for presentation and display. The approach stems from the observation that there is a stepwise progression in the way signals (from the environment and the system under consideration) are extracted and transformed into data, and then analyzed and abstracted to form representations (e.g., indications and icons) on the user interface. In physical environments such as aerospace and process control, many system components and their corresponding data and information are interrelated (e.g., an increase in a chamber's temperature results in an increase in its pressure). These interrelationships, when presented clearly, allow users to understand linkages among system components and how they may affect one another. Organization of these interrelationships by means of an orderly structure provides for the so-called "big picture" that pilots, astronauts, and operators strive for.

This report begins with an analysis of an aviation incident involving a modern airliner, where the flight crew had difficulties understanding the physical interrelationships that existed among several engine and fuel system indications provided on the cockpit display. Analysis of the incident highlights some of the limitations in the design of information systems with respect to organization of information and user understanding of automation processes. We then analyze the map of the London Underground to understand successful examples of simplification and abstraction, integration of information, and nonlinear organization of the display to help viewers better understand the system as a whole. The next section describes the application of these concepts to the design of a graphical display for a statistical analysis of pilot-automation interaction. The last section describes the design of an experimental engine display for a research helicopter that integrates information from engine parameters and organizes them in the context of other subsystems. In Appendix A we provide the technical background for a statistical technique (canonical correlation analysis) used to analyze deviations from expected patterns in pilot-automation interaction. In Appendix B we detail a new approach for transformation of signals, and creation of an alphabet to represent sensor signals in order to foster detection of anomalies in data streams. We conclude with several inferences about information organization and offer some insights for those interested in pursuing the challenge of developing a theory for information integration and organization.

Keywords: *interface design, abstraction of data, integration of information, organization of information.*

On Organization of Information: Approach and Initial Work

Asaf Degani, Charles Jorgensen, David Iverson, Michael Shafto & Leonard Olson

NASA Ames Research Center, Moffett Field, CA 94035-1000

1. Introduction

Modern control and information systems, such as automated devices, monitoring and decision support systems, and information-gathering tools, contain and provide extensive amounts of data that are available for analysis and display. In aerospace applications, for example, information about the state of the vehicle is vital, as end users (pilots, dispatchers, astronauts, and mission controllers) are usually isolated and removed from the actual working of the machine or the system. Modern aircraft and spacecraft, such as the Boeing B-787, Airbus A-380 and A-350, the current space shuttle, and the new Orion spacecraft, are wired with thousands of sensors, sending multitudes of signals about the state of the craft's machinery and the surrounding environment. Similarly, onboard automatic control systems generate thousands of signals about internal events and computations.

With the introduction of Integrated Vehicle Health Monitoring (IVHM) technology, even wider sensor coverage will be available, allowing for meticulous computation and analysis of vehicle data. In the not-so-distant future, it is even foreseeable that with nanotechnology many components, down to critical bolts, will also transmit signals about their state. This capability will not be limited to hardware and software; it will also include physiological information (e.g., about astronauts' health on long-duration missions). While it is quite clear that future systems will be sensor laden, providing huge amounts of data for processing and display, it is still unclear how to best provide this wealth of data to aid users in monitoring the system, understanding and predicting its behavior, and making appropriate decisions. Given the limited display "real estate" in modern cockpits and contemporary approaches to

information organization—e.g., the "one-sensor, one-indicator" paradigm, the assignment of subsystems to separate menu-based display pages, and the focus on alphanumeric representation to convey data and information—user interfaces are quickly becoming the bottleneck of information flow, with potentially deleterious effects on efficiency and safety.

The approach advocated here stems from the assertion that we, as humans, have the innate capability to take in very large amounts of sensory cues, abstract and integrate them without much conscious effort, and then internally organize the information into a whole—so as to understand the state of the world around us, predict and anticipate its course, and then take appropriate actions.

The Party Scene

You slowly step out of your car and walk reluctantly into the large hotel. It's late and you're tired from your day's work—but an obligation is an obligation. Inside, you wander randomly till you gravitate toward the music and arrive at the crowded ballroom, and although you had intended to just say "hi" to the host and leave quickly, you end up thinking to yourself "this is a very pleasant party. I'm going to stick around." And although you really hardly know any of the guests, you end up walking straight to small group of people over by the window.

But wait a minute! How did you make the decision to stay in the room, and how come you ended up joining that particular group—all without much deliberation?

It may seem obvious, and we take it for granted, but the quick decision was the result of a rather complex progression, much of it below the level of your conscious awareness. Let's go back to the doorway, to the moment of entry: Light waves of various frequencies are emitted from the scene, along with airwaves, airborne molecules of different compositions, and many other physical quantities such as temperature and humidity that you perceive, integrate with other cues, and then interpret.

The light waves are picked up as colors, shapes, and movements. The airwaves are interpreted as sounds, and the airborne molecules become odors that we can smell. Then these separate bits combine with information from our memory, and the shapes become a room with bodies and faces and loaded tables. The colors form into articles of clothing and various foods. The sounds form strings of speech, laughter, and music. Rich smells confirm the identity of the foods and the wines. Progressively these fragments come together to identify the scene as a lively party, with genuine smiles and laughter that indicate a pleasurable event. The gentle movement of the people by the window, their open and relaxed body language, and their welcoming faces attract you. You decide to stay, and walk toward them.

The process you just went through had a certain progression. You first sensed and extracted physical quantities such as light and sound frequencies, molecule airflows, and temperatures from the ballroom scene. You then interpreted, or abstracted shapes, sounds, and smells out of the many cues sent to your spinal cord and brain. Next, you somehow integrated additional pieces of information such as the relationship between a broad open smile and an open-handed gesture, and added some stored information from your long-term memory about their potential connotations in the given context.

And finally, you were able to organize it all into a holistic interpretation of the scene (i.e., yes, this is a pleasant party) which allowed you to make decisions (yes, I'm staying despite my earlier plans, and I'm heading toward the group by the window). Although extremely fast and seemingly effortless, this process was actually very complex.

A Conceptual Framework

The amount of information that is available to us in our everyday interaction with the physical world is incredibly large, yet most of us seem to deal with it casually. Our spinal cord and brain can apparently abstract and integrate many of the cues from our senses and memory to form a workable picture of the world and, when necessary, initiate appropriate action. We have been doing this for millennia and are, generally, well adapted to interact with a familiar physical world successfully.

We are less well adapted to interact with technological worlds, because many of the causal interactions that allow us to sense, feel, and intuitively understand our surroundings are absent. We are usually removed from the physical environment and hence we can't feel, for example, the speed of the air around an aircraft, we don't see the flaps, and there is no way to touch the engine. We only see outputs from sensors and computer-generated events, usually presented visually in an alphanumeric format and discrete manner. Furthermore, in the context of virtual information worlds such as the Internet, what we perceive through user interfaces is not only intangible, but also incompatible with the regularities of the physical world (Degani, Shafto, & Kirlik, 2006). Our perception and understanding of these remote and virtual worlds is only as good as the representation and presentation power afforded by visual displays and user interfaces. Nevertheless, we need to interact with these worlds, work in them, and form a picture about their behavior that we can act on. Hence the problem we must address concerns not only the difficulty of understanding the intangible world of technology, but also the ever-increasing amount of signals, data, and information that such technological worlds afford. In his book *Things That Make Us Smart*, Norman, (1993, Ch. 3) argues that the real power of human intelligence comes from devising external aids that enhance our own abilities. The main assumption behind the research presented in this report is that when the output from technological systems is organized in a way that matches the way we, as humans, internalize and organize cues, it will be possible for us to take in and better understand large amounts of abstract data. By devising sophisticated representations and holistic presentations, we can, in fact, enhance our abilities to comprehend very complex technological worlds.

To this end we propose here a conceptual framework for considering a continuum for representation and presentation of information. The idea is that sensors, databases, and user interfaces to technological worlds can be designed to create and reconstruct for us what our senses and bodies intuitively do. The system must sense and extract physical quantities into signals, process and transform them into data, and then abstract the data into meaningful information. It then must take the abstract representation of information and integrate and organize it for presentation. How this technological process evolves, such that it is compatible with the way humans form an integrated representation of the physical world, and how best to present this information, are the main questions this research is attempting to explore.

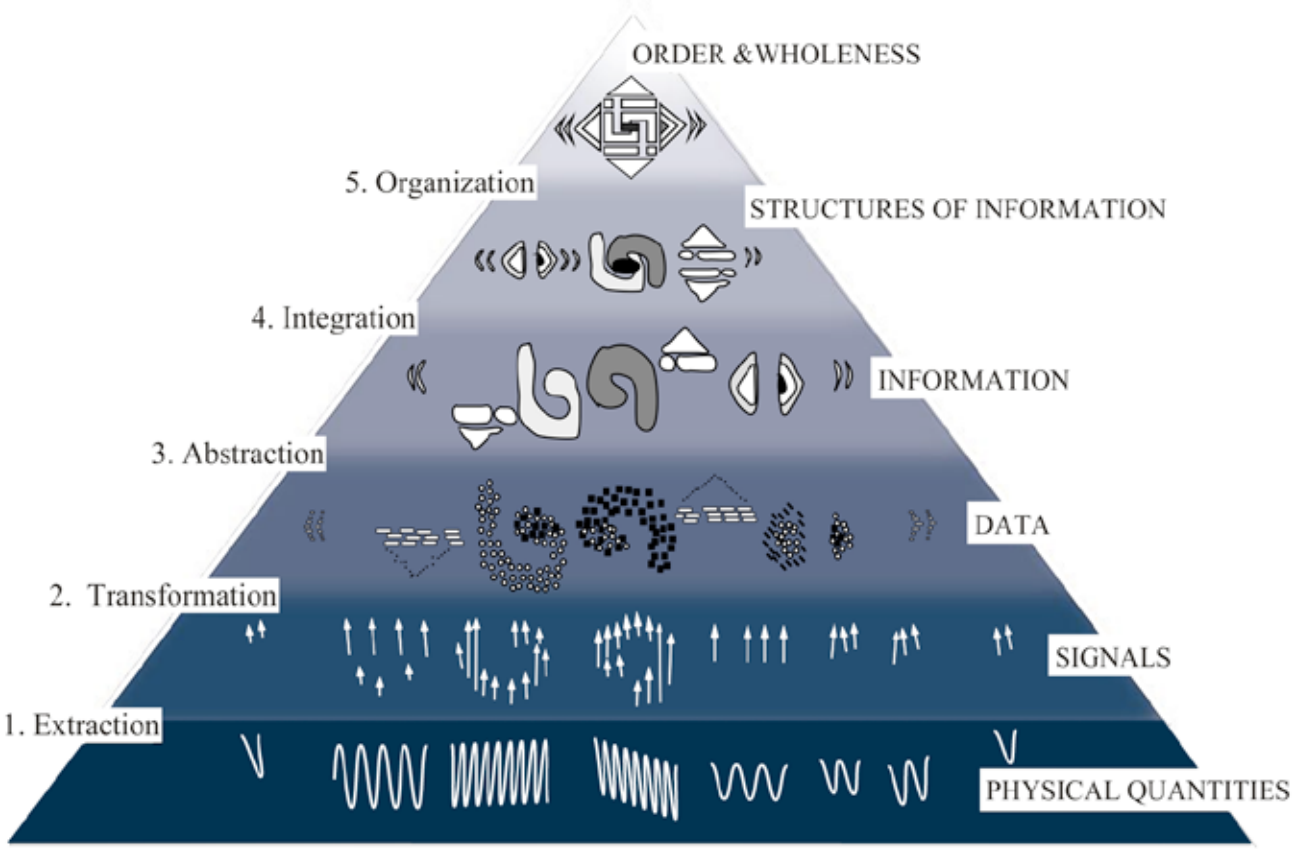


Figure 1-1. A hierarchical arrangement of the six levels involved in representation and presentation of information.

In Figure 1-1 we provide a hierarchical framework for this technological process (Degani, Shafto, & Olson, 2006b). There are six distinct levels in this framework—physical quantities, signals, data, information, information-structures, and wholeness. And there are five transformations that take place between them: (1) *extraction* of physical quantities and turning them into signals; (2) *transformation* and processing of signals and recording them as data; (3) *abstraction* of the data into meaningful information; (4) *integration* of information from several sources into a coherent structure to present interrelationships and compositions; and (5) *organization* of these information structures to create order and provide a sense of wholeness. Let us step through these transformations one at a time.

(1) Extraction of Physical Quantities.

The environment and technological system under consideration emit multitudes of physical quantities. Some of these quantities can be observed, while others are unobservable and beyond our reach. Observable physical quantities are extracted via sensors using measurable characteristics of the physical environment, such as electromagnetic or thermodynamic quantities. In the case of computer software, content, instructions, and internal events can also be extracted. In both cases, the outputs from sensors and computers become signals.

(2) Transformation of Signals.

In this step signals are processed and transformed into a feature set, which can be numbers, alphabets, glyphs, names, etc. Outputs from computer systems (e.g., electrical current) are also transformed into a feature set, which can be binary code, on/off switching, numbers, alphabets, etc. The selection of these feature sets and all subsequent processing must be done judiciously so that they faithfully represent the signals and physical quantities. Thus, the processing and transformations imply that only parts of the signal will be turned into features (e.g., numbers, alphabets, etc.), while the rest is ignored. These features are then recorded to become data.

(3) Abstraction of Data.

Data are stored and made available for computation, manipulation, and display. We argue, however, that from the perspective of human interaction with technological systems, data is not necessarily information. For data to become information it must be (1) relevant for the task and (2) meaningful and well suited to the users who need to perform the task. One way to consider that which is relevant is Gregory Bateson's assertion that "the elementary unit of information is a difference which makes a difference" (1972/1999, Part V, Chapter 5). So, for example, the fact that the temperature in a conference room slowly fluctuates ±1 degree around the set value of 72 degrees Fahrenheit is certainly a difference, but it makes no difference (and is of little immediate consequence) to the occupants of the room. However, if the temperature in the air conditioning ducts is rapidly rising to 120 degrees, then this is a difference which makes quite a difference, because it may indicate fire in the ducts. Fluctuations of ±1 degree around 72 degrees are data; a rapid increase to 120 degrees is information. The point is that information is a quantitative difference (e.g., +48 degrees) that also makes a qualitative difference. The determination of a qualitative difference requires a judicious consideration. It depends on the task and may change rapidly.

The second requirement, abstraction of the data such that it becomes meaningful and well suited to the users who need to perform the task, concerns communication. The term "information" comes from the Latin word "informare" which means "knowledge communication." For data to become information it must not only be relevant, but also communicated in a way that is meaningful to users and matches their perceptual and cognitive abilities. But as any good communicator knows, just to provide meaningful information is not enough—it must also be well integrated and organized.

(4) Integration of Information.

For information to be useful beyond being a set of individual, albeit meaningful and well-represented pieces of information (e.g., indications of an engine's speed, temperature, and pressure), it must be linked and integrated. Since in technological domains users are often physically isolated from the system under consideration, and with increased automation they are further removed from the control aspect of the system (e.g., airline pilots no longer fly aircraft manually during long flights), users are losing the ability to kinesthetically "feel" the system (Norman, 1986). This loss of ability tends to reduce users' capacity to integrate information through their body, and diminishes their ability to understand the situation and anticipate what might happen next. Hence, with increased automation and the resulting inescapable user isolation, it becomes even more important to present integrated information (e.g., direct cause and effect, negative or positive correlation, side effects, trends, etc.) about relationships that exist within a technological system. Such integrated information can help foster better situational

awareness of the system's behavior and the environment (Bennett, Toms, & Woods, 1993). One of the main reasons for visualizing such relationships is to produce patterns that the eye can perceive and comprehend (Card, Mackinlay, & Shneiderman, 1999, p. 25). Once such normal patterns are established, deviations from normality can also be perceived and recognized.

(5) Organization of Information Structures.

The objective at this level of the hierarchy is to organize information in an orderly manner and thus provide a holistic view for the user. The basic assumption is that we all have an inherent capacity to see wholeness and that it is technically feasible and operationally suitable to support this capacity (e.g., by presenting patterns). The focus is on hanging the "structures of information" from the previous level (e.g., clusters of indications) on an underlying structure in a way that reveals interrelationships and conveys patterns of information. In such an underlying structure, a small set of basic geometrical elements can generate a very complex field with manifold interrelationships (Degani, 2008; Lu & Steinhardt, 2007). We discuss and illustrate on such underlying structure in section 5 of this report.

To conclude, this hierarchical structure of levels allows us to consider the measured progression from signals to representations and presentation of information. In terms of analysis, the hierarchy allows us to see that the upper levels can be broken down into lower-level elements and how each level supports the others in the progression toward wholeness. In terms of synthesis, the pyramid-like structure allows us to evaluate the necessary cohesion, from level to level up the hierarchy, for achieving wholeness. Similar hierarchies have been successfully used in the past to better understand relationships between parts and the whole. See for example Alexander's work on formal approaches to urban planning (1964), Miller's (1978) hierarchical analysis of living systems, and Jan Rasmussen's (1986, Ch. 4) analysis of how operators of nuclear power plants reason about the behavior of the plant—from the very concrete sensor readings and indications all the way to conceptual understanding of the plant as a whole.

In the following section we anchor the above conceptual discussion within an operational context: We use the hierarchical framework to analyze an aviation incident in order to better understand problems of information organization and their impact on users' understanding of a complex system.

2. The Framework In Context

The incident concerns the emergency landing of Air Transat Flight 236, an Airbus A330-200 aircraft, as a result of a fuel leak. We analyze this incident from four aspects: (1) the representation of data (from sensors and computer systems), (2) the need to present such data in a way that helps the pilots see important interrelationships, (3) the critical link between cockpit automation and information presentation, and (4) the overall issue of information management in modern cockpits.

The incident took place on August 23, 2001. The aircraft, a twin-engine airliner with 13 crew members and 293 passengers on board, inbound from Toronto, Canada to Lisbon, Portugal, experienced a serious fuel leak while flying over the Atlantic Ocean. The flightcrew did not become aware of the leak for quite some time. After losing almost all of their fuel, the crew diverted the aircraft toward the island of Terceira in the Azores. Within 25 minutes of the diversion, the right engine quit due to fuel starvation; shortly thereafter, the left engine quit. The crew piloted the powerless aircraft for 19 minutes, gliding from an altitude of about 34,000 feet for some 65 nautical miles and finally making a safe landing at Lajes airport (Government of Portugal, 2004).

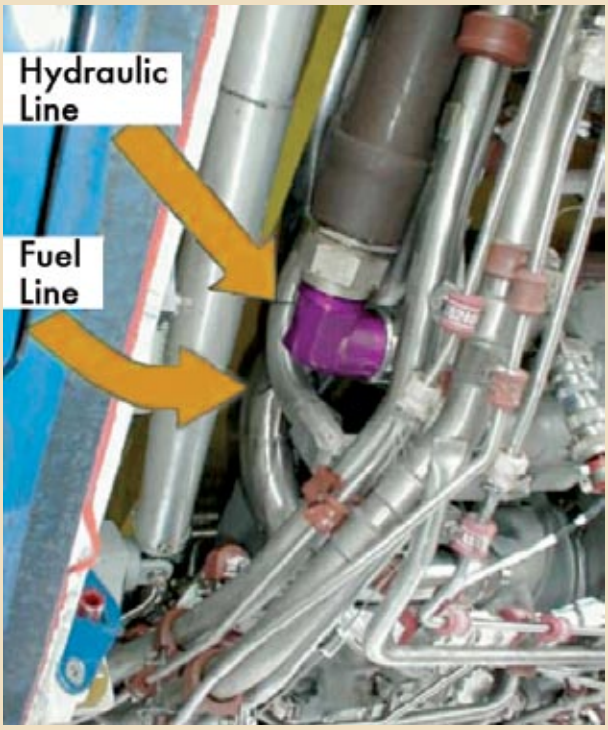


Photo: Government of Portugal

Figure 2-1. The hard contact between the hydraulic and fuel lines (adopted from Government of Portugal, 2004, p.17).

The incident investigation revealed that the aircraft was dispatched with sufficient fuel for the Atlantic crossing (five tons more than was actually needed, reserves included). But about four hours into the flight, a serious fuel leak developed inside the right engine. The fuel leak was caused by a rupture in the high-pressure fuel line, as a result of hard contact with an adjacent hydraulic line (see Figure 2-1). The reason for the hard contact was a type mismatch between the fuel line and the hydraulic line, each belonging to a different version of the Rolls Royce RB211 Trent engine. The incompat-

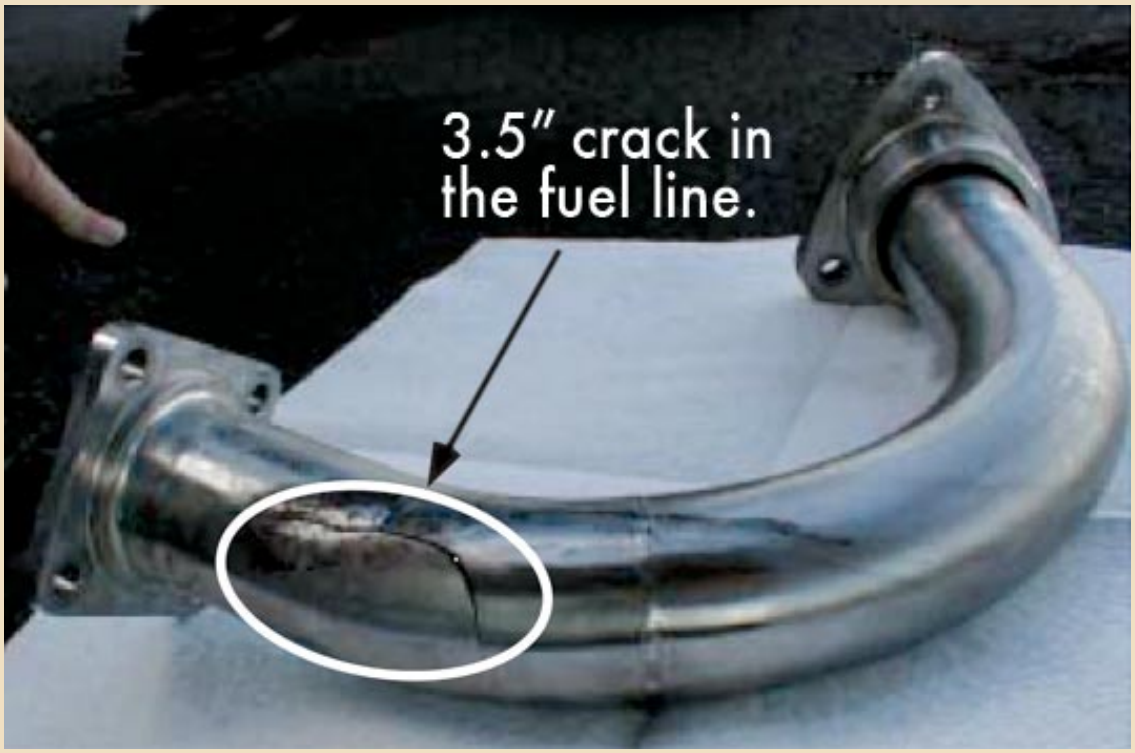


Figure 2-2. The crack in the fuel line.

ible fuel and hydraulic lines were installed by Air Transat technicians several days before the flight and the problem was not detected during a quality control inspection. During the flight, the hard contact between the hydraulic line and the fuel line—compounded by in-flight vibration and normal pulsations of hydraulic fluid in the line—punctured the fuel line. The resulting crack was approximately three inches long, spread to a width of 1/8 of an inch (see Figure 2-2). Post-incident analysis of the aircraft's fuel system data showed that the rate at which the fuel leaked through this crack reached a maximum of about 13 metric tons (approximately 28,000 pounds) of fuel per hour.

From an information representation and presentation perspective, what interests us here are questions regarding extraction, transformation, abstraction, integration, and organization of information. Namely, of the many signals available about the state of the engine and fuel system, what was captured (extraction) and how did these signals and resulting data correspond to what was actually occurring in the engine compartment? How was the data abstracted and represented to the pilots, and what data was not represented? What was the relationship (integration) between the various pieces of information and indications available to the crew? How was the information about the engine presented (organization) with respect to the fuel system (as well as other systems)?

We begin our analysis with the factual information and its physical manifestation. The leak in the fuel line, deep inside the right engine compartment, began 3 hours and 46 minutes into the flight. Since there is no cockpit indication for such a fuel leak, the crew was initially unaware of the developing situation. Twenty-five minutes after the onset of the leak, the flight crew observed three unusual indications about the right engine:

- The oil temperature was low (about half of the value seen on the left engine);
- The oil pressure was high (almost twice that seen on the left engine);
- The oil quantity was relatively low.

The crew had no knowledge of the meaning of these unusual indications, and there was no reference to such an abnormal combination of indications in the aircraft manuals. In a quandary, the pilots contacted Air Transat's maintenance control center in Montréal, but technicians and engineers there had no explanation for such an abnormal combination, nor could they find any reference to it in their maintenance manuals.

Hence, the three oil indications, albeit abnormal and eventually critical, were operationally vague; no one could make sense of what caused the unusual values, nor could anyone suggest the nature of the emerging relationship among the three indications and what that meant for the engine as a whole. After watching the oil indications for some time and trying to develop a working hypothesis, the captain concluded that the unexplained combination of abnormal values reflected some kind of a sensor error ("computer error" was the actual term used).

In retrospect, the abnormal oil indications were indeed related to each other and also to the fuel leak. The position of the leak was downstream of the fuel/oil heat exchanger. The fuel/oil heat exchanger is a unit that takes cold fuel from the wing tanks and preheats it by running hot oil along the fuel line. The fuel is made warmer so that it will burn more efficiently; the hot oil, returning from the engine, is cooled down by the fuel and sent back into the engine. Because of the crack, no back pressure existed in the line and the fuel was gushing through the heat exchanger almost freely. As a consequence, it was overcooling the engine oil and hence the low oil temperature value seen by the flight crew.

The abnormal oil pressure (high) and oil quantity (low) readings were also consequences of the overcooled oil: when oil is cooled, its viscosity increases and thereby the oil pressure increases. The oil pressure sensor, located at the outlet of the oil pump, was registering the increased pressure and transmitting this information to the cockpit display. The oil quantity reading was again a factor of the increased viscosity. Oil quantity is measured at the oil reservoir. Due to the increased viscosity, the oil was flowing very slowly back to the reservoir. Therefore, at any point in time, there was less oil inside the reservoir than normally expected. Post hoc analysis showed that the oil level was about 1.5 liters below the (reservoir) full line, about 1 liter less than expected after 5 hours of flight.

None of these intricate relationships between the oil indications were portrayed in the cockpit. Current cockpit information systems do not show relationships between aircraft systems, e.g., fuel and oil, and most avionics displays do not provide any information about the fuel/oil heat exchanger unit itself. In modern aircraft, fuel information appears on one screen page or one group of gauges in one location on the panel, while engine (and oil) information appears on another.

Shortly after the flight crew recognized the abnormal indications, the Engine Electronic Centralized Aircraft Warning System (ECAM) advised the crew of a developing fuel imbalance between the right and left fuel tanks, each one separately supplying fuel to the engine on its respective side. While the total amount of fuel is always displayed on the main (engine/warning) display, individual tank quantities can only be read on the fuel system page. Figure 2-3 is a picture of the primary display, called the

Engine/Warning Display, and below it is the more detailed Engine parameters page. When the pilots switched to the fuel page, they saw that the right wing tank had less fuel than the left tank. Such fuel imbalance either hints at a difference between the engines' fuel consumption or a fuel leak somewhere in the fuel tanks or fuel lines. The corrective procedure, contained in the flight operation manual, calls for considering the possibility of a fuel leak, then opening the fuel cross-feed valve and turning OFF the wing fuel pumps on the affected, "lighter" side (in this case, the right side). The logic here is that by moving fuel from the left tank to the right engine, the imbalance (of more fuel in the left-wing tank than in the right-wing tank) will eventually be corrected, because both engines will be drawing fuel from the fuller tank. Once the fuel level in the tanks is equalized, the procedure instructs the crew to turn ON the right wing's pumps and then close the cross-feed valve to restore the system to its normal status whereby each wing tank supplies fuel to the engine on its respective side.

Unfortunately, the crew performed the procedure from memory and did not refer to or follow the written FUEL IMBALANCE procedure from the flight operation manual (see Figure 2-4). According to the airline's standard operating procedures, pilots are required to read aloud and follow the procedure from the book, step by step. The crew, in a rush and under stress, believed that the fuel imbalance was an incidental event unrelated to the engine oil indications. The written procedure, however, had a clause at the very beginning, stating "Do not apply this procedure if fuel leak is suspected. [Instead] Refer to FUEL LEAK procedure." By not following the procedure from the book, the crew missed an important hint that had been placed there by well-intentioned engineers and test pilots to guard against inadvertent consequences in a fuel leak condition.

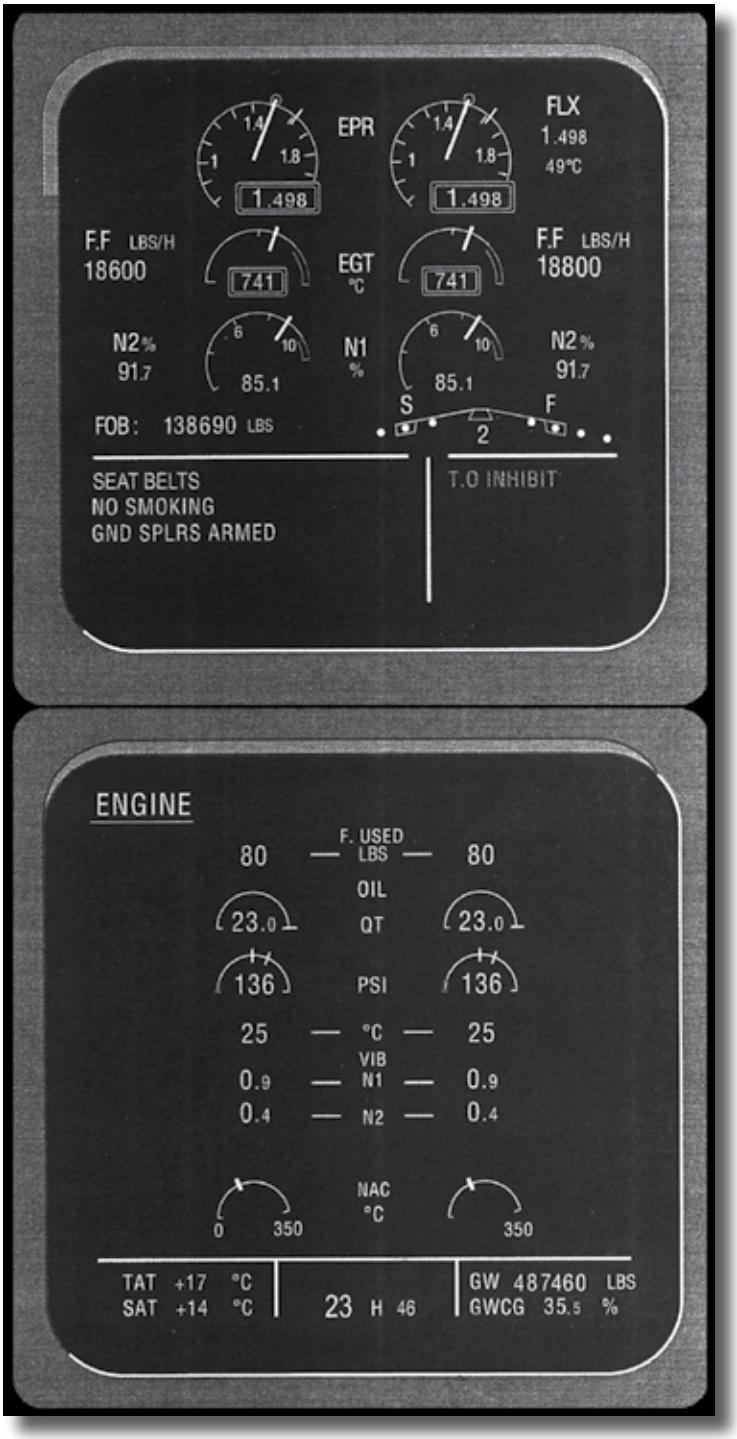
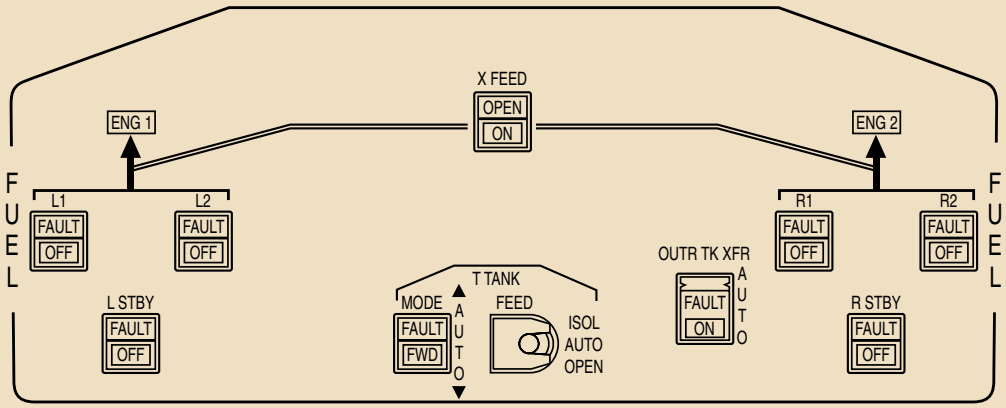


Figure 2-3. The Engine/Warning Display and the dedicated Engine Display page (Pratt & Whitney engines).

This turned out to be a pivotal point in this incident, because by trying to correct the fuel imbalance, the pilots sent precious fuel from the left wing to the leaking right engine, unknowingly further depleting the remaining fuel in the aircraft. Following the crew action to cross-feed the engines, the fuel situation deteriorated and the workload of the two pilots (now considering a diversion) became so high that they had limited mental capacity to re-examine the situation and the consequences of their actions.

Since the quantity of fuel in the tanks constantly decreases as a flight progresses and is a function of the amount of fuel initially loaded, it is difficult to acquire an intuitive feel about the expected fuel in the tanks at any given moment during the flight. To verify that the remaining amount of fuel in the tanks is indeed within reasonable bounds, one can either manually calculate the fuel values or compare the values seen on the displays to the values written in the flight plan forecast (pre-computed by dispatch



FUEL IMBALANCE

CAUTION
Do not apply this procedure if the fuel leak is suspected. Refer to FUEL LEAK procedure.

– WING X FEED . . . . . ON

■ **If the WING X FEED valve is open :**

- **On the lighter side :**
  – (ALL) FUEL PUMPS (STBY then NORM) . . . . . OFF
- **When fuel balanced :**
  – PUMPS (NORM then STBY) . . . . . ON
  – WING X FEED . . . . . AUTO

■ **If the WING X FEED valve is failed closed :**

– OUTER TK XFR . . . . . ON

– BANK ANGLE . . . . . 3 DEG WING DOWN ON LIGHTER SIDE
Fuel transfer only occurs if bank angle is at or above 2 to 3 deg.
Modulate the fuel imbalance through the bank angle.

– RUDDER TRIM . . . . . USE
Use rudder trim to get constant course and neutral stick.

Figure 2-4. A schematic of the fuel panel and switches (note the cross feed switch, "X FEED," in the center of the panel), and the fuel imbalance procedure (Government of Portugal, 2004, p.52).

and given to the crew before the flight). Current aircraft display systems do not provide any integrated information, or any warning to the pilot, that can hint at a fuel leak (e.g., comparing current fuel quantities vs. amount of fuel initially loaded in relation to amount of fuel burned by the engines).

Nevertheless, there are several locations in the flight management computer which provide integrated value of fuel expenditure over distance/time: (1) in the Flight Plan B page – "Estimated Fuel On Board and Tail Wind component" at each waypoint, (2) in the Vertical Revision page – "Estimated Fuel On Board and Extra [saved] Fuel" at each waypoint, and (3) in the Fuel Prediction and Performance page – "Time and Estimated Fuel On Board at destination." In the context of this incident, the latter value, which provides a single value as to the estimated amount of fuel at destination, was probably the most striking source of information about the loss of fuel, because the value would be zero or less. However, this information, which is buried deep within the flight management computer pages, was not part of the regular cockpit scan and probably went unheeded.

While trying to equalize fuel levels in the two wings, the crew continued to focus their attention on the abnormal oil indications, assuming that their major problem was the engine, the fuel imbalance being secondary. The fuel-flow indications, measured as the fuel enters the engine, were normal and were not affected by the leak. The crew also continued to believe that these abnormal indications were somehow all related to a computer error. Because system-related information is organized in modern cockpits by means of a computer-like screen with several embedded pages (e.g., engine, fuel, hydraulics, etc.), the need to switch repeatedly among pages may have limited the flight crew's capacity to consider the health of the aircraft as a whole.

With the cross-feed valve open, the crack kept spraying tons of precious fuel into the air. However, in the middle of the night, the resulting fuel vapor, which would have been visible during daylight from the aircraft windows, went unnoticed. By 05:45, the fuel on board was reduced to below the minimum required to reach Lisbon, and the crew initiated a diversion to the Azores Islands (located about 1000 miles west of Lisbon). By 05:48, the crew advised Santa Maria Oceanic air traffic control (ATC) that the fuel on board was down to 7.0 tons. At 05:59, during an intense dialog with their company maintenance control center, the crew reported that the fuel quantity was further reduced to an alarming level: 1 ton in the right tank and 3.2 tons in the left tank.

By that time, the early morning sun began to shine the ocean with bright yellow light. But the situation of Flight 236 was only getting worse by the minute. At 06:13, when the aircraft was at 39,000 feet and 150 miles off the Azores, the right engine stopped. At 06:15, the crew reported to ATC that the fuel on board was down to 600 kilograms and that ditching the aircraft at sea was a possibility. At 06:23, the First Officer declared "Mayday" (emergency) with Santa Maria Oceanic Control, and at 06:26, when the aircraft was 65 nautical miles from the Azores, at an altitude of 34,000 feet, the left engine quit. From that moment on, the aircraft became a glider.

The crew executed the ALL ENGINE FLAME OUT procedure and began a shallow descent toward Lajes Air Base, a military airfield on Terceira Island. With the benefit of the early morning light ahead of them, the crew was able to orient themselves and begin a gradual glide toward the runway. Assisted by radar vectors and flashing of the runway lights, the aircraft arrived about 8 miles off the approach end of the runway at approximately 13,000 feet. The Captain advised the tower that he was conducting a left 360-degree turn in order to lose altitude.

Figure 2-5. Air Transat Flight 236 at Lajes Air Base (August 24th, 2001).

The captain and his copilot demonstrated exceptional piloting skills to lose altitude and still make the runway. At 06:45, the aircraft crossed the runway threshold at about 200 knots, touched down hard and bounced back into the air. The second touchdown was further down the runway and then maximum braking was applied. Shortly afterwards the aircraft came to a stop about three quarters of the way down the runway. There was fire in the tires and the captain called for evacuation of the aircraft through the emergency exists. All 306 passengers and crew members managed to evacuate the aircraft safely (Figure 2-5). The aircraft suffered structural damage to the fuselage and main landing gear.

Analysis

It is important to remember, while reading such incident reports, that the crew was completely unaware of all the information that is readily available now following months of intense investigation, testing, and analysis. The flight crew, operating under extreme stress, was confronted with multiple indications that did not lead to any conclusion and had to entertain dozens of different hypotheses about what was possibly going on. All of this while piloting a crippled aircraft at night, planning and executing a diversion, and getting ready for the dreaded possibility of ditching the aircraft in the middle of the ocean.

Our focus here is to try to understand how signals, data, information, and their structures are currently represented and presented in modern cockpits. We are interested in finding out, in the context of this incident, where the confusion sprang from and what kind of internal disorder in the way information is organized contributed to this incident. By understanding the sources of this disorder, we can begin to develop better approaches for information organization.

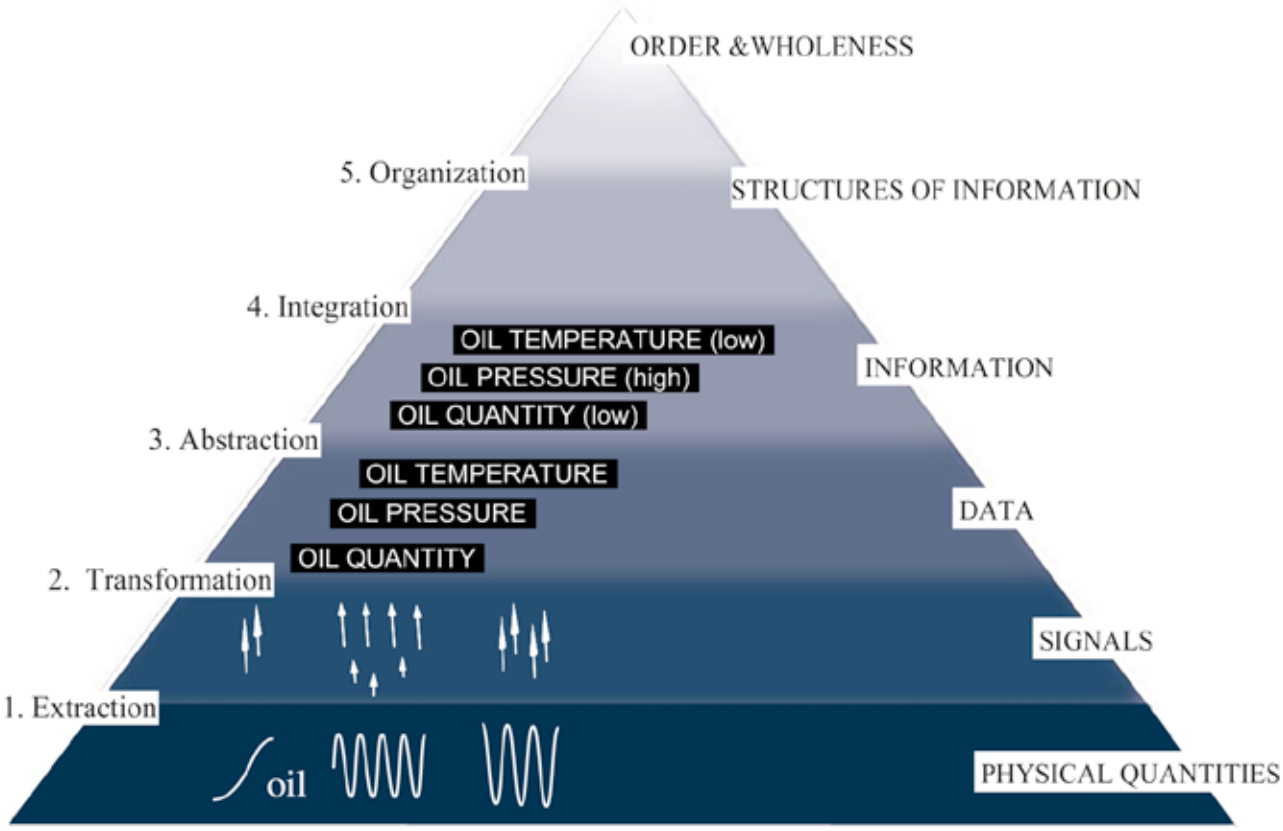


Figure 2-6. Analysis of the oil indications.

Figure 2-6 is a representation of the oil indications of Air Transat 236 within the hierarchical structure of information introduced earlier. The crew saw three indications on the right engine (oil temperature, oil pressure, and oil quantity) that evoked concern. To them, the data became information—a difference that makes a difference. Note that the elevation of oil data to information occurred in the pilots' mind. Furthermore, there was no way to understand any of the relationships among the three indications, nor was there a presentation form to support such integrated understanding.

Hence the indications did not really qualify as meaningful and supportive information, because in many ways they only misguided and confused the crew further. Progressing up the hierarchy, note that the integration and organizations levels are empty. This is because the aircraft's cockpit displays did not provide any presentation of information structures (e.g., combining oil information with fuel information), nor was there any display feature in place to provide the pilots with a holistic picture of the situation.

It is also possible to use this hierarchy to illustrate fuel indication (Figure 2-7). Data about the amount of fuel on board at any given time, and engine fuel consumption (Fuel Flow) were available on the main cockpit displays. Computation of these two parameters over time would have revealed a situation where the amount of fuel drawn from the tanks was greater than the amount of fuel consumed by the engine. This information, however, was not presented in the vicinity of the engine fuel

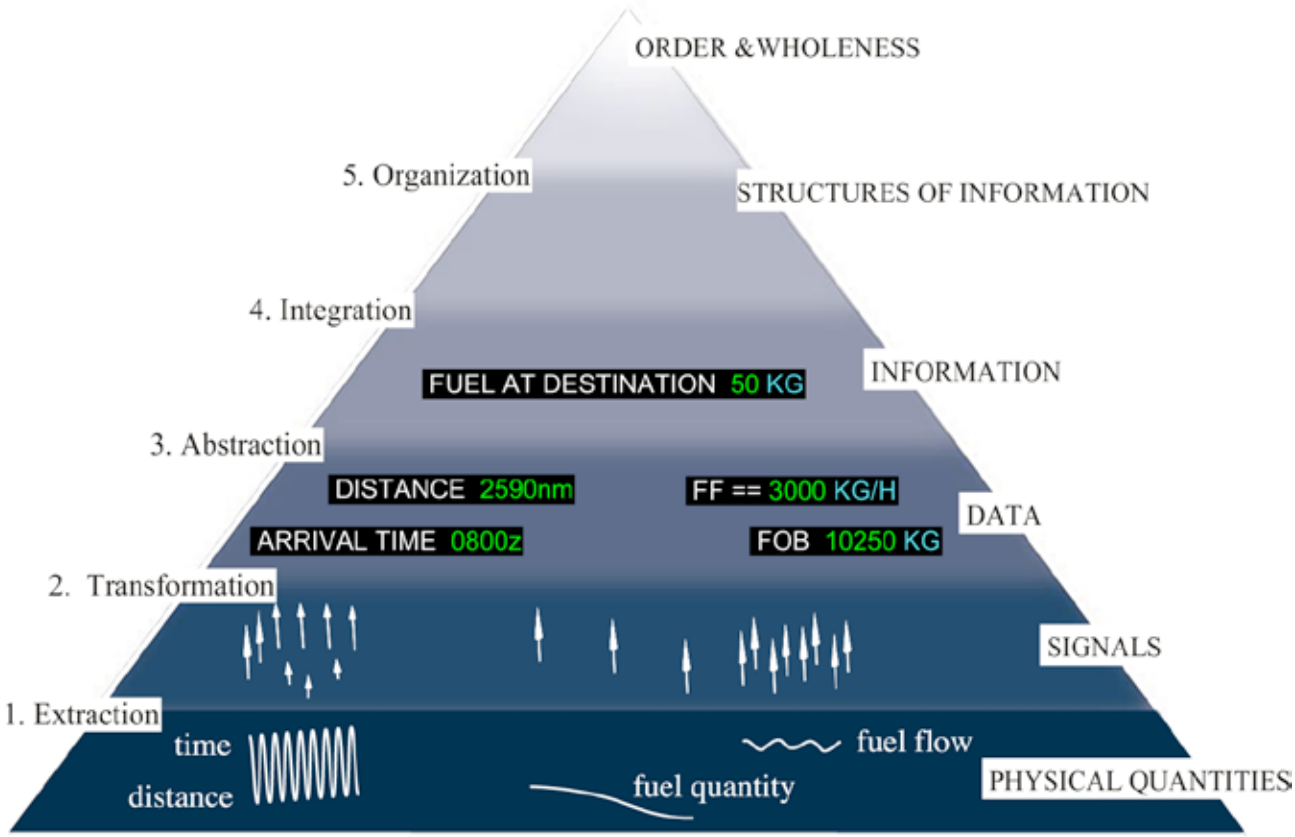


Figure 2-7. Analysis of the fuel anomaly.

consumption rates (Fuel Flow) or fuel quantities for each tank and overall Fuel On Board indication. There were no positive indications alerting the crew that the amount of fuel on board was not going to be sufficient to bring them to their destination. Nevertheless, as mentioned earlier, integrated information about estimated fuel on board at each waypoints and at destination was provided inside the flight management computer.

The point is that providing information on one or even several pages within a computer system, such as the Flight Management Computer, does not necessarily mean that it will be attended to. In such menu-driven displays with multiple pages, overall system understanding and situation awareness can be severely degraded because pilots cannot see all pertinent information in a single scan, nor can they see how separate pieces of data and information are related. Furthermore, when it comes to dealing with an emergency involving complicated contingencies on several systems, pilots have to flip between multiple display pages, perform manual calculations, and look at various sources of information in order to try to understand the situation. There is little support for integration and organization of information in modern cockpits, and therefore it is still the task of the human pilot to infer meaning from the data, integrate available information, and then organize it into a whole. The goal of the research presented here is to better understand how to integrate and organize information in order to help pilots better perform their duties. One objective is to present, in an abstract form, inter-

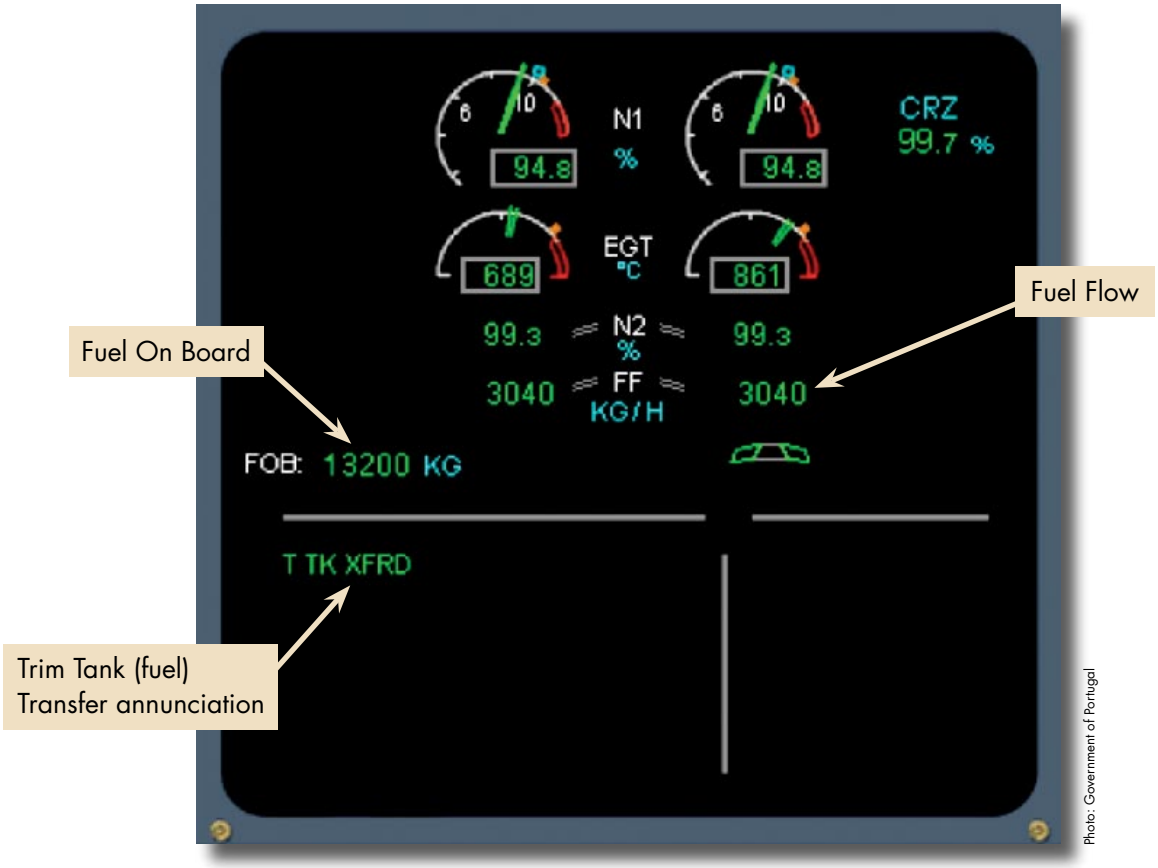


Figure 2-8. The Engine/Warning Display page. The lower left area, which is used to display event information (memos), indicates that fuel from the trim tank in the tail of the aircraft is transferred to the main tanks in the wings. (Adapted from Government of Portugal, 2004, p. 19).

connectivity and holistic relationships that are commonly only available to designers and engineers (e.g., power plant engineers, electrical engineers, and mechanical engineers) who are very familiar with the internal working of the respective aircraft subsystems.

One contributing factor to the pilots' inability to see the whole in this particular incident has to do with the design of the automated fuel transfer system. The system, as the name implies, transfers fuel between tanks so as to constantly adjust the center of gravity of the aircraft. Fuel transfers between the tail trim tank and the wing tanks occur intermittently during the flight and the system is programmed to initiate, perform, and terminate them automatically. The crew gets a text annunciation about the onset of transfer (see Figure 2-8), but there is no indication as to the actual amount of fuel transferred, its distribution, or the rates at which these fuel transfers are taking place. If the crew of Flight 236 had been aware of the increasing rate of fuel transfer and the large amounts of fuel involved, they would have gained a valuable cue about the continual loss of fuel. As can be seen in Figure 2-9, all the pertinent data about the transfers were available, but the form of representation on the Engine/Warning Display page was too abstract. It was masking important data and information from the crew.

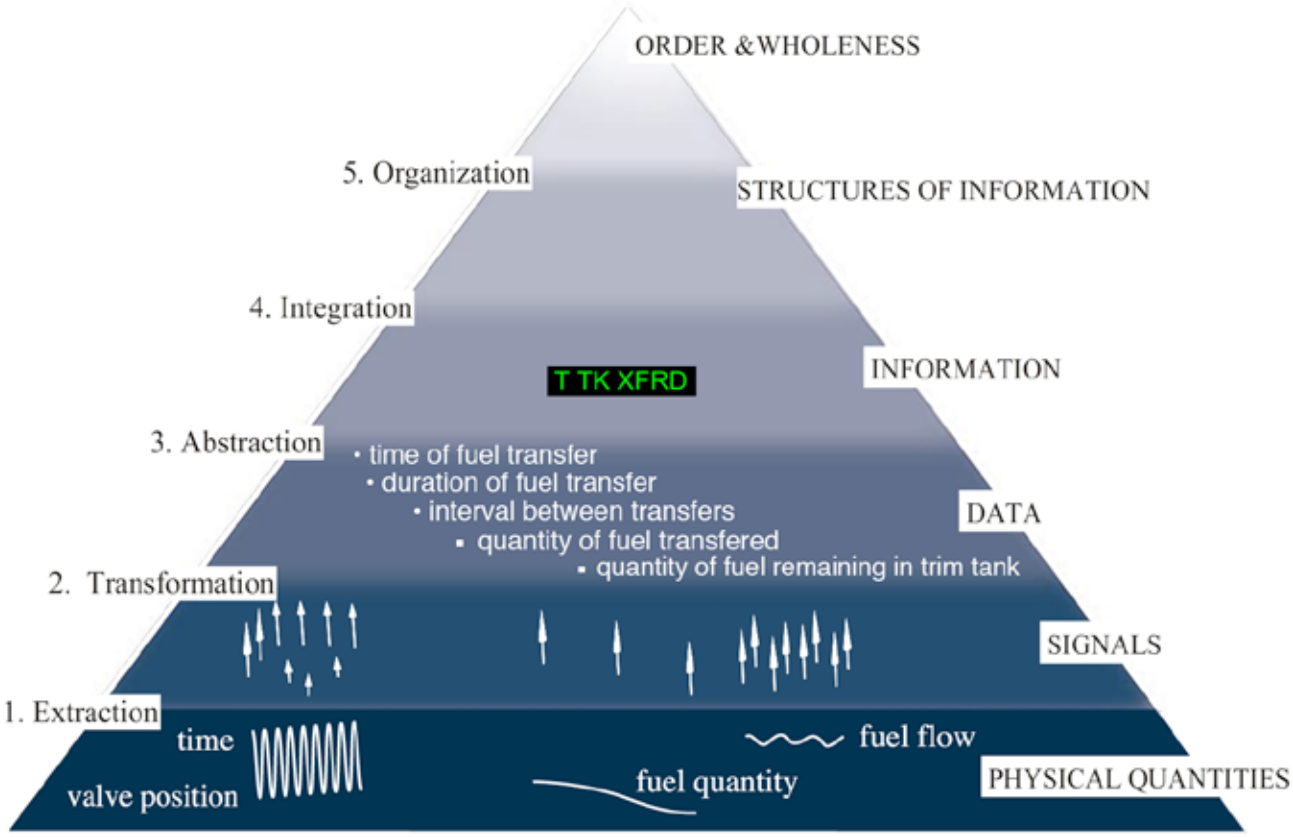


Figure 2-9. Analysis of the fuel transfer indication.

In older generation aircraft, such as the DC-10 and earlier versions of the Boeing 747, it was the role of the flight engineer to manually transfer fuel. By actively opening and closing valves and keeping a fuel log, the flight engineer was kinesthetically aware of the fuel transfer situation. With automation, several choices are available to designers with respect to representation and presentation of information. One common option is to automate, yet provide the crew with all related data. Another option is to provide the pilots with limited data about the process (to allow monitoring), and then augment that data with additional information when the situation becomes abnormal (cf. Billings 1997; Parasuraman, Sheridan, & Wickens, 2000).

In the design of the automated fuel transfer in this aircraft, the crew received only a memo annunciation "T TNK XFRD," and no data to monitor or information to alert them of any abnormality. Hence while the fuel transfer system relieves the pilot from the manual task of opening and closing fuel valves and manifold manipulation, it also removes him or her from remaining aware of what is going on with one of the most important systems in the aircraft. But instead of helping the pilots become aware of the situation by means of better information representation and presentation, most current designs take the opposite approach of minimizing information or not providing it at all, under the assumption that the automation can take care of itself (Degani, 2004, Ch.17).

One of the important lessons from the ongoing research on human-automation interaction is that with increased automation more information is needed, not less (Norman, 1990). What's needed is not necessarily reams of data piped out to the user interface, but rather a better and more sophisticated way to organize this data and produce meaningful information. If indeed we want users to properly supervise the working of the machine, automation design and information design must go hand in hand, not be treated as separate design problems.

Summary

The purpose of this section was to highlight some of the shortcomings of information design in modern cockpits. The Air Transat Flight 236 incident is an important case study because all the necessary signals in the system were available; all the physical quantities were known. Yet, as a result of design choices concerning data abstraction and information presentation, the crew was unaware of the developing situation until a true emergency took place. The fuel system is critical and leaks are a known risk. Change in quantity over time can be calculated such that rate of fuel flowing out of the tank can be compared to the rate of fuel flowing into the engine. Any difference between these rates can be displayed as a possible leak. Design choices were made to not provide data and information on trim tank fuel state and only a limited indication about transfer activity. There was no integration of oil state and fuel state, and no integration of fuel flow out of tank with fuel flow into the engines.

Current approaches to information design, which primarily focus on abstraction (with limited integration and organization capabilities), did not help the crew to identify, analyze, understand, and take appropriate action to rectify the problem, nor did it help them to alleviate the ambiguity and stress they encountered. The truth of the matter is that we are currently very limited in our understanding of what an integrated display actually means, what constitutes its building blocks, and how to go about putting them together. Likewise, we are also very limited in our understanding of what kind of data and information must be presented about the workings of automated systems and how to best organize it.

The state of affairs in our understanding of information abstraction, integration, and organization cries out for a theory. No such theory has been proposed, and it appears, given our limited knowledge of the basic principles of integration and organization, that we are quite far away from achieving one. Therefore, one aspect of this report is to explore other domains such as graphical design, architecture, and art where concepts of abstraction, integration, and organization have been developed and successfully applied. In the following section, we discuss these concepts in the context of a twentieth-century graphical map design with which we are all familiar.

3. A Tale of Two Maps

Map creators have always wrestled with the problem of how to pack large amounts of information, sometimes multidimensional and layered, within a limited space while at the same time achieving clarity and efficient organization (Imhof, 1982). Map designers spend considerable amount of time developing schemes to exclude irrelevant details, yet present complexity in a meaningful and clear way. One famous example is the Underground train network in London. The one shown in Figure 3-1 dates back to 1933. Since then, tracks have been added, stations have opened and closed—but the abstract and clean graphical format has stayed the same.

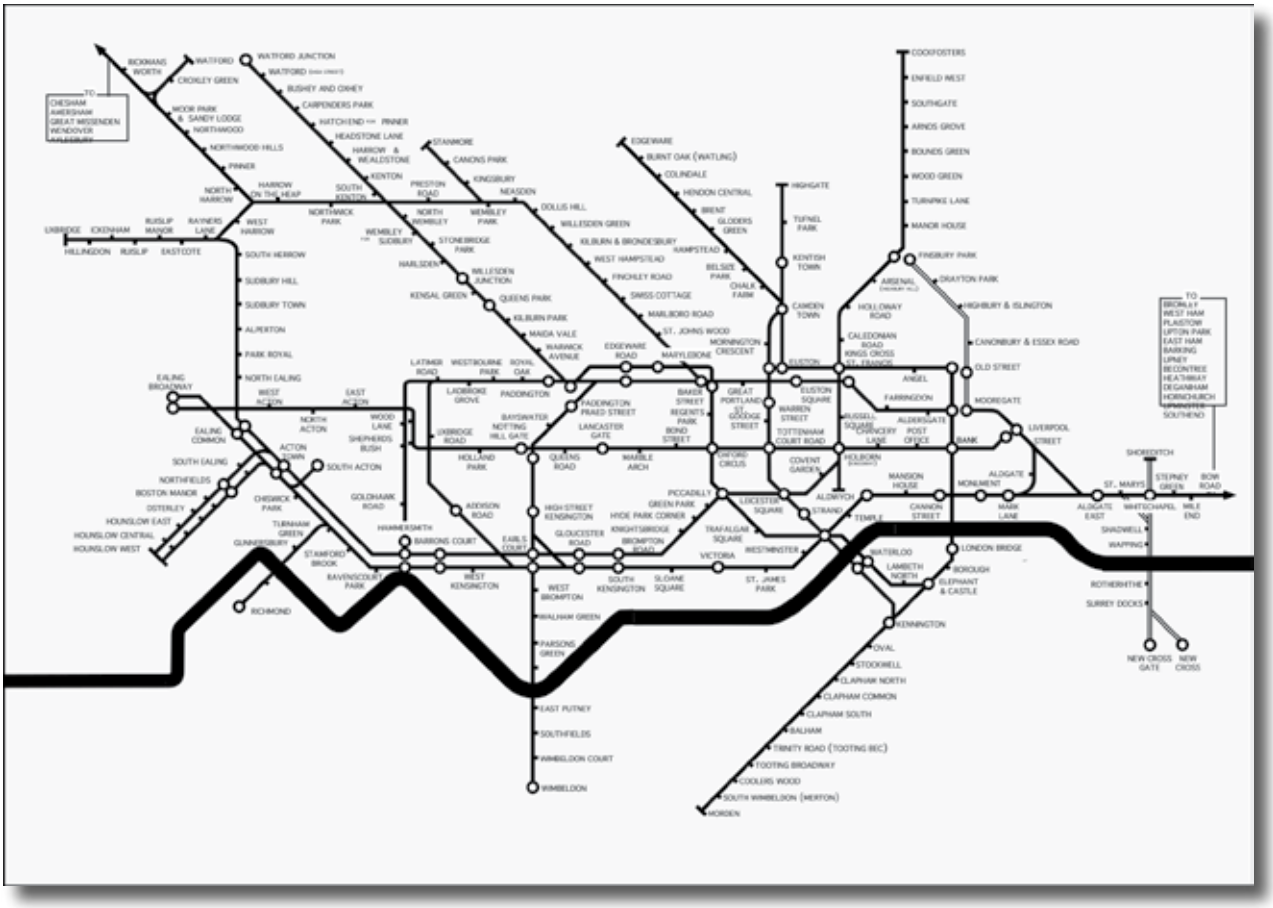


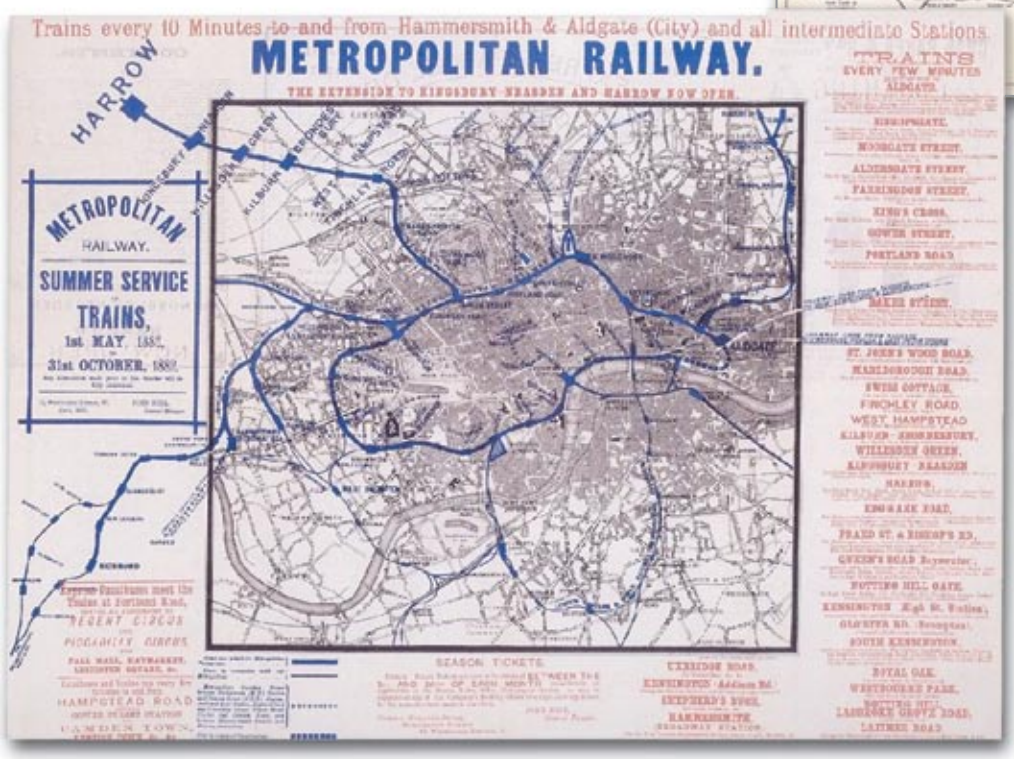
Figure 3-1. Abstract depiction of the London Underground train network (circa 1933).

Figure 3-2 is the geographical depiction of the same Underground network. Here you can see all the twists and turns of the downtown train lines, the sprawl of lines to the suburbs, and the real curves of the Thames River. These features were all stripped down to the basics. What was left was then conveyed in an abstract and schematic form. The result was an instantly clear and comprehensible map that would become an essential guide to London and a template for transport maps the world over.

For more than 50 years, since the inception of the London train and trolley system in 1870, the presentation of the network was primarily geographical (see Figure 3-3). The reason for the switch to a more abstract representation was, interestingly enough, economically driven. In the late 1920s and early 1930s, London Transport, which operates the Underground, was losing money. Survey after survey demonstrated that train commuters had difficulties understanding the network, let alone navigating their way with the geographical map. When the abstracted map was introduced in 1933, it was a big hit with Londoners, as it catered to their information needs rather than to geographical details. The map was quickly nicknamed "the Diagram," primarily because it resembled an electrical wiring diagram. This was not coincidental, as Mr. Harry Beck, the originator and designer of the abstracted diagram, was an electrical draftsman for London Transport (Garland, 1994; Leboff & Demuth, 1999). To appreciate the design of the London Underground diagram and better understand how it was achieved, we will consider and analyze it along the concepts of abstraction, integration, and organization discussed earlier.

Figure 3-2. Geographical depiction of the London Underground train network (circa 1933).

Figure 3-3. Older geographical depictions of the London Underground train network. The larger one, circa 1882, provides a detailed description of the surface streets. The smaller one (top right), is from around 1874. It only provides the stations, few geographical landmarks (Hyde Park and a portion of the Thames River), and a limited number of major buildings (Albert Hall, British Museum, and Marble Arch). Both maps follow the geographical model. (Leboff & Demuth, 1999; p.10).



Abstraction

The London Underground system is made up of several train lines, regular stations, and interchange stations (i.e., stations that have more than one line passing through them). Unless the origin and destination station are on the same line, travelers and commuters have to switch lines at an interchange station, sometimes as often as two or three times during a ride. Therefore, in order to get from one station to another in the maze of the Underground, it is important to have a very clear understanding of lines and interchange stations. References to the outside world and geographical "truth" are somewhat less important.

Beck recognized that the main purpose of the Underground diagram is to help travelers navigate their way between stations *within* the train network. He understood that from the commuters' point of view, the network is an abstract "world." While in the underground, all that travelers see are stations and dark tunnels; they cannot see neighboring stations or any geographical references. Their only reference to the "concrete world" outside is via the stations' names and the orientation of the network as a whole.

Beck's understanding of this subtle, yet powerful, distinction between the true geographical world and the abstract world of the underground represented a major breakthrough in presentation of railway networks. As Beck began to formulate his abstract world, he realized that by escaping the binding constraints of the geographical world, he could do a much better job of representing lines, regular stations, and interchange stations in a way that would help commuters better navigate the network. This abstract jump in the way station and lines are represented as nodes on clean straight lines as well as the overall grid presentation seems rather trivial nowadays, but it was new and very innovative at the time, and has since been incorporated into all train and subway maps around the world (e.g., New York, Berlin, Paris, St. Petersburg, Tokyo, and Athens).

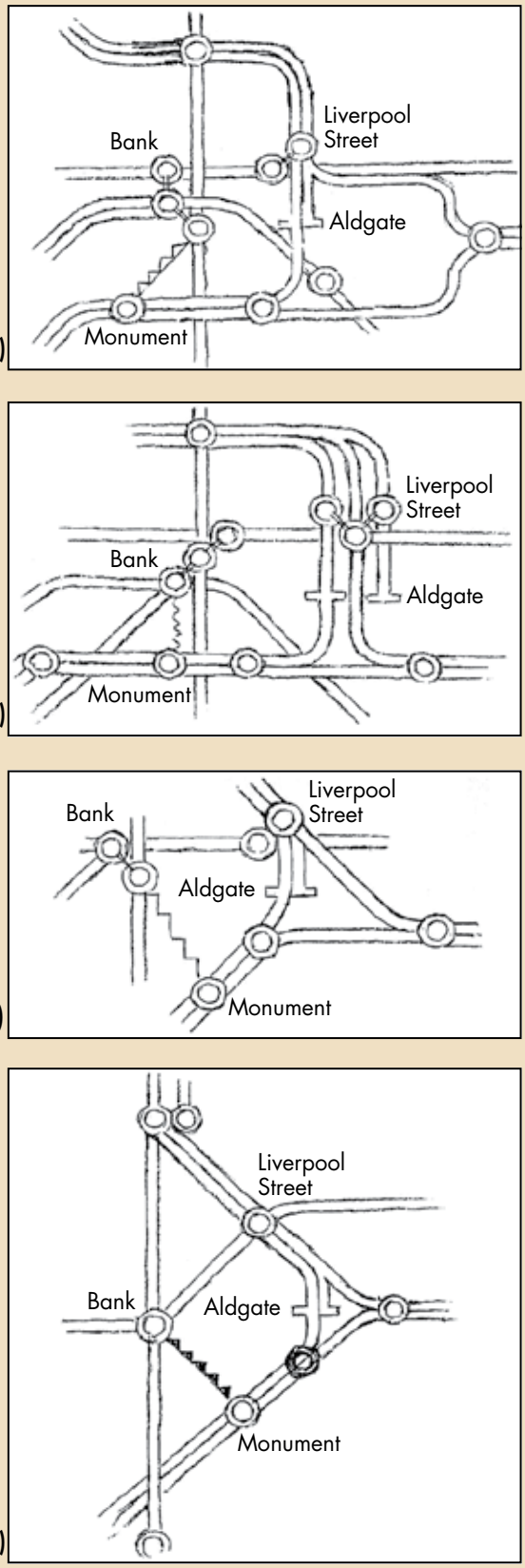
Integration

Beck's initial work on his abstract Diagram was devoted to deciding how to best lay out the eight lines that made up the network. He then turned his attention to stations. Ordinary stations (i.e., those that have only one train line pass through) were arranged relative to other stations on the same line. Here Beck introduced the small square 'tick' as a way to represent an ordinary station. This allowed him to pack more stations together while preserving clarity because the tick could appear on either side of the line, depending on where he decided to place the station's name. Interchange station posed a much greater challenge. Here Beck had to make multiple adjustments to the lines in order to show interconnectivity. This task of integrating abstract representations of lines and stations into a coherent "structures of information" consumed much of Beck's efforts in creating the initial design of the Diagram.

Throughout the decades there were more than a few revisions and modifications to the Diagram presentation, many of them involving heated debates among Beck, his colleagues, and supervisors at London Transport. Most of the changes (and also the debates) focused on how to improve the presentation of interchange information. (The lines, with few exceptions, stayed pretty much the same). For example, consider the area on the eastern part of downtown London consisting of the Liverpool Street, Aldgate, Monument, and Bank stations. When new lines were constructed in the early 1960s, much of the graphical design work focused on the best geometrical structure to present the complexity of this particular area (termed the Aldgate Triangle due to the shape created by the lines passing through this area in the original Diagram). Figure 3-4 depicts four sketches, done by Harry Beck sometime in 1965, showing different solutions for the Aldgate Triangle.

What makes this integration task complex is that several lines here share stations and tracks, and the Bank and Monument stations are connected through an escalator. In sketches (a) and (b), Beck is trying to move away from his original triangular design, which some believed was confusing. Using a more square-like organization he tries to portray that within Liverpool Street station, for example, there is a dedicated ramp for each line and tunnels in between (sketch b). In the sketches (c) and (d) he kept the triangular presentation and gave up on showing the details of the ramps, favoring a single circle to represent all ramps (d). Note also the various representations for the escalator between Monument and Bank, and how he tries to detail the vertical relationship (the escalator goes up from Monument to Bank) in the sketches. The resulting four different clusters, made of individual pieces of information arranged in a coherent way, are an example of what we defined earlier as "information structures."

Figure 3-4. Four different solutions to the "Aldgate Triangle" integration problem, circa 1965 (adopted and reprinted from Garland, 1994). The inset with the red border shows the initial solution used by Beck in the 1933 Diagram.



Over the years, this particular area received much attention, with other designers offering a variety of presentation solutions to the difficult integration problem of how to best show relationships between regular stations (e.g., Aldgate), interchange stations and their ramps (e.g., Liverpool Station), and lines. Clearly, the flexibility to try many presentation solutions to the Aldgate Triangle problem (and many other similar integration problems in the network) is made possible by the use of abstraction to represent lines and stations. Had the form stayed geographical (as in the maps prior to 1933), the range of possible presentation solutions would have been much smaller. Evidently, a well-designed abstract representation can excel, in terms of clarity and ability to pack lots of details and organize information coherently, far beyond any concrete representation. It affords the designer freedom to mold the space.

Organization

One of the most striking aspects of the Diagram is the way Beck organized all lines either vertically, horizontally, or diagonally. On careful examination, we find that there is an underlying aperiodic grid that supports the entire network (Figure 3-5). Although some areas are periodic (e.g., the five vertical lines in the east side, and the four horizontal lines at the bottom), the aperiodicity has a certain rhythm to it which is rather appealing. This underlying grid is not rigid, but rather relaxed; many times Beck deviates from the grid. Yet we intuitively feel the sense of order it casts on the design. (Note, for comparison, the rather similar rhythmical grid used by Piet Mondrian in his *Composition with Red, Blue, and Yellow* shown in the inset of Figure 3-5).

Beck used the vertical and horizontal grid intersections as a guide to determine the beginning and end of diagonal lines (see the blue circles in Figure 3-5). The formation of such a repetitive grid, covering the entire space, is an important property of many profound architectural designs. It allows the designer to hang, so to speak, the entire design (columns, walls, windows, etc.) on a "hidden" structure, which,

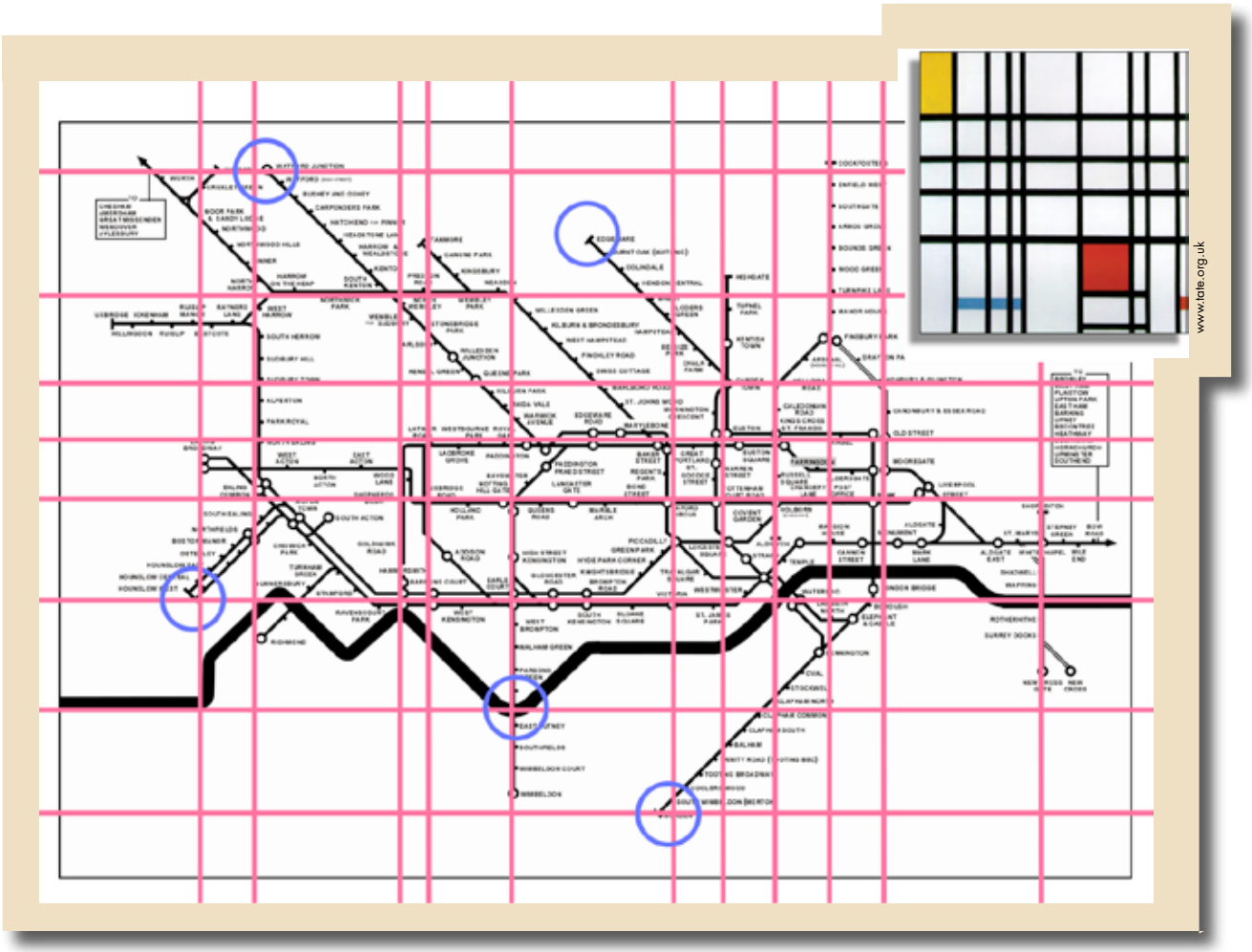


Figure 3-5. The underlying grid and Piet Mondrian's *Composition with Red, Blue, and Yellow*.

in most celebrated designs, is aperiodic. It appears in great architectural sites such as the Alhambra Palace in Granada, Piazza San Marco in Venice, and Montserrat Abbey, near Barcelona, Spain. It helps the builder/architect ground the design, yet conveys a relaxed sense of symmetry and order that viewers appreciate and intuitively understand (Alexander, 2002b, Chapter 15).

In addition to the organization of the lines along a strict structure (vertically, horizontally, or diagonally) and the relaxed underlying grid, there is an aspect of the map that is somewhat hidden. This aspect has to do with the topology, or a purposeful deformation, of the known geographical space. In the course of laying out the organization of the map as a whole, Beck extended one of his predecessors ideas of expanding the central, downtown, area to improve cartographic clarity. Beck realized that he needed considerable amount of (map) space to work the details of interchange stations and specific areas (such as the Aldgate Triangle and Kings Cross/Euston complexes) in the downtown area. On the other hand, in the suburbs, where there is usually only one line, there was hardly any need to integrate information, and even relative distances between stations could be easily compromised. Recognizing these two types of information needs, he used them to his advantage by making the map nonlinear and not to scale: The central area in downtown London is enlarged, while the lines going to the suburbs are heavily compressed. This clever distortion makes it easy to see the details in the downtown area with little negative effect in the outer suburbs areas, where the details are straightforward.

Summary

There is no better summary to this section than Harry Beck's own description of his initial conception and design process:

> *Looking at the old map of the Underground railways, it occurred to me that it might be possible to tidy it up by straightening the lines, experimenting with diagonals and evening the distance between stations. The more I thought about it the more convinced I became that the idea was worth trying, so, selecting the Central London Railways as my horizontal base line I made a rough sketch. I tried to imagine that I was using a convex lens or mirror, so as to represent the central area on a larger scale. This, I thought, would give a needed clarity to the interchange information. (Garland, 1994, p. 17)*

With that touch of ingenious insight, months of detail work on the initial design, and decades of continual improvement, Henry Beck created the most celebrated graphical design of the twentieth century (Garland, 1994). By some accounts, Beck, who was not commissioned to develop the design and did it in his spare time, was never actually paid for the job; others mark the price at five guineas. Whatever it was, it is a meager price given that the Diagram is reproduced over 60 million times each year by companies other than London Transport. Overall, Beck's design approach paid well over the years for London Transport. It turned out to be flexible enough to effectively accommodate all the new stations and lines that were added since 1933 (Figure 3-6). It will probably be used well into the twenty-first century, as it currently supports the prototype of a future Underground network, including several lines that will be added in the future.

When analyzed from this report's perspective, the Underground map begins to reveal to us some of the characteristics of "good" information presentation. Abstraction is used to represent data in a simplified form that adheres to an underlying grid. Separate elements of information, such as lines, stations, and platforms, are cast together to create "structures of information" that are then reused in other locales. These modular units of composition, repeating themselves throughout the design, support the sense of harmony that emanates from this design. Finally, intentional deformations of space, such as enlarging the downtown area and compressing the suburbs, are used to organize information in a way that accommodates the user's information requirements. But most importantly, this example highlights the revolutionary step and courage, on the part of the designer, to escape the "concrete" and familiar form of presentation when it becomes constraining.

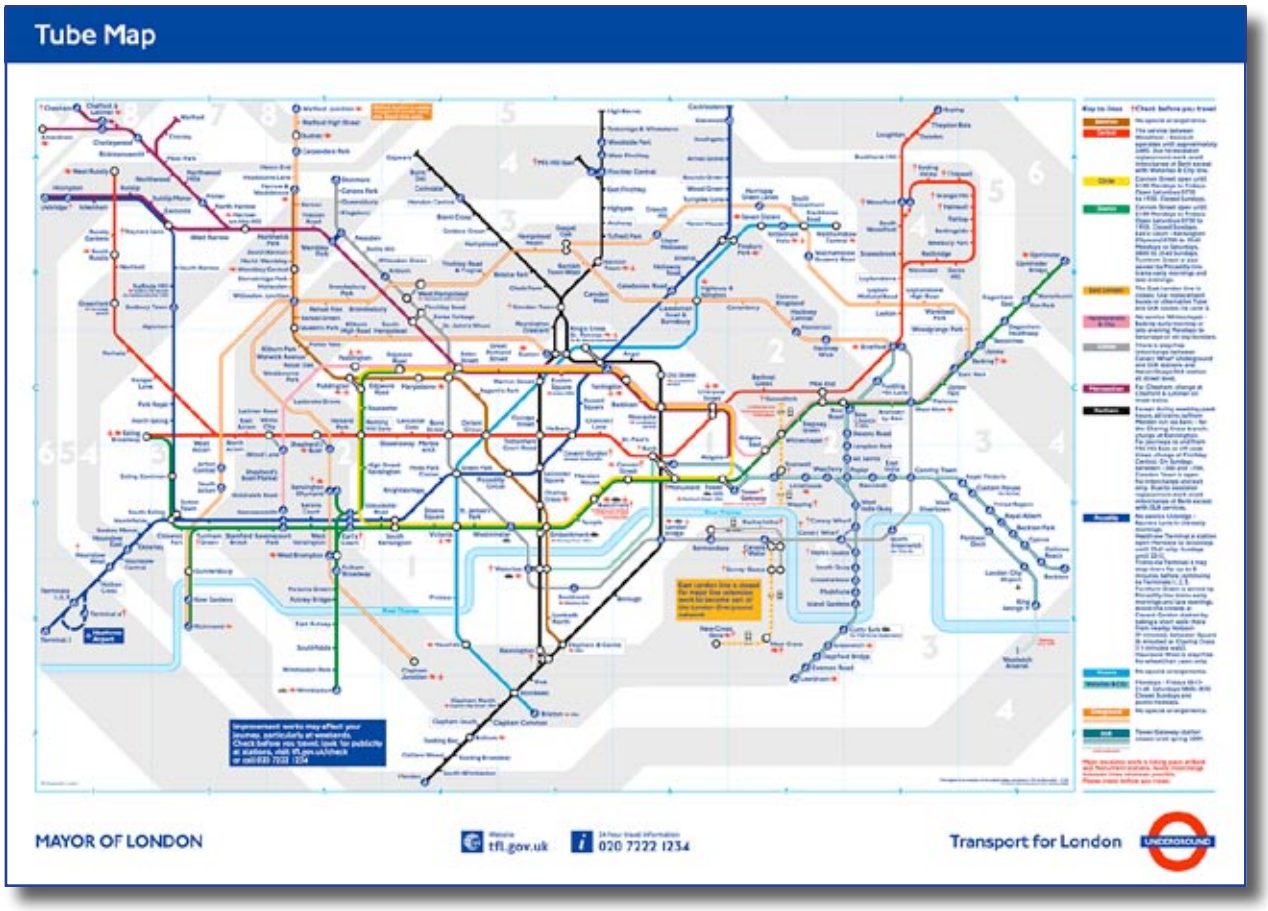


Figure 3-6. The current (2008) Underground Diagram.

4. Analysis of Pilot-Autopilot Interaction

In the previous section we analyzed a well-known example, showing how abstraction, integration, and organization schemes were used to organize information. The schemes used in the Underground example were intuition based; there was no systematic method to achieve abstraction and integration. We could only observe Beck's results and analyze them to try to understand what he attained and the kind of heuristic design processes he employed.

In this section we focus our attention on two statistical methods for achieving abstraction. Using these methods we attempt to gain better understanding of abstraction in the context of large datasets and then use this understanding as a springboard to develop ways to represent individual elements of information and then present them together in an integrated manner. We use data from a field study of pilot interaction with the automated flight control system of a Boeing 757/767 aircraft. The research goal was to quantify the relationships between ATC commands and pilots' mode selection. We first describe the operational environment and then discuss the type of data collected in this study. The study's objectives were to (1) identify which (environmental) factors trigger and prompt pilots' selection of automation modes, and (2) reveal the pattern of interdependency between environmental factors and mode selection.

Field Study Data

Sixty commercial flights between six city pairs (Atlanta-Washington, Washington-Cincinnati, Cincinnati-Atlanta, Atlanta-New Orleans, New Orleans-Dallas/Fort Worth, and Dallas/Fort Worth-Atlanta) were observed during regular revenue flights. In the course of these flights, three types of data were recorded from within the cockpit: (1) aircraft control modes, (2) aircraft flight data, and (3) data about the operational environment. Sitting in the jump seat, an observer recorded every change in the aircraft control modes, either manually initiated (e.g., the pilot selected a new mode) or automatically initiated (e.g., the autopilot switched to a new mode), along with all settings related to the flight control system status (e.g., waypoints and altitudes selected by the pilot). Aircraft flight data such as altitude, distance and bearing from the airport, and critical speeds were also recorded. Every noticeable change in the operational environment (e.g., a new instruction, or clearance, from air traffic control, switching from one ATC facility to another, changes in weather, instructions from the airline's control

center) were recorded as well as pilots' comments relating to their tasks. Figure 4-1 is a data collection sheet used during a departure and climb out from Atlanta's Hartsfield airport.

Figure 4-2 is a spreadsheet presentation of the same data, organized in a way that mimics the observer's note-taking strategy: For every observable change that took place, either on the ATC side or the pilots' side, the observer recorded the aircraft's control modes, relevant flight data, and the state of the operational environment. This, in a way, was like taking a snapshot of the operational environment and corresponding cockpit configuration. Overall, the dataset consisted of 1665 such snapshots, each one containing 30 parameters. About half of the parameters had to do with the operational environment and the other half with pilot responses via the flight control system.

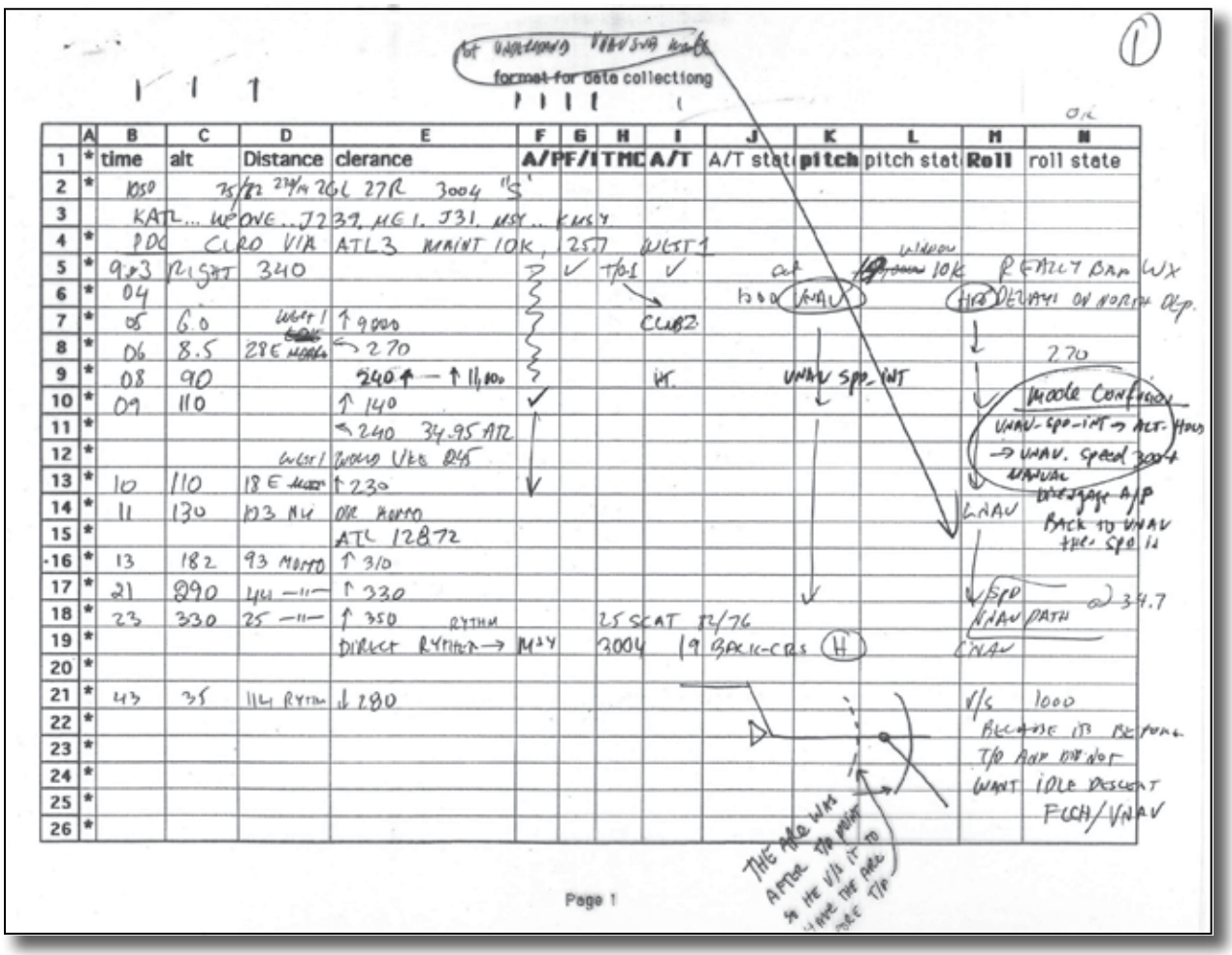


Figure 4-1. Data collection sheet and observation notes taken during takeoff and climb from Atlanta International (ATL) to New Orleans Airport (MSY) onboard a Boeing 757 aircraft.

Linear Regression Analysis

The analysis focused on testing for any causal relationships that might exist between the state of the operating environment and the pilots' responses and actions. Therefore, for the purpose of the statistical analysis, the state and demands imposed by the operational environment were considered as the independent variables (X's), and the actions taken by the flight crews through their interaction with the automatic flight control system were considered as the dependent variables (Y's). One approach to test for any the linear relationship between X's and Y's is to use a linear regression.

Linear regression is one of the most widely used techniques in statistics. In the area of human-machine interaction and decision-making research, many authors have shown that regression models are quite useful for understanding and accounting for the variance in human action and judgment (Dawes, 1979; Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Goldberg, 1968). Regression techniques can be used to identify the most important factors (among the selected set of candidates) that contribute to making decisions and taking actions.

The purpose of the regression analysis was to identify those factors which trigger and prompt the selection of automation modes by pilots. Specifically, we used regression to identify the relationship between the set of independent variables (e.g., the aircraft altitude, phase of flight, ATC clearance) and the dependent variable (configuration of the automated control system). Since the dependent variable of interest, mode configuration, is a vector of Pitch and Roll modes, as well as whether the Autopilot is engaged or not, we abstracted this information to a single ordinal variable, derived from a matrix

Location	time	alt	Phase	ATC_Facil	Dist	Airpor	Clearance	Later	A/P	F/D	A/T	TMC	Prev_	pitch	Prev_Rol	Roll
ATL-MSY	13:02	1000	Climb	Tower_TO	0	KATL	cleared for takeoff	clear	OFF	ON	ON	TO_1	TO	TO	TO	HDG_HLD
ATL-MSY	13:03	2000	Climb	Departure		KATL	turn right to a heading of 340	turn	OFF	ON	ON	CLB_1	TO	TO	HDG_HLD	HDG_SEL
ATL-MSY	13:04	3000	Climb	Departure		KATL		n	OFF	ON	ON	CLB_1	TO	VNAV	HDG_SEL	HDG_SEL
ATL-MSY	13:05	6000	Climb	Departure	7	KATL	climb to 9000 feet	n	OFF	ON	ON	CLB_2	VNAV	VNAV	HDG_SEL	HDG_SEL
ATL-MSY	13:06	8500	Climb	Departure		KATL	left to heading of 270	turn	OFF	ON	ON	CLB_2	VNAV	VNAV	HDG_SEL	HDG_SEL
ATL-MSY	13:08	9000	Climb	Departure		KATL	left to heading of 250	turn	OFF	ON	ON	CLB_2	VNAV	VNAV	HDG_SEL	HDG_SEL
ATL-MSY	13:09	9100	Climb	Departure		KATL	climb to 14000 feet	n	ON	ON	ON	CLB_2	VNAV	VNAV	HDG_SEL	HDG_SEL
ATL-MSY	13:09	95000	Climb	Departure		KATL	left to a heading of 240 degree	turn	ON	ON	ON	CLB_2	VNAV	VNAV	HDG_SEL	HDG_SEL
ATL-MSY	13:10	10000	Climb	Enroute		KATL	Atlanta Center on 128.72	n	ON	ON	ON	CLB_2	VNAV	VNAV	HDG_SEL	HDG_SEL
ATL-MSY	13:10	10500	Climb	Enroute	17	KATL	climb to 23000 feet	n	ON	ON	ON	CLB_2	VNAV	VNAV	HDG_SEL	HDG_SEL
ATL-MSY	13:11	11000	Climb	Enroute	25	KATL	direct MORRO	direc	ON	ON	ON	CLB_2	VNAV	VNAV	HDG_SEL	LNAV
ATL-MSY	13:12	13000	Climb	Enroute		KATL	Atlanta Center 128.72	n	ON	ON	ON	CLB_2	VNAV	VNAV	LNAV	LNAV
ATL-MSY	13:13	18200	Climb	Enroute	35	KATL	climb to 31000 feet	n	ON	ON	ON	CLB_2	VNAV	VNAV	LNAV	LNAV
ATL-MSY	13:21	29000	Climb	Enroute		KATL	climb to 33000 feet	n	ON	ON	ON	CLB_2	VNAV	VNAV	LNAV	LNAV
ATL-MSY	13:23	33000	Climb	Enroute	84	KATL	climb to 35000 feet	n	ON	ON	ON	CLB_2	VNAV	VNAV	LNAV	LNAV
ATL-MSY	13:24	33100	Climb	Enroute	103	KATL	direct RYTHM direct MSY	direc	ON	ON	ON	CRZ	VNAV	VNAV	LNAV	LNAV
ATL-MSY	13:43	35000	Climb	Enroute	145	KATL	descend to 28000 feet	n	ON	ON	ON	CRZ	VNAV	VS	LNAV	LNAV
ATL-MSY	13:50	28000	Descent	Enroute	91	KMSY	descend to 25000 feet	n	ON	ON	ON	CRZ	VS	VS	LNAV	LNAV
ATL-MSY	13:51	27000	Descent	Enroute		KMSY	descend to 24000 feet	n	ON	ON	ON	CRZ	VS	VS	LNAV	LNAV
ATL-MSY	13:51	27000	Descent	Enroute		KMSY	frequency change to Houston Cen	n	ON	ON	ON	CRZ	VS	VS	LNAV	LNAV
ATL-MSY	13:52	26500	Descent	Enroute		KMSY	descend to 11000	n	ON	ON	ON	CRZ	VS	FLCH	LNAV	HDG_SEL
ATL-MSY	13:53	26200	Descent	Enroute		KMSY	turn 15 degrees to right for tr	turn	ON	ON	ON	CRZ	FLCH	FLCH	HDG_SEL	HDG_SEL
ATL-MSY	13:54	26000	Descent	Enroute		KMSY	leaving 16000 feet, direct MSY	direc	ON	ON	ON	CRZ	FLCH	FLCH	HDG_SEL	HDG_SEL
ATL-MSY	13:56	16000	Descent	Approach	45	KMSY	frequency change 120.6	n	ON	ON	ON	CRZ	FLCH	FLCH	HDG_SEL	HDG_SEL
ATL-MSY	13:57	12000	Descent	Approach		KMSY	descend to 3000 feet	n	ON	ON	ON	CRZ	FLCH	FLCH	HDG_SEL	LNAV
ATL-MSY	13:58	7000	Descent	Approach		KMSY		n	ON	ON	ON	GA	FLCH	FLCH	LNAV	HDG_SEL
ATL-MSY	13:59	5000	Descent	Approach	26	KMSY		n	ON	ON	ON	GA	FLCH	FLCH	HDG_SEL	HDG_SEL
ATL-MSY	14:02	3400	Descent	Approach		KMSY		n	ON	ON	ON	GA	FLCH	FLCH	HDG_SEL	LOC
ATL-MSY	14:03	2800	Descent	Approach		KMSY		n	ON	ON	ON	GA	FLCH	FLCH	LOC	LOC
ATL-MSY	14:04	2000	Descent	Tower_Lan	14	KMSY	clear to land	clear	OFF	ON	OFF	GA	FLCH	FLCH	LOC	LOC

Figure 4-2. The data from Figure 4-1 in a spreadsheet format. Each row represents a unique snapshot in time, triggered by some change either on the environment side or pilot interaction.

of 45 operationally plausible mode configurations of the system (see Figure 4-3), rank-ordered in correspondence with the level of automation. We used two subjective criteria to rank level of automation: (1) the precision with which the automation follows the predetermined flight path; and (2) the degree of human-machine involvement in controlling the flight path.

In this abstraction scheme, low values were assigned to a combination of pitch and roll modes that were highly automated (e.g., VERTICAL NAVIGATION / LATERAL NAVIGATION mode was ranked 1). High value was assigned to a combination of modes in which the autopilot was not engaged and the aircraft flown manually (MANUAL ROLL / MANUAL PITCH mode was ranked 45). The assigned rankings, which provided for an estimate of the level of automation employed by the flight crews at every instance, allowed us to regress independent variables (X's) on this single dependent (Y) variable.

The linear regression model was initially built using half of the dataset (30 flights) in accordance with standard statistical processes (Neter, et al., 1990, chap. 12). The results identified the significant (independent) factors that contribute to reducing the variance around the predicted Y (R^2adjusted = 0.55, $p < 0.001$). The regression analysis of the model-building set was then validated on the remaining 30 flights (R^2adjusted = 0.64, $p < 0.001$). The stable and highly predictive results indicate that the regression model captured the important characteristics of operational environments that trigger and prompt pilots' mode selection. We then applied this model to the entire data set of 60 flights (see Figure 4-4).

The results indicate that 59% of the variance in the level of automation selected by the

Pitch Modes \ Roll Modes	Heading Hold	Heading Select	Lateral Navigation	Localizer Mode	Manual Roll
Takeoff Mode	29	27	8	20	44
Altitude Hold	32	31	18	26	43
Vertical Navigate (Spd_int)	16	13	2	6	34
Vertical Navigate	15	12	1	5	33
Vertical Navigate (Des_now)	17	14	3	7	35
Flight Level Change	30	25	10	21	39
Vertical Speed	28	23	9	19	40
Glide Slope	24	22	4	11	37
Manual Pitch	42	41	36	38	45

Figure 4-3. Ordinal ranking of the 45 mode configurations of the aircraft autopilot system. Dark shades denote high levels of automation and light shades denote low levels.

R squared = 59.6% R squared (adjusted) = 59.0%
s = 6.625 with 889 - 13 = 876 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	56677.4	12	4723.12	108
Residual	38452.1	876	43.8951	

	Variable	Coefficient	Standard-Error	t-ratio	probability
	Aircraft Altitude	-0.0002	0.001	-5.86	≤ 0.0001
phase	Descent	5.0622	0.739	6.85	≤ 0.0001
ATC facility	Departure	-15.0313	0.982	-15.3	≤ 0.0001
	En-route	-15.0103	1.306	-11.5	≤ 0.0001
	Approach	-10.8755	1.162	-9.36	≤ 0.0001
	Tower	-14.8360	1.385	-10.7	≤ 0.0001
clearance	"Turn to Heading	7.4598	0.666	11.2	≤ 0.0001
	"Direct to"	-2.1719	0.816	-2.66	0.0079
	"Cleared for the Approach"	8.4038	1.471	5.71	≤ 0.0001
	"Descend to"	2.8143	0.655	4.30	≤ 0.0001
	"Cross at"	1.1818	1.256	0.941	0.3471
	"Expedite your Climb"	7.3282	1.815	4.04	≤ 0.0001

Figure 4-4. Linear regression analysis results (60 flights).

pilots can be explained by four main factors: (1) *the aircraft altitude*, (2) *the phase of flight*, (3) the type of *ATC facility* supervising the aircraft, and (4) the type of ATC *clearance* issued by this facility. Aircraft altitude turned out to be a highly significant factor for predicting the level of automation selected by the pilots. The "Descent" phase of flight was another significant factor. All four ATC facilities—Departure, En-route, Approach, and Tower (at the landing airport) —contribute to the decision to switch modes. Of the three elements in a clearance (lateral, vertical, and speed), only the lateral and vertical elements directly relate to mode selection. In the lateral element, we identified three significant clearance types: "turn to a heading," "direct to (a waypoint)," and "cleared for approach." In the vertical element, there were three significant clearance categories: "descend to an altitude," "cross (a waypoint) at an altitude," and "expedite your climb." (For additional details on this and other aspects of the regression analysis, such as residual analysis and outliers, see Degani, 1996, Ch. 7).

In summary, the regression analysis allowed us to abstract a set of meaningful pieces of information from the data. It identified four main factors (aircraft altitude, phase of flight, ATC facility, and ATC clearances) that prompt mode transitions. The findings provided evidence for a strong unidirectional relationship between the structure of the operating environment and aircraft's mode configuration. One implication of this finding is that any design (or modification) of an automated flight control system must be done in light of the (future) ATC environment the aircraft will be flown in. This may be important in the context of a future air traffic environment, which will be more condensed with traffic and inevitably require more information sharing between the cockpit and ATC. Therefore, by presenting the very structure of the environment, we may foster better decision-making and consequential actions on part of the flight crews.

Canonical Correlation Analysis

The linear regression analysis applied to the field study data is relatively simple and its statistical processes are commonly used. However, there was a price we had to pay. To use it, we had to abstract the details of the aircraft mode configuration to a single composite Y variable, thereby limiting the amount of raw information that enters the model (see Walker & Catrambone, 1993, for a detailed treatment of this issue). This is indeed a serious limitation, as we have a genuine interest in each of the variables that makes up this composite. Furthermore, the results of the regression analysis, albeit precise in quantifying the contribution of each factor to reduce variance, do not portray any relationships between the X's (beyond interaction effects, which were not found in this particular regression analysis). If we are interested in how factors come together to trigger and prompt specific mode selections, regression analysis keeps us in the dark.

Regression analysis was developed to tease out main effects (or *factors* in our terminology here). It was developed and used primarily to confirm or refute the null hypothesis, and as such is not concerned with integration of information, or how factors come together, at various strengths, to prompt a certain mode selection. In other words, regression analysis is not about identifying patterns of information. Fortunately, there is a well-established statistical analysis, called Canonical Correlation, which does reveal patterns and integration of factors.

Canonical Correlation Analysis (CCA) is a type of multivariate linear statistical analysis, first described by Hotelling (1935; 1936). CCA requires a set of independent variables X (input or causal variables) and a set of dependent variables Y (output or effect variables). Standard methods of linear algebra are used to estimate weight vectors (u and v), such that the correlation coefficient between X_u and Y_v is as large as possible. As such, the vectors u and v may be thought of as representing patterns of X and Y relationships. In theory, there could be many different pairs of weights (u_i, v_i), each one representing different pairs of patterns. In practice, it is common to have only several meaningful patterns. (For further discussion of CCA theoretical foundation and a literature review of recent applications, see Appendix A).

There is a two-fold objective for a canonical correlation analysis: The first objective is to identify and understand the causal relationships between the independent variables (environmental demands, in our case) and a set of dependent variables (pilots' mode selection) by revealing the set of statistically significant patterns in the dataset. The second objective is to use the resulting patterns as a way to detect deviations from established norms and to understand their operational meaning. In the context of this study, deviations from the nominal pattern of pilot-autopilot interaction may indicate an operational situation that is becoming abnormal and potentially unsafe (Shafto, Degani, & Kirlik, 1997). The following is an example of such an abnormal and potentially unsafe situation:

One of the flight crews mistakenly used an inappropriate vertical mode, FLIGHT LEVEL CHANGE, during the last phase of the approach (below 2,000 feet) into New Orleans. This problem occurred as the pilot expected to intercept the glide slope signal and switch from FLIGHT LEVEL CHANGE mode to GLIDE SLOPE mode for vertical guidance, as is commonly done. However, due to unusual winds, the aircraft was given a clearance to a runway that only had a back-course localizer approach and no glide slope signal. The crew failed to recognize the meaning of the new situation, and did not switch

to VERTICAL SPEED mode, which is the appropriate mode for a back-course localizer approach. Since the flight crew was using the FLIGHT LEVEL CHANGE mode through the early phases of the approach, they continued using this mode all the way down to 600 feet and only then recognized the problem. At that point they quickly disengaged the autopilot altogether, took manual control of the aircraft, and continued to hand-fly the aircraft throughout the reminder of the approach and landing. (The use of FLIGHT LEVEL CHANGE is not recommended below 2,000 feet because at low speeds the autopilot may increase the aircraft's pitch attitude to maintain the selected speed, thus leaving the aircraft in a perilous state of slow speed and high pitch attitude, while the aircraft is close to the ground). Operationally speaking, there was a period of time, from 2000 feet to 600 feet, in which the aircraft was configured improperly. Had some other unexpected factor (e.g., wind shear, pilot incapacitation, go-around) occurred, the crew would have been faced with a combination of factors that would have made recovery more difficult.

During the 60 flights, the observer recorded 22 situations that were considered anomalous. They involved cases where the pilots used an inappropriate mode (as discussed above) and cases where the pilots were surprised by the behavior of the automatic flight control system (e.g., the aircraft did not automatically level off as expected and in some cases failed to comply with altitude restrictions that were programmed into the flight management computer). About half of the 22 atypical cases showed up as statistical outliers. All the other outliers that were identified by the statistical analysis were primarily due to safe, yet irregular configuration (e.g., one aircraft was dispatched with an inoperative autothrottle system, resulting in a set of infrequent, yet perfectly safe, mode combinations).

Abstraction and Representation of Data

Traditionally, the results of canonical correlation analysis are presented by means of numerical tables, not unlike the regression table presented in Figure 4-4. However, a tabular format, with only numerical values, is rather limited in presenting patterns. Since one of the objectives of the research outlined and described in this report is to develop principle approaches for information presentation, we wanted to explore the use of information abstraction, integration, and organization concepts, as discussed in earlier sections, in order to present patterns of information.

Using the correlations of the X canonical variate with each of the original independent variables, and of the Y canonical variate with each of the original dependent variables, we organized the variables in a sunburst presentation. All the independent variables ($X_1, X_2, X_3 \ldots Xm$) were placed on the right side of the circle and all the dependent variables ($Y_1, Y_2, Y_3 \ldots Yn$) on the left. The sunburst circular structure was chosen because it is a commonly used symbol of wholeness (Tucci, 1961). The circular arrangement also has the benefit of implicitly suggesting that all variables are connected (at the center and around the rim) and are equal in terms of importance (whereas a vertical or horizontal layout usually implies a rank order).

A canonical correlation analysis of the data collected identified four sets of statistically significant patterns $R^2 = 0.90, 0.77, 0.65, 0.52$; $p<0.001$ (sets of patterns below $R^2 = 0.5$ were considered weak and discarded). Figure 4-5 shows the first set $R^2 = 0.90$, which is made up of two sub-patterns, one depicted by white bars and the other depicted by black bars.

The white sub-pattern in Figure 4-5 indicates that, for all independent variables (X's), when:

- the Air Traffic Control facility is "departure" control, and
- the vertical clearance is "climb to altitude,"

then the most likely mode/settings selected by pilots will be:

- auto-throttles ON (engaged)
- pitch mode in VERTICAL NAVIGATION, and
- thrust mode is CLIMB-2.

The black sub-pattern indicates that when:

- current aircraft altitude is high (above the average of 13,000 feet),
- the phase of flight is "descent,"
- the Air Traffic Control facility is "approach" control,
- and the vertical clearance is "descend to altitude,"

then the corresponding modes and settings selected by the pilots are most likely to be:

- autopilot ON (engaged),
- pitch mode in FLIGHT LEVEL CHANGE,
- thrust mode in CRUISE,
- and GO-AROUND armed.

The product of a canonical correlation analysis is a pattern pair (as detailed above). Note that although the two patterns share the same R^2 value, they are different. In fact, the relationship between the pair can be thought of as a contrast. The white bars describe a pattern of interaction during initial climb (to altitude). The pattern presented with the dark bars describes an equally strong pattern during approach (to landing). Technically speaking, it is the statistical strength of the two contrasting patterns, each pulling opposite the other, that drives up the R^2 value.

The contrast between the two patterns in Figure 4-5 is also operationally meaningful—different modes are used in climb (VERTICAL NAVIGATION) vs. descent (FLIGHT LEVEL CHANGE). During climb, the autothrottles are mostly ON (engaged) and the autopilot is not a consistent factor (sometimes it is OFF and sometime ON), whereas in the descent, the autopilot is commonly ON (engaged) and autothrottle is not a consistent factor. This unique pattern of pilot-autopilot interaction, where there is a contrast between autopilot and autothrottle use, is a reflection of the Standard Operating Procedure (SOP) of the airline that requires autothrottle engagement during the initial climb, yet leaves autopilot use to pilots' discretion. The pattern during the early descent is also reflection of a recommended practice (not yet a mandated one) to use the autopilot in the busy terminal area so as to allow the pilots more time to scan the outside scene for other traffic and also deal with the fast pace of ATC clearances directed at them. (See Degani & Wiener, 1994 for a detailed discussion of standard operating procedures and practices, recommended or otherwise, in airline operations).

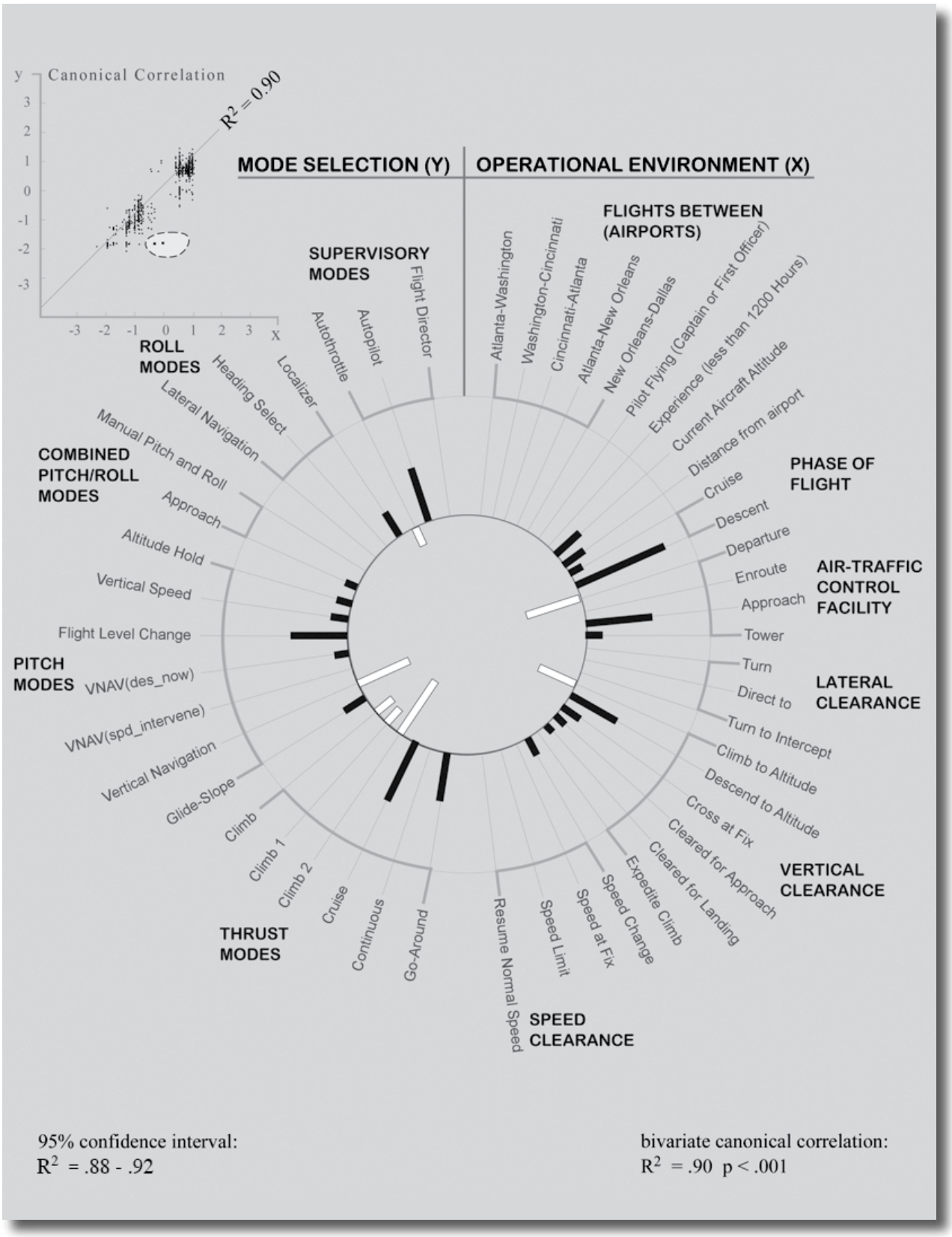


Figure 4-5. The first set of patterns (R^2 = 0.90).

Integration and Presentation of Information

The array of environmental factors such as the "descent" phase, "approach" control, and "descend to altitude" clearance and the rest of the black bars in Figure 4-5, gives a sense of the environmental structure, as a whole, that is being experienced by the flight crew in a particular situation (in this case, early descent). Likewise, on the mode selection side, we get a feel of the corresponding configuration of the aircraft, FLIGHT LEVEL CHANGE mode is commonly used, but also ALTITUDE HOLD and VERTICAL SPEED to a lesser degree. This kind of information, sliced for a given situation, such as early descent and initial climb in Figure 4-5, takes us to a higher level of understanding of the environmental factors and autopilot modes at play. Going back to the autopilot and autothrottle contrast finding discussed earlier, note that unlike linear regression analysis, where such a inverse relationship (termed "interaction effect") would only become significant if it was consistent throughout all flights phases. Canonical Correlation reveals such patterns for a specific phase or situation, thus giving us a more sensitive description of the relationship between environmental factors and mode selection.

From an information organization point of view, the selected structure to house the information (sunburst in our case) must also allow the trained eye to become attuned to those environmental factors and modes that are not active. The fact that a certain factor or mode is not contributing to the pattern is important information and the graphical presentation must support it. This is the reason for forming an outer ring, above the black bars, so as to give an upper line of reference. After working with such sunburst presentation for some time, one begins to see not only the shape created by the bars, but also the shapes created by their absence.

This method of using "empty space" to provide additional information is commonly used by artists, designers, and architects. The two top pictures in Figure 4-6 are of the Temple of the goddess Hera, located in Italy, south of Naples, at Paestum (Greek Poseidonia). Dating back to around 550 BC, the huge temple (80 x 170 feet) was built in Doric style. The beauty of the Doric colonnade takes its form not only from the physical (stone) shape of each individual column, but also from the elegant vase-like shape that emerges between the columns (filled in with dark pen in the lower sketch in Figure 4-6). While it is difficult to scientifically determine how naïve viewers perceive such empty space, it is still a useful space into which additional features could be embedded in order to enhance the design, aesthetically in this case. Such space is referred to as "positive space" (Alexander 2002a, Ch. 5), and from our point of view is just valid a space to pack information into a limited display area. This method of packing additional information is one of the 15 properties for organizing space (as will be discussed in the next subsection).

The canonical correlation analysis performed on the data provided a set of common, or nominal, patterns of interaction. These patterns arise from two main factors: the regularity in the way ATC operates and manages traffic; and the standard way, shaped by the airline's philosophy and policies of operations, by which pilots respond to the demands from the environment. From the cockpit's side, any deviations from these statistically significant patterns represent an operational configuration which is different from the nominal way of operating the aircraft. Such *deviations* can be either nonstandard but completely safe pilot actions, or nonstandard actions that may have safety implications. From the ATC side, any deviations from the pattern may hint at either a nonstandard, yet perfectly safe, way of managing

Photo: QT Luong/terragalleria.com

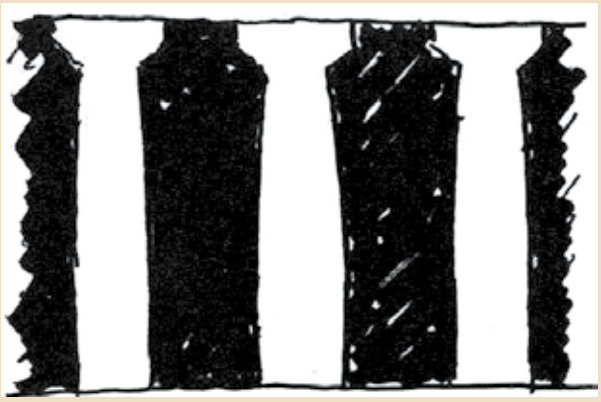


Illustration: Christopher Alexander

Figure 4-6. Doric columns in the Temple of Hera (Paestum, Italy).

traffic or a nonstandard clearance that may have safety consequences. Deviations can also reflect a changing condition at the airport (e.g., runway closure, change of navigation aids) or some new ATC procedures that affect traffic management.

Deviations from the nominal patterns of interaction are detected by computing the statistical patterns of X and Y relationships. For each individual pattern, we compute and present a bivariate plot of Xu_1 and Yv_1 (upper left corner of Figure 4-5). Deviation from the regression line can be identified statistically and visually. For example, note the broken line that encircles the two data points in the bivariate plot; both points correspond to non-nominal mode configurations. (Upon examination, the two configurations turned out to be safe, yet quite irregular for the phase of flight).

Organization of Multiple Patterns into a Whole

Thus far, the way we presented information has reached the level of information structures (per Figure 1-1). We used the sunburst structure to arrange all the factors and modes around a circle. We then used the inner ring as a reference line for both the black bars (protruding outwards) and the white bars (inwards). The outer circle is there to aid the eye see the negative space formed by the inactive factors and varying strength of the active factors. A rather simplistic form of integration, mostly juxtaposition of bars, allowed us to present a pattern of information within the sunburst structure in a rather efficient and clear manner.

Nevertheless, the $R^2 = 0.90$ pattern presented in Figure 4-5 is only one of four sets of patterns identified by the canonical correlation

analysis. And while it is possible to present each set of patterns separately (as in Figure 4-7), there are several problems with such a presentation approach. Assuming there is only one display, users will have to continuously switch between display pages in their effort to see all the patterns as a whole. And even if the four patterns are arranged close by on four dedicated displays, it is still difficult to make out

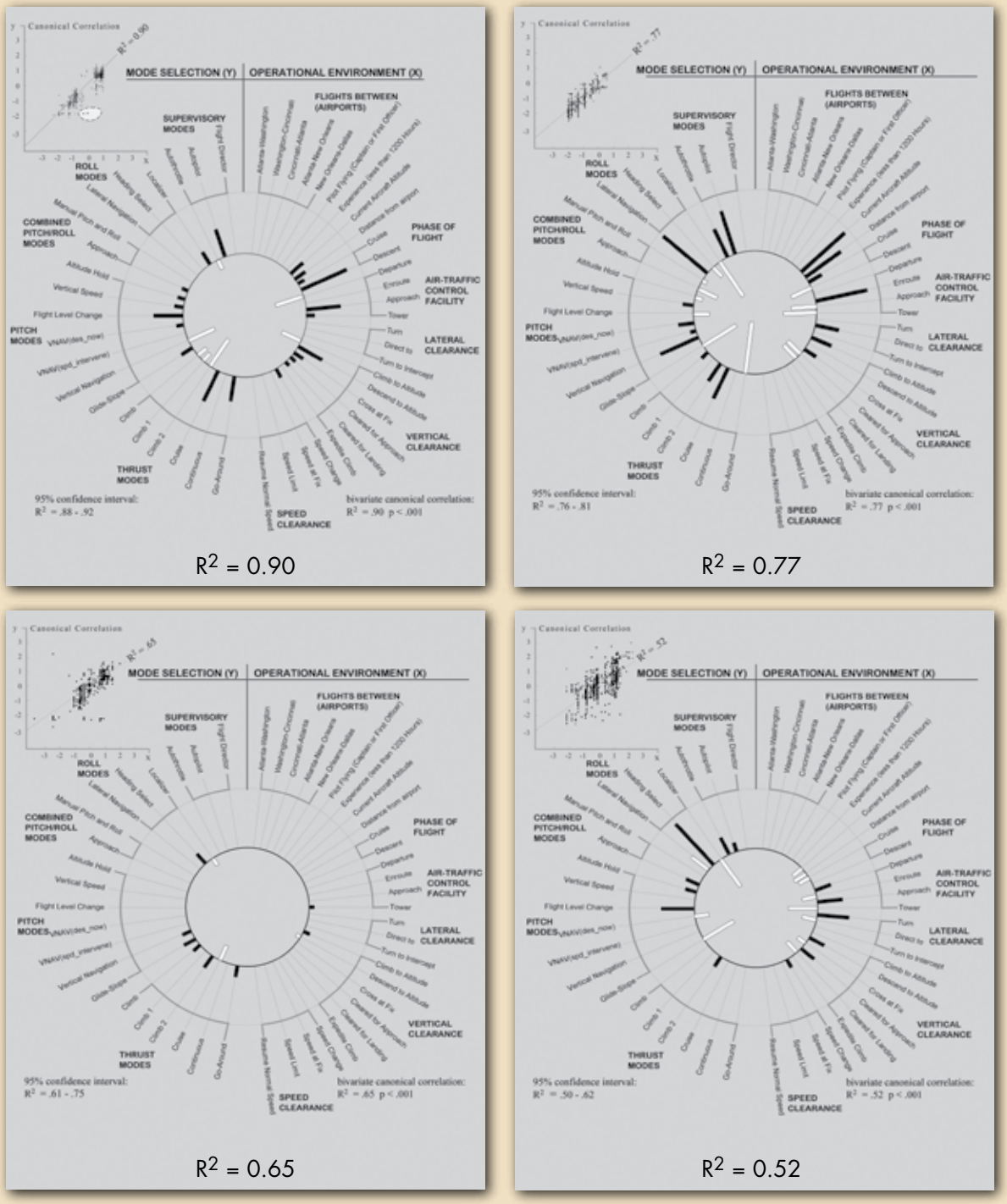


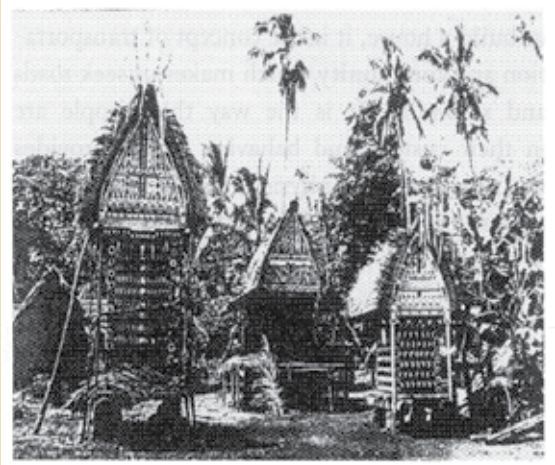
Figure 4-7. First, second, third and fourth sets of patterns.

the relationships among the different patterns. That is, how is a single factor active, or not, in several patterns, and how do sets of factors behave across patterns? This becomes even more critical when the analysis is computed dynamically and the resulting patterns change over time (e.g., deteriorating weather, tactical changes in ATC choices due to a partial system breakdown, airport construction, etc.). In such situations, we may want to see which patterns stay constant and which change. Do some patterns emerge at the expense of others?

To address this limitation, we decided to combine all four pattern sets within a single presentation (Degani, Shafto, & Olson, 2006a). Here our intent is to rise beyond the "structure of information" level to the next level—"order and wholeness." In our effort to create such a composite of patterns, we used the approach and concepts described in Christopher Alexander's *Theory of Centers* (Alexander, 2002a). The theory describes 15 heuristic properties that help create local integration and global, or holistic, organization in any given design space. Alexander's term *center* does not refer to a physical "center" per se, but rather to a coherent geometrical structure. In his work, spanning over 40 years of research, Alexander demonstrates that many visually powerful artifacts (e.g., religious sites and shrines, palaces, ancient prayer rugs, and sophisticated tilework) contain several, if not all, of these 15 basic properties (Alexander, 2002a, Ch. 4-5).

One important property in Alexander's theory is "levels of scale." It highlights how the existence of local coherent structures, at different sizes, all interlinked to create a whole, has a strong visual and emotional impact on the observer. Thus, with respect to the four sets of patterns, each with different level of statistical significance (R^2 = 0.90, 0.77, 0.65, 0.52), it became advantageous to pack them as concentric rings ordered according to their statistical significance level. This particular organization of the sets was inspired by a plan view of a typical Trobriand Island village in Papua New Guinea (Figure 4-8). The concentric organization of the village's houses and store huts corresponds to the social structure in the village (Alexander, 2002b, p. 343).

The circle of houses typical in Trobriand society. The circle was one of the major patterns which defined their culture



Physical character of the Trobriand houses around the circle. Here is another major pattern which defined their culture.

Photo: Kenneth Costigan

Figure 4-8. Plan view of a typical Trobriand village (adopted and reprinted from Alexander, 2002b, p. 343).

Figure 4-9 shows how the "levels of scale" property and the concentric organization were applied to pack in the four set of patterns. For ordering the circles, we used another one of Alexander's properties, called the void. This property defines a unique center that is placed at an important location, many times in the middle of an architectural space or an artifact, so as to draw the eye inward and beyond. A deep burial nook, high cathedral ceilings, and prayer niches are designed to impart a sense of infinity. Similarly, some oriental prayer rugs have a prayer niche representation in the middle of the rug to draw the eye into the abyss-like *void* (see Figure 4-10). The color of the void is usually dark, sometimes jet black. In Figure 4-9 the rings in are organized so as to portray that as the size and features associated with each ring decrease, so does their statistical significance (from $R^2 = 0.90$ to the lowest $R^2 = 0.52$). The implicit message is that as patterns become weaker and less significant, they collapse into the void.

Another property in Alexander's theory is *boundary*. Boundaries serve to tie a given geometrical form with its surroundings, and amply exist in nature (e.g., a beach surrounding a lake, reeds on the sides of a brook), in art (e.g., heavy brush strokes and/or shading around an object), and in architecture (e.g., the trim around windows and doors). They can also serve to create a sense of continuity. For example, wood shingles on the roof of a cottage help create a sense of continuity between the man-made structure and the surrounding redwoods. In Figure 4-9, the environmental factors' and modes' labels form a boundary between the inner world of data and information (values, significance, etc.) and the outer, operationally relevant, context.

One of the most striking property in Alexander's theory of centers is *interlock*. This property describes situations where coherent geometrical forms are hooked into others, creating fusion and coupling between two forms (Alexander, 2002a, p.195). A powerful aspect of interlock is that when coupling is created, another geometrical form emerges. A dovetail (carpentry detail) is an example of an interlock between two pieces of wood, where the dovetail shape emerges as separate form. With interlocks, additional information can be represented with little display real-estate penalty. In Figure 4-9 note the overlap between black and white bars of the same parameter (e.g., VERTICAL NAVIGATION mode), indicating a strong effect that shows up independently in adjacent rings. The degree of overlap corresponds to the strength of the effect (e.g., "Flight Level Change" mode between the white bars of the second pattern and the black bars of the third pattern).

Other properties from Alexander's theory employed in Figure 4-9 include *contrast* (between black and white bars), *gradients* (in the magnitude of bar sizes, which, for the purpose of representation, were abstracted into three categories—strong, weak, and none), and *Alternating patterns*. The latter property is present in the ray-like spokes that guide the reader's eye as the rings become smaller and shrink into the void. Figure 4-10 lists all the fifteen properties in Alexander's theory (and provides a visual representation for some). For a deeper understanding of these properties, see Alexander (2002a, Ch. 1-5).

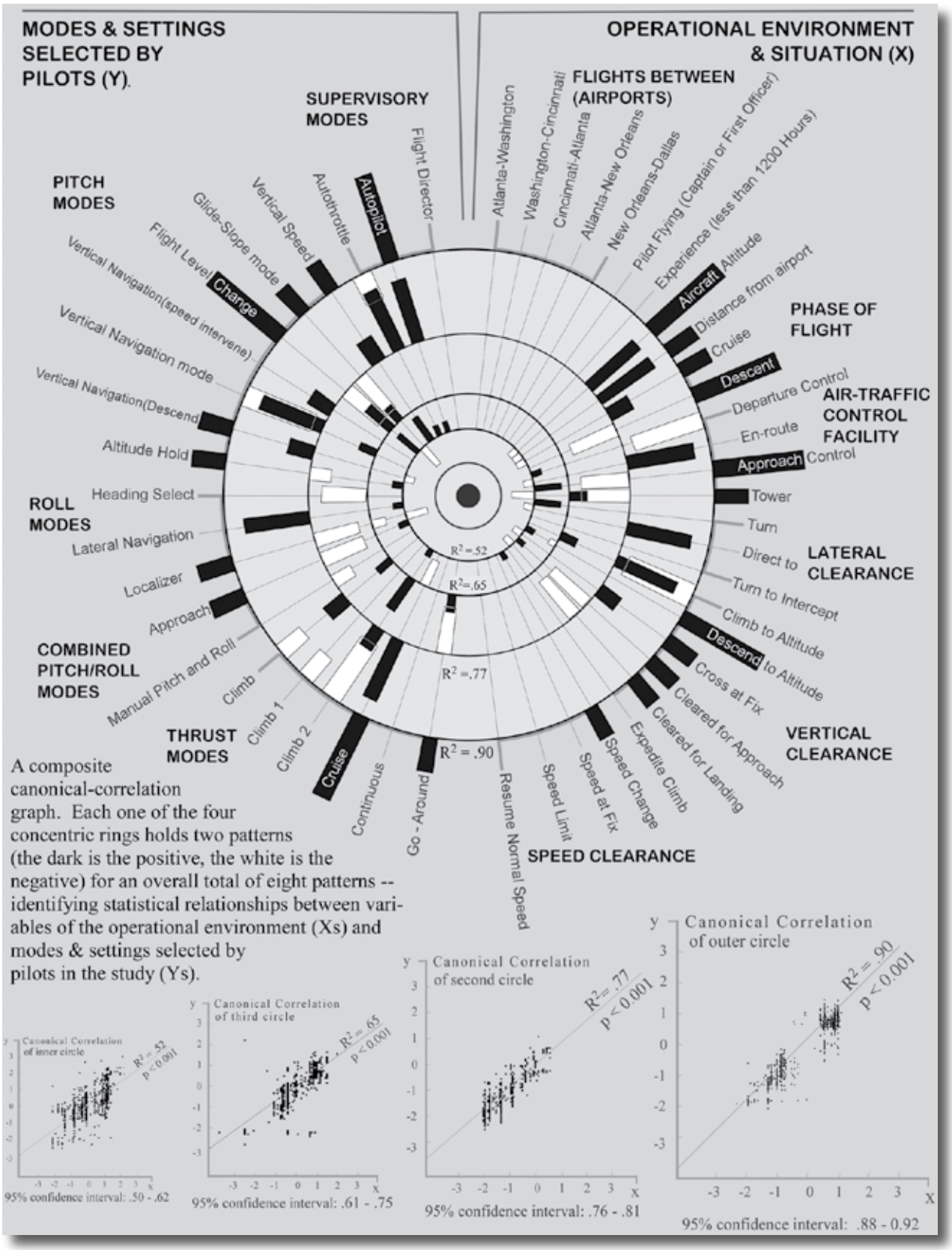


Figure 4-9. A composite canonical-correlation graph.

The organization of the patterned information into a whole allows us to pull out many interesting operational findings. We note how the "Aircraft Altitude" and "Distance from Airport" appears as an active pair in three patterns. Naturally, they are related—when altitude is high, distance from the airport is great. The pair "Approach" control and "Tower" also affect the unique and highly stable mode switching and configurations that take place as the aircraft nears the landing. Likewise, the ATC clearances "Clear for the Approach" and "Clear for Landing" is yet another pair which is operationally dependent, comes in sequence, and has a short time interval.

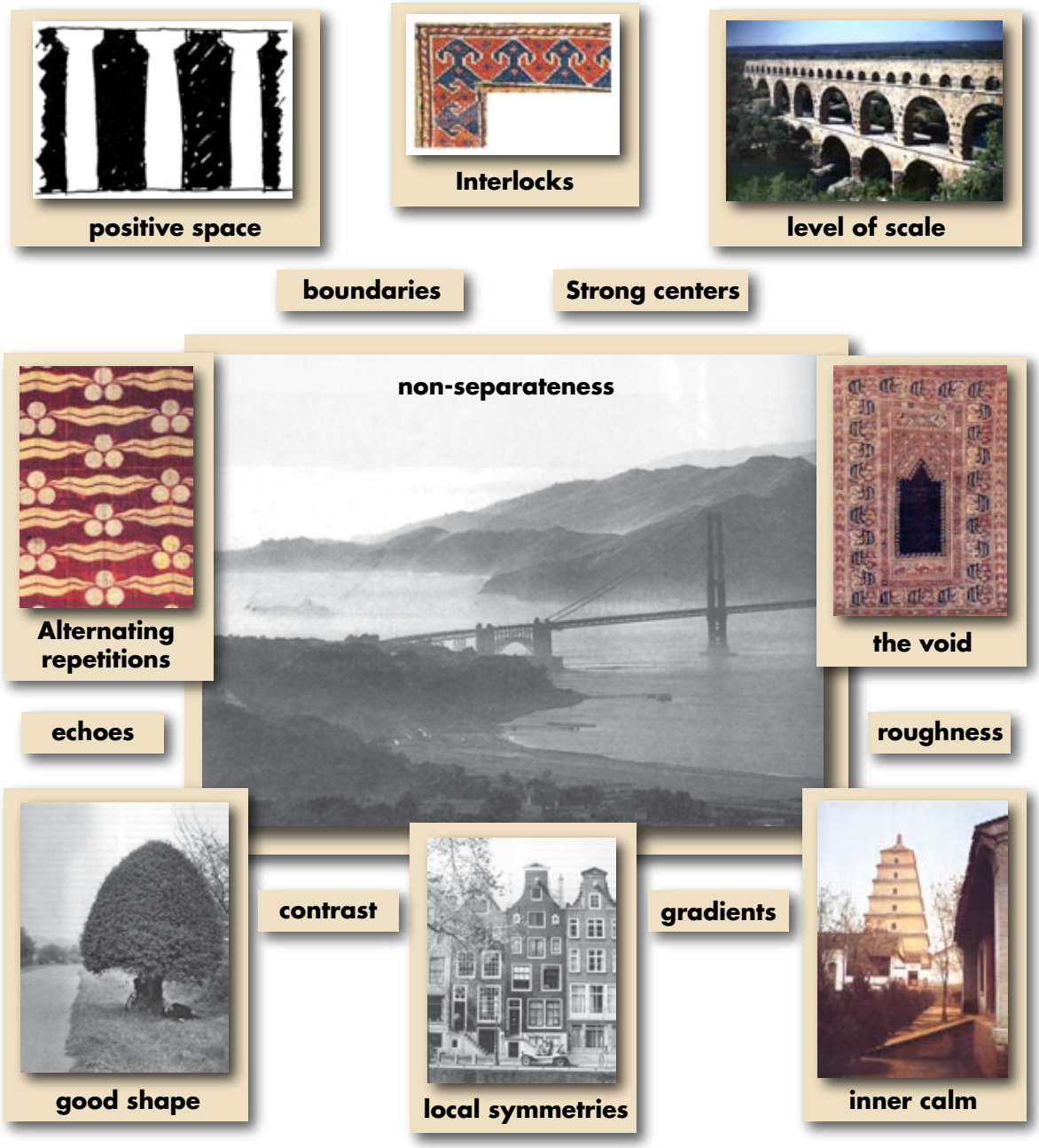


Figure 4-10. Alexander's 15 properties of a coherent space (2002a).

On the mode side, we see that CONTINUOUS THRUST is hardly used (as it is not a very fuel-efficient thrust mode). The "GO-AROUND" thrust mode is indeed armed in all configurations that involve descent and landing, as is required by standard operating procedure (and hence any deviation from that should be easily detected). CLIMB-2, a fuel-efficient thrust mode, is seen in four patterns, whereas CLIMB-1 (a less efficient climb mode) is active in only one pattern (white bars of $R^2 = 0.90$). The VERTICAL NAVIGATION mode, which is extensively used during climb and cruise in order to save fuel, is seen in three patterns. ALTITUDE HOLD is seen primarily during descent patterns, revealing that air traffic control rarely holds aircraft at various intermediate altitudes during climb. FLIGHT LEVEL CHANGE and GLIDE-SLOPE are temporally related, and the switch from the former to the latter takes place during the approach phase (recall the earlier discussion of the inappropriate use of the FLIGHT LEVEL CHANGE by continuing to use it at low altitude and not switching to GLIDE-SLOPE). As for autopilot and autothrottle use, we see that during some approach and landing patterns the aircraft is flown manually, while at high altitude both autopilot and autothrottle are engaged.

One final note regarding organization: Initially, before the four sets were organized into a single composite, we arranged the factors around the rim according to clusters. For the environmental side (X's) we started, quite arbitrarily, with the cluster of airports at the top of the sunburst structure, followed by pilot information and flight location (altitude and distance from airport); then phases of flight, and clearance type. As for the Y's, the arrangement of clusters of related modes was also done arbitrarily. For all practical purposes, we viewed the arrangement of the X and the arrangement of the Y variables as two distinct, unrelated, design issues. Once we moved up a level into considering the four sets of patterns as a whole, it became rather obvious that X and Y variables should be arranged in a way that captured relationships between related pairs (e.g., "clear for landing" (X) and GLIDE-SLOPE mode (Y)). We chose to represent such relationships by placing pairs of factors and modes on the same radial or nearby (e.g., "cross at fix" – VERTICAL NAVIGATION). Practically speaking, it turns out that it is quite difficult to achieve radial pairings for all related factors and modes because it breaks up the clusters, but it's definitely possible to arrange pairs on near-radials. Furthermore, it became just as advantageous to extend this arrangement scheme to sets of factors and modes (e.g., "Vertical Clearances" and "Pitch Modes") which correspond operationally. Naturally, not all factors modes and clusters have a related pair; for example, there is nothing on the modes side that corresponds to airports and nothing on the environmental side that can be paired with a thrust mode.

The above observation raises an interesting issue with respect to creating structures that reveal patterns. On the one hand we strive for a theory and methodology that will allow us to make such design choices (e.g., the radial pairing) a priori, as opposed to just stumbling upon them. At the same time, it is quite clear that no theory can anticipate the emerging patterns that the eye can detect, revealing to us new and unanticipated relationships. So what is the role of the designer in such a world of patterns? The design and theoretical effort, we believe, must first and foremost focus on creating fundamental structures and order in the way information is presented. From there on, the eye and human intelligence must be let go to roam freely (MacEachen, 1995).

Summary & Conclusions

In this section, we began to employ concepts of abstraction, integration, and organization as well as Alexander's Theory of Centers to construct an integrated display of statistical information. The findings from the statistical analysis point to a very strong relationship between mode selection and the structure of the operating environment. So strong is this relationship that one can predict, with a high degree of reliability, an aircraft's mode configuration just by noting a few environmental factors (Degani, 1996). This is, in part, because pilots in this study used only a limited set of mode configurations in response to a very structured and standard way that air-traffic control manages aircraft. Using canonical correlation analysis we were able to go further and reveal four sets of patterns that portray the strong relationship between the structure of the environment (and the standard way ATC manages aircraft) and corresponding mode configuration used by the pilots. Such patterns of interaction, running crisscross in the data, help us quantify the interaction (e.g., reveal which modes are used or not) and identify deviations from standard way of operating.

Throughout this section we dwelt on two statistical methods: linear regression and canonical correlation. Together, these two methods are the backbone of the abstraction approach described here. We tried to detail the methods and their application, because we believe that in order to develop sophisticated displays for representation of data and presentation of information, one must develop deep understanding of the underlying mechanics of abstraction. Only then does it become possible to design structures to hold the data and information and apply some of the more artistic concepts to create visual forms to portray the results.

The creation of the statistical pattern display can be explained according to the levels of the pyramid of Figure 1-1: Of the many measurable physical quantities available, only a subset was extracted by the observer and written down as abbreviated text on a notepad (see Figure 4-1). Similarly, audio signals from ATC frequencies were extracted and written down as abbreviated text on the notepad. After the completion of all flights, the notes were checked for errors and missing data. Once corrected, all signals were transformed into "Data."

The data were then arranged in a spreadsheet format and made ready for analysis (see Figure 4-2). Using two statistical analyses, Linear Regression and Canonical Correlation, the data points were abstracted to show main effects and patterns. In the case of the regression analysis, the four main effects identified by the analysis—aircraft altitude, phase of flight, ATC facility, and ATC clearance—were represented in numerical form (regression coefficient, standard error, t-ratio). As for the canonical correlation analysis, we abstracted the numerical (structured correlation) value of each variable and represented it as one of three distinct classes (strong, weak, or none).

Leaving the results of the regression at the information level only, we moved up to the structures of information level with the results of the canonical correlation. Here we employed a sunburst structure to contain all the variables (environmental factors and modes) and used geometrical forms (bars) to represent their statistical strength. Each variable was integrated with related variables to form a cluster along a sunburst structure, and, when applicable, linked by means of a radial line to corresponding variable on the other side of the sunburst. The rays emanating from the sunburst outline a (radial) grid which serves to unite the space. By representing the active variables (and their strength) on the sunburst structure, we provided for a pattern on each side of the circle. The combined overall pattern of environmental factors and automation modes is a presentation of their interplay.

The next level, order and wholeness, was achieved by organizing all pattern sets within a unified presentation. The packing of all four patterns into a single presentation allows us to identify relationships between sets of patterns. We can see pairs of factors or modes that are always active together and we can also see regions where there is high intensity of activity (i.e., where bars of a certain cluster are juxtaposed or where interlocks exist between similar variables of different pattern sets). Just as important, we can now also see regions where the intensity of activity is low or absent, indicating parameters, mostly on the environmental (X) side, that do *not* contribute much to pilots' responses. For example, the fact that the "flights between airports" region is not active for all pattern sets reveals that in the context of this dataset, there is nothing of significance about the idiosyncrasies of any particular flight between airport pairs. This observation suggests that the patterns are independent of geographical region and airport. Hence we can assume that they are general.

There are important implications of the above finding regarding the lack of airport effect. It suggests that when indeed a specific airport (or a cluster of airports) emerges, it may be due to some kind of out-of-ordinary situation. Such information can prompt an inquiry into the kind of policies and procedures, for example, used by ATC at the specific airport, or may hint at some unique weather-related, runway-related, or traffic flow problems. Along the same line, any deviation from a consistent pattern of environmental factors or mode selection is indicative of an event worth exploring. It may just be a non-standard event or it may be a situation that has no safety implication, or a situation that can deteriorate into an unsafe behavior. Currently, no such monitoring and display systems are utilized in any type of aircraft operations.

It is conceivable, however, that in the future, systems that monitor both the operational environment and flight crew's actions will be developed and deployed. Providing such systems within the cockpit, and also at airline and ATC control centers, may allow for online monitoring to detect deviations from a consistent pattern of environmental factors or mode selection. Some non-standard deviations may only require a cursory query while others (e.g., those that can potentially deteriorate into an unsafe situation) may warrant action. Although there is still a considerable amount of work to be done, both in terms of the statistical analysis and presentation (very simple for the cockpit, more involved for a control center) to develop such a monitoring system, the approach discussed here opens up the possibility of a statistical monitoring and alerting system that can detect non-standard situations while at the same time collect extensive information about how aircraft are managed and flown.

Similarly to the way many airlines currently analyze their aircraft flight data recorders to monitor performance and identify system-wide trends (e.g., FOQA and APMS), a system such as the one discussed in this section can provide for online computations of interaction patterns and early identification of deviations. This will require comparison of the aircraft's current configuration in the context of: (1) all aircraft in the same fleet; (2) aircraft on specific routes and airports; and (3) possibly even data from aircraft of other airlines flying to the same airport. For exploratory work along these lines, see Maille, Ferryman, Rosenthal, Shafto, & Statler (2006) and Statler (2007). Finally, while the approach presented here was applied to automation modes and the ATC environment, we believe that it can also be extended to other aircraft components and systems such as engines, hydraulic, pneumatic, electrical, etc.

5. An Integrated Engine Display

In the following section, we focus our attention on the presentation of engine parameters for a research helicopter. We first describe the engine data that were collected, and then discuss the application of the Inductive Monitoring System (IMS) for abstracting data into information. Taking a side track from the helicopter, we explain how the inductive monitoring system was used to analyze telemetry data from the space shuttle Columbia (STS-107) accident. We then discuss an old tradition, dating back to medieval times, which developed and employed integration and organization techniques to produce highly sophisticated geometrical forms. We analyze one such artifact, dating from the fifteenth century, and focus our attention on the arrangement of geometrical forms to define a visual field. In the context of the helicopter example, we show that it is possible to take the geometrical forms of this mathematically inspired fifteenth-century tilework and use it to embed engine parameters as well as information produced by the Inductive Monitoring System. We end with a brief discussion of how to extend the arrangement of engine parameters in order to embed other helicopter subsystems (hydraulic, pneumatic, etc.). The objective of the design is to geometrically present the multiple interrelationships that exist among various systems in the helicopter and help pilots develop a better (i.e., more integrated and holistic) understanding of the craft's behavior.

Data

Data was collected during engine run-ups and flights of NASA's Rotorcraft Aircrew Systems Concepts Airborne Laboratory (RASCAL) helicopter. Using special data buses that were installed on the helicopter, engine data was collected over a period of 18 days. There were several flights per day and each flight contains several data collection periods, for a total of 181 data collection periods. Altogether there were 520,941 data points. Four engine parameters were recorded for each of the helicopter's General Electric T700-701 turbo shaft engines:

- Power Turbine Speed (Np) (in % of rated value)
- Gas Generator Speed (Ng) (in % of rated value)
- Engine Torque (in % of rated value)
- Fuel Flow rate (in pounds of fuel per minute)

In addition, the speed of the helicopter's rotor was also recorded (in % of rated value). Figure 5-1 shows a snapshot of the data. The sampling rate is 4 Hz (i.e., one sample every 250 milliseconds).

As typical for experimental data collection, the signals coming out of the sensors and data buses were noisy and included periods of sensor failures. We use the following criteria for cleaning the data for analysis: (1) data collection periods in which parameter values oscillated between two extremes, suggesting a sensor failure, were removed; (2) values that were zero, indicating a sensor failure, were also removed; (3) short data collection periods (1-5 minutes long) that were not contiguous with other data collection periods, probably indicative of some kind of filing mismanagement, were eliminated. Of the original 520,941 data points, 178,293 were eliminated in accordance with the above criteria, resulting in a data set of 342,648 individual values.

During the design of aerospace engines, considerable knowledge about the idiosyncratic behavior of the engine is obtained through engineering analysis, simulations, testbed runs, and test flights. Two important aspects of such knowledge are systematic documentation of the multiple relationships among engine parameters, and how these relationships can be used to predict engine abnormalities and failures. Since many engine parameters are related to each other, they form a tightly knit web of information that can be used as a measure of the overall health of the engine. When we compute higher-order relationships such as those between pairs of parameters, triplets, quadruples, quintuples, etc., the web becomes exceptionally useful in its sensitivity to detect deviations from expected relationships. These deviations can point to (both known and novel) anomalies that are not easily detected when data are monitored piecemeal.

Photo: Dominc Hart

Left Engine				Rotor Speed	Right Engine			
L-Np	L-Ng	L-ff	L-Tq		R-Np	R-Ng	R-ff	R-Tq
100.589	82.733	10.375	49.806	100.698	85.366	11.063	18.758	100.464
100.589	82.624	10.313	51.877	100.698	85.151	11.063	18.348	100.366
100.589	82.624	10.375	52.427	100.581	85.257	11.063	18.348	100.366
100.472	82.733	10.438	52.013	100.581	85.577	11.188	18.485	100.366
100.472	82.838	10.438	52.978	100.464	85.471	11.250	18.758	100.269
100.589	82.948	10.500	53.670	100.581	85.577	11.313	18.895	100.366
100.589	82.948	10.438	49.806	100.698	85.577	11.250	18.895	100.366
100.589	82.838	10.438	48.837	100.581	85.577	11.188	19.032	100.366
100.589	82.838	10.438	47.873	100.581	85.577	11.188	18.895	100.464
100.589	82.733	10.375	47.044	100.581	85.471	11.188	18.895	100.366
100.706	82.733	10.375	52.564	100.698	85.366	11.125	18.758	100.366
100.706	82.624	10.250	50.220	100.698	85.366	11.125	18.622	100.464
100.472	82.518	10.250	50.080	100.581	85.257	11.125	18.622	100.366
100.472	82.733	10.438	50.771	100.464	85.366	11.125	18.758	100.269
100.472	82.838	10.438	56.568	100.464	85.577	11.313	18.895	100.269
100.589	82.948	10.438	54.084	100.464	85.686	11.313	19.032	100.366
[illegible]	82.948	10.438	52.705	100.698	85.686	11.313	18.895	100.366
[illegible]	82.838	10.375	53.256	100.581	85.577	11.188	18.895	100.464
[illegible]	82.733	10.375	51.463	100.698	85.471	11.188	18.758	100.464
[illegible]	82.733	10.313	48.701	100.581	85.366	11.125	18.622	100.464
[illegible]	82.624	10.375	49.943	100.581	85.366	11.125	18.758	100.366
[illegible]	82.733	10.313	50.494	100.581	85.366	11.125	18.622	100.366
[illegible]	82.733	10.438	51.048	100.581	85.471	11.188	18.622	100.366
[illegible]	82.838	10.438	48.978	100.581	85.577	11.188	18.758	100.366
[illegible]	82.948	10.438	48.837	100.581	85.577	11.250	18.895	100.366
[illegible]	82.838	10.438	52.705	100.581	85.577	11.188	18.895	100.366
[illegible]	82.948	10.438	52.013	100.581	85.686	11.250	18.895	100.366
[illegible]	82.948	10.438	54.912	100.698	85.686	11.250	18.895	100.464
[illegible]	82.838	10.313	54.634	100.698	85.577	11.125	18.622	100.464
[illegible]	82.624	10.250	51.736	100.698	85.471	11.125	18.485	100.464
[illegible]	82.624	10.250	53.392	100.581	85.471	11.125	18.485	100.464
[illegible]	82.624	10.250	48.978	100.581	85.471	11.125	18.622	100.366
[illegible]	82.733	10.375	53.119	100.581	85.471	11.125	18.758	100.366
[illegible]	82.838	10.375	53.256	100.581	85.577	11.250	18.622	100.366
[illegible]	82.733	10.375	52.978	100.581	85.577	11.188	18.485	100.366
[illegible]	82.838	10.375	55.189	100.698	85.577	11.250	18.895	100.464
[illegible]	82.733	10.250	53.670	100.698	85.471	11.125	18.622	100.464
[illegible]	82.518	10.188	54.634	100.698	85.366	11.063	18.485	100.464
[illegible]	82.518	10.188	57.396	100.581	85.257	11.063	18.348	100.366
[illegible]	82.518	10.250	47.732	100.581	85.366	11.125	18.485	100.366
[illegible]	82.733	10.250	54.498	100.581	85.471	11.188	18.622	100.366
[illegible]	82.733	10.250	53.947	100.464	85.577	11.188	18.758	100.366
[illegible]	82.733	10.313	62.643	100.464	85.577	11.188	18.758	100.366
[illegible]	82.838	10.313	61.678	100.581	85.577	11.188	18.758	100.366
[illegible]	82.838	10.313	50.080	100.581	85.577	11.188	18.622	100.366
[illegible]	82.733	10.250	60.021	100.698	85.471	11.125	18.622	100.366
[illegible]	82.733	10.250	50.634	100.698	85.366	11.125	18.485	100.464
[illegible]	82.624	10.250	54.220	100.581	85.366	11.125	18.348	100.366
[illegible]	82.733	10.250	56.291	100.464	85.366	11.125	18.622	100.366
100.472	82.733	10.375	52.978	100.464	85.471	11.188	18.622	100.269
100.472	82.838	10.438	62.092	100.464	85.577	11.250	18.895	100.269
100.472	83.053	10.500	56.709	100.581	85.792	11.375	19.032	100.366
100.589	83.053	10.500	23.989	100.581	85.792	11.313	19.032	100.464
100.706	82.948	10.438	28.958	100.698	85.686	11.188	18.758	100.366
100.589	82.948	10.438	39.036	100.581	85.577	11.188	18.895	100.464
100.589	82.948	10.438	39.036	100.581	85.577	11.188	18.758	100.366
100.589	82.948	10.438	36.415	100.581	85.577	11.125	18.622	100.366
100.589	82.838	10.375	36.552	100.698	85.471	11.125	18.485	100.366
100.589	82.838	10.375	36.966	100.581	85.471	11.125	18.348	100.366
100.589	82.838	10.375	35.310	100.698	85.471	11.125	18.485	100.464

Figure 5-1. The photo at the top shows the helicopter, a modified UH-60, on the flight line at NASA Ames Research Center. In the photo of the cockpit the current engine parameters' displays are surrounded by the green circle. On the right is a portion of the data.

Data Abstraction

Many high-risk industries are currently exploring anomaly detection systems in an attempt to help users better monitor and understand the behavior of complex systems. Using both archived and real-time data, anomaly detection techniques can identify unusual events (e.g., outliers) in streams of data. This is commonly done by analyzing "good" archived data (to characterize nominal system performance) and then comparing the results to incoming real-time data. Such comparisons can alert users when system behavior differs from nominal performance, and also relate these observed anomalies to known failures in the system.

One such system for anomaly detection is the *Inductive Monitoring System* (IMS) developed by NASA to monitor the behavior of aerospace systems (Iverson, 2004). The IMS takes input either directly from the system under consideration or from high-fidelity simulations. Initially, nominal data are processed and training knowledgebases are generated from them. (The term "training" is to signify that the IMS is being trained as to the expected values of the system under consideration). Numerical techniques are then used to characterize system behavior by identifying all regions of the nominal state space (Bradley, Mangasarian, & Street, 1997). Specifically, the IMS algorithm employs two clustering techniques to identify these regions: K-means clustering (Bradley & Fayyad, 1998) and density-based clustering (Ester, Kreigel, Sander, & Xu, 1996).

The basic data structure of the inductive monitoring system's algorithm is a vector of parameter values. Depending on the sampling rate, a new vector is input into the system at the selected time interval. Figure 5-2 is a vector of left and right engines parameters and rotor speed for the RASCAL helicopter. Theoretically, each such vector defines a point in N-dimensional space. The intention behind the processing of the data is to create a representation that allows us to analyze all parameters of related vectors together.

The steps in the IMS algorithm are as follows: First, the training data is formatted into a vector and then the training knowledgebase is created. This knowledgebase contains clusters that define the ranges of "allowable" value for each vector's parameters. The vector of high values and the vector of low values in a cluster can be thought of as corners defining a minimum bounding rectangle in N-dimensional

|---- Left Engine ----||---- Right Engine ----|								
Left Engine Np	*Left Engine Ng*	*Left Engine Fuel-Flow*	*Left Engine Torque*	*Right Engine Np*	*Right Engine Ng*	*Right Engine Fuel Flow*	*Right Engine Torque*	*Rotor Speed*
100.589	82.733	10.375	49.806	100.698	85.366	11.063	18.758	100.464

Figure 5-2. The vector of left and right engines parameters and rotor speed.

space. Points that fall inside this rectangle are considered to be within the system's nominal operating range. The high and low ranges for each parameter in the cluster are considered as allowable ranges, provided that other parameters are within the ranges specified in that cluster (cf. Hamscher, 1991).

Creation of the knowledgebase begins with an empty cluster. The first training vector becomes the initial cluster, and each subsequent vector is compared to the existing clusters to find one that is closest to it. The distance between a given vector and a cluster can be measured using a variety of metrics, but the standard Euclidean distance has proven effective for most applications. The point that represents a cluster can be computed in many ways. One option, based on the K-means clustering method cited earlier, measures the distance from the centroid of the cluster, computed by forming a vector from the averages for each parameter in the cluster.

As in density-based clustering (Ester, Kreigel, Sander, & Xu, 1996), a threshold value, ε, defines the maximum allowable distance between a cluster and vector to determine if the vector should be incorporated into the cluster. (The value of ε is defined based on experience or heuristics). If the vector is close enough (distance less than or equal to ε), the cluster parameter intervals are expanded to include the new vector. If the distance between the new vector and the closest cluster in the database is greater than ε, a new cluster is formed and incorporated into the knowledgebase. This process is repeated until all of the training data has been processed and incorporated into clusters. The result is a knowledgebase of clusters that characterize system performance in the normal operating envelope.

Once the training knowledgebase is in place, the next phase concerns the analysis and identification of anomalies. Here we take in real-time data that need to be monitored for anomalies and format the input data into a vector. The next step involves query of the training knowledgebase to locate the cluster that best matches the incoming vector. The algorithm then finds the distance of the vector from the nearest cluster. If there is no cluster in the knowledgebase that contains the new input vector, the system is behaving in an unexpected manner, indicating a possible anomaly.

The objective of the analysis is to provide the user (e.g., a pilot, astronaut, or mission controller) with an idea of how far the system behavior is deviating from nominal operations. In situations where the input vector is not contained within a known cluster, the distance between the vector and the closest point in the bounding rectangle of the nearest cluster is reported. The implication of the distance between the vector and the bounding rectangle depends on what is defined as an acceptable tolerance range: An input vector close to the acceptable range warrants extra vigilance, while an input vector beyond the tolerance calls for immediate attention.

In many situations, it is quite advantageous to scale or normalize the data values *before* they are inserted as vectors into the knowledgebase. For instance, a parameter value can be represented as percentage of the maximum value or in terms of standard deviation. Likewise, the output data can be scaled so that the distances between vectors and clusters represent a percentage of the maximum distance in the knowledgebase. Such derived variables provide a more meaningful representation of a given parameter value, allow us to compare parameter values of different units (e.g., pressure and temperature), and provide a more comprehensible measure of how much the system is deviating from nominal operations.

Application to Aerospace Systems

One of the first studies with the inductive monitoring system involved an analysis of the temperature sensors in the wings of the Space Shuttle *Columbia*. The ill-fated mission, STS-107, came to a disastrous end during reentry to earth's atmosphere on February 1, 2003. The probable cause of the accident was a breach in the thermal protection system on the leading edge of the left wing. The breach was most likely caused by a piece of insulating foam that separated from the shuttle's external tank and struck the left wing approximately 82 seconds into the flight. The first physical indication of the consequence of the damage was noticed by mission controllers during the re-entry, 17 days after the launch (Gehman et al., 2003, pp. 64-67).

A post hoc analysis of the *Columbia* temperature sensors inside the wings was conducted using the inductive monitoring system described above (Iverson, 2004). The knowledgebases for the launch and ascent phase were generated from data collected during five previous *Columbia* flights and then compared to telemetry data from STS-107. The analysis focused on four temperature sensors in the left wing of the shuttle and four sensors on the right wing (Figure 5-3). Since ambient temperatures differed on each flight, the data was normalized to the most centrally located sensor, "Brake Line Temperature B," by expressing the other sensors' values in relation to the value of this sensor, thus yielding a vector size of three.

In addition, monitoring was further focused by comparing sets of data in a moving four-second window (typically 12 to 16 data points). The data sets used for training and analysis covered the time period from launch to just before the Main Engine Cut Off (MECO) point, which is about eight minutes into the ascent. The left wing knowledgebase contained 502 clusters and the right wing knowledgebase contained 240 clusters.

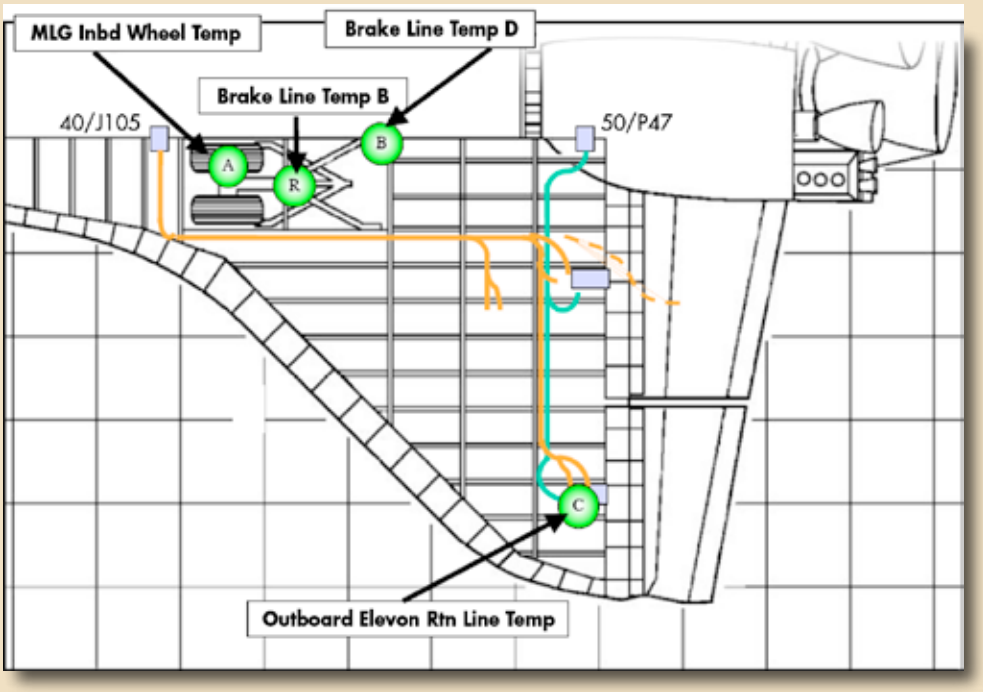


Figure 5-3. Columbia's left wing and location of the four temperature sensors used in the analysis.

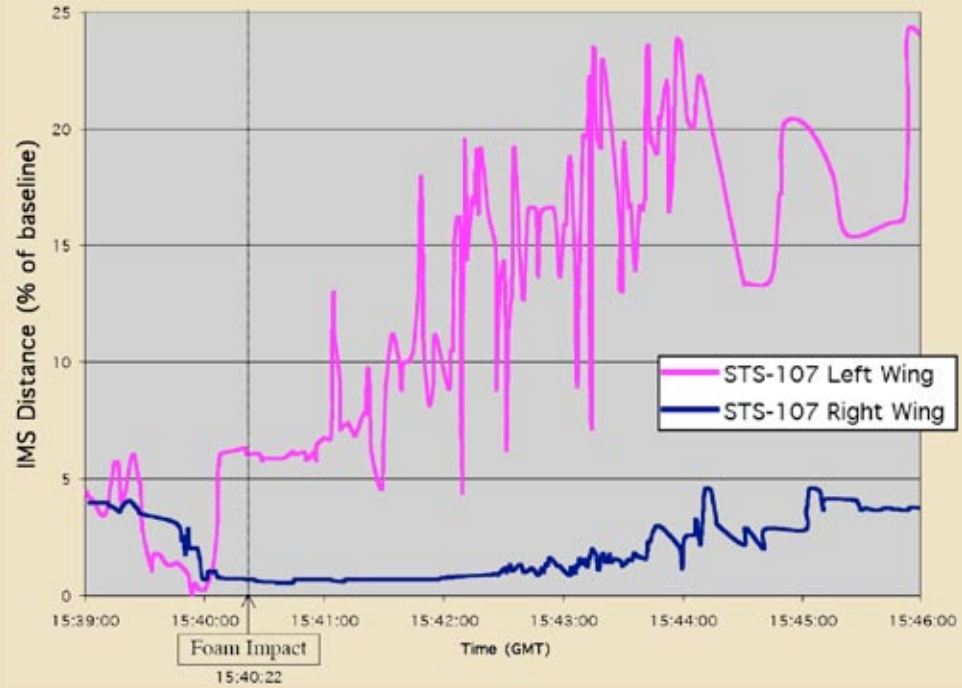


Figure 5-4. Results of the IMS analysis.

The results of the analysis are presented in Figure 5-4. The horizontal axis represents time, beginning at the moment of liftoff; the vertical axis represents a measure of deviation from nominal behavior. The deviation is measured as the distance of the input vector from the closest (nominal) cluster computed as a percentage of the maximum distance in the vector space covered by the training data. Results for the left wing are marked by a magenta line; results for the right wing are in blue. The vertical line at time 15:40:22 hours shows the moment the foam struck the wing. The left and right wing results track each other fairly well until shortly after the foam impact, at which point they begin to diverge: The deviation values for the right wing are within five percent of maximum distance; the left wing values increase to nearly five times those of the right wing (Iverson, 2004).

The post hoc analysis of the *Columbia* accident data suggested the merit of using a data mining tool, such as the IMS, to abstract information from the data. We began by normalizing the values from four sensors as a way to reduce noise (from ambient temperature). We then abstracted the resulting three values into a single vector. Vectors from previous uneventful launches were fed into the IMS and represented as points in an abstract space. Vectors from STS-107 were then fed into the system in order to see how much their points' deviated from expected. The results showed significant and continuous deviations from expected values as well as from values observed on the right wing. The results are quite striking in the sense that it became possible to pick up meaningful information that was previously concealed in the data.

It is important to note here that this analysis was post hoc; it took place after the accident occurred. The analyst knew that a breach occurred during liftoff and that temperature sensors are best suited to reveal the problem. However, the analysis does suggest the possibility of using a methodology such as IMS to detect novel situations where standard detection schemes that do not take into account combi-

nations of values and historical data are just not sensitive enough to pick up subtleties in the data. If we wish to use the IMS to deal with both known and novel situations, then we should ask which signals best represent the system under consideration for such an analysis. Likewise, the abstraction process of deciding which set of sensors to combine into a vector must be done based on some judicious criteria. Currently there are no systematic and rigorous methodologies to address these extraction and abstraction decision; both are done ad hoc and through trial and error. In the following subsection we describe some of the considerations in creating such sets in the context of the helicopter data.

Abstraction of Engine Parameters

As discussed earlier, the inductive monitoring system is designed to deal with sets, or composites, of parameters; it is possible to compute IMS values for pairs, triplets, quadruples, quintuples, and so on. When there are strong interrelationships between parameters in a composite, the IMS results are very sensitive to any deviation from expected values and can indicate an anomaly long before it manifests in individual parameters. If, for example, we know that parameters *a* and *b* are statistically correlated, then it may be advantageous to create a composite from them and monitor for anomalies. However, simply adding parameter *c* to the composite of *"a & b,"* when *c* is independent (i.e., uncorrelated with both *a* and *b*) does not improve the quality of the representation. But when *c* is indeed related to *a* and *b*, the triplet *"a & b & c"* is more sensitive to deviation than all of its pairs *("a & b," "a & c," "b & c")*. Special composites can be created to reflect a combination of parameters that correspond to a known failure or a unique operational situation that should be noted by the pilots. As mentioned in the previous subsection, not every parameter within a composite is necessarily treated similarly; weights can be placed on parameters to reflect their relative contribution to, for example, a known engine failure such as compressor stall.

For the helicopter example, we have four parameters for each engine and one parameter, rotor speed, which relates to both engines. Analysis of the data for both engines revealed meaningful interrelations between the following parameter pairs:

- "Rotor Speed" & "Power Turbine Speed (Np)"
- "Gas Generator Speed (Ng)" & "Fuel Flow"
- "Gas Generator Speed (Ng)" & "Engine Torque"
- "Engine Torque" & "Fuel Flow"
- "Power Turbine Speed (Np)" & "Gas Generator Speed (Ng)" (moderate relation)

If we were to arrange the five parameters as an ordered list, such that adjacency corresponds to a meaningful pair interrelationship, then the following is one suitable order: "Rotor Speed," "Power Turbine Speed (Np)," "Gas Generator Speed (Ng)," "Engine Torque," and "Fuel Flow."

If indeed we decide to put together the engine parameters in this order for display, then it may be possible to present the abstracted IMS results for each parameter pair somewhere in the space between each (adjacent) pair. We can represent the situation when IMS results are abnormal with some geometrical shape and/or color corresponding to the severity of the abnormality. Naturally, the approach can be extended to triplets, quadruples, quintuples, and so on, depending on the number of parameters.

To present all the parameters in a coherent manner we first need to address how to represent individual IMS values in graphical form. Second, we need to determine the set of possible composites. In the case of the helicopter engine parameters, we have 10 possible pairs, 10 triplets, 5 quadruples, and 1 quintuple to choose from. And this combinatorial list is only for one engine and does not include unique combinations (e.g., with different weights) that are in place to reflect special type of failures. Third, we need to address the problem of how to visually present the relationships between raw parameter values and composite IMS. Last, we are faced with the difficult problem of how to organize information about IMS pairs, triplets, quadruplets etc., in a way that reflects their interrelations. As we have seen before in section 4, when numerical values are represented in tables, it is quite difficult to identify relationships and observe patterns. And while we can use a graph format such as the one used in Figure 5-4 to present IMS results, it does not solve the problem of how to show relationships between the many graphs.

Hence we now focus our attention on the problem of how to organize sets of information that are highly interconnected, so as to reveal how they are related to one another and observe the resulting patterns. Specifically, in the context of this helicopter example, we need to address the problem of how to organize raw parameters and the abstracted IMS results. Furthermore, the engine is not a stand-alone system; it is connected, directly or indirectly, to other systems such as hydraulic, pneumatic, electrical, etc. As will be discussed in the following subsections, the search is for the properties of a geometrical structure that contain order.

An Old Tradition with a New Application

There used to be a business maxim that "information is knowledge," but we are slowly finding that although one may have access to considerable amounts of data (via the Internet, for example), real knowledge can only be achieved when this data is relevant and well organized, otherwise attention is spread and interconnections are missed (Simon, 1971). When we talk about organization, we are really talking about creating order. If there is order in the way data is abstracted, integrated, and organized, then it is then possible to comprehend the behavior of the system under our control, as well as relationships to other subsystems and the environment.

To proceed deeper into this issue of order, we first begin by classifying two main attributes of information. As we discussed earlier in section 1, one aspect of information is how signals of the object under consideration are extracted, transformed, and then abstracted—that is, the process of picking up raw signals and deciding what is relevant and what is not. This process is referred to as correspondence (Brunswick, 1956; Goldstein, 2006; Mosier & McCauley, 2006). Technically speaking, it involves the establishment of a mapping function between the object under consideration, through the physical quantities it emits, the signals received and data collected, and the abstracted representation (Card, Mackinlay, & Shneiderman, 1999, pp. 17-23; Degani & Heymann, 2002). We have seen a heuristic mapping approach in the case of the London Underground Diagram, statistical mapping in the pilot-autopilot analysis, and a computational mapping with the Inductive Monitoring System. A similar mapping function has been developed to allow for a formal analysis of user interfaces, where the user interface is considered as an abstract representation of the underlying state space of the machine (Degani, 2004; Heymann & Degani, 2007; Oishi, Tomlin, & Degani, 2003). It is possible within that framework to develop criteria for and formally judge the correctness of the mapping.

The process of integrating abstract representations of information and then organizing them to communicate a sense of wholeness is referred to as coherence (Brunswick, 1956; Mosier & McCauley, 2006). Coherence relates to the relationships among individual pieces of data and information presented to the viewer. In the context of this research, coherence relates to structures of information. It can be thought of a measure of the "glue" that binds geometrical forms together, as a jigsaw puzzle, to create a unit (or a "center" in Alexander's terminology). But unlike correspondence, where a mapping function can be created mathematically, the problem of coherence is rather elusive. We currently do not have a methodology to create coherence, nor do we understand its building blocks. As such, coherence is still in the realm of intuition. Nevertheless, coherence has been explored and created in the past. Powerful geometrical patterns with sophisticated coherence can be found in Australian Aboriginal art, North-American Indian practical designs, textile work from Central Africa, Indo-Tibetan mandalas, Meso-American architecture, and old Oriental rugs (Figure 5-5). The intent behind such designs was not only to mold the space in a perceptually coherent and aesthetical way, but sometimes also to use the resulting patterns to transmit messages (e.g., maps and directions, cultural motifs, and philosophical ideas) to the viewer.

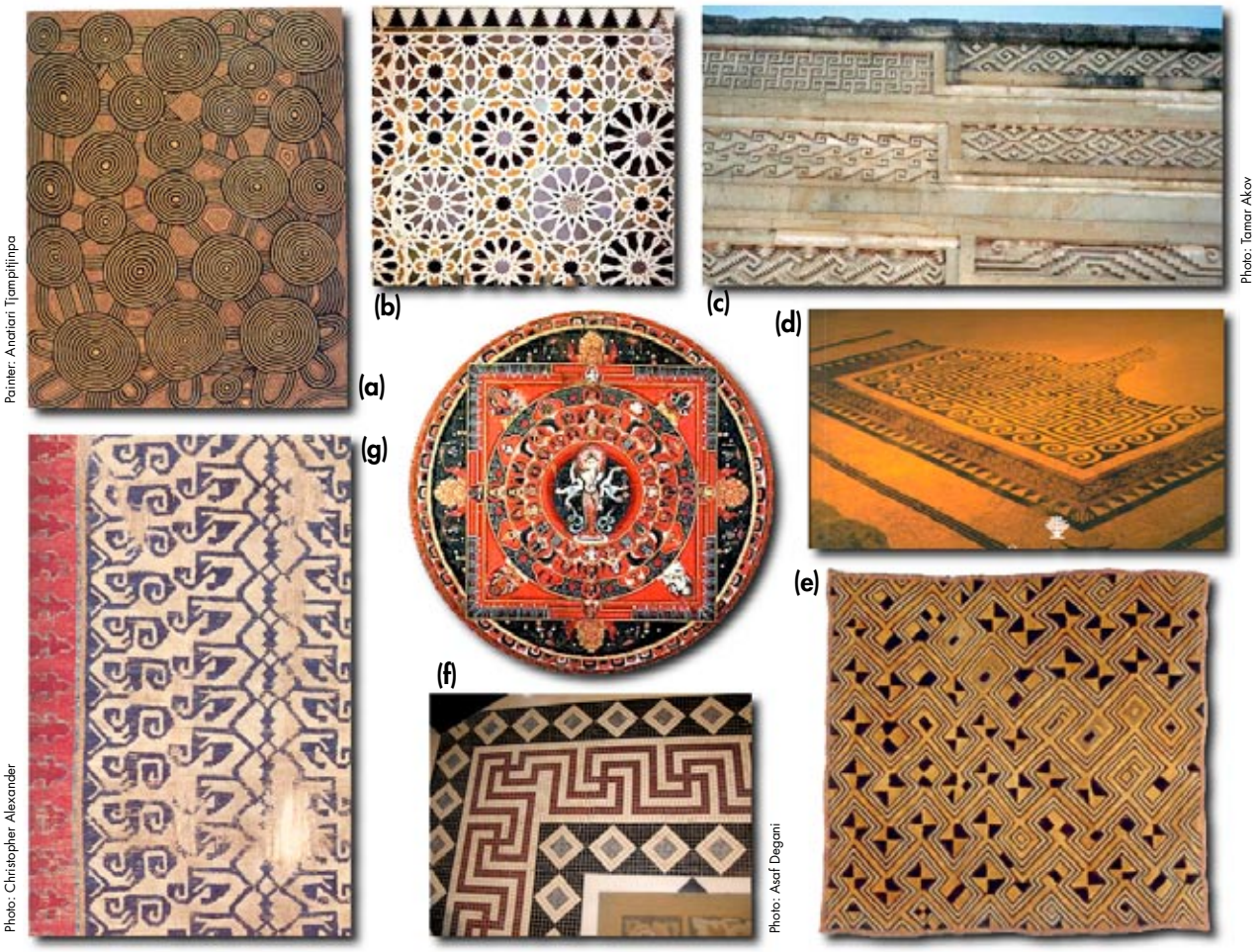


Figure 5-5. Pattern designs from various cultures: (a) Aborigine map of water holes and soaks in the desert; (b) tilework from the Alhambra Palace, Granada, Spain; (c) Zapotec stonework from Mitla, near Oaxaca, Mexico; (d) mosaic floor tile of a Roman Villa, Jerusalem; (e) African textile designs (Showa); (f) "running dog" pattern from the Hotel Grande Bretagne, Athens, Greece; and (g) *Seljuk Carpet with Infinite Repeat of Dragons*, Central Anatolia, Turkey. In the center is a Tibetan Buddhist mandala. (Photos not to scale).

While the essence of coherence is hard to pin down, Alexander's work and theory suggest four major aspects of coherence (2002a). One, coherence appears to spring up from multiple interrelations that are somehow brought up to the surface. A coherent structure portrays a lively intensity between small and large elements of the design. Two, the presentation of the many relationships in the design is based on some underlying grid. We may not necessarily see the grid consciously, but we definitely feel its presence. Three, information is commonly presented in layers, and there are always multiple messages embedded the design. Four, there is permeating sense of wholeness—a feeling that the space is intense, multidimensional, and visually complete.

These aspects of coherence, albeit subtle and hard to quantify, can be beneficially applied to presentation of modern information. The ability to display relationships among parameters (e.g., sensor data), for example, is a necessity during normal operations and becomes vital during an emergency (Heymann, Degani & Barshi, 2007; Mosier & McCauley, 2006). Order is a must for any organization of cockpit information, and pilots strive to see all the cockpit systems as a whole. In this respect, the geometrical forms in medieval tilework that adorn monuments from Spain to Central Asia and all the way to Eastern India are some of the best known examples for organizing space (Necipoglu, 1995). These geometrical forms, cleverly integrated with each other to create dense compositions on wall surfaces of public buildings, palaces, and shrines, were created by designers and architects who have also been competent in geometry and mathematics (Ozdural, 1995). The realization that many of these artifacts are mathematically inspired and correspond to an underlying geometrical structure makes them quite appealing for the objectives of this research.

The roots of this tradition are believed to stem from a culture that flourished in Central Anatolia for centuries (Necipoglu, 1995). Original work involved Neolithic wall paintings, later converted to rug designs (Alexander, 1993). By the eighth century, many of the abstract geometrical forms began to show up as wall decoration. Delicate brickwork, stonework, terra cotta, and paint were used as revetment for buildings, with virtuoso flourishes of light and shadow creating fields of stars and sunburst-infused patterns (Michaud, Michaud, & Barry, 1996, p. 32). From around the eleventh century, fired tilework with radiating colors was used to cover the façade of important buildings. The knowledge of how to organize space and integrate geometrical forms using mathematical and geometrical principles was initially developed somewhere around the tenth or eleventh century and reached its peak during the fifteenth century. The resulting spectacular geometrical designs can be found in public spaces, palaces, and shrines throughout the Middle East and Central Asia (see Figure 5-6).

How these intricate and geometrically sophisticated designs were created still puzzles art historians, mathematicians, and architects. Presentations of infinity with Penrose-like tiling (Lu & Steinhardt, 2007) and new classes of symmetry that were previously unknown have been identified in these designs. Some patterns contain up to five different levels of sub-patterns, all interconnected geometrically to create a multidimensional presentation. In these formations, there is heavy reliance on geometrical arrangements that are highly integrated and appear whole to the human viewer (stars, sunbursts, flowers, leaves, etc.). While the architectural designs of this tradition were previously venerated primarily for their aesthetic quality, modern mathematicians and art historians ask questions concerning the mathematical structures that underlie these designs, as well as the meaning and messages that medieval architects and mathematicians tried to impart (Grünbaum, 2006; Necipoglu, 1995). Many modern scholars believe that one important cultural and philosophical message that these artifacts transmit to the viewer concerns interconnectivity between elements, reflecting the notion of "unity in multiplicity" which informs many of the architectural elements of this tradition. (Necipoglu, 1995 Ch. 5; Michaud, Michaud, & Barry, 1996).

(a)

Photo: Roland & Sabrina Michaud

Figure 5-6. Medieval Islamic patterns: (a) Tilework from the Buyuk Karatay Madrese, Konya, Turkey (13th Century); (b) Friday Mosque, Kerman, Iran (16th Century); (c) "Tiled" (Sircali) Madrese, Konya, Turkey (13th Century); (d) fourteenth-century tile, terracotta, and paint work from Natanz, Iran; (e) Timurid shrine of Khwaja Abdullah Ansari in Gazarghah, near the town of Herat, Afghanistan; (f) tile pattern on a spandrel from the Darb-i Imam Shrine in Isfahan, Iran (15th Century); (g) portal from the Darb-i Imam Shrine in Isfahan, Iran (15th Century) (Photos not to scale).

(b)

Photo: Roland & Sabrina Michaud

(c)

Photo: Roland & Sabrina Michaud

(d)

Photo: Roland & Sabrina Michaud

(e)

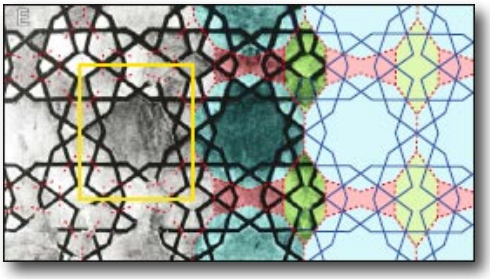


Illustration: Peter J. Lu.
(Taken from Lu & Steinhardt, 2007)

(f)

Photo: K. Dudley & M. Elliff

(g)

Photo: K. Dudley & M. Elliff

Analysis of a Fifteenth-Century Tilework

Figure 5-7 is a picture of a tilework dating back to the fifteenth century. It shows a wild and powerful composition, replete with stars and flowers that are connected via interlocks. As discussed earlier, it appears that it is possible to employ this structure in order to organize information such as engine parameters. To better understand how order and organization was achieved, we start with a brief analysis of the tilework using Christopher Alexander's Theory of Centers. Throughout the analysis, we identify centers, or "perceptible region[s] of coherent space," in the design, and consider how they relate to and are linked to one another (Salingaros, 1995). In his writings, Alexander argues that every human-made artifact has such centers, some strong like an altar in the middle of a cathedral, and some weak like a mundane office window; some big like a tree-lined boulevard, and some minute like a rug detail. Proximal and overlapping centers can reinforce and support one another to create a sense of wholeness and interconnectedness. Centers can also support one another from a distance—for example, by employing similar patterns in such a way that the eye picks up the association. Centers, of any kind, are important for the purpose of representing and presenting information; they are building blocks of any integrated design.

Figure 5-7. Fifteenth-century tilework from Gazarghah, near the town of Herat, Afghanistan. Interconnections are highlighted on the right.

The large half flower mid-way down the right side forms a main center (Figure 5-8 (a)). It has six petals, each with a brown-orange anther in its middle. Extending beyond the dark blue petals are white four-sided polygons (quadrilaterals), appearing like a garland around the main flower (Figure 5-8 (b)). A similar flower formation appears in the lower left side, and a small portion of another flower can be seen in the top left corner. There are two additional flower formations in the design (middle left side and lower right corner), also in light blue but with smaller and somewhat less elaborate petal formations and no anthers. The space between the flowers is filled with five-pointed stars, made up of a black pentagon in the center and five white kite-shaped quadrilaterals (Figure 5-8 (c)). Each star is interlocked with its three neighbors by means of these white quadrilaterals. The white quadrilaterals also serve to interlock the stars with the flowers. So are these white quadrilaterals a garland on the flower or part of the star? The answer is, of course, that they belong to both.

In the middle of each star is a blue circle. The blue circle attracts the eye to the center of the star. Because of the emotional connotation of stars as unfathomably remote celestial objects, we may associate the blue circle as a representation of a "void"—something that is beyond our concrete understanding yet acts as an attractor (into the infinite). Going back to more earthly matters, the blue circles of the stars can also be viewed as an additional (outer) garland for the main flower. Note that two of the blue circles in the middle of the design also serve as an outer garland for the smaller flower (on the left, opposite the main flower). As such, the two blue circles chain link the main flower (on the right) with its smaller neighbor (on the left). Likewise, the two six-pointed stars, located between the main flowers connect them via the joint black pentagons and the use of the brown-orange anthers (which is identical to the ones seen on the flowers' petals). Inside the six-pointed star is a rhomboid formed by the two black pentagons and the two quadrilaterals (with the brown-orange anthers appearing like two eyes, see Figure 5-8 (d)). The two side edges of the rhomboid point to the two main flowers and the two bottom/top edges point to the smaller flowers; thus the rhomboid serves as an intersection between the four flower centers that dominate the design. This is an example of how two main centers (flowers, in this design), that are physically apart can connect through the use of a variety of intermediary forms (blue circles, black pentagons, and rhomboids).

Beyond the garlands and the stars, note how three of the five white quadrilaterals of each star come together as a triangular ray extending from the main flower (shown in light yellow in Figure 5-8 (e)), thereby making each of the flowers appear from a distance like a radiating star, and the blue circles discussed earlier as voids now come to represent, perhaps, planets circling the radiating star. This constant "double duty" of geometrical forms (e.g., one time representing a flower and the next representing a radiating star) is what makes such abstract artifacts appear to us infused with many lively objects (or information, from our perspective). The design does not change its form from a flower to a star, but rather it is our eyes and internal interpretation of certain geometrical forms, which manifests this double meaning.

How such sophisticated integration of geometrical forms is conceived and created is far from trivial. So far no manuscripts, theoretical treatises, or manuals explaining the process of creating such geometrical designs have emerged. All that we have today are actual built examples that survived and several books and scrolls that contains designs, but with no textual explanation (Necipoglu, 1995). We can only analyze current designs and try to understand, in a reverse engineering kind of a way, how the integration of geometrical forms was done.

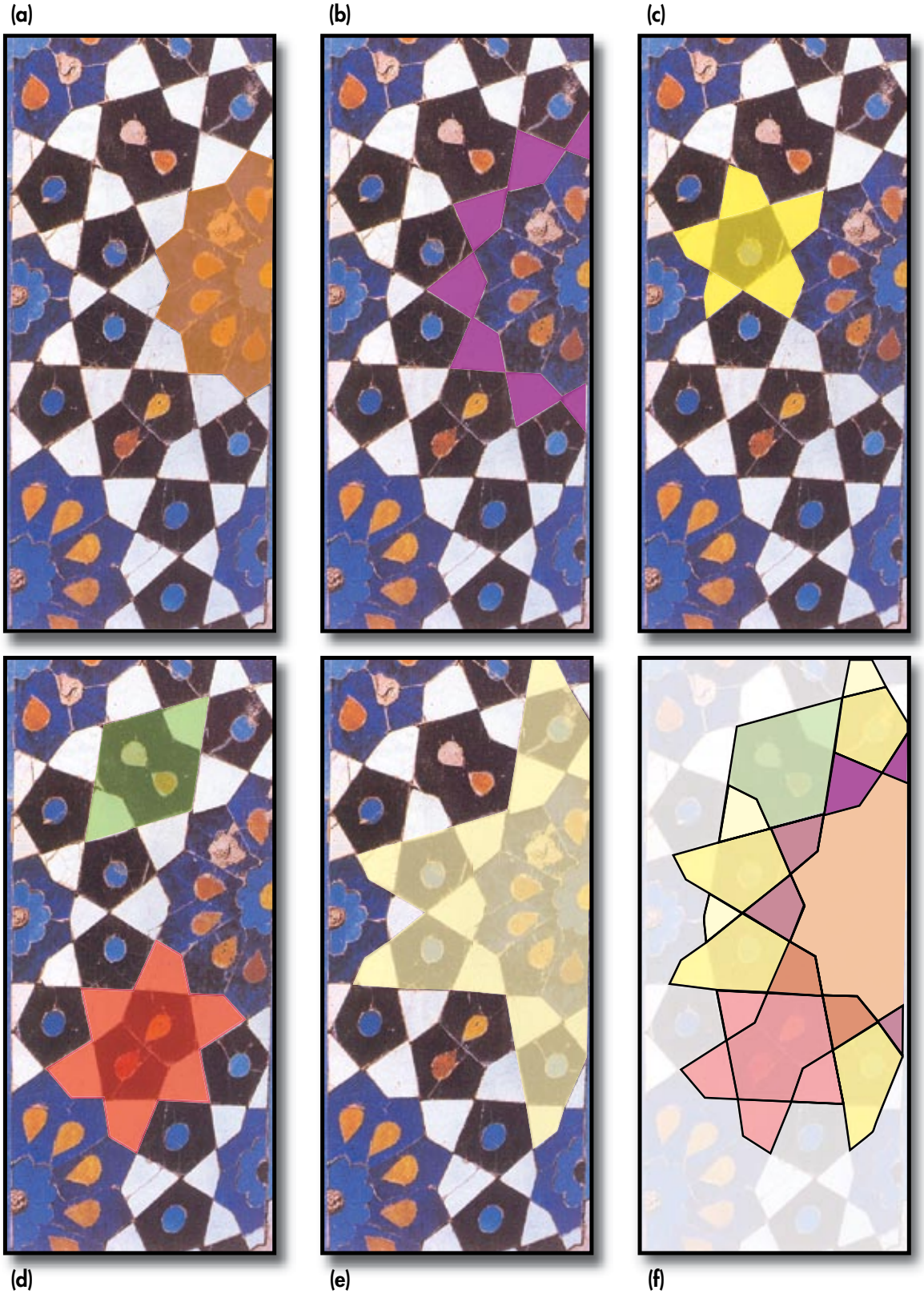


Figure 5-8. Some of the geometrical forms embedded in the design: (a) flower, (b) garland, (c) five-pointed star, (d) six-pointed star, (e) large star. The last plate, (f), shows how all the forms come together in an integrated, puzzle-like, manner.

Any analysis of such information-rich abstract geometrical forms must confront, at a basic level, an issue that all artists and art historians are concerned with: What makes us recognize a representation of a bird, fish, writing, or leaf, when what we see often bears little resemblance to the real thing (Grabar, 1992)? While there is no definite answer to this fundamental question, we see that a sophisticated artifact such as the one discussed here not only portrays objects that we can recognize (e.g., flower and stars), but also integrates them coherently. In this context, Salingaros (1995) argues that an abstract design with dense field of centers and interlocks can actually convey more of the essence of a living thing than a photograph of the same physical objects. If indeed this is the case, then it is perhaps possible to apply some of the integration techniques observed in medieval architecture and those discussed in Alexander's theory in order to house large amounts of information.

Like many similar medieval artifacts from this tradition, the tilework of Figure 5-7 contains many cultural and philosophical messages. One message, which is consistent with the prevailing philosophy of the mathematicians and architects whom designed it, is that celestial formations (e.g., stars) are no different in their inherent structure than earthly formations (flowers). From this perspective, a star and a flower are qualitatively similar; both ostensibly share the same structural form that we, as humans, so intuitively understand. This philosophical point of view is represented in the medieval saying of "as is above [in the heavens] so is below [on earth]." The tilework of Figure 5-7 may be an attempt to artistically portray this philosophy by showing that flowers and stars are linked, and that what appears initially as a flower can also appear as a star.

The essence of any sophisticated visual design relates to how its abstract geometrical forms correspond to physical reality (or the message it tries to convey), and how these geometrical forms are bound together in a coherent way. When one constructs inanimate objects that are full of interlocks and positive spaces, as well as many of the other properties discussed earlier in section 4, the resulting field of geometrical forms appears very rich and intense (see e.g., Figure 5-9). Beyond the aesthetic quality and emotional reverberations that such a tilework design produces, it can also embody large amounts of messages (or information, from our perspective). How this is done is still in the realm of art, not science.

Figure 5-9. Extending the original tilework to create a field.

From Tilework to Cockpit Displays

We now turn our attention on incorporating principles of tilework arrangement into the process of organizing engine parameters for the purpose of display. Based on several discussions with pilots, engine technicians, and engineers, it became clear to us that the first stratum of information should include the raw parameters. Only after raw information is embedded within the design should one begin to amplify with abstracted information (e.g., higher-order relations). The idea is that there will always be a way to see the link to individual raw parameters (and perhaps even back to signals in more elaborate designs).

We begin by placing the "raw" engine parameters in the middle of the main flower. Each one of the four engine parameters—Power Turbine Speed (Np), Gas Generator Speed (Ng), Torque, and Fuel Flow—is represented by a petal. Naturally, we place the parameters of the left engine on the left side of the flower and the parameters of the right engine on the right in order to achieve symmetry. Rotor speed, which is a product of both engines' outputs, is represented on top like a keystone supported by the two sets of parameters. Figure 5-10 shows the progression from the abstract flower formation of the tilework to an arrangement for placing engine parameters. Note that we added an artificial parameter, labeled Xx, to match the six-petal structure of the actual tile's flower. With proper modifications to the tile structure it is possible to arrange for a smaller (e.g., five) number of parameters, as will be shown later. Naturally, it is also possible to modify the structure to hold a larger (e.g., seven, eight, etc.) number of parameters.

The values for each parameter are arranged such that there is a lower bound and an upper bound. The anthers of the flower can be used to provide rate information (differentiation) such that when power is advanced, some anthers grow together outwards. Likewise, when power is reduced, the anthers shrink

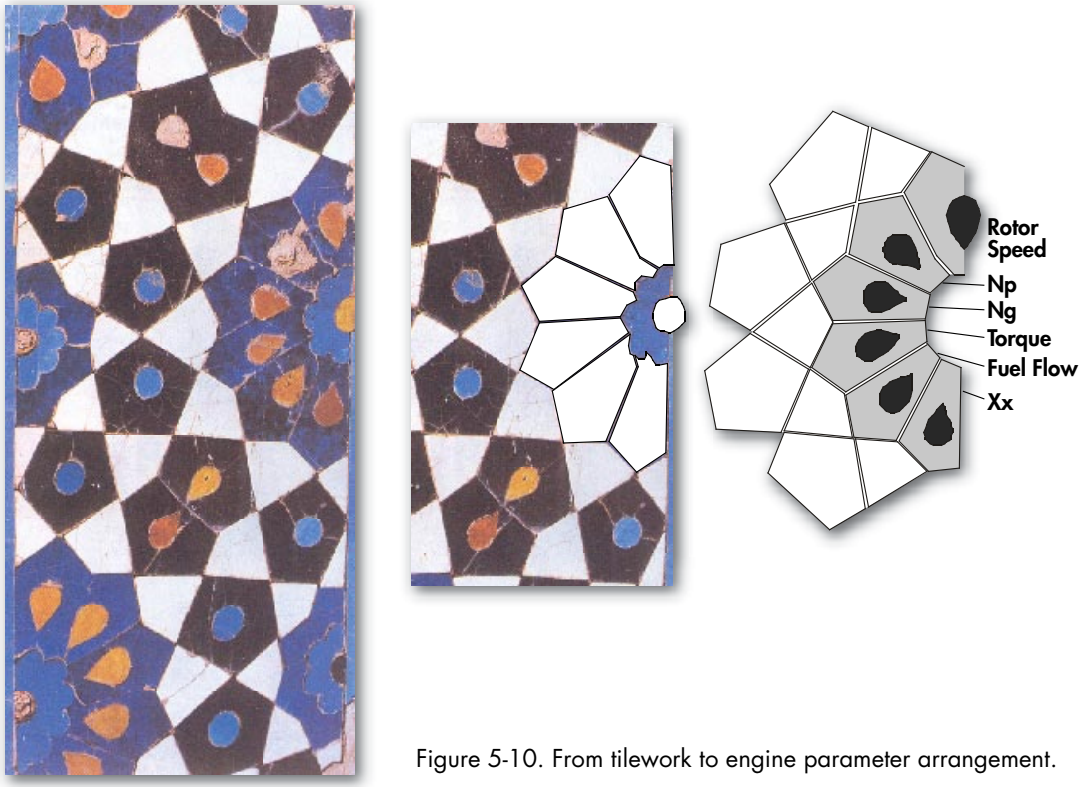


Figure 5-10. From tilework to engine parameter arrangement.

or retract inward toward the center of the flower. Our eyes perceive the growing anthers and petals as a unit, a whole. In the display approach explored in this report, we make the following assumptions as to the potential implications of this Gestalt phenomenon: (1) information which is initially represented as separate elements comes together as a unit in our perception because of their "proximity" and "common fate"—two established Gestalt principles; (2) the unit fosters better understanding of the interrelations that exist among the parameters; (3) the unit and its resulting pattern(s) provide for faster information processing; (4) any deviations from the unit's overall pattern can be quickly detected; and (5) the implications of a deviation, in the context of other parameters, can be grasped and understood intuitively.

Arrangement of Engine Information: Composites

Next comes the organization of composites computed by the Inductive Monitoring System (IMS) discussed earlier. We begin with the simplest case where we combine all six parameters—Rotor Speed, Power Turbine Speed (Np), Gas Generator Speed (Ng), Torque, Fuel Flow, and Xx—into a single composite. We have one composite for the left engine and one composite for the right engine (overlapping on Rotor Speed). The top drawing of Figure 5-11 shows one possible design solution for this

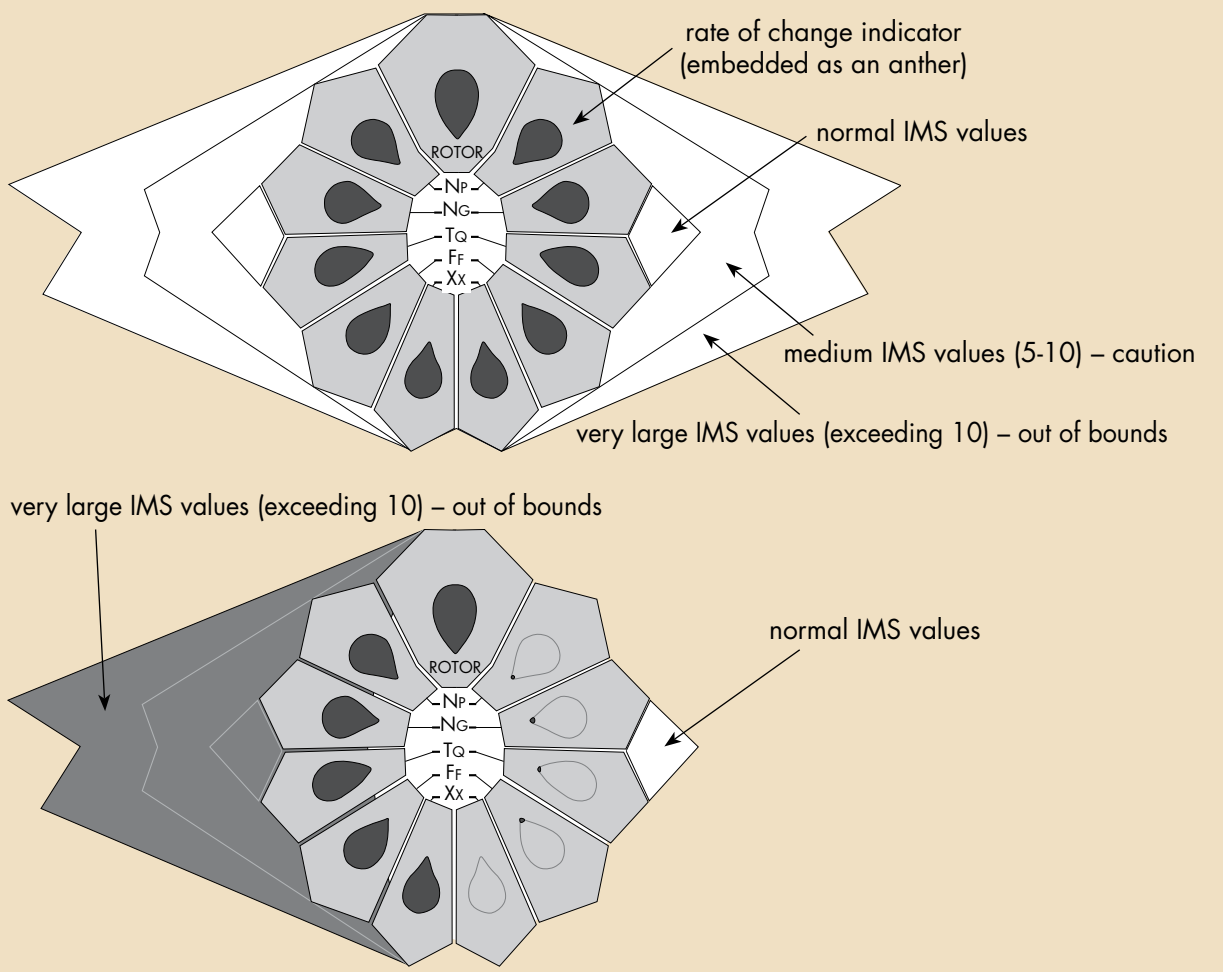
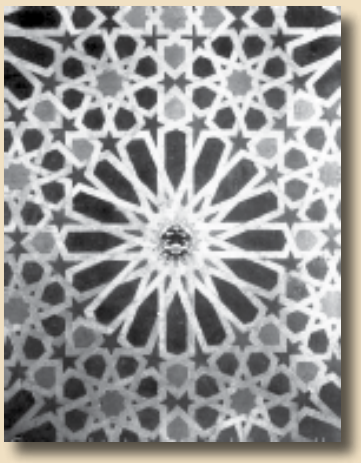


Figure 5-11. Top: Raw engine parameters presentation with a sextuple composite.
Bottom: The left engine parameters are deviating from expected and the right engine is working normally.

sextuple composite. Here lines are extended from the bottom and top of the floral arrangement to represent the composite. The quadrilateral polygon between Ng and Torque represents the case where the resulting IMS values are within the normal range. The second polygon is used to represent values in the cautionary region of the operational space (5-10 range). The third polygon represents a serious abnormality (exceeding 10). The bottom drawing shows a case where the right engine is working normally and the left engine, as a whole, is deviating from expected values.

As discussed earlier, it is possible to create composites of varying sizes. When composites are arranged together, such that interrelationships between composites are presented in a structured manner, the network of relationships between engine parameters can emerge, providing localized information about impending problems in the engine as well as the state of the engine(s) as a whole. The difficult question is how to visually arrange the composites in a coherent way. The arrangement in Figure 5-12 is one possible design that employs the "garland" around the main flower as space to embed information about pairs of parameters (e.g., the composite of "Rotor Speed & Np"). Similarly, the pentagons between the quadrilateral polygons serve as geometrical containers to information about triplets of parameters (e.g., the composite of "Rotor Speed & Np & Ng"). Finally, the arrow-like hexagon between the pentagons contains the composite values for quadruples (e.g., "Rotor Speed & Np & Ng & Torque"). This organization of the engine parameters and their composite has been developed and implemented as an experimental software system. Much of the initial development effort has focused on the development of an underlying coordinate system with a potential to accommodate a variety of geometrical forms. Ongoing work focuses on how to meaningfully represent the results from the Inductive Monitoring System and how to present this information.



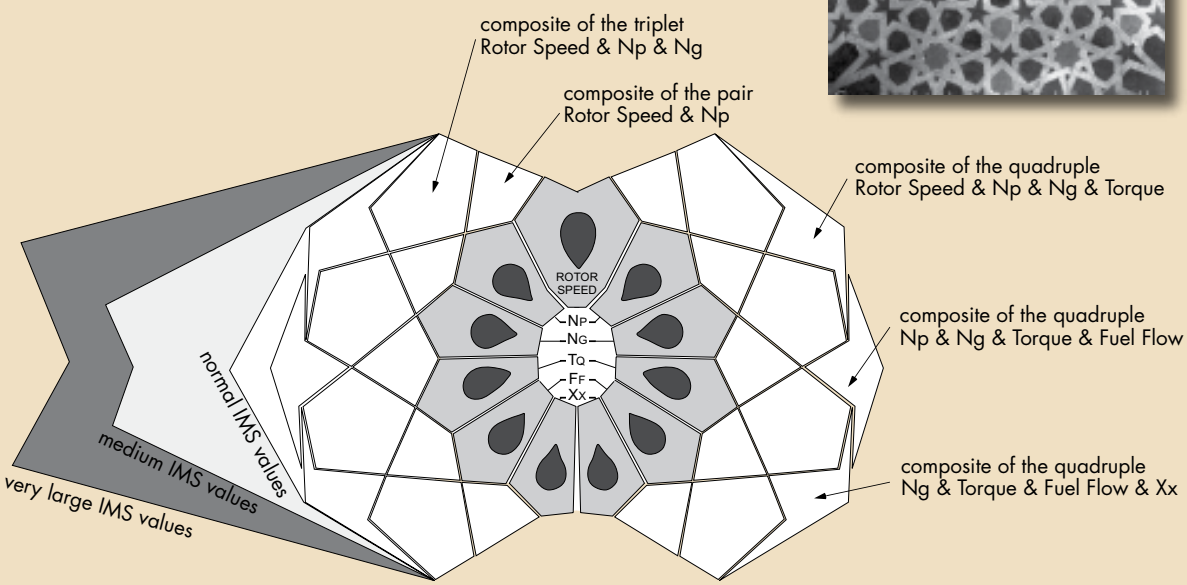


Figure 5-12. Pairs, triplets, quadruples, and composite of all six parameters. The inset is a tile arrangement from the Alhambra Palace in Granada, Spain (13th-14th Century).

Arrangement of Engine Information: Linking Related Subsystems

The most serious operational problems, and those that pilots have less success in solving, are in the interactions between subsystems. Individual parameters are rarely the key to understanding the problem. In the previous subsection we noted that the capacity to portray multitudes of intricate interrelationships, which is part of the philosophical and artistic motivation behind the design and construction of many surviving examples of medieval tilework, provide us with a coherent structure to show relationships among engine parameters. This structure can also be extended beyond the engines to show relationships among subsystems in the aircraft. As discussed earlier, subsystem information in modern cockpits (i.e., engine, hydraulic, pneumatic, electrical, etc.) is commonly arranged piecemeal, many times on a single multifunctional display page. While this approach allows us to represent many separate subsystems using a single display, the disadvantage is that we can only view one subsystem at a time. Important relationships between subsystems cannot be presented and the pilot has to integrate information from various displays in order to diagnose and understand the situation.

In most aircraft, the main engines provide power to subsystems by mechanical linkages to their respective pumps, motors, electrical generators, etc. So if, for example, the left engine drives the hydraulic system, it may be advantageous to present this link. Figure 5-13 illustrates an attempt to integrate subsystem parameters into the engine presentation by harnessing the geometrical links between two flower centers (discussed in the previous subsection). There are several ways to link

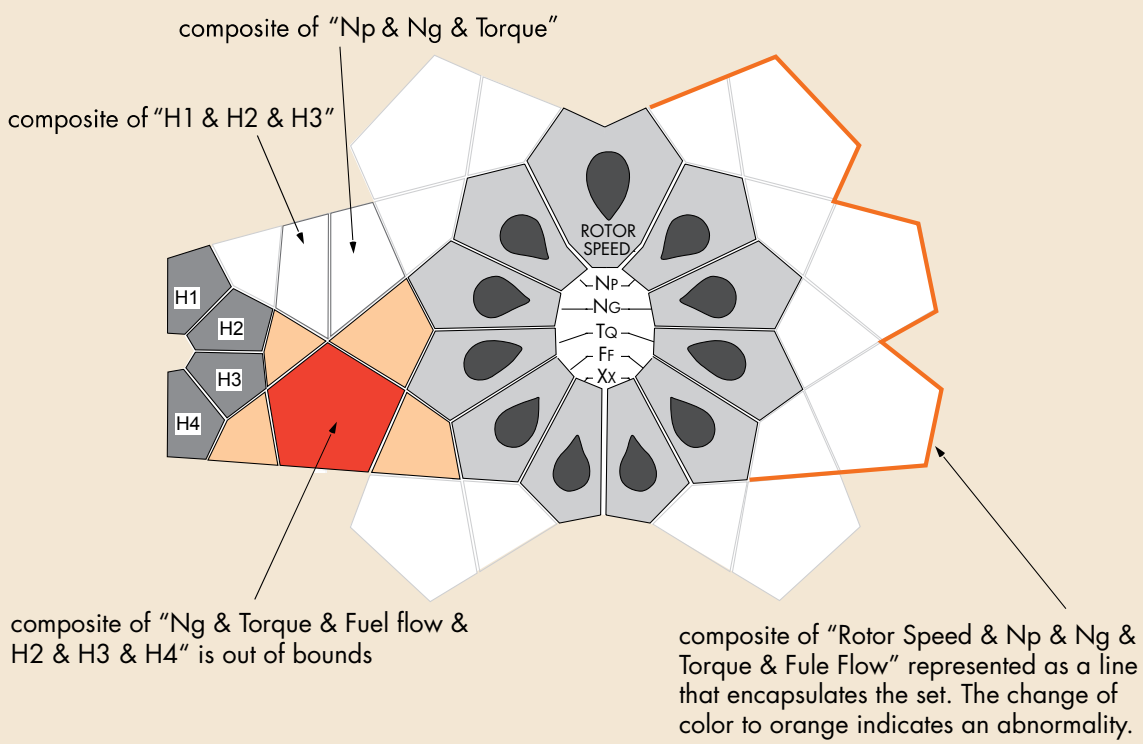


Figure 5-13. Integration of the left engine parameters and their composites with another subsystem.

the two systems. One approach is to split the pentagon that is shared by the engine (main flower) arrangement and the hydraulic system (smaller flower). Specifically, the triplet of "Np & Ng & Torque" occupies half the pentagon on the engine side, and "H1 & H2 & H3" on the hydraulic side. Another option is to use the pentagon to represent the composite of all six parameters ("Np & Ng & Torque & H1 & H2 & H3").

It is also possible to use color to present information about the composites. Figure 5-13 shows a situation where the composite of "Ng & Torque & Fuel Flow & H2 & H3 & H4," represented by the pentagon between the engine and the hydraulic system, indicates that it's IMS value are abnormal. Related composites, such as the pairs "Ng & Torque," "Torque & Fuel Flow," "H2 & H3," "H3 & H4," are somewhat affected and turn light orange. The emerging (partial star) formation highlights a localized problem in the left engine and hydraulic system.

In situations where there is no dedicated polygon to represent a composite, it is possible to use the borderlines between polygons. Information can be represented in the color and/or thickness of the line. For example, in Figure 5-13, we encapsulate the quintuple "Rotor Speed & Np & Ng & Torque & Fuel Flow" of the right engine with a continuous line. We use here the orange color to indicate an abnormality in this quintuple composite. The use of the borderlines can be extended beyond a single composite to show relationships between polygons, such that the polygon corners and lines become nodes and edges of a network.

This ability to employ lines that run along and between geometrical forms as conduits of information can also be used to link parameters that are not necessarily geographically collocated. Other solutions for linking parameters as well as subsystems include symmetry and geometrical presentations that use similar, perhaps repeating, patterns with the intention of helping the eye see the link.

Finally, the concept of linking up two related systems (e.g., engine and hydraulic), to achieve integration, can be extended to sets of subsystems. Figure 5-14 shows one possible structure of information where the engine indicators are integrated with three related subsystems. The structure is a replica of the tiling structure from Figure 5-9. Naturally,

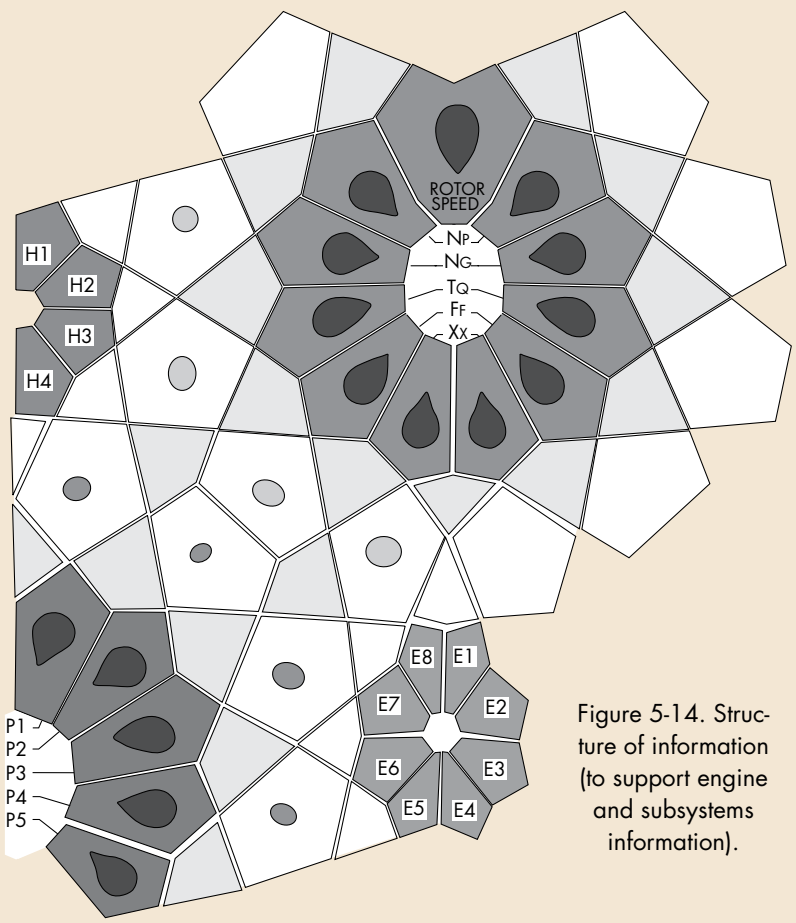


Figure 5-14. Structure of information (to support engine and subsystems information).

the structure can be expanded and Figure 5-15 shows how this presentation structure can be used to support eight different subsystems. It is important to note here that any organization of information to create such a holistic view can only be achieved if the field is held by an orderly structure. Looking at the arrangement of the tiles in Figure 5-9, the structure appears complex. But as we shall describe next, the complexity is born out of a simple, yet quite sophisticated, tiling structure that lies hidden underneath.

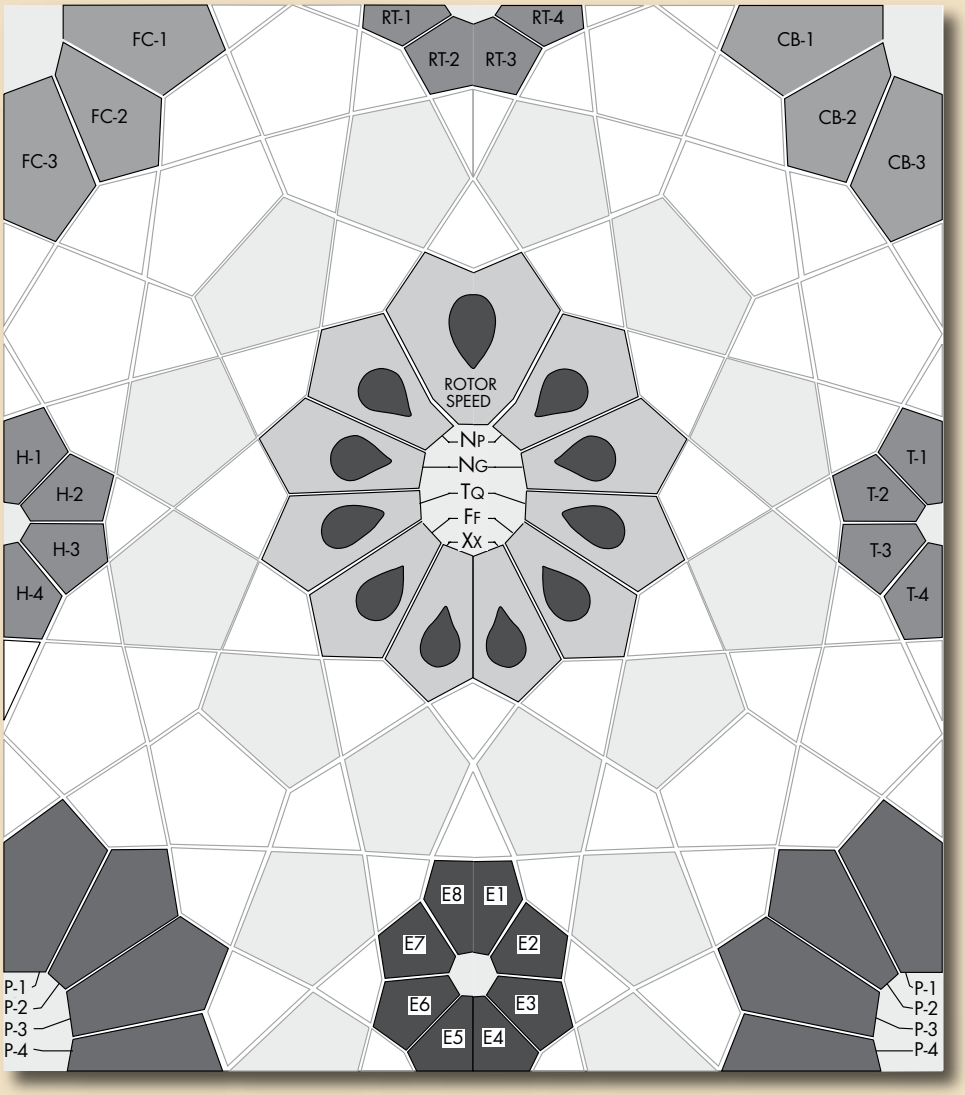


Figure 5-15. Organization of the engine with eight subsystem to create a whole.

Below the Lines

The tilework arrangement of Figure 5-16(a) is based on a ten-sided polygon (decagon) arrangement. The twelve-sided (dodecagon) tiling arrangement that we have used so far to organize the helicopter information is related to this decagon-based arrangement: note the main flowers, the five- and six-pointed stars between the flowers, the rhomboids, the pentagons, and the quadrilaterals (of the garland). But unlike the geometrical forms in Figure 5-16(b) that have a relaxed symmetry about them (e.g., note the highlighted star at the bottom of the Figure), the geometrical forms in Figure 5-16(a) are more strict (e.g., all the pentagons are equilateral).

Underlying this particular decagon-based arrangement are three decorated polygons (decagon, hexagon, bowtie) seen in Figure 5-17. These three basic polygons are part of a family of five specific equilateral polygons with special decorative motifs, called "girih tiles," which underlie many decagon-based arrangements found in architectural sites throughout the medieval Islamic world (Lu & Steinhardt, 2007). In their seminal paper, Lu & Steinhardt have also shown that some girih tile patterns can be mapped to arrangements of two unique quadrilateral shapes, called "kite" and "dart." These two geo-

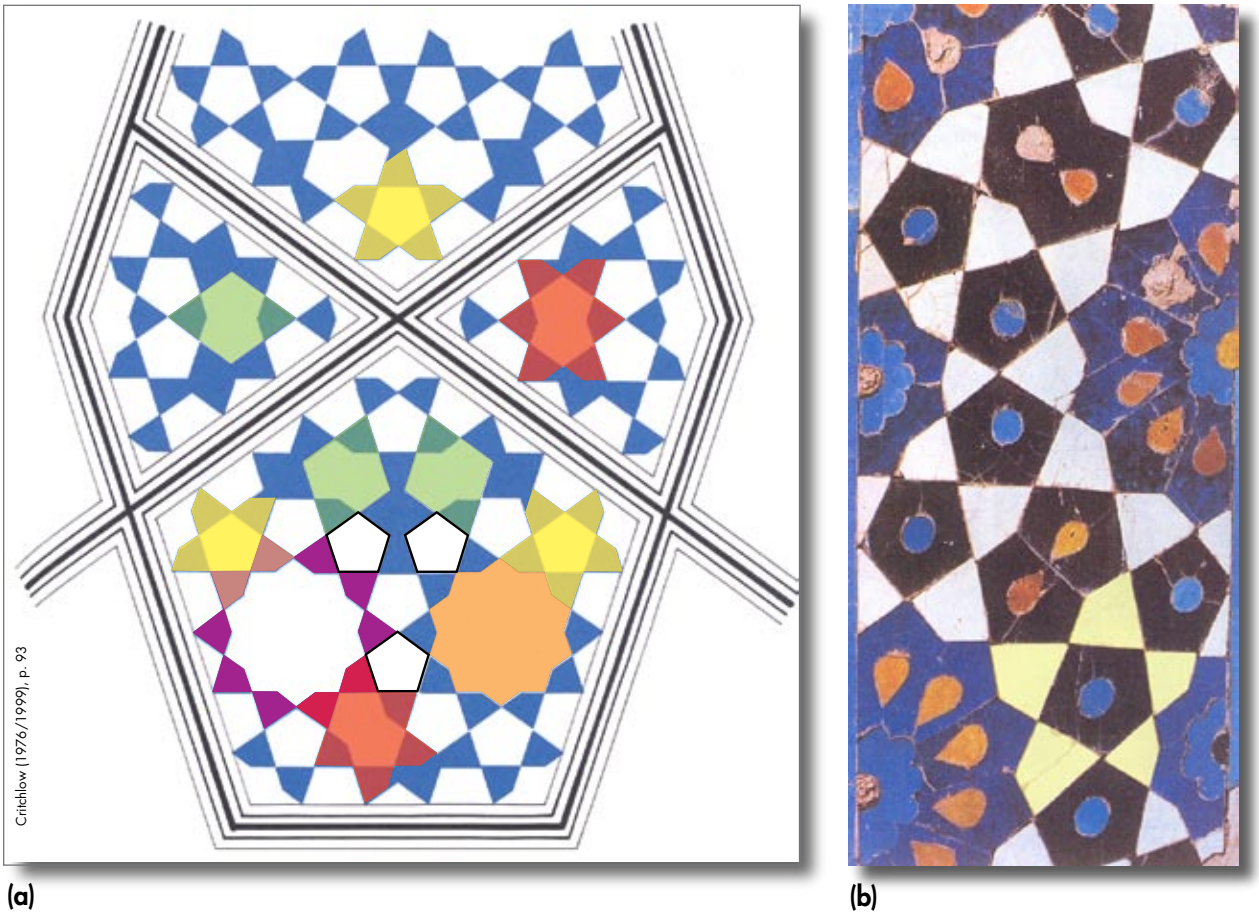


Figure 5-16. Two related tile arrangements: (a) is a decagon-based and (b) is a dodecagon-based. (Highlighted in (a) are the main flower, five-pointed stars, six-pointed stars, rhomboids, pentagons, and quadrilaterals seen previously in Figure 5-8).

metrical forms can be employed to create extremely rich tiling arrangements (Figure 5-18). In fact, some kite and dart arrangements, referred to as Penrose tilings, are quasiperiodic—the resulting pattern possesses special symmetry forbidden in periodic patterns, and never repeats. In other words, the pattern is infinite.

Lu and Steinhardt (2007) have shown that the tile decoration on the Darb-i Imam shrine in Isfahan Iran, dating to the fifteenth century (see Figure 5-6(g) and 5-6(f)), includes several complex quasiperiodic tile arrangements, just as in the case of the Penrose tiling. This finding is a testament to the mathematical sophistication of medieval mathematicians and architects, who must have understood, theoretically or otherwise, complex concepts of geometrical arrangements and notions of infinity, back

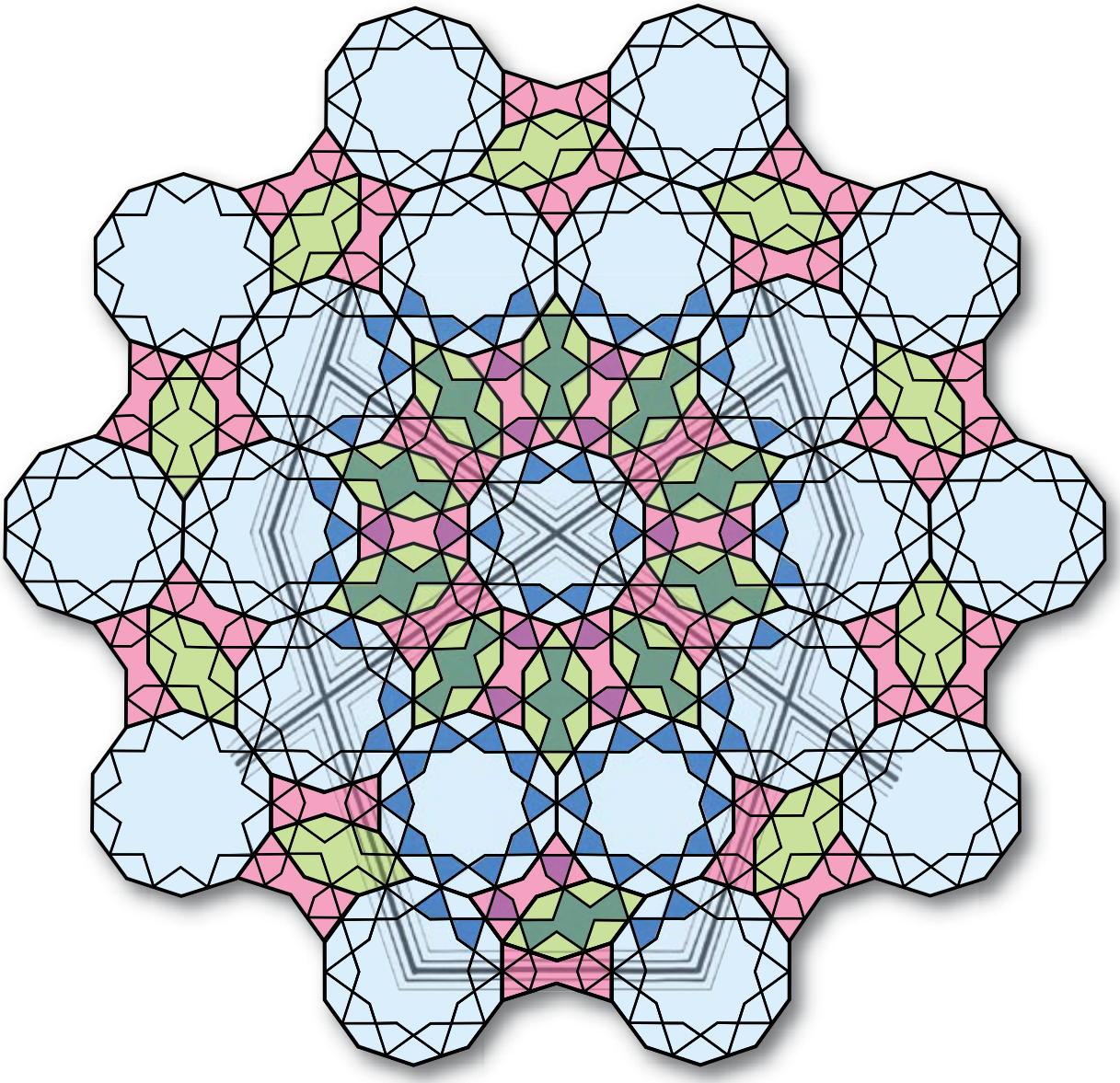


Figure 5-17. The three basic girih tile shapes (decagon, hexagon, and bowtie) that can be used to generate the decagonal tile arrangement of Figure 5-16 (a).

in the fifteenth or sixteenth century. Unfortunately, no textual information about the underlying theory and practical design methods that contributed to these mathematical and architectural achievements is known to exist today. In the context of the proposed approach to information visualization discussed in this report, Lu & Steinhardt's findings hint at the possibility of using tiling arrangements to create intricate information displays that can house large amounts of information.

To conclude, the tile arrangement we have chosen for the helicopter display embodies a unique mathematical sophistication that appears to be fundamental for any comprehensive attempt to organize space in order to achieve wholeness. The arrangement of Figure 5-15 contains an underlying structure that can unfold in an orderly manner. We believe that it is technically feasible and operationally suitable to employ such unfolding structures to organize extremely large amounts of information. It may be that the existence of such sophisticated underlying structures is indeed what generates the sense of order and wholeness that emanates from many medieval tile arrangements. We believe that in-depth understanding of this order, which is both mathematically rigorous and aesthetically pleasing and harmonious, is an important prerequisite for developing modern techniques for organizing information.

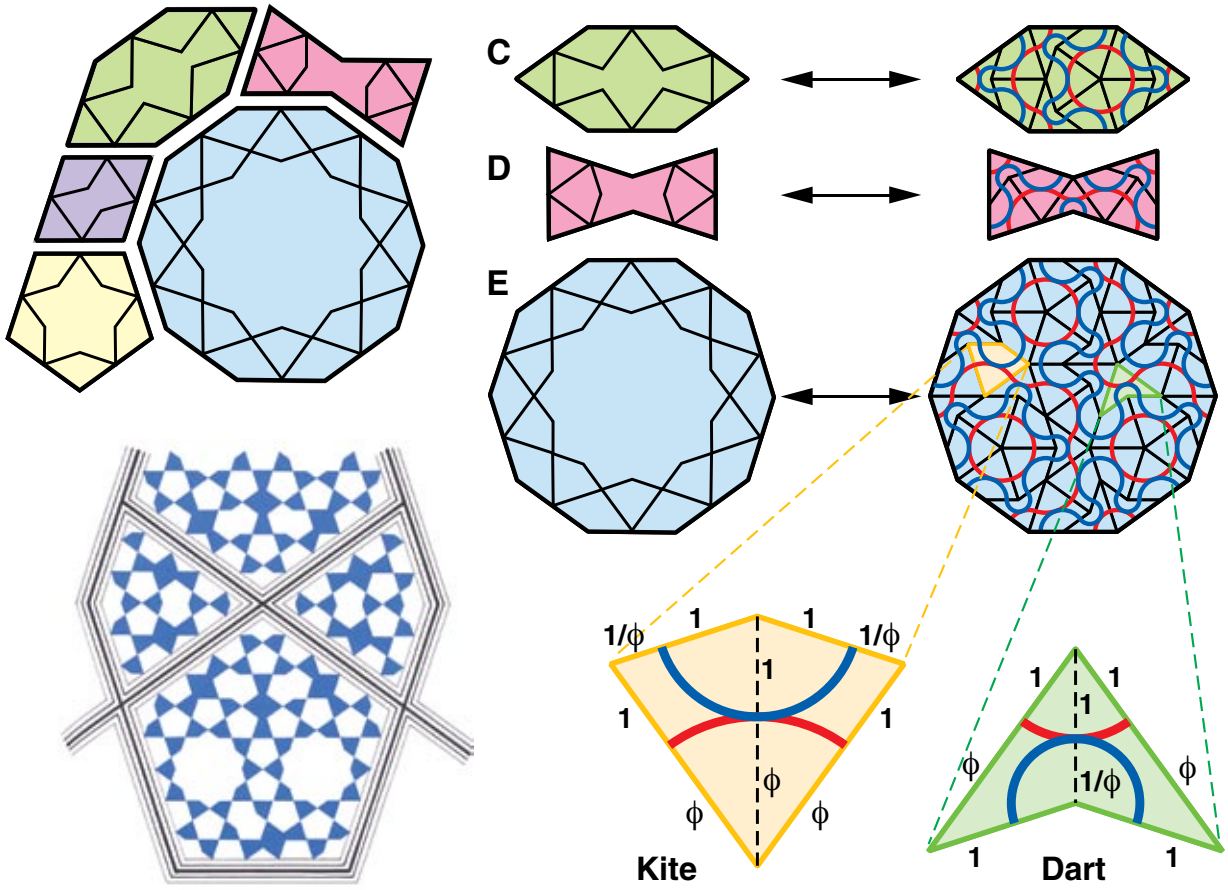


Figure 5-18. The five basic girih tile shapes (decagon, pentagon, rhombus, hexagon, and bowtie) described in Lu & Steinhardt (2007), and Penrose's darts and kites.

Summary

Much of the beauty of the tilework of Figure 5-7 comes from its many perceptible regions of coherent spaces (or centers), overlapping geometrical forms, interlocks, and the existence of a profound underlying structure. Every geometrical form in the tilework is linked to others to create an intense field of patterns. In such arrangements there is never any leftover space—every region of the field, from the smallest polygon to the largest, is "working" to make the field intense. We believe that these characteristics of medieval architecture can be used represent abstraction (e.g., composites of pairs, triplets, etc.), present integration (by knitting together the composites into a structure), and organize these structures of information in an orderly manner to provide a holistic view.

To apply some of the abovementioned principles for information organization, we need to view geometrical forms as containers of information. In this context, tiling arrangement that can unfold and expand, without necessarily conforming to a predefined grid, may provide us with a structure to view and better understand complex sets of information. To this end, the mathematical mapping between signals and abstract geometrical forms needs to be worked out thoroughly, as well as how these abstract geometrical forms relate to one another (e.g., in clusters, patterns, or a given tiling arrangement). While such designs are quite difficult to create from scratch and the theories, methodologies, and techniques used to build them are lacking, hundreds of medieval Islamic geometrical patterns and structures have survived and are documented. Likewise, books and lectures about the philosophical school of thought that contributed to the ideas behind these artifacts have survived. Furthermore, the existence of centuries-old traditions, in almost every ancient culture, that created integrated and holistic artifacts (Jung, 1955/1972) provides us with a broad foundation and many successful examples to learn from.

6. Conclusions

…The devil and a friend of his were walking down the street. They saw ahead of them a man stoop down and pick up something from the ground, look at it, and put it away in his pocket. The friend said to the devil, "what did that man pick up?" "He picked up a piece of Truth," said the devil. "That is a very bad business for you, then," said his friend. "Oh, not at all," the devil replied, "I'm going to let him organize it."

—Jedu Krishnamurti (1929/1996, p. 1)

The goal of the research outlined and discussed in this preliminary report is to further understand how to organize information. The main objective is to develop a principled approach for abstraction, integration, and organization of information in order to allow users to understand a system as a whole. We believe that the key for organizing information, in the context of human interaction with technological systems, concerns three main issues:

(1) the correspondence between the system under consideration and its representation,

(2) the coherence among representational elements and clusters of information,

(3) the intrinsic order (e.g., of the underlying structure) that supports the overall presentation.

Throughout this report we have tried to show that there are serious problems in the way information is currently organized, thereby limiting users' ability to monitor, understand, and supervise technological systems. Modern information systems provide us with multitudes of data, and the amount of data will only grow in its size and complexity in both the near and far future. Yet despite much advancement in data analysis and processing, we are very limited in our understanding of how to best organize this data and information. In this report we have also tried to argue that without a theory, it will be very difficult to develop good solutions for these problems. Today we have no standards for measuring the quality of information organization, let alone methods for achieving quality.

In considering such a theoretical approach, we believe that it may be prudent to draw from existing work in perception, graphical information presentation, human factors, computer graphics, and mathematics, as well as art and architecture. Brunswik (1956) and Julesz (1971) describe the strengths and limitations of human visual perception in the context of the environment, and the results of their work

have been successfully extended to technological environments. Fishwick (2008) and Tufte (1983; 1997) illustrate the crucial role of creative insights and concepts for increasingly dense information displays that lie on the border between art and engineering. Bertin (1967/1983; 1977/1981) provides a theoretical basis for visualizing and plotting of numerical and statistical data. Norman (1986, 2007) and Hollnagel & Woods (1999) made considerable headway applying psychological concepts of information and turning them into engineering design guidelines. Telea (2007) and Diehl (2007) portray recent efforts to systematize research from the past three decades in computer graphics and to develop implementation techniques for information presentation. In the context of mathematics, Drewes (2006) and Ehrig et al. (2006) suggest promising new foundations in category theory and collage grammars, which seem to have the potential to formalize the structure of information displays and to express rules for their composition. In Appendix B of this report we discuss an approach for processing analog signals with an eye towards its graphical representation and integration among streams of data. Finally, in the last two sections of this report we have tried to show that some of the fundamentals for a theory of information organization have been addressed in art and architecture, and that it is possible to borrow and extend many of the concepts of integration developed in the Middle Ages to address modern problems of information organization.

One motivation behind the research described in this report is to begin considering a theoretical approach for organizing information in the context of the kind of monitoring, decision-making, and supervisory control tasks that characterize aerospace and similar technological domains. In light of the extensive use of automation in current and future technological domains, we strongly believe that a theoretical approach for information organization must go hand in hand with a theoretical approach for human-automation interaction. The design of any automation aid cannot be successful without a thorough (i.e., theoretically based) consideration of how information about the working of the machine is communicated to the user (e.g., mode changes, parameter settings, expected outcomes, approaching boundaries of operation and subsequent transfer of control). For both approaches to work together, a common framework is needed. Clearly, this business of developing theories and methodologies for organizing truth, or in our case, information, is a diabolic challenge. But it's also a beautiful one.

Acknowledgements

The work described in this paper was conducted as part of NASA's Integrated and Intelligent Flight Deck project, Aeronautics Research Mission Directorate. The authors would like to thank Immanuel Barshi for continual sustenance, extensive reviews, and writings that were incorporated into this report. Michael Heymann, Sandy Johan, Loukia Loukopoulos, Peter J. Lu, and Robert Mauro provided valuable reviews, comments, and insights throughout the research and writing process. Michael Feary provided oversight and support to the project from its inception. Sandy Johan developed the Java-based environment to visualize integrated structures of information as well as the routines to generate IMS results in order to create a computer-based simulation of the experimental helicopter display described in section 5. Tom Esposito lent his artistic talents to this project by creating many of the figures and producing the report. Specials thanks to Gail Bates for her invaluable help in bringing the age-old artistic and mathematical foundations of this research into light.

Observations and Inferences

We list herein some of the inferences that can be extract from this report. While it may be difficult to substantiate these inferences, we believe that they can provide some insights and possible leads for those interested in pursuing this topic further:

- There is a logical extension from art and architecture to interface design and information representation and organization, as all deal with the arrangement of a field. Therefore, some of the concepts, methods, and tools that are applicable to art and architecture can be extended to interface design and information presentation.
- Geometrical forms and patterns—as seen in art, architecture, and ornament design—are designed to be viewed and interpreted by humans as messages. Some of these messages are more pronounced (e.g., maps) and some more subtle (cultural motifs).
- Geometrical properties of a field, such as interlocks, positive space, levels of scale, the void, boundaries, gradients, alternating patterns, etc. (Alexander, 2002a, Ch. 5), provides a language for describing the elements of an integrated field and may hold the key for developing a systematic approach for constructing integrated displays.
- The amount of information in geometrical forms can grow when we recognize that whenever two or more basic elements interlock, additional elements are created. The same applies to layers of information that are embedded in a field.
- For a field to accommodate interrelationships and become coherent, it must rest on a underlying structure. When the underlying structure and apparent geometrical forms are well integrated and organized in an orderly manner, a sense of wholeness permeates.
- Structures that have wholeness in them (e.g., a flower motif, stars and constellations, tree-like organizations) can house considerable amounts of information, because we are attuned to the wholeness that such structures provide. In turn, deviations and/or collapse of this wholeness can be quickly identified.
- A highly integrated design allows the viewer to consider multiplicity of interrelationships, some of which may be beyond what the designer anticipated yet still true to the system being represented. This feature of an integrated design and human capacity to detect patterns (and deviation from thereof) is very beneficial for identifying unpredictable situations.

Appendix A. Canonical Correlation Analysis: A Literature Review

Canonical correlation analysis (CCA) subsumes many other linear least-squares statistical methods (Ang, 1998), including multiple regression and discriminant analysis. It has been extended in various ways to repeated-measures data (Cumming & Wooff, 2007) and to nonlinear analysis (Hardoon, Szedmak, & Shawe-Taylor, 2003; Huang, Lee & Hsiao, 2006). Canonical correlation analysis also has a strong conceptual connection to artificial neural networks, because it closely mirrors the many-to-many mapping of input nodes to output nodes.

Makarenkov and Legendre (2002) describe polynomial Redundancy Analysis and polynomial Canonical Correspondence Analysis. These techniques differ from standard canonical correlation analysis in two respects: Categorical data are used as dependent measures, and polynomial regression functions are used to represent nonlinear patterns. Kernel CCA (KCCA) has also been proposed as a general method to extend canonical correlation analysis to high-dimensional, nonlinear representations (Hardoon et al., 2003).

Vía, Santamaría, and Pérez (2005, 2007) discuss how canonical correlation analysis can be extended to three or more sets of variables—for example, X, Y, and Z with estimated weights u, v, and w—to maximize an appropriate correlation metric among X_u, Y_v, and Z_w. Kim, Wong, and Cipolla (2007) have generalized canonical correlation analysis between two sets of vectors to Tensor CCA (TCCA). Presumably, it should be feasible to combine these separate lines of work and extend TCCA to more than two sets of tensors (Savarese, DelPozo, Niebles, & Fei-Fei, 2008).

The state of the art in graphical representation of multivariate correlational data is presented in Kerren, Ebert and Meyer (2006) and in Yeh (2007a, b). Kerren et al. provide a diverse set of chapters on human-centered interactive graphic display, including theoretical issues and advanced examples from geography, computer science, and medicine. Yeh describes the SAS Constellation Diagram and its applications to visualization of multivariate correlation in pharmaceutical and other kinds of clinical data. The Constellation Diagram generalizes concepts from several sources, such as Friendly (2002), in order to provide a flexible, principled framework for adapting complex displays to specific analytical purposes. Examples include fMRI and liver function data, as well as neural network visualization formatted in a variety of ways, such as node-link diagrams, tree diagrams, and color-coded matrix formats.

Further steps toward the theoretical integration of the kinds of complex displays described by Yeh (2007a, b) have been taken by Seo and Shneiderman (2006), using the rank-by-feature framework. This framework begins with the recognition that displays of three or more dimensions are difficult to interpret, whereas two-dimensional correlational displays, such as scatter plots (Friendly, 2002; Yeh 2007a, b) are either limited to two variables at a time, or require supplemental annotation, such as color coding, to represent implicit relations among more than two variables.

The rank-by-feature framework (Seo and Shneiderman, 2006) uses a principled interactive control structure to enable the user to organize multiple one- or two-dimensional displays in order to capture complex multivariate relations. The key elements of the interactive control scheme allow the user to perceptually identify interesting features, and enable structured exploration of those features by drilling down to lower-level details without getting lost in an unsystematic browsing process. Principal elements in the underlying theory are concerned with the design of this control scheme according to "Graphics, Ranking, and Interaction for Discovery (GRID) principles—a set of principles for exploratory analysis of multidimensional data, which are summarized as: (1) study 1D, study 2D, then find features (2) ranking guides insight, statistics confirm" (Seo, 2005, p. 67).

In summary, Canonical Correlation Analysis and its various extensions constitute an important class of examples of information abstraction and integration. CCA and related methods can potentially handle any number of sets of variables, and each set can, in turn, contain any number of variables. These methods have proven useful in simplifying the tasks of identifying and interpreting complex data sets. The number of variables requiring attention may be drastically reduced, and composite variables can be linked back to the original observed data if desired. New possibilities are provided for identifying trends and norms, as well as for quickly highlighting important deviations or outliers. Recent technical advances have enabled imaginative, dynamic displays of complex data sets such as multivariate time series. At the same time, we are just beginning to understand the theoretical principles underlying the design of graphical structures to organize and display complex multivariate relations, directing the user's attention rapidly and accurately to the most important information while hiding irrelevant details and avoiding confusion, clutter, and distraction.

Appendix B: Signal Transformation To Support Abstraction And Integration

The focus of this report has been on abstraction and integration methods for representing information in organized structures of information to form a whole. However, we did not address the issue of how signals are transformed into data and represented in a way that is usable for abstraction. In designing integrated displays, where multiple streams of data must enter the displays, definition of physical and temporal relationships between signals is critical for abstraction and integration to take place.

To turn signals into data so it can be then used to create information, one needs to identify a meaningful change. Traditionally, for analog data, the task is performed numerically via statistical pattern recognition and signal processing. But as the dimension of information increases, the complexity of the representation task can become prohibitive. This appendix describes one technique for creating a graphic alphabet for analog signals that facilitates such representation. The cornerstone of the approach is based on the realization that in the context of data-rich information environments, the methods for processing analog signals and the methods for representing their information content must go hand in hand. Thus, we first describe an approach to characterizing a single signal's content and then discuss how to represent that characterization using graphic alphabets. Finally we propose an approach to combine the content of multiple analog signals.

Signal Analysis and Alphabets

Traditional signal analysis methods are based on Shannon's (1962/1948) model of data recording which uses discrete sampling of a signal at a given frequency. Numerical descriptions of analog waveforms are obtained by taking $2^*\Omega^*\tau$ samples (Shannon Numbers), where Ω is the signal bandwidth and τ is the signal length over time period. This time-based signal sampling approach is founded on the theorem that if a function $f(\tau)$ contains no higher frequencies than Ω cycles per second, it is completely determined by its ordinates as a series of points spaced $\tau\Omega$ seconds apart.

To use this model in practice, analog signals must be sampled in a series of equally spaced points no less than two times the fundamental frequency (the Nyquist frequency). A variety of transforms such as the Fast Fourier Transform (FFT), Wavelet, or other coding schemes are then applied to these sampled signals to form feature sets used for analysis or display. These sets are then presented to machine classifiers to recognize and highlight phenomena of interest in the data. As noted by Lupu et al. (2003), traditional signal sampling strategies have three requirements. First, they require measurement and recording of the signal amplitudes; second, they require regular samples from a signal regardless of the signal's temporal behavior; and third, they must determine a correct sampling rate for the information content, otherwise they can distort or miss critical events entirely.

Current information displays found in cockpits and mission control utilize threshold or smoothed versions of signal parameters. It is incumbent upon the user to recognize both regularities and abnormalities. The underlying assumptions made for signal sampling and their transformation into features can both limit and enhance users' ability to recognize patterns. Determining on a case-by-case basis what information must be abstracted, and then presented to the user, is yet another important design challenge that directly affects the ultimate system performance. What is less obvious, given the overwhelming use of interval-based signal sampling, is that the features used can be badly distorted by the sampling logic. Hence current information displays may miss important temporal information, fail to capture complex interrelationships between data streams, or add extraneous artifacts such as aliasing. Often, either the smoothed signal sample or a subset of indirect features, displayed as frequency plots, are used to present sensor data and information. This is accomplished in cockpits with analog gauges or the use of indicator thresholds such as warning lights to mark designer-chosen important changes. What users do not receive is a continuous information pattern (close to the original signal) from which they can directly infer system behavior. Another problematic aspect of current information displays is that each sensor is presented individually and not in relation to others and the whole.

The approach advocated in this report is to redefine a signal's features as a finite alphabet from which information is inferred, compressed, and arranged to create a geometrical form (and eventually a pattern). However, unlike standard sampling approaches, the method discussed here can be inverted to reproduce a close approximation of the original raw signal. In the case of verbal information, the signal alphabet used in English is acoustically based. The alphabet symbol strings form our usual text-based communication model and the sound which represents the word is approximated by the letters used. Creating such an alphabet for analog signals, given tremendous source variability, is another issue entirely. However, if it were possible to create signal-specific alphabets for an analog data stream quickly, then the display symbols used to represent that data could be used in many ways. For example, symbol frequencies could be tabulated in a display histogram where changes in the underlying information change the histogram over time, or, as discussed in section 5, the symbols themselves might become graphic components which could be assembled into an integrated pattern of many sensors.

Here we employ an alternative sampling strategy derived from observations made by Licklider & Pollack (1948). During studies of bandwidth compression in telephone communications, they demonstrated that even if all amplitude information from an acoustic speech signal was removed from a waveform (i.e., turning the waveform into a binary signal that preserved only the zero crossing points of the original signal), random word intelligibility scores of 97.9 % were achieved. The result motivated considerable interest in finding the mathematical limits for characterizing signals solely through the use of zero-crossing information.

In the literature of complex signal functions, this topic is referred to as the study of Entire Functions (EFs). It was recognized early on that large classes of practical signals can be thought of as the output of some undefined polynomial of infinite degree which are analytic over the complex plane. Specifically, entire functions are a generalization of polynomials of infinite degree which arise naturally in many areas of science and engineering (Requicha, 1980). If one treats an entire function polynomial (given assumptions about limited bandwidth) as a function made up of strings of complex numbers, then through approximation theory one can regard the real and imaginary zeros crossings of such a function as its information-bearing attributes (Bond & Cahn, 1958). This reasoning produces a different set of native signal description features than that used in traditional signal sampling. In the context of this report, it means defining analog data streams from multiple aircraft sensors in terms of this new feature set.

Approach

The transformation, or coding scheme, presented here was inspired by one entire function (EF) approach proposed by Holbeche, Hughes & King (1986). In their initial paper and in later work by other authors, complex analog signals were represented using sets of derived combinations of real and imaginary zero-crossing points. They referred to these combination features as Time Encoded Signal Processing and Recognition, or TESPAR, codes (King & Phipps, 1999). In later refinements, a signal was first processed to find its real zero-crossing points. The signal behavior between pairs of real zero crossings was then re-analyzed to approximate the signal polynomial's imaginary roots between those crossings. A direct computation of these roots becomes intractable for the real-time or large-scale signals that are found in aircraft and spacecraft signals.

Although a direct solution of the entire function polynomial is computationally intractable, the main finding of the above mentioned papers was that some of the imaginary zeros can be approximated rather than solved directly. This is done by tabulating the number of negative minima and positive maxima sign changes occurring within each real zero-crossing interval. Such approximation captures most of the imaginary root content needed to reproduce the original signal polynomial. Thus, the recoded signal maintains the majority of the information content from the original signal, but does so in a very compact way. By creating a new alphabet corresponding to the recurrence of these real and imaginary root sequences, we can then use this alphabet to recode a new signal (see Figure B-1), much as our phonetic alphabet is used to represent words (Jorgensen & Sharma, 2009; Sharma & Jorgensen, 2009).

To explore this coding scheme for analog signals, we generated Matlab routines to build custom TESPAR alphabets. We used the following method to quickly generate sensor-specific alphabets: in order to create a custom signal alphabet, one must extract many signal-crossing event combinations and

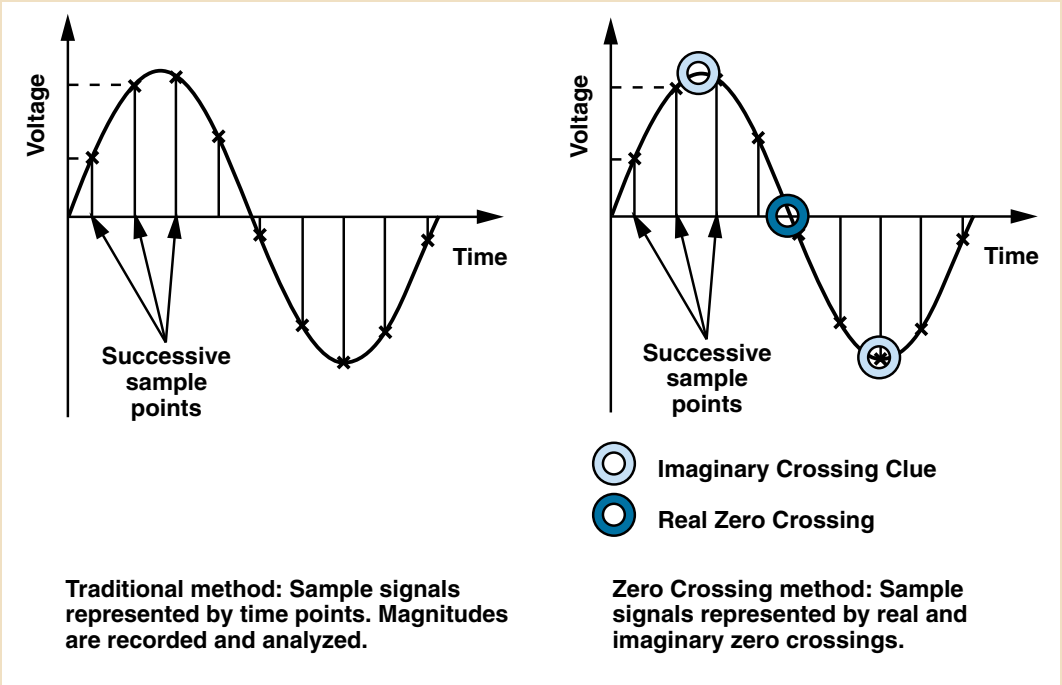


Figure B-1. Traditional and Zero Crossings methods.

form a density distribution of their occurrences. This involves recording real and imaginary zero signal positions and calculating the number of sign changes between real zero crossing points, as is illustrated in Figure B-2. In this Figure, the X axis corresponds to the number of elapsed clock cycles between real zero crossings and the Y axis corresponds to the number of sign changes which occur in a sampled signal between real zero crossings. Each point on the graph represents a signal segment. The density typically signifies a larger number of events that have a small number of sign changes between real zero crossings. At the same time, there are many differing numbers of clock cycles on the real axis, depending upon the analog data's inherent frequencies and number of underlying signal components.

The obtained density for a signal (or set of signals) is summarized by creating a smaller fixed-size alphabet (relative to the total number of possible combinations). These are in effect sub-areas of the

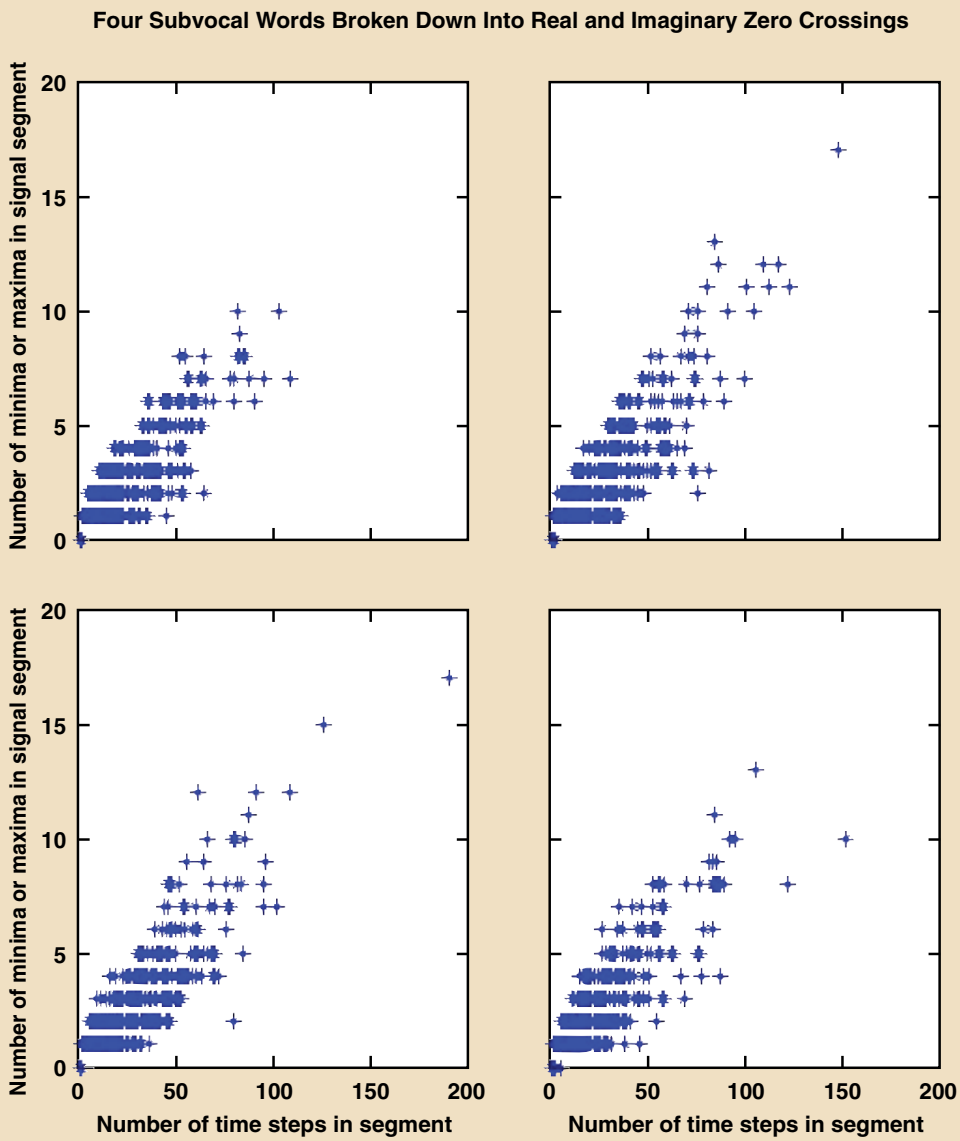


Figure B-2. Example of zero-crossing densities of four different signals.

distribution. These areas were initially coded simply by numbers; later on we represented them as geometrical shapes. Once obtained, this alphabet can either be specific for a given sensor or generic for several similar sensors. Initial tests were performed using human biometric signals recorded by silver chloride surface electrodes. We obtained over 70 percent recognition of new signal instances through the use of a previously generated TESPAR alphabet when combined with a scaled conjugate gradient neural network classifier based on the signals alphabetic codes. The ability to do such pattern recognition mathematically confirmed that we had captured important characteristics of the original signal in an extremely compact form.

Signal-specific Coding

We sought improvement of our first results by designing a generic coding scheme for the alphabet generation process. The available literature on TESPAR codes mentioned above does not state the process by which an alphabet is derived. Consequently, we explored a number of approaches for selection and automatic generation of alphabets. The first approach used a Matlab parsing code to build tabulated lists of real and imaginary zero combinations for different sample signals. For example, a typical electromyography (EMG) waveform produced over 1600 signal event combinations. Removing duplicated combinations from the list dropped the number of potential alphabet codes to 372, a reduction by a factor of four. The resulting density function was plotted again and re-analyzed for other ways to reduce the set.

In the first approach, we used a tree classifier to determine how the TESPAR density distribution could be best grouped into varying levels of detail, corresponding to alphabet elements. However, such classification treated each area of the density function as a polygon of fixed size and did not take into account non-homogeneous spatial coverage within the frequency distribution. It did, however, permit us to select an arbitrary level of alphabet detail and hence generate an alphabet of predetermined size. In the second approach, a linear regression analysis was used. Because the number of imaginary zero crossings was much lower than the real number of sampling intervals, each scalar level of imaginary crossing magnitude was used as a feature, and the bounds for point inclusion for that feature were the confidence interval range determined by the regression fit bounds (Figure B-3). Nevertheless, although the two approaches created classifier features, both lacked sensitivity to the original signal density after duplications were removed.

Our final approach was to use a K-means clustering algorithm to find K reference points in the original density (Figure B-4). When a new signal was presented, an alphabetic label associated with each TESPAR signal event (real and imaginary root combination) was matched to a coordinate of the nearest K–mean point. This method had the advantages of fully covering the space, taking into consideration the density (i.e., more fine-grained where there were more points), fitting the nonlinear shape of the distribution, and permitting pre-selection of the alphabet size (a major advantage for later use with any pattern classifiers).

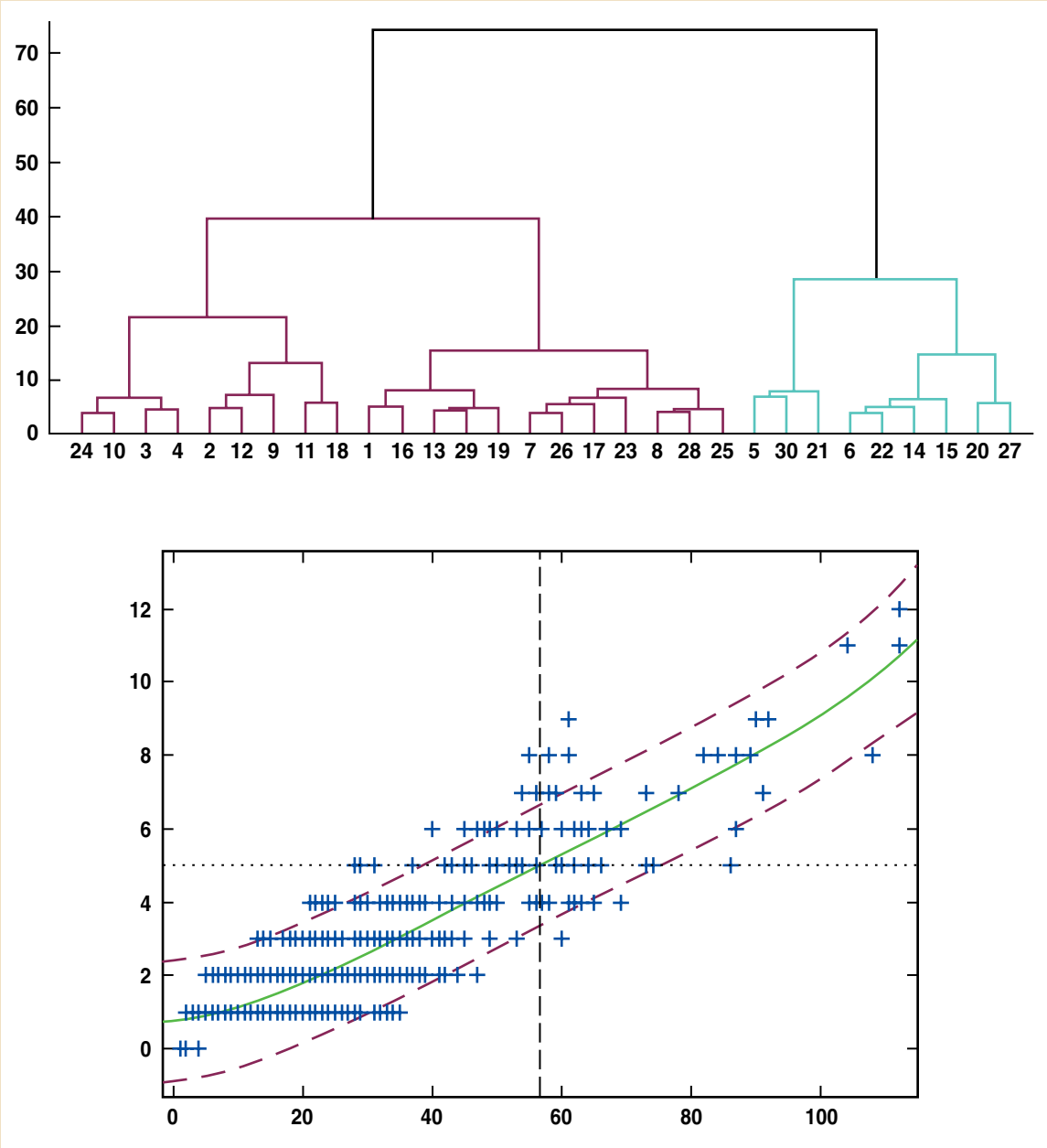


Figure B-3. Creating a signal-specific TESPAR alphabet:
A tree classifier (top) and a regression fit (bottom).

Representation of Signal Alphabets

Once signal information was recoded into strings in a smaller, fixed alphabet, we needed a systematic way to represent an analog signal alphabet sequence as geometrical forms. Figure B- 5 presents an approach to the coding of a unique geometric alphabet from a generic analog signal. Given an arbitrary raw analog signal, the raw signal is passed through a filter recoding the original signal into patterns of real and imaginary zero crossings. Based on the alphabetic mapping developed for that signal, a TESPAR number code corresponding to the real and imaginary crossing combinations is associated with a geometrical figure. The figures are produced using geometrical structures that we call a Box Code and a Circle Code.

Figure B-5 illustrates the use of a box code to represent alphabets. Each box code frame comprises a certain number of points arranged in a square array with a single center point. For example, the simplest usable box code could have five points. The figure generation points are placed at each corner and the center. A line between some subset of these points draws one possible character inside the frame. For example, to create one alphabet, some subset of characters of length three is taken from the frame. In this case, the removal of duplicates and possible retraces that overlap previous paths results in a maximum number of 31 unique characters for this particular box frame. The set can then be used to represent a subset of TESPAR real/imaginary signal combinations appearing in the density plot such as seen in Figure B-2 of size 31 or less.

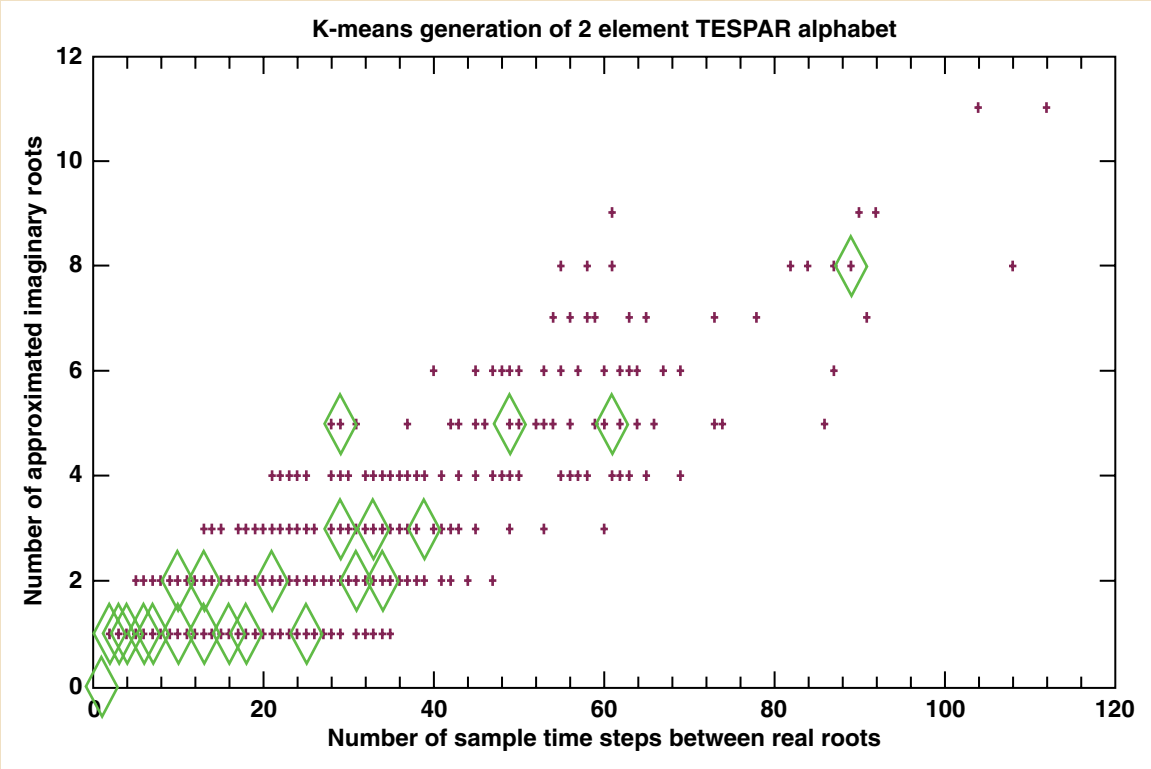


Figure B-4. K-Means Clustering.

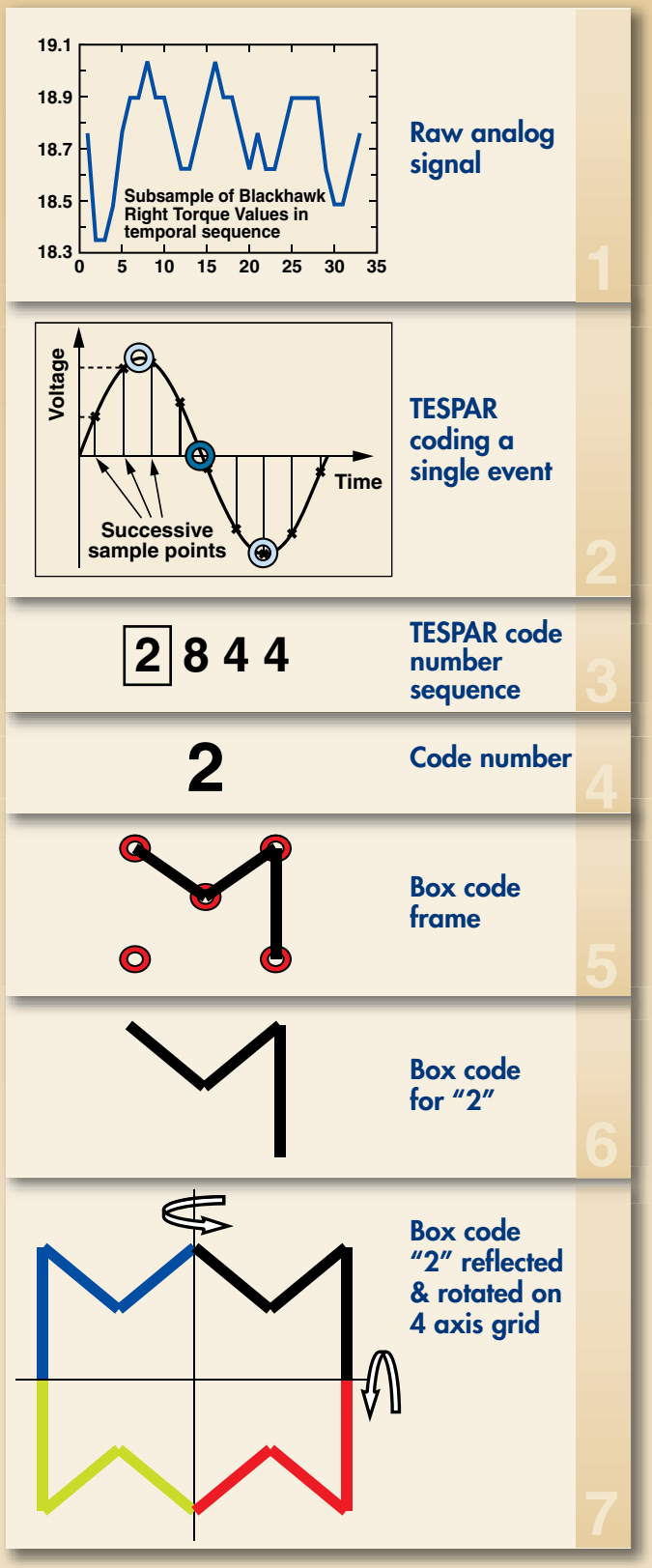


Figure B- 5. Mapping of signal segments into box codes.

Creating Geometrical Structures from Signal Alphabets

Once a graphic alphabet is created, the next step is to design geometrical structures for housing it. Two important concepts are used to generate a two-dimensional representation from a one-dimensional signal—reflection and symmetry (Figure B-5). Both help the human eye to easily detect patterns.

To test this approach, we used signals from accelerometers distributed along a Space Shuttle main wing, in order to detect foam impacts during launch (see the discussion in section 5). The sample results came from Mission STS-114—the first "Return to Flight" mission following the Space Shuttle Columbia accident—launched on July 26, 2005. In our analysis of these analog signals, we employed direct translation of the signal (i.e., without using a TESPAR code) into a two-dimensional pattern. We did this by taking a normalized value of the raw signal at each point and transforming it into a parameter of a radial geometric function. Hyper-geometric and less complex cosine functions both produced radial symmetric patterns. These patterns proved to be highly responsive to small changes in input conditions and temporal variation.

Figure B-6 shows first-, second-, and third-order interactions of three laterally located accelerometers in the shuttle's left wing sampled at 20 KHz, about 18 seconds into ascent. The radial patterns were generated by passing a normalized, time-stamped value of the wing accelerometer readings at that time into cosine function arguments. During a launch sequence, these dynamic radial patterns

changed in size, symmetry, and number of nodes as a function of acceleration changes (see Figure B-7). An intermittent sensor failure observed during launch resulted in a solid filled circle (upper left green circle). A value of less than one G (a shake) produced a distorted figure (upper right). Higher G's resulted in more radial nodes being produced (lower left). Sensor differencing produced similar effects.

The analysis identified an intermittent sensor problem that was not seen in the raw data by inspection and other computational methods that were used to analyze this dataset. By simply organizing the sensor values such that first-, second-, and third-order interactions can be visualized, the static output of a faulty sensor was easily spotted. The results motivated us to study the applications of graphic symmetries to represent the information content of a signal as well as simple local magnitude changes. To do

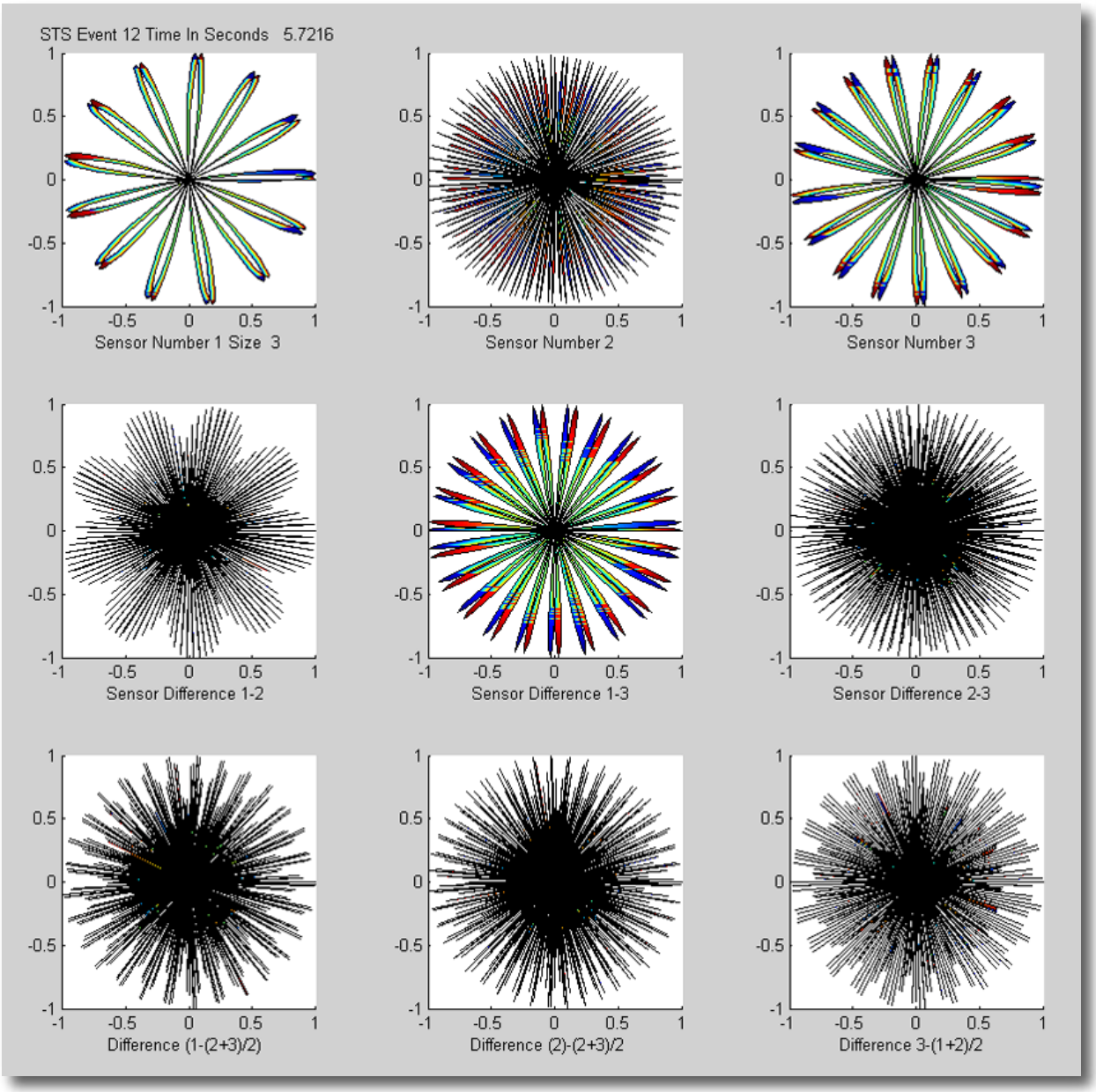


Figure B-6. Cosine display of shuttle accelerometer data.

that, however, it was necessary to develop the methodology to capture information content of a signal, preserve temporal sequence of the signal, and still produce a composite picture that could be used in a higher-order structure which later became the TESPAR approach described above. What remained to be developed was the method to combine the alphabet elements into a 2-D composition.

A General Approach for Creating 2-D Compositions

Figure B-8 illustrates one approach for combining information content as well as temporal magnitude changes using the TESPAR graphic alphabets. Since each TESPAR alphabet character matches an area of a signal's information content (its real and imaginary roots over an interval), we can represent a

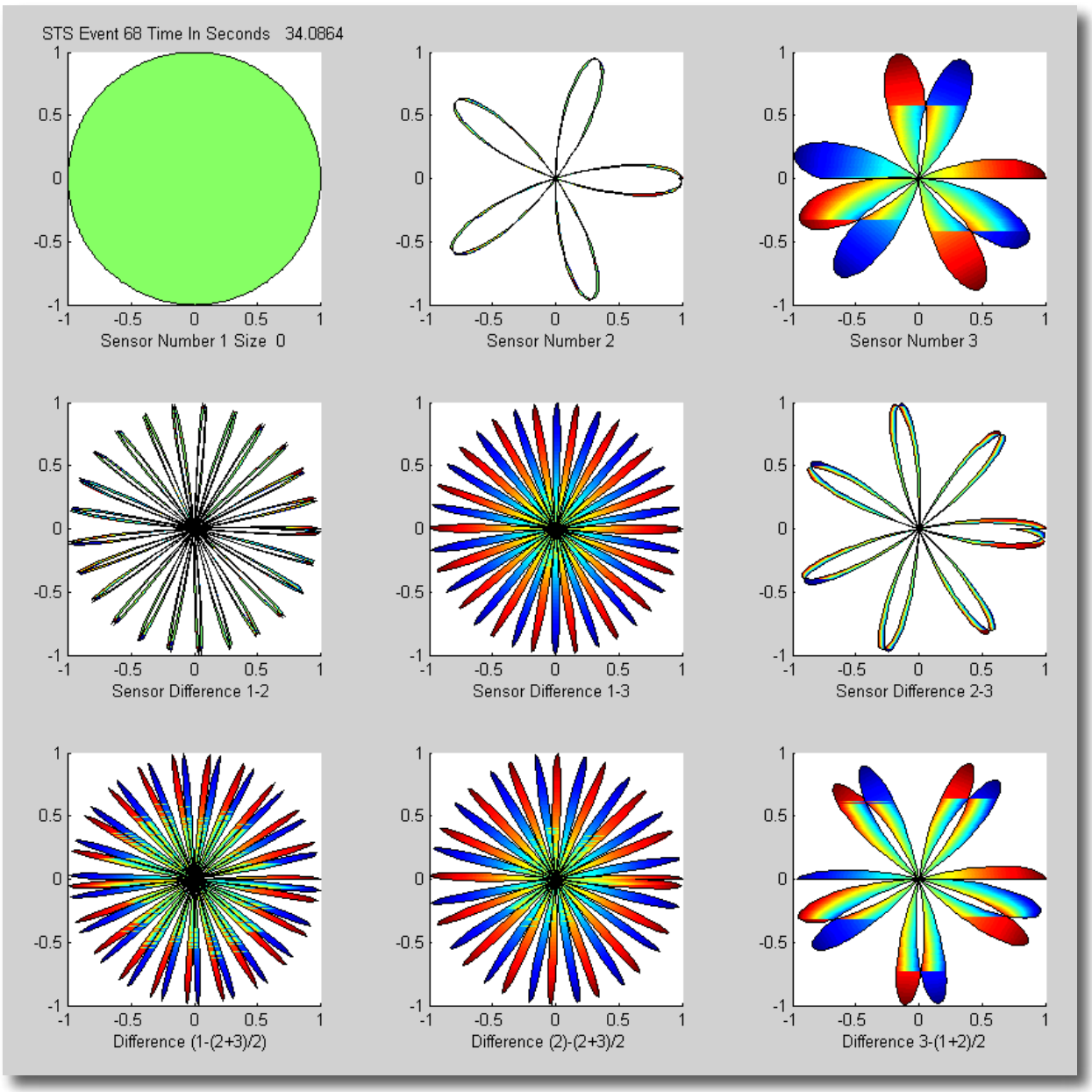


Figure B-7. Distortions and sensor failure (upper left).

window for a large time period of the signal by combining visual shapes of TESPAR characters. Then, by using some of the properties for organizing a field and creating wholeness, as discussed in sections 4 and 5, the combined figure can be copied and reflected into two-dimensional symmetric patterns.

The arrangement presented in Figure B-8 is somewhat analogous to a kaleidoscope where mirrors reflect and combine multiple pieces of glass into a symmetric shape. The idea required the creation of an algorithm to connect signal alphabet sequences of set lengths into patterns. The same pattern was then repeated along multiple axes and manipulated in terms of size and location to produce a single figure over some time window. To do this, we generated a Java-based graphic display tool for the signals.

We can choose any number of reflected axes. In this case, we arbitrarily selected four axes. As for the weaving of the alphabet characters, each box code symbol in Figure B-8 is assigned a start and end point. The is moved as close to the beginning of the starting axis as possible. In this case, the character is a blue triangle. The pattern is then duplicated on the second through nth axes, each one rotated appropriately (90 degrees in the case of a 4-axis structure). For example, the second character looks like a left-facing "staple." This staple is placed so that its last graph point is as close to the reference axis as possible without intersecting the first character on any other axis. This results in the staple pattern being slightly offset relative to the first triangle's placement. This axis is then duplicated for all four directions and the process continues. For a very long string, the algorithm produces a spindly figure. The apparent visual density can be improved in our visualization tool by increasing the number of axes, or having each character grow proportionally in size as it moves out from the axis center to fill space between axes. Figure B-9 illustrates one result of a similar process.

The upper left frame of Figure B-9 is the first 14 letters of the box code alphabet created in Figure B-5. It is printed sequentially from upper left (the first character) to lower right (the fifteenth character) along a 135-degree axis. The upper right frame shows a repeating signal. It has a signal period of three TESPAR events long that repeats three times. It is plotted along six axes instead of four to show the effect of increasing radial density. The lower left frame illustrates the effects of dramatically increasing figure

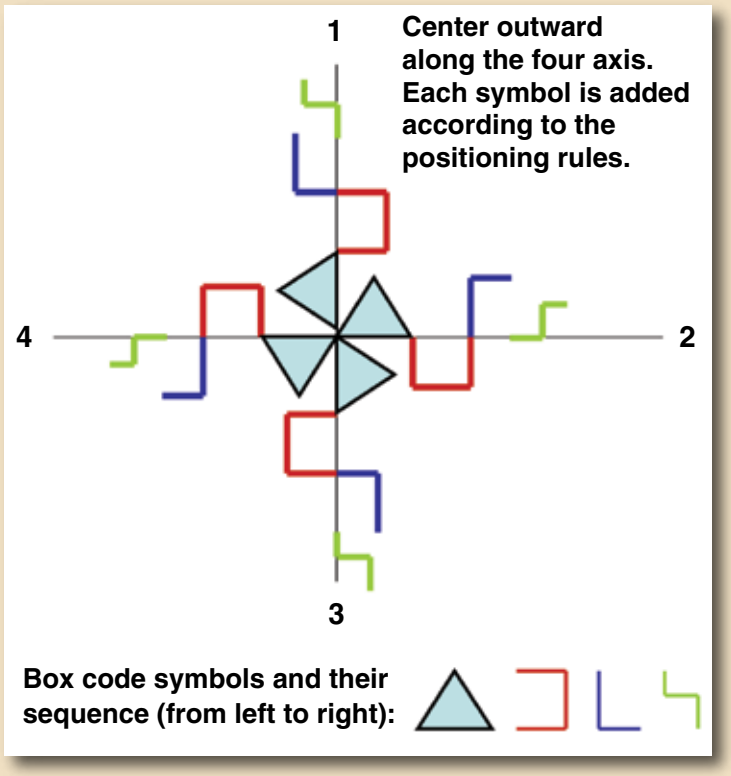


Figure B-8. Creating a two-dimensional shape from a character string.

density by increasing the axis number to 60. The signal string begins to take on a distinct pattern-like signature as the axes are increased and the eye tends to fill in detail at the character junctions. The lower right figure uses the fourth through sixth TESPAR characters in another three-character repeating pattern. This pattern is shown with 30 axes. The center of the figure looks distinctively different than the first figure for the other signal pattern. The need to fill in the pattern as the characters move outward is evident.

The above examples do not show additional graphic modifications available to a display designer which we have used in follow-on work, such as the use of color (e.g., to define temporal aspects of a pattern element), space filling (to create a fully connected figure between axes), or connection (to smoothly unify multiple analog signals presented in a common display window). Neither do they illustrate the use of circle codes alone or in conjunction with box codes to mix angular and smooth perceptual elements. They do illustrate, however, first steps in moving from a pure analog signal to a field-like presentation.

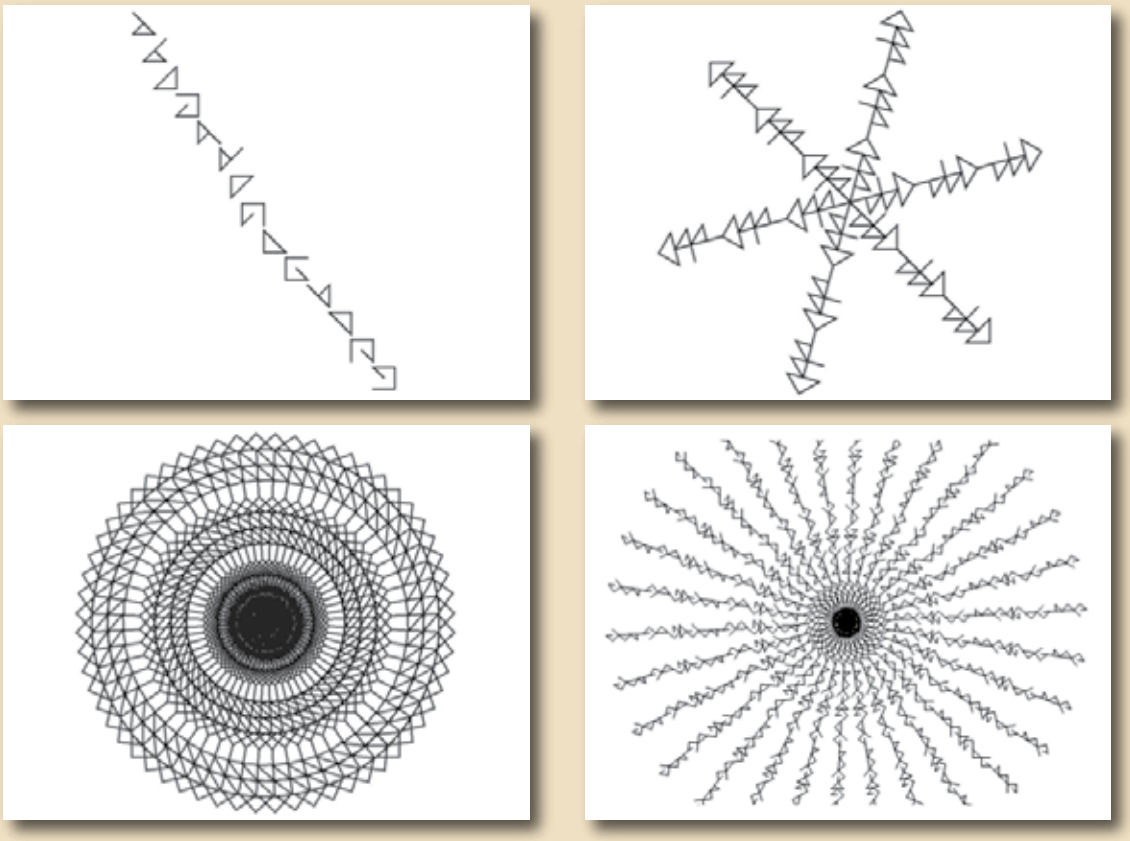


Figure B-9. TESPAR code sequences in multiple axes.

Summary

In this Appendix we discussed an alternative approach for transforming analog signals for the purpose of visual displays. The technique uses a unique descriptive feature set (real and imaginary zero crossings) to characterize complex signal information content in fixed-length alphabet symbols. The approach maintains control over information loss, which differentiates it from many visual display concepts. These feature alphabets are integrated to create structures of information. The technique allows the following gains over traditional analog display methods: First, we can describe the signal's information content that may not be well differentiated by a standard frequency representation. Second, we have control over the dimensionality of the feature set we use. Third, we use an analog signal recording approach that is not based on arbitrary sampling frequencies but rather on zero crossing points determined by the function itself. Fourth, we generate an efficient, compact visual representation that permits input into fixed-size feature classifiers, and then combines features to create a whole (examples include neural networks or support vector machines for automated recognition of signal events). Finally, the system can be implemented in a parallel architecture with high efficiency and return reduced computational costs while still maintaining a large degree of the original raw signal information. We completed the process by showing how the signals could be arranged in strings of invertible visual characters in which the eye can rapidly detect variations in time.

References

Alexander, C. (1964). *Notes on the Synthesis of Form*. Cambridge, MA: Harvard University Press.

Alexander, C. (1993). *A Foreshadowing of 21st Century Art: The Color and Geometry of Very Early Turkish Carpets*. New York: Oxford University Press.

Alexander, C. (2002a). *The Phenomenon of Life*. Berkeley, CA: The Center for Environmental Structure.

Alexander, C. (2002b). *The Process of Creating Life*. Berkeley, CA: The Center for Environmental Structure.

Ang, R. (1998). Canonical correlation analysis as a general linear model: An effective instructional tool. *Perceptual and Motor Skills, 87*, pp. 1315-1323.

Bateson, G. (1972/1999). *Steps to an Ecology of Mind*. Chicago, IL: University of Chicago Press.

Bennett, K., Toms, M., & Woods, D. (1993). Emergent features and graphical elements: Designing more effective configural displays. *Human Factors, 35*(1), pp. 71-97.

Bertin, J. (1967/1983). *Semiology of Graphics: Diagrams, Networks, Maps* (W.J. Berg, Trans.). Madison, WI: The University of Wisconsin Press.

Bertin, J. (1977/1981). Graphics and Graphic Information Processing (Section B: Graphic Constructions). Reprinted in S. Card, J. Mackinlay, & B. Shneiderman, 1999, *Information Visualization: Using Vision to Think*. San Francisco, CA: Morgan-Kaufmann. pp. 62-65.

Billings, C. (1997). *Aviation Automation: The Search for a Human Centered Approach*. Mahwah, NJ: Lawrence Erlbaum.

Bond, F. & Cahn, C. (1958). A relationship between zero crossings and fourier coefficients for bandwidth limited functions. *IRE Trans. Information Theory, IT-4*, pp. 110-113.

Bradley, P., Mangasarian, O., & Street. W. (1997). Clustering via Concave Minimization. In M. Mozer, M. Jordon, & T. Petsche (Eds.), *Advances in Neural Information Processing Systems, 9*, pp. 368-374. Cambridge, MA: MIT Press.

Bradley, P. & Fayyad, U. (1998). Refining initial points for K-means clustering. *Proceedings of the International Conference on Machine Learning*, pp. 91-99.

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. (2d ed., rev. & enl). Berkeley, CA: University of California Press.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Information Visualization: Using Vision to Think*. San Francisco: Morgan-Kaufmann.

Critchlow, K. (1976/1999). Islamic Patterns: An Analytical and Cosmological Approach. Rochester, VT: Inner Traditions International.

Cumming, J. & Wooff, D. (2007). Dimension reduction via principal variables. *Computational Statistics & Data Analysis, 52*, pp. 550–565.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, pp. 571-582.

Degani, A. (1996). *Modeling human-machine systems: On modes, error, and patterns of interaction*. Unpublished doctoral dissertation. Atlanta, GA: Georgia Institute of Technology.

Degani, A. (2004). *Taming HAL: Designing Interfaces Beyond 2001.* New York: St. Martin's Press (Palgrave Macmillan).

Degani, A. (2008). Some properties of multilayered patterns (off information): Insights from medieval architecture. Poster presented at the *Human Factors and NextGen Conference: The Future of Aviation*. University of Texas, Arlington. May 28-29.

Degani, A. & Heymann, M. (2002). Formal Verification of Human-Automation Interaction. *Human Factors, 44*(2), pp. 28-43.

Degani, A., Shafto, M., & Olson, L. (2006a). Canonical correlation analysis: Use of composite heliographs for representing multiple patterns. *Proceedings of the Fourth International Conference on the Theory and Application of Diagrams*. Springer-Verlag: Lecture Notes in Computer Science 4045. pp. 93-97. June 28-30, Stanford, CA.

Degani, A., Shafto, M., & Olson, L. (2006b). Data Abstraction and Integration of Information for Display: Analysis of Pilot-Automation Interaction. *Proceedings of the 11th International Conference of Human-Computer Interaction in Aeronautics*. pp. 164-167. September 20-22, Seattle, WA.

Degani, A., Shafto, M., & Kirlik, A. (2006). What makes vicarious functioning work? Exploring the geometry of human-technology interaction. In A. Kirlik (Ed), *Adaptive Perspectives on Human-Technology Interaction*. New York: Oxford University Press.

Degani, A., & Wiener, E. (1994). *On the Design of Flight-Deck Procedures*. NASA Technical Memorandum #177642. Moffett Field, CA: NASA Ames Research Center.

Diehl, S. (2007). *Software Visualization: Visualizing the Structure, Behaviour, and Evolution of Software*. Springer-Verlag.

Drewes, F. (2006). *Grammatical Picture Generation: A Tree-Based Approach*. Springer-Verlag.

Ehrig, H., Ehrig, K., Prange, U., & Taentzer, G. (2006). *Fundamentals of Algebraic Graph Transformation*. Springer-Verlag.

Einhorn, H., Kleinmuntz, D., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review, 86*(5), pp. 465-485.

Ester, M., Kreigel, H., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the 2nd ACM Special Interest Group on Knowledge Discovery and Data Mining*, pp. 226-231.

Fishwick, P. (2008). *Aesthetic Computing.* Cambridge, MA: MIT Press.

Friendly, M. (2002). Corrgrams: Exploratory displays for correlation matrices. *American Statistician, 56*, pp. 316–324.

Garland, K. (1994). *Mr. Beck's Underground Map*. Harrow Weald, Middlesex: Capital Transport Publishing.

Garner, W. (1974). *The Processing of Information and Structure*. Potomac, MD: Lawrence Erlbaum.

Gehman, H. et al. (2003). *Columbia Accident Investigation Board Report, Vol. 1*, NASA and U.S. Government Printing Office.

Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgment. *American Psychologist, 23*, pp. 483-496.

Goldstein, W. (2006). Introduction to Brunswikian theory and methods. In A. Kirlik (Ed.) *Adaptive Perspectives on Human-Technology Interaction*. New York: Oxford University Press.

Government of Portugal (2004). *Accident Investigation Final Report of the Ministry of Public Works, Transportation, and Communications: All engines-Out Landing Due to Fuel Exhaustion, Air Transat, Airbus A330-243, Lajes, Azores, Portugal, 24 August 2001*. (Available from the Transportation Safety Board of Canada).

Grabar, O. (1992). *The Meditation of Ornament*. Princeton, NJ: Princeton University Press.

Grünaum, B. (2006). What symmetry groups are present in the Alhambra? *Notices of the American Mathematical Society, 53*(6), pp. 670–673.

Guo, D. (2007). Visual analytics of spatial interaction patterns for pandemic decision support. *International Journal of Geographical Information Science, 21*, pp. 859-877.

Hamscher, W. (1991). ACP: Reason maintenance and inference control for constraint propagation over intervals. *Proceedings of the 9th National Conference on Artificial Intelligence*, pp. 506-511.

Hardoon, D., Szedmak, S., & Shawe-Taylor, J. (2003). *Canonical correlation analysis: An overview with application to learning methods*. Technical Report CSD-TR-03-02. Department of Computer Science, Royal Holloway, University of London.

Heymann, M., Degani, A., & Barshi, I. (2007). Generating procedures and recovery sequences: A formal approach. *Proceedings of the 14th International Symposium of Aviation Psychology*, pp. 252–257. April 22–25, Dayton, OH.

Heymann, M., & Degani, A. (2007). Formal analysis and automatic generation of user interfaces: Approach, methodology, and an algorithm. *Human Factors, 49*(2), pp. 311-330.

Holbeche, R., Huges, D., & King, R. (1986). Time Encoded Speech (TES) descriptors as a symbol feature set for voice recognition systems. *IEE International Conference on Speech Input/Output: Techniques and Applications*, pp. 310-315.

Hollnagel, E. & Woods, D. (1999). Cognitive systems engineering: New wine in new bottles. *International Journal of Human-Computer Studies, 51*, pp. 339-356.

Hotelling, H. (1935). The most predictable criterion. *Journal of Educational Psychology, 26*, pp. 139-142.

Hotelling, H. (1936). Relations between two sets of variants. *Biometrika, 28*, pp. 321–377.

Huang, S., Lee, M., & Hsiao, C. (2006). *Kernel Canonical Correlation Analysis and its applications to nonlinear measures of association and tests of independence*. Institute of Statistical Science, Academia Sinica, Taipei, Taiwan. (http://www.stat.sinica.edu.tw/syhuang/stat.htm)

Imhof, E. (1982). *Cartographic relief presentation*. Berlin, Germany: Walter de Fruyter.

Iverson, D. (2004). Inductive System Health Monitoring. *Proceedings of the 2004 International Conference on Artificial Intelligence (IC-AI'04)*.

Jorgensen C., & Sharma, S. (2009). Visual coding of analog signals for advanced display development. Paper accepted to the *2009 International Conference on Modeling, Simulation, and Visualization Methods (MSV'09)*. July 13-16, Las Vegas, NV.

Julesz, B. (1971). *Foundations of Cyclopean Perception*. Chicago, IL: The University of Chicago Press.

Jung, C.G. (1955/1972). *Mandala Symbolism*. Princeton, N.J.: Princeton University Press.

Kerren, A., Ebert, A., & Meyer, J. (2006). Human-centered visualization environments. *Lecture Notes in Computer Science 4417*. Springer-Verlag.

Kim, T., Wong, S., & Cipolla, R. (2007). Tensor canonical correlation analysis for action classification. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Minneapolis, MN.

King, R., & Phipps, T. (1999). Shannon, TESPAR, and approximation strategies. *Computers and Security, 18*(5), pp. 445-453.

Krishnamurti, J. (1929/1996). *Total freedom*. San Francisco: Harper-Collins.

Leboff D., & Demuth, T. (1999). *No Need to Ask: Early Maps of the London Underground Railways*. Harrow Weald, Middlesex: Capital Transport Publishing.

Licklider, J., & Pollack I. (1948). Effects of differentiation, integration, and infinite peak clipping upon the intelligibility of speech. *Journal of the ASSA, 20*(1), pp. 42-51.

Lu, P. J. & Steinhardt, P. J. (2007). Decagonal and Quasi-Crystalline Tilings in Medieval Islamic Architecture. Science 315, p. 1106.

Lupu, E., Feher, Z., & Pop, P. (2003). On the speaker verification using the TESPAR coding method. *Proceedings of the IEEE, vol. 9*, pp. 173-176.

MacEachen, A. (1995). *How maps work: representation, visualization, and design*. New York: Guilford Press.

Maille, N., Ferryman, T., Rosenthal, L., Shafto, M., & Statler, I. (2006). *What Happened and Why: Toward an Understanding Of Human Error Based on Automated Analyses of Incident Reports*. NASA TP 2006-213466.

Makarenkov, V., & Legendre, P. (2002). Nonlinear redundancy analysis and canonical correspondence analysis based on polynomial regression. *Ecology, 83*, pp.1146–1161.

Michaud, R., Michaud S., & Barry, M. (1996). D*esign and Color in Islamic Architecture: Eight Centuries of the Tile-Maker's Art*. New York: Vendome.

Miller, J. (1978/1995). *Living Systems*. Boulder, CO: University Press of Colorado.

Mosier K., & McCauley, S., (2006). Achieving coherence: meeting new cognitive demands in technological systems. In A. Kirlik (Ed.), *Adaptive Perspectives on Human-Technology Interaction*. New York: Oxford University Press.

Necipoglu, G. (1995). *The Topkapi Scroll – Geometry and Ornament in Islamic Architecture*. The Getty Center for the History of Art and the Humanities. New York: Oxford University Press.

Neisser, U. (1982). *Memory Observed: Remembering in Natural Context*. San Francisco: W. H. Freeman.

Neter, J., Wasserman, W., & Kutner, M. (1990). *Applied Linear Statistical Models: Regression, Analysis of Variance and Experimental Designs*. Homewood, IL: Richard D. Irwin, Inc.

Norman, D. (1986). Cognitive engineering. In D. Norman & S. Draper, *User Centered System Design*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Norman, D. (1993). *Things That Make Us Smart*. Reading, MA: Addison-Wesley Publishing Company.

Norman, D. (1990). The "problem" with automation: Inappropriate feedback and interaction, not over-automation. *Phil. Trans. Research Society London, B 327*, pp. 585-593.

Norman, D. (2007). *The Design of Future Things*. New York: Basic Books.

Oishi, M., Tomlin, C., & Degani, A. (2003). *Discrete abstraction of hybrid systems: Verification of safety and application to user-interfaces*. NASA Technical Memorandum #212803. Moffett Field, CA: NASA Ames Research Center.

Ozdural, A. (1995). Omar Khayyam, mathematicians, and conversazioni with artisians. *The Journal of the Society of Architectural Historians, 54*(1), pp. 54-71.

Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics. Part A: Systems and Humans, 30*, pp. 286-297.

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. New York: North-Holland.

Requicha, A. (1980). The Zeros of Entire Functions: Theory and Engineering Applications. *Proceedings of the IEEE, 68*(3).

Salingaros, N. (1995). In defense of Alexander. *HALI: The International Magazine of Antique Carpet and Textile Art. 78,* pp. 67-69.

Savarese, S., DelPozo, A., Niebles, J., & Fei-Fei, L. (2008). Spatial-temporal correlations for unsupervised action classification. *Eleventh IEEE International Conference on Computer Vision*. Rio de Janeiro, Brazil.

Seo, J. (2005). *Information Visualization Design for Multidimensional Data: Integrating the Rank-By-Feature Framework With Hierarchical Clustering*. Ph.D. Thesis. Department of Computer Science, University of Maryland.

Seo, J., & Shneiderman, B. (2006). Knowledge discovery in high-dimensional data: Case studies and a user survey for the rank-by-feature framework. *IEEE Transactions on Visualization and Computer Graphics, 12*, pp. 311–322.

Shafto, M., Degani, A., & Kirlik, A. (1997). Canonical correlation analysis of data on human-automation interaction. Proceedings of the *41st Annual Meeting of the Human Factors and Ergonomics Society*. Albuquerque, NM. San Diego, CA: Human Factors and Ergonomics Society.

Sharma, S., & Jorgensen C. (2009). Visual coding of monotonic signals and closed functions for the development of advanced displays. Paper accepted to the *2009 International Conference on Modeling, Simulation, and Visualization Methods (MSV'09)*. July 13-16, Las Vegas, NV.

Shannon, C. (1962/1948). The Mathematical Theory of Communication. In C. Shannon & W. Weaver (Eds.), *The Mathematical Theory of Communication*. Urbana, IL: The University of Illinois Press.

Simon, H. (1969). *Sciences of the Artificial*. Cambridge, MA: MIT Press.

Simon, H. (1971). Designing organizations for an information-rich world. In M. Greenberger (Ed.), *Computers, Communications and the Public Interest.* Baltimore, MD: The Johns Hopkins Press.

Statler, I. (Ed.). (2007). *The Aviation System Monitoring and Modeling (ASMM) Project: A documentation of its history and accomplishments, 1999-2005*. NASA TP 2007-214556.

Telea, A. (2007). *Data Visualization*. Wellesley, MA: AK Peters, Ltd.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
5/1/2009	Technical Memorandum	1/1/2007-5/1/2009

4. TITLE AND SUBTITLE

On Organization of Information: Approach and Early Work

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Asaf Degani, Charles Jorgensen, David Iverson, Michael Shafto, & Leonard Olson

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

292487.01.01.02

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

NASA Ames Research Center, Moffett Field, California, 94035-1000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington D.C. 20546-0001

10. SPONSORING/MONITOR'S ACRONYM(S)

NASA

11. SPONSORING/MONITORING REPORT NUMBER

NASA/TM–2009-215368

12. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified — Unlimited Distribution: Nonstandard
Subject Category: 12, 90
Availability: NASA CASI (301) 621-0390

13. SUPPLEMENTARY NOTES

Point of Contact: Asaf Degani, Ames Research Center, MS: 269-4, Moffett Field, CA 94035-1000
(650) 604-0013, adegani@mail.arc.nasa.gov (asaf.degani@gmail.com).

14. ABSTRACT

In this report we describe an approach for organizing information for presentation and display. The approach stems from the observation that there is a stepwise progression in the way signals (from the environment and the system under consideration) are extracted and transformed into data, and then analyzed and abstracted to form representations (e.g., indications and icons) on the user interface. In physical environments such as aerospace and process control, many system components and their corresponding data and information are interrelated (e.g., an increase in a chamber's temperature results in an increase in its pressure). These interrelationships, when presented clearly, allow users to understand linkages among system components and how they may affect one another. Organization of these interrelationships by means of an orderly structure provides for the so-called "big picture" that pilots, astronauts, and operators strive for.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Asaf Degani
Unclassified	Unclassified	Unclassified	Unclassified	105	19b. TELEPHONE *(Include area code)* (650) 604-0013