21

Emulation of Haptic Feedback For Manual Interfaces

by

Karon E. MacLean

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1996

Author ....
Department of Mechanical Engineering
November 17, 1995

Certified by ....
William K. Durfee
Associate Professor and Director of Design Education
Department of Mechanical Engineering, University of Minnesota
Thesis Supervisor

Accepted by ....
Ain A. Sonin
Chairman, Departmental Graduate Committee

Emulation of Haptic Feedback For Manual Interfaces

by

Karon E. MacLean

Submitted to the Department of Mechanical Engineering
on November 17, 1995, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

In designing manual user interfaces, haptic feedback (the touch sense, whereby humans perceive body positions and forces) is often neglected in part because an iterative design strategy requires hands-on manipulation of a prototype interface. Physically interactive prototyping tools will depend on the computer-controlled emulation of haptic phenomena, a technique which has not yet been mastered.

The *fidelity* of a haptic emulation to its real or virtual target is a nonlinear and largely unknown function of many "emulation variables" — characteristics of hardware, control and environment model. Deducing this function's dependencies is a prerequisite for emulating haptic interfaces at a specified fidelity, in turn a requirement for haptic interface prototyping tools.

The governing objectives of my research were to develop a conceptual framework and methodology which relates haptic emulation performance to emulation variable values; and to test the methodology's usefulness by quantifying the influence of an important set of emulation variables. To this end I developed a haptic emulation testbed, a library of haptic emulations, techniques for identifying haptic properties of real manual control devices, and carried out human subject experiments to test emulation variable influence on perceived emulation fidelity to real targets.

The experimental hardware and control setup, designed with an emphasis on versatility and performance, consisted of a one degree of freedom linear actuator equipped with sensors and interface handle, variable physical damping and a PC-implemented haptic environment controller.

The experiments were based on subjects comparing an emulation to a set of real devices; in each trial the emulation targeted one real device, the subject chose the one real device which felt most like the emulation, and emulation variable values were changed for the next trial. The subjects' ability to discern which real device was targeted was thus related to emulation variable values. Emulation variables tested were sampling rate, physical damping and virtual environment expression.

Results were threefold, comprising (a) insight on hardware, control and modeling issues pertaining to the creation of high fidelity emulation of haptic feedback; (b) specific experiment results re influence of the emulation variable tested; and (c) a critique of the experiment methodology.

Thesis Committee:	Prof. Woodie C. Flowers, Head Prof. Thomas B. Sheridan Dr. David Zeltzer

Thesis Supervisor: William K. Durfee
Title: Associate Professor and Director of Design Education
Department of Mechanical Engineering, University of Minnesota

I want to reach that state of condensation of sensations which constitutes a picture.

— Henri Matisse

Acknowledgments

In completing this work, I received help and support from more people than I can name here. I thank my advisor, Will Durfee, for his technical guidance and personal friendship; I aspire to be as fine a teacher someday. The other members of my committee (Tom Sheridan, Dave Zeltzer and especially Woodie Flowers) supplied extra attention after Will left MIT, above and beyond what was required or expected. Former Dean Frank Perkins, who made heroic efforts to keep me funded, writes the book on not passing the buck. His sincere concern with the well-being of all the students he comes in contact with earns deep gratitude from those who have exhausted all other alternatives.

My fellow inhabitants of the Newman lab were a rich and well-utilized source of feedback on work, weight-room camaraderie, banter (a pale shadow of the real thing, I'm told — the south mezzanine only looks like a locker room) and an infallible reference on comic book superheros and Howard Stern. Thanks particularly to Justin and Joe for never acting like they were tired of answering my questions; I wish they had asked more of me.

Norm Berube taught me to machine and tried to teach me to slow down. His competence and sour humor made shop time the best part of the week even when I was looking for something to throw at him. May ground hogs never eat Norman's daffodils again (they'll be sorry if they do).

The mighty and united femmes of MEGAwomen added incalculably to my stay here: Jeannie, Louise, Heather, Naomi and others. Most of all Li, my regular accomplice in brief and dubious bids for escape and provider of premium raw material for my short stories. My only regret is that I didn't tape-record more of our conversations over microwave-boiled pasta and cold Ragu.

I am indebted to those who served as experiment subjects (sorry, I can't give your real names) and thesis readers (Will, Rich, John and Lou), possibly the two most drudgerous ways to earn mention in someone's thesis acknowledgments.

Family and friends away from MIT have been there through my hard times and sometimes even succeeded in reminding me that there is Life Outside. I seem to be leaning more in every way on Mom and Dad now than I have since I was a teenager; the last year would have been hard to survive without them. Two nephews (James and Andrew) have arrived since I've been here, and two more will come the month I leave. Andrea, dear friend and emotional double, gets credit for the time we've made in our over-booked lives to take care of each other. She is the hard one to leave behind. Rich, far away on Vancouver Island, has been waiting longer for me to finish than I think he ever bargained for. He ought to be canonized for putting up with me all this time and hardly ever complaining (in any case, he probably should see a doctor about it).

My housemates, although they haven't seen me for a year, apparently continue to believe in my existence. To whichever of you shoveled a parking spot free of snow for me last night, a special thanks and a warm dry spot by the fire in Housemate Heaven.

The final thanks go to my most constant and faithful friend of all, who is at this very moment urgently requesting that I finish this and join her outside to romp in the new midnight snow. MIT will be a more hygienic place without you, Zeebo, but a lot duller.

Financial support for this research was provided collectively by Brit and Alex d'Arbeloff, the Ijaz Engineering Sciences Fellowship and the Silicon Valley Group. Without their philanthropy the work might still have happened but it wouldn't have been nearly as much fun.

To my parents.

Contents

List of Figures

Part I

The Context

Chapter 1

Introduction

1.1 Have You Ever Felt Like Whoever Made That Thing Forgot About You?

Then keep on reading ...

Everyone has a favorite:

the option-loaded remote control;

the unprogrammable VCR;

the driving column control stalk that turns the windshield wipers on whenever you signal for a right-hand turn;

the digital alarm clock that won't be turned off when it shatters your Sunday morning slumber;

the answering machine that keeps broadcasting its message after you've been slow to reach the receiver, and there's no apparent way to shut it up short of pulling its plug ...

Why, when any ordinary person in Boston, MA can send a letter to Kathmandu in the space of five minutes and thaw a frozen chicken in the same time, does it require the processing power of a computer and the patience of a saint to do battle with many of the so-called labor saving, entertainment and lifestyle enhancing devices on the market today? When scientists are taking close-up pictures of Mars, why do air traffic controllers retire at age 45 with ulcers and high blood pressure, and how can Three Mile Island be explained?

The Bottom of the Problem

The disease which gives rise to these symptoms is the ineffective design of user interfaces. The source of the disease varies from case to case.

Some say that technological advances have led to cheap, commonplace device complexity: it costs next to nothing to design a microprocessor-controlled device with fifty functions, but trying to make it obvious what those functions do, let alone how to use them, is often difficult and consequently expensive when it can be done at all.

This is a good point, but it does not explain why simple devices with only one or two buttons or knobs can be as irritating or confusing as those with fifty.

Some interfaces are hard to use because they are complicated; others because they cannot be easily reached, or the control labels are misleading, or the buttons are too close together, or the display is unreadable, or because the feel of a control is inconsistent with what the control is supposed to do.

It is important to note that the problems at issue run deeper and more broadly than the home electronics industry. However, the consumer products market is rich with familiar and graphic examples, and they also provide a logical problem attack point by virtue of their manageable sizes and diversity of form and function. As the possibilities of a rapid prototyping approach to user interface design are explored in the ensuing discussion and a plan proposed for its implementation, consumer products will be drawn upon as case studies.

User Interfaces

The term *user interface* implies the entire set of factors which together comprise a person's total contact with a machine or device; the study of human interfaces is aimed at improving the machine's overall ease-of-use by identifying and optimizing these factors. They range from human-machine communication links (modes of information and power transfer to and from the user, optimization of which requires attention both to ergonomic and cognitive issues), to purely comfort-creating considerations such as chair back support and computer screen glare. In general, a given human interface component will have a role both in communication and in the user's bodily comfort.

Cut "That Idiot" Some Slack: User Interfaces Are Hard to Design

User interfaces are hard to design because although some guidelines exist, it is often difficult to draw on these general rules for aid in either the development process or the specification of required quality and characteristics of given interfaces. Further, the "goodness" or "badness" of an interface is a highly subjective and multi-faceted rating, determined by such factors as anatomical variations and personal or cultural biases. Often the only way to determine a prospective interface's adequacy is to build it and use it.

When one can't get to a solution from analysis or past experience, one iterates. Doing this rapidly and inexpensively enough so that it is cost-effective in a market sense is called ***rapid prototyping***, and there are mushrooming industries which cater to the rapid prototyping of a multitude of engineering challenges.

1.2 Haptic Feedback In Manual Controls

The haptic (touch) sense — its name comes from "haptos", Greek for "hand" — is a fundamental avenue of sensory feedback which is almost universally overlooked in the commercial design of manual controls.[1] Aspects of user-machine interfaces which act upon the haptic sense through physical contact and the consequent transfer of power between operator and device — manual control elements which are routinely grasped, twisted, pushed or pulled — are often crucial to the device's intuitiveness

---

[1] The most complete dictionary definition of "haptic" I have located is simply "the science concerned with the tactile sense"; "tactile" is defined as "Relating to touch or to the sense of touch" [185]. The inadequacies of these formal definitions reflect the youth of this field. Five years ago (1990) several different terms were being used to represent a single concept — for example, "kinesthetic", "proprioceptive", "force feedback", "tactile", "physical", etc.. Only in the last 3–4 years has "haptic" come to be accepted as the standard means of referring to real and virtual environments which act on the touch sense, as opposed to vision or audition; and those psychophysical characteristics which are influenced by such environments. Whether or not this is the correct psychophysical definition, it has come to have an explicit meaning to those in this field.

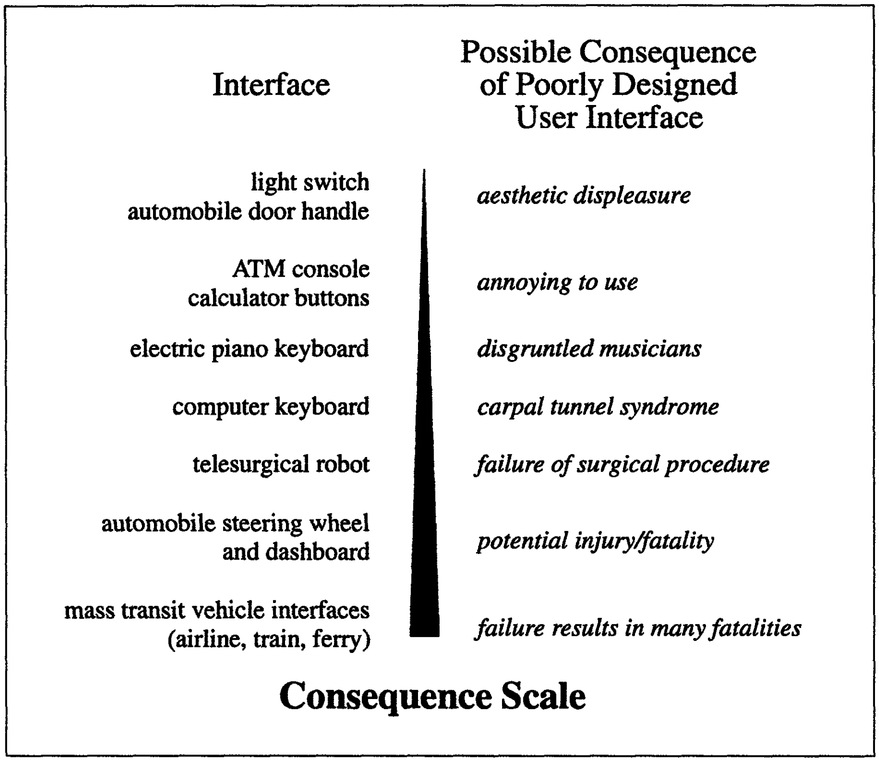


Figure 1-1: Representative user interfaces and what can happen when they fail.

and to its ease and comfort of use.[2] In short, haptic feedback in a user interface matters when there is a manual component of interaction and the user's primary visual attention is occupied elsewhere.

The left column in Figure 1-1 lists some examples of when haptic feedback is important. At one end of the scale is a light switch or a door handle which merely annoys you. At the other end, things get more serious.

Regardless of its consequence, haptic feedback is rarely designed carefully if at all, and therefore many of these consequences are unfortunately a part of our everyday lives.

Information Transfer

In general, manual controls permit bidirectional communication, in that they provide both a forward path through which the user commands the machine, and a return sensory feedback path by which the user can discern something about the state of the machine. At the lowest level, this might be whether the control action was recognized; for instance, a discernible click signaling engagement when a button has been fully depressed.

Such feedback provided directly to the user's hand is often referred to as "force feedback" or "haptic feedback", and in many cases has advantages over visual or auditory feedback. It unloads other sensory feedback paths for simultaneously executed tasks which do not involve physical contact (looking through the windshield while driving a car) and sometimes is simply the most natural and logical feedback mode.

---

[2] The term "haptic human interface" refers to those aspects of any complete human interface which involve human-machine communication and physical (most often manual) contact with the machine, with the concomitant transfer of mechanical power. Examples are pushing a button, turning a knob, moving a mouse, reaching out to pull down a lever.

Ergonomics

Haptic feedback influences not only the conveying of information to the user but also how comfortable and intuitive a machine is to use; the task of improving the latter is potentially able to benefit from the same strategies and tools as the design of haptic-aimed feedback.

In whole-workspace control environments (for example, auto and airplane cockpits), the complete and large-area involvement of the user brings into effect biomechanical considerations such as reach, natural directions of motion, individual anatomical differences and subtle variations in motor control in the different regions of an individual's workspace. All of these can be optimized only through careful attention to biomechanical and sometimes haptic factors. A lever operated at arm's reach will be difficult, uncomfortable and possibly dangerous to use if it is either too stiff or not stiff enough or has more settings than the operator can resolve with his reduced coordination at that distance.

Multiple Sensory Modes

Important as they sometimes are, haptic user interfaces do not stand alone; manual controls that are felt but never seen or heard are rare. Furthermore, a designer often has a choice in how to present multiple pieces of information to the user.

Whether one is deciding what the dominant feedback mode will be, or designing a control that relies predominantly on one feedback mode but has significant characteristics in other sensory channels as well, all channels ought to be considered or at minimum their interactions understood well enough to avoid design blunders.

Rapid Prototyping of Haptic Interfaces: Not Yet

Haptic-dominated interfaces are hard to design in part because each user and each situation is different; this makes it difficult to predict what will be best in a given design task. Further, the computer aided design (CAD) tools which today streamline the design of the more visual aspects of human interfaces do not work when touching is involved.

Thus an iterative design strategy is required. Iteratively building an actual interface mockup which is realistic in haptic characteristics is rarely either rapid or inexpensive on the scale of the product's eventual market value. The result is that either a disproportionate amount of time is spent in concept-to-market interface development, and/or the interface is not very good. This costs everyone: the designer and manufacturer in extra costs and lost time, revenue and competitive edge; the customer in increased expense, irritation, inefficiency and sometimes even catastrophic damage or loss of life.

Thus the question:

> What quality of multisensory feedback in the final interface will be required for the user to perform the required task?

raises the question of how realistic the prototype has to be in order to serve its intended purpose. Stated in another way:

> What level of abstraction in the simulation of the final interface will afford sufficient information to enable good design decisions?

1.3 A New Approach to Designing Haptic Interfaces

This thesis grew from the premise that the same concepts of rapid prototyping which have revolutionized other aspects of design could be productively applied to the haptic aspect of user interface design:

> Create a *virtual* haptic interface and use it to make iterative design decisions based on the input of prospective users *prior* to building a physical mockup.

The concept of a "virtual haptic interface" requires definition for some readers. As in the more commonly encountered visual and auditory expressions of virtual environments, "virtual" merely implies some manner of representing something that isn't really there — a scene, an object, a situation. The representation may be literal, as in a flight simulator; it may be abstract, as in a virtual environment for the exploration of molecular models; or it may be somewhere in between — for example, a virtual microsurgery interface wherein a literal environment is transformed in some manner in order to augment the surgeon's perception and/or her manual dexterity. The better the representation is, the more the user is likely to make the leap of faith that representation is equivalent to the real thing, in either or both a qualitative (seems "real") or functional (improves the user's ability to perform a task) sense.

Thus, "virtual haptic interface" simply refers to doing this in the realm of what can be felt, most commonly but not exclusively with the hand. The proposed virtual haptic design tools will:

- be based upon a *force display* which conveys key aspects of the interface.
- permit *rapid iteration* on those key aspects; it must be quick and easy to modify the display's "feel".
- be *accessible* to potential users, so that the designer may conveniently test the prototype on a representative user or customer base.
- rely on integrated *multisensory feedback*, so that feel, sound and appearance may be simultaneously emulated.

For this research, we focus the definition a bit more: here we are interested in representing haptic "things" artificially: things which do, have or potentially will exist somewhere in the real world. They are things which we wish to eventually build tangible mechanical production versions of; they are similar to the devices we can find in our automobile or home, buy at the hardware store, or build in the machine shop.

I use the term *emulation* in reference to this particular sort of virtual environment, in order to emphasize the omnipresent relation to be maintained between the virtual environment and a specific real one.

Emulating Haptic Feedback

A *haptic emulation* is defined for the present purpose as a synthesized illusion of the feel of a haptic environment which may be arbitrary *but extant in the real world.* In this work, the haptic environments of greatest interest are manual control devices with usually one or occasionally two degrees of freedom; thus those types of environments will be most often used for examples.

One mechanism able to create this illusion to some degree is a high bandwidth servo motor, computer controlled to reproduce the specified characteristics of certain classes of physical systems (Figure 1-2).

While this approach to creating haptic virtual environments is at present the only one in general use, it is a young technology with a plethora of performance ceilings and inconvenient tradeoffs of which the reader will be hearing much more.

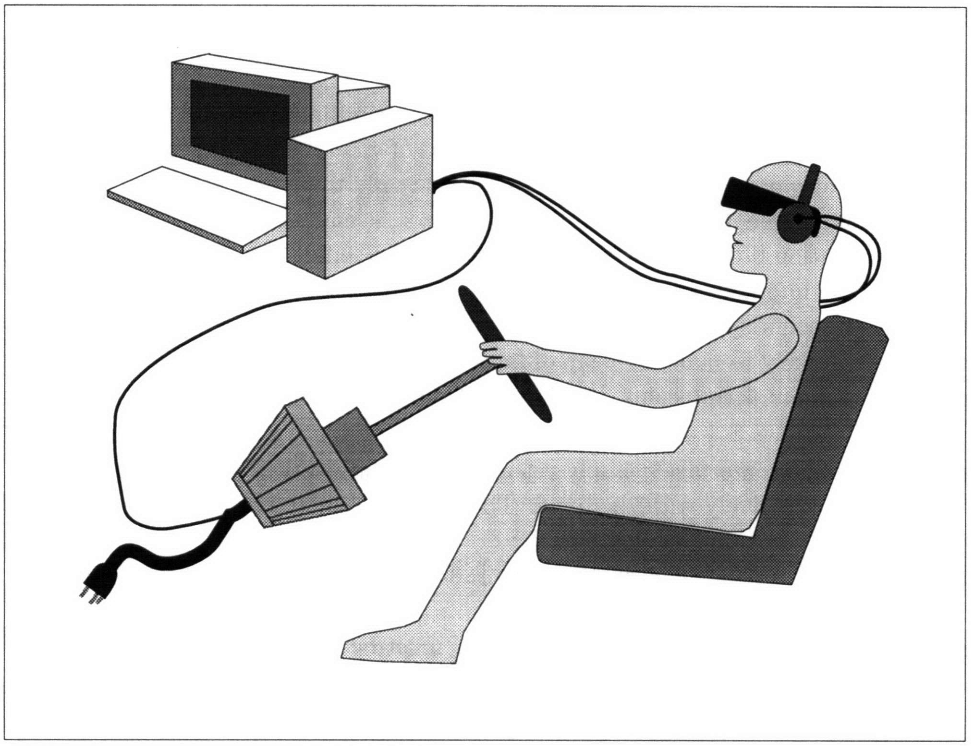


Figure 1-2: Form of haptic design tools: a manual force display which can be computer-controlled to feel like an arbitrary haptic environment. Ideally, this haptic display will be integrated with emulations of the design prototype the auditory and visual modalities.

What Must Happen First

Such tools are some distance in the future, for the prototyping of any but the simplest haptic interface. One of the purposes of this research was simply to ascertain *how* far off, because this was not previously known.

A representative sample of the key issues to be resolved along the way is listed here. Note that these issues are largely culled from the technological challenges encountered in the pursuit of furthering the single application of haptic feedback (product prototyping); someone researching a different application will certainly have a different list which overlaps this one but includes other items.

1. Construction of haptic feedback hardware:
   - What hardware specifications correspond to what sort of performance in what kinds of tasks?
   - What is the best type of controller to use, in what situations?
   - What is an effective *method* for ascertaining hardware and controller specifications in the most general manner possible?
2. Emulation of haptic environments:

   Benchmarks:

   "Realism": Will it pass as a "real" environment? This is the haptic "Turing Test".[3]

---

[3] In the field of Artificial Intelligence, a computer interface would pass the Turing Test if a real human, through

    "Fidelity": Does it feel like a *particular* real device?

    "Resolution": How subtle a difference between two real things can a given emulation system convey?

    "Functionality": setting aside "quality" considerations, what characteristics of haptic feedback best aid a user in performing a given task?

  Haptic system identification: modeling and parameterization of the types of real environments which are to be recreated virtually.

  "Tricks": ad hoc strategies for emulation of difficult (nonlinear, discontinuous) features.

3. Perception:

   - *Simple environments* (continuous and described by a small number of parameters): Does human haptic perception correspond to what we can measure in engineering terms of stiffness, damping, inertia, etc.; or does it follow another scale, and if so what is that scale, and is it always the same?
   - *Complex environments* (discontinuous and/or multiparametered): What, out of their many choices, do people notice? Does everybody notice the same thing?

     Does all the complexity need to be emulated — what can be left out and what can be merely approximated?

     When is qualitative correspondence to the real world critical, and when is it an unnecessary flourish — and is it ever actually counterproductive?
   - *Multisensory augmentation*: to what extent is the sight or sound of a real interface important in its emulation; and further, to what extent can non-haptic sensory feedback be used to substitute for absent or limited haptic representations?
   - *The "Monet" effect*: Some haptic environments will always be difficult to emulate using "photorealistic" modeling approaches, e.g. because they are unstabilizable or unmeasurable.

     In these circumstances, is it possible to create emulations by combining achievable but not necessarily physical-model-based elements which, by intricacies of the user's perceptual hardware, blend into a convincing illusion of something unrelated to its discrete elements when perceived as a whole?

4. Link to manufacturing: as soon as we wish to build something that has been emulated, we require an efficient correspondence between the emulation and the manner in which it will be physically realized. This includes:

   - *Mapping* of emulation strategies back to something real, in quantitative terms.
   - *Specification of emulation model in manufacturing terms*: the terms will be specific to the type of product and method of manufacture.
   - *Influence of prototyping method on manufacturing step*: can features of the final product be modified in order to make them more emulable, for the benefit of the overall process?
   - *Translation* of research prototype emulation systems into a form of *practical commercial value* — including both hardware and, perhaps more importantly, creation of libraries of feature emulations and model-to-manufacturing-process conversions.

---

conversation or other interaction with the interface, could not tell whether it was human or computer. Likewise, passing a "haptic Turing Test" means that a representative subject is unable to tell the difference between the real device and the emulated one; zero leap of faith is required to believe.

Clearly, anyone proposing to enter this field has a great deal to choose from.

Perhaps the hardest problem of all is that of finding a commonality in the relentlessly specific answers to these questions, solutions which will emerge as the field matures. As one first begins to dig, every situation seems to be a special case: every haptic feature needs a different modeling strategy, every parameterization problem is unique, and every emulation application will no doubt mandate a somewhat different hardware system for optimal performance.

We need a coherent structure in which to place these answers so that with time and experience, we may begin to predict them instead of having to solve each new problem on an ad hoc basis; and to easily locate the answers which have already been found by others. The research described here shows that this will not be easy.

1.4 Overview of This Thesis

1.4.1 The Question

> What are the critical components or *emulation variables* of high-fidelity haptic emulations, and how may they be systematically explored and analyzed?

And,

> When does "high fidelity" matter?

We care about high fidelity because of the application: virtual product prototyping. Only in response to a quantitative likeness to the eventual real thing can the prospective user give the designer useful feedback. How close the likeness needs to be is an unknown, and will depend on the design task.

The concept of breaking the emulation down into components is an obvious approach to the study of the complex multiparametric input-output nature of any emulation system and any emulation task; and for the same reason, a systematic method of component study is required.

"Emulation Variables" are members of an arbitrarily long list of the modeling, control and hardware characteristics which influence virtual environment performance independently or through interaction with other emulation variables (Figure 1-3).

hardware	motor inertia transmission backlash motor force saturation **physical damping** drive train stiction
virtual environment model	structure model parameterization **model expression**
servo control	**latency** gains structure (PID, "smart", etc.)

Figure 1-3: Representative list of emulation variables. The variables indicated by bold type are the ones explicitly tested in the course of this research.

...

With this starting point, a problem statement may be formulated. First we require a recognition of the size and nature of the space encompassed by these questions, and of some of the other complications inherent in their analysis:

- The critical emulation components will differ from one emulation task to another.
- The optimum values of the critical emulation components will also differ.
- There is an infinite number of possible emulation tasks.
- Because it will not be cost-effective to always build emulation systems that are high performance in every aspect (and often that will be impossible anyway), tradeoffs in emulation variables — "By using less of ___ and more of ___, can I get the same overall performance?" — will be crucial to understand.
- Performance is difficult to quantify.

Thus, we are talking here not just about conducting a simple set of experiments and trying out a few variations in the hardware and controller. We require a whole framework in which to think about this challenge, because one doesn't yet exist; what we have is a disparate, albeit rapidly growing, collection of specific bits of information. Within the new framework we need a methodology, a set of experiment techniques, with which we can gradually build up the framework.

The problem statement is to propose a framework and a methodology and then to test and critique those experimental tools. To completely answer all of the questions posed above, or even completely answer a small number of them, is a huge task. The work here describes makes a start by proposing and testing one way in which it may be done, and offers a thorough analysis of the success and failures of that method which will provide the groundwork for the next iteration.

1.4.2 The Process

Emulation Variable to Emulation Performance Mapping

The conceptual structure proposed here is a mapping between emulation variables to emulation performance (Figure 1-4). This hypothetical map may be entered from either end — hardware/control specifications defined and resultant performance predicted, or desired performance predicted so that requisite emulation variables may be specified.

Of course, for this concept to be useful it must incorporate real data reflecting the quantitatively measured influence on output by the emulation variables of interest. Further, we require a working definition of "performance"; most probably, we will require more than one, since high performance will mean different things in different situations. Finally, we need to decide which emulation variables are interesting; that is, influential.

The process of creating and filling this database comprise what I refer to as "methodology". It is based on:

1. A versatile emulation testbed.
2. The use of this testbed in a set of experiments where by their responses, human subjects indicate emulation performance as a function of emulation variable.
3. The analysis and critique of the experiment data, not only to understand the specific factors tested but also in the larger sense of determining the methodology's strengths and weaknesses.

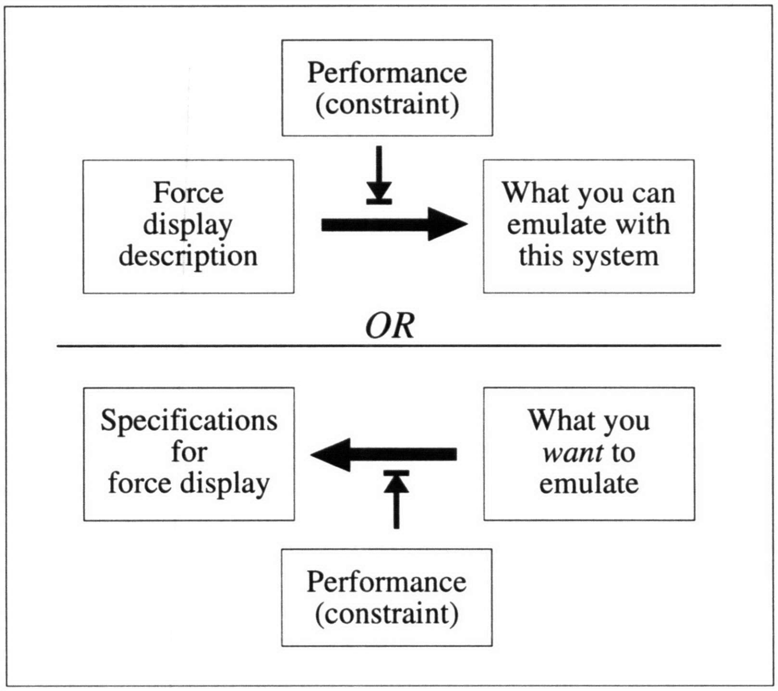


Figure 1-4: Proposed mapping between emulation variables and emulation performance.

Versatile Haptic Emulation Testbed

The target characteristics for the emulation testbed were:

- high bandwidth actuation
- high actuator force saturation
- variable physical damping
- full sensor set
- endpoint (handle) force sensing
- high digital sample rates
- action (linear or rotary) and range of motion representative of an interesting class of real devices

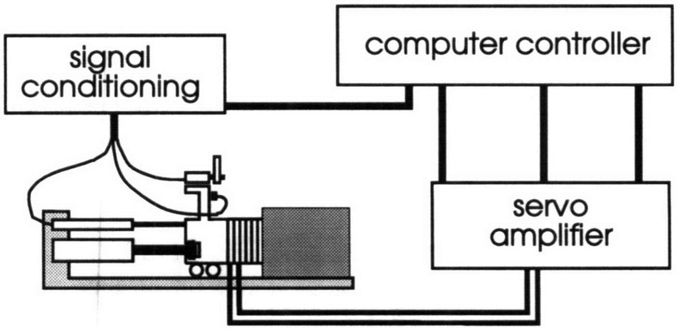


Figure 1-5: The versatile haptic emulation testbed built for this thesis.

– stiff transmission (direct drive if possible)

Most manual control devices, the potential emulation target set for this emulator, have only one degree of freedom (DOF); thus a single DOF emulator was specified. As a result other specifications, such as the stiff transmission, high output force and endpoint force sensing, were more easily met.

The testbed designed and built for these experiments is a high-performance, single degree of freedom actuator instrumented with force, position and acceleration sensors and controlled by a 60 MHz Pentium PC (servo rates of up to 10 KHz). The actuator is a linear-acting voice coil motor with relatively high force saturation (60 N), a force slewing bandwidth of about 100 Hz and an 8 cm range of motion. Power is supplied by a pulsewidth modulated servo amplifier, and physical damping can be added through a detachable external module.

The stiff, direct drive single DOF interface is ideal for the emulation of one-dimensional linear-acting manual control devices such as slider switches and detents; it can also be used to approximate the action of some limited-action rotary devices, such as toggle switches. The high force saturation allows emulation of high-impedance systems, such as walls and hard detents. The nearly full sensor set (velocity is obtained by differencing position) permits the implementation of a variety of control structures as well as closed loop (high fidelity) force and position servos.

Human-Subject Experiments

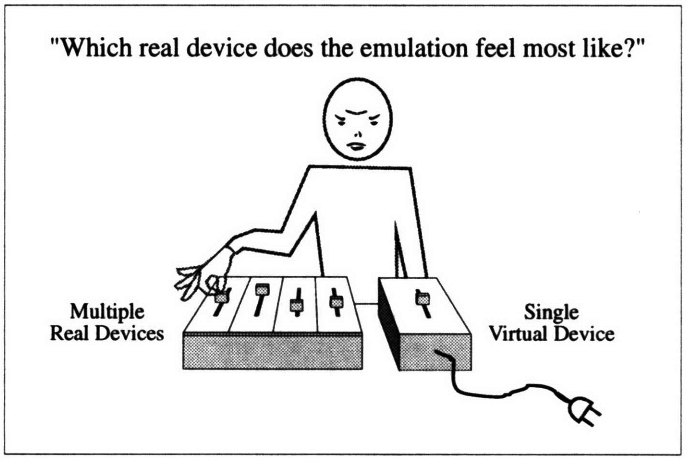


Figure 1-6: Human subject experiments were based on an experiment unit (a single trial) in which a subject compared a single emulation to a set of real devices.

The purpose of the human-subject experiments was to quantify how emulation variable modulation influenced the average subject's ability to recognize real devices through information conveyed in the emulation. The experiments were based on a unit (a single trial) which consisted of a comparison between the emulation and a set of real devices, one of which was targeted by the emulation (Figure 1-6). The real devices differed from each other only in the haptic characteristic being tested; here, this characteristic was some variant of mechanical impedance.

In an experiment trial, a subject felt the single emulation and the set of real devices and told the investigator which real device the emulation felt most like. In successive trials, the emulation variables were set at different levels according to a factorial variation. By comparing the responses given for different combination of emulation variable settings, the effect of a single variable on the subject's ability make a positive identification could be determined.

Fidelity and *resolution* in the emulation were thus demonstrated by the frequency with which subjects made "correct" choices, and by the degree of haptic variation in the real device set.

1.4.3 The Outcome

The results of building and using the emulation hardware and of carrying out and analyzing the experiments were threefold:

The Art of Emulating Haptic Things

A wealth of insight was gained into the hardware, control and modeling issues pertaining to the creation of high fidelity emulation of haptic feedback. These are documented in Chapters 4–7.

Experiment Results

The experiment design, detailed in Chapter 8, encompassed the testing of emulations of three distinct families of real devices (rolling masses, slide switches and toggle switches) on six subjects, in a total of 18 experiment sessions.

The results of these experiments are presented in detail in Chapter 9 and interpreted in the first part of Chapter 10.

The emulation variables tested were sampling rate, physical damping and virtual environment expression (impedance versus admittance). As expected, the results of the analysis were anything but cut and dry, and yielded more in the way of qualitative insights than in hard numerical data. The analysis highlights include:

- An emergent distinction between "quality" emulations versus "functional" ones.
- Importance of the role of interactions between emulation variable effects.
- Patterns of individual differences and consistencies in perception and attention.
- A study of the distortion in subjects' perceptions of impedance distributions.
- Relation of the length of time taken to make a decision to different aspects of performance.
- The minor effect of virtual environment expression, and the camouflaged effect of sampling rate.

Validation of Methodology

The entire process of emulation creation, experimentation and analysis used here is critiqued in the latter part of Chapter 10. The aspects of the methodology analyzed there include:

- Adequacy of the emulation system.
- Problems in modeling and emulation of real devices.
- Adequacy, power and appropriateness of the experiment model.
- The distinction between Quality and Function, as emulation characteristics.
- Method and value of obtaining information about subjects' perception of the real device set used in the experiments (as opposed to just the comparison responses, described above).
- Influence on subject response of the type and distribution of the real device set.
- Protocol administration.
- Usefulness of various analysis elements.

To summarize, this thesis concludes that this iteration of the experiment methodology was for the most part cleanly implemented and yielded valid results. Second, those results consisted of informative and quantitative information regarding the specific emulation variables tested. Finally, a great deal was learned about how to do things better in the next iteration.

1.5 Guide to the Casual Reader

In many places, this thesis contains a level of detail which the casual reader does not require. To understand the work without spending days at it, I suggest passing through the document with this level of care:

- Read 2.1–2.3 for orientation in the field, and 3 for description of the problem addressed and approach taken by this thesis.
- Skim 4–7 for any clarification desired in the hardware, control, virtual environment modeling and construction of the real devices used in the experiments.
- Read 8.1 for an overview of the experiment design; the remainder of 8 may be omitted.
- Read 9.1 for a summary of the experiment results, and skip the rest (but feel free to glance through the cool plots!)
- Read carefully 10 and 11, which extract and interpret the key thesis results.

Chapter 2

Background

In order to properly demonstrate the importance of the work described here, we first examine the current knowledge and practice in a series of topics which are relevant to the present research. These are:

2.1 Definition and illustration of user interfaces

2.2 Review of what industry is doing now in terms of haptic design

2.3 Rapid prototyping and the role it can (but does not yet) play in haptic design

2.4 Related research

2.1 What is a User-Machine Interface?

The purpose of this section is to categorize different kinds of user interfaces and discuss the design challenges inherent in each. We are here most interested in those interfaces in which manual feedback plays a role. Appendix A contains a list of such interfaces; many items on the list might benefit from more careful haptic design. Note that many are typically operated while primary visual attention is occupied with a higher level task.

A Classification of User Interfaces

User interfaces vary widely in their purpose, appearance and mode of interaction, as well as in the success of their design. I present this classification to place the work done here in the larger context of possible design applications, to justify my choice in focus and emphasize the scope of the design challenge.

For the purposes of this discussion, user-machine interfaces may be sorted into a few major groups which are described below. All but the first category may be classified as predominantly physical in nature in that the primary input and output modes involve direct physical contact with interface devices which obey the laws of physics, and accommodate the transfer of power between the user and the machine or device.

As with any attempt to classify a diverse population, this sorting is imperfect; there are many examples of interfaces which overlap categories or do not clearly fit into any of them. For example, according to this classification a computer keyboard considered by itself qualifies as a panel interface and a computer mouse as a hand-held tool even though the combination of screen, keyboard, mouse and software is a computer interface. Conversely, microprocessors are becoming so commonplace that a backyard grill technically qualifies as a computer. When trying to describe a telephone that might be a Garfield doll or a handheld, programmable cellular computer, these distinctions begin to lose their meanings.

Nevertheless, the classifications are useful at a more abstract level to facilitate discussion and analysis.

Graphical computer interfaces: Applications range from word processors and databases to programming and hardware control environments. These tools share an emerging trend towards icon-based interfaces which aim to render human-computer communication intuitive and transparent. Communication in such interfaces is almost invariably symbolic, accomplished via monitor, keyboard, mouse or other input devices.

Panel interfaces: Interaction may be intermittent, to adjust a setting or program a sequence (the user approaches or reaches out to the panel and then withdraws) or continuous, as in a keyboard. Examples are computer keyboards, stereos, VCRs, remote controls, light switches, car heaters and microwave ovens. Faucet handles and volume control knobs are components of panel interfaces. Complexity ranges from binary on/off switches to controls comprising hundreds of multi-function buttons or knobs. Aesthetics and the availability of a large number of functions can be important selling features in products of this type, occasionally pre-empting issues of ease-of-use in the consumer's priorities.

Hand-held tools and devices: The descriptor "hand-held" here refers to a subset of all interface devices which involve hand contact: objects which the operator grasps and sometimes supports, and which he may move freely in at least one and generally all degrees of freedom. Well-designed interfaces for hand tools such as electric drills, Walkman-style radios marketed to joggers and steering wheels approach the state of an extension of the user's own body, with the best interface "disappearing" altogether. In general, the challenge is less to clarify or simplify complex controls than to optimize highly ergonomic and often poorly defined and subjective factors such as "heft" and "grip", as well as the location and kinesthetic/tactile feedback for a small number of basic controls.

Virtual environments: An environment that might include visual, audio, kinesthetic and/or tactile components is created around the user with the aid of binocular computer graphic displays, stereo headphones, emitter-detector position sensors for locating a subject's spatial location and orientation, exoskeletal position-sensing gloves, and force-reflecting joysticks or other mechanically interactive devices. The intent of virtual environments is to provide the user with multi-sensory input modes and feedback, to enable communication with computer and hardware in direct and intuitive language. For instance, the user might indicate what he/she wants to see in the graphic display by walking around the room, turning her head or gesturing with her hand.

Virtual environment interfaces are based on ambitious but rapidly developing technology. Although they hold great promise, even the most sophisticated current implementations are still afflicted by slow update times, disorienting time delays and sub-state-of-the-art graphic images constrained by enormous realtime data processing requirements, by clumsy body-mounted hardware, and by primitive command interpretation. However, these performance ceilings are rising. A vast improvement has been observed during the period of this research.

Panel Interfaces

The research described here focuses on the problems and limitations pertaining to the design of panel interfaces. This choice arises out of multiple considerations:

1. Although desirable characteristics have been cataloged in general terms for haptic control interfaces (Section 2.2) the design process itself is still far from ideal for interfaces whose efficacy cannot be readily evaluated through a computer image or a three-dimensional mockup.
2. Panel interfaces are the most commonly encountered interfaces in today's society, and their inadequacy has become the source of an increasing discontent on the part of the consumer.
3. It is necessary to limit the research to a manageable scope.

However, the methodology developed and tested here will be useful to a much wider community than the designers of panel-type interfaces: any haptic virtual environment application will benefit, as well as the developers of robotic teleoperation systems, complex and large-scale interfaces, and custom workplaces for the disabled.

2.2 Design for Touching: Current Practice

The logical starting place for a study seeking to improve the way something is created is to examine current practice. The observations in this section have been drawn from visits and/or interviews with representatives of the following product and industrial design firms: IDEO (previously David Kelley Design, Matrix and ID2), Herbst LaZar Bell, Inc., Design Continuum, Fitch Richardson Smith, GVO Inc., and the product development divisions of several large corporate firms including Apple, Raytheon, Eastman Kodak, CIBA-Corning, Texas Instruments, Casio and Terradyne.

Who Designs User Interfaces

It is important to know who designs interfaces and to understand their objectives, constraints and expertise, because (a) these designers will be the market for the interface design tools envisioned here; and (b) there is much to be learned from them. Designers in a number of applications are concerned with making better haptic user interfaces:

- *Product designers* require cost-effective rapid prototyping technology which will let them iterate quickly to the best "feel" in an arbitrary conventional or novel manual control, at minimal development and manufacturing expense.
- Creators of advanced *teleoperated robots and virtual environments* are often willing to pay a high price to get the best possible total sensory feedback, but we don't in general know yet what that means.
- Designers of *complex, large-scale interfaces* such as may be found in commercial aircraft, chemical process and nuclear power control and even automobile dashboards need a way to systematically optimize efficiency through choice of the location and operational details of many controls competing for space and a user's attention.
- Those designing *custom* workplaces for disabled individuals and high-end products need to make the most of their client's limited and highly specific control and feedback channels, or the ability to pay for exactly what they prefer.

The details of the actual data and design tools that each will finally use differ, but all will be based on the same fundamental rule: get the most information back to the user at the lowest cost and complexity.

Evolved Practices of "Good" Interface Design

Through experience, convention and in a few cases the study of biomechanical factors, a body of experience has evolved regarding "good" and "bad" practices in specifying the ergonomics of physically controlled devices. They have been compiled and to some extent quantified as "interface design guidelines", generally taking the form of qualitative recommendations. Often they are simple common sense, in other cases matters of personal preference over which designers disagree. For the most part these guidelines have arisen through research in the military sector relating to guidance of demanding interfaces such as those in high performance aircraft.

It is beyond the scope of this document to repeat these guidelines in detail; others have already done so [7, 27, 44, 80, 130, 144, 193, 201, 227, 228, 226]. However, note that some general conventions and recommendations do exist for the design of some aspects of control device dynamics as well as for their ranges of motion, effort levels and resistance profiles, and for their size, placement and labeling. Examples are:

Intrinsic and extrinsic feedback — the manner in which information is provided to the user regarding system state.

Tracking control input mode — e.g. compensatory or pursuit tracking works better depending on the tracking task.

Controller impedance characteristics — compliance, damping and inertia is often good, depending on the situation; deadspace, backlash and static friction are usually bad.

Control/display ratio — the kinematic or force scaling between user signal and resultant action varies depending on fineness of control required.

Control/display layout and operation flow — where the control should be displaced depends on the way it will be used.

It must be noted that although these guidelines are valuable in a first cut, they should never be used thoughtlessly because (1) they are not valid in every situation; (2) even when valid, they are often not sufficiently specific, and (3) often tradeoffs must be made due to multiple controls, space and cost constraints. For example, conventions and aesthetic preferences vary with culture, time, age group, income level and a multitude of other factors [2, 44, 122, 132, 149]. Japanese disagree with Americans not only on which side of the road to drive, but also on how much leg room a car should provide, what language controls should be labeled in, and perhaps on what an attractive dashboard looks like. These differences in perspective apply to any kind of interface, including panel controls. Likewise, blind adherence to rule-of-thumb recommendations will not be likely to produce an optimal interface for either a Japanese *or* an American; nor will these recommendations stand up to complex multi-control design problems. Physical prototyping and testing with an appropriate user base are required to solve problems of this nature.

Current Practice: Myopia?

Except for a few isolated cases, little attention is given to explicitly haptic aspects of user interface design in industry. There are many reasons for this. To begin with, most consumers don't buy something explicitly because it feels better; at least, not knowingly (maybe a light switch, but probably not a car where they must worry about mileage and airbags).[1] Partly because of the indirect

---

[1] A notable exception to the claim of designers overlooking interface haptics is the recognition by auto manufacturers that customers are influenced by the feel of the door closing. The customer is probably not aware of the degree to which this feature shapes his perception of the entire vehicle. It is this sort of subconscious user concern which I feel designers would in general benefit by addressing.

nature of the customers' reaction to quality of haptic feedback, its contribution is not recognized by marketing and design.

However, there are a few things to keep in mind here which tend to make this attitude shortsighted. First, "feel" does contribute to the overall sense of product "quality", the intangible part of the difference between a Yugo and a Mercedes Benz. Second, the concept that the consumer doesn't explicitly consider feel in purchase decision might be changing. Good haptics are added value: for some items, particularly mature products such as automobiles, this trait will separate competitors which are matched in more visible features — such as airbags and mileage.

Finally, it is worthwhile to distinguish the private from the "scientific" consumer — for instance, the customers who buy airplane cockpit and astronaut training simulators. The scientific consumer will tend to be better informed about the contribution that haptics make, and that constitutes another market.

Primary Bottleneck: It's Hard. However, the main reason that hands-on characteristics are not well designed is that they are not easily prototyped. A case in point is an anecdotal justification supplied by a U.S. auto-industry executive in charge of the human interface design of a new line of luxury automobiles: he claimed that it was not cost-effective for him to design interfaces at all, since his engineers could more cheaply copy the interface innovations and refinements developed by other (i.e. Japanese) companies.

Although that company might be a tough nut to crack for other reasons, it is reasonable to believe that when haptic prototyping becomes cheap and easy, adding the value will come at a lower cost and hence become more commonplace among those entities truly interested in innovation and competitive edge.

2.3 The Rapid Prototyping Paradigm

Benefits of Rapid Prototyping

The general premise of rapid prototyping is that "in order to win, we must gain intimacy with the final thing as early in the design cycle as possible".[2] The touted benefits of rapid prototyping include aiding the designer's visualization of and interaction with products still in conceptual phases, accelerating the design cycle, enhancing the quality of and more richly interpreting market data, and improving interdisciplinary and interdepartmental communication. In each of these functions, rapid prototyping provides the service of *preperception*: insightful foresight is afforded by the ability to examine, handle and test three-dimensional models which are realistic in at least some important subset of the projected end result's attributes.

Preperception processes can highlight potential problems at any number of points along the design/manufacturing line, as well as constitute a positive step towards product improvement. Issues of manufacturability, aesthetics and human factors all too often do not surface until the product is far along in development and much of its design committed, resulting in extra costs for late redesign and retooling, increased manufacturing costs or an inferior product.

Likewise, the accuracy of market information increases with the use of realistic prototyping: test-user responses are most meaningful when there is a minimum of subjectivity and maximum of variety in what the tester is asked to react to. For instance, a firm might actively invest in the *feel* of its product's controls by building and user-testing large numbers of mechanically realistic models.

Prototyping techniques must be used with two caveats:

---

[2] Woodie Flowers, MIT

1. *The prototype must capture and convey the essence of the characteristic under examination.* To decide the best size and shape for a new computer workstation, a designer can quickly construct a few foam mockups which on close visual inspection are scarcely distinguishable from the real thing. A foam mockup would be useless, however, in optimizing the compliance and travel of the same workstation's keys, since, the "illusion of reality" disappears when finger touches foam.[3]

2. *Models used to assess subjective characteristics such as aesthetics and human factors are only as effective as the objectivity of testing applied.* If the right questions are not asked of the right user base, the best preperception processes imaginable will lead to inconclusive answers.

It is useful to turn the latter observation around and state it more generally: the realm of design issues (e.g., failure analysis, aerodynamics, interference of mechanical assemblies, visual aesthetics, and physical human interface factors) needs to be mapped directly to available and appropriate prototyping technologies (computational stress-strain modelers, wind tunnel analysis, stereolithography, foam, wood, polyurethane and computer solid-body modeling). When viewed this way, it is clear that there are no prototyping techniques available for physical interface prototyping.

Rapid Prototyping in General

A burst of new technologies make it increasingly more reasonable to imagine an array of processes that let the designer, market analyst and manufacturing engineer visualize an end product earlier. In general, models lending insight into analytical, geometric and visual questions may be constructed relatively quickly and inexpensively.

Analytical models usually utilize three-dimensional computer simulation tools which permit ever more detailed and accurate solid body modeling, material property and stress/failure analysis. For geometric models, the paired physical prototyping technologies of stereolithography and polyurethane molding augment capabilities which were recently limited to those available from numerically controlled machining, and have drastically reduced the time required to make molds for complex parts. Available for visual prototyping are either graphical computer representations (solid body models) or methods which produce physical (as opposed to graphic) prototypes which are not meant to be touched [5, 6, 209, 210].

Various media are available for the prototyping of some types of human interfaces. Much work has or is being done both to classify the ergonomics of computer interfaces and to provide tools for their prototyping [202, 206]. Computer tools are also being generated and improved upon to aid in the prototyping of some aspects of more physical interfaces: for instance, solid body modeling has in some design firms supplemented or even replaced the role of rendering and physical modeling of three dimensional products. Stereolithography, as well as more traditional techniques such as foam, clay and wood modeling, permits the quick representation of the exterior surfaces of three dimensional handheld objects although it does not convey correct material properties and/or mass distributions. Virtual environments are in a sense prototypes themselves, a developing technology that is not yet ready to perform as a prototyping medium. It is beyond the scope of this article to comment in depth on the vast effort that is currently being directed in this area.

Finally, interface layout can be modeled to a certain extent through computer simulation, since a graphical model that is addressable through mouse and keyboard theoretically contains all the information necessary to determine whether controls are arranged in a logical and understandable manner. Computer tools have been developed which specifically address this task; for example,

---

[3] Remark by Apple executive with respect to Macintosh prototypes.

Apple's Hypercard permits the two-dimensional computer simulation of certain kinds of panel interfaces.

Rapid Prototyping of Haptic Interfaces

What is not available are techniques which permit one to rapidly model things which are to be touched, particularly those which are intended to provide any degree of tactile feedback to the user; i.e., accurate and precise simulation of the feel of moving electromechanical parts. Hypercard-like utilities cannot aid the designer in deciding how big or how springy a button should be, nor how many detents a dial should have and how stiff they should be; neither can they help with layout of large three-dimensional control interfaces, such as automobile or airplane cockpits, which bring into play physiological ergonomic factors. And certainly, there exists no tool which permits modeling and rapid modification of *all* aspects simultaneously.

A danger to which panel control interface design is particularly prone is that of making needless and often unconscious assumptions about communication mediums which may detrimentally influence or constrain the final design. For example, presupposition of a lever versus a pushbutton control might influence which variables are chosen to be controlled, when in fact this choice should be made on the basis of ease-of-use factors and the two control devices might be equally easy to implement. Conversely, a poorly considered initial choice of control variables could force the interface design to be unnecessarily confusing, complex or inconvenient.

2.4 Related Research

To the best of my knowledge, there has been no academic research by other groups nor publicized industrial effort to systematically surmount the unique challenges presented by haptic user interface design.

There are, however, many technological capabilities required to achieve this goal. The following summarizes the foci of several relevant fields of research, some quite mature, upon which the haptic design project will build.

The Human Operator

Human Factors Design

Human Factors Overview Works: A number of general references for human factors design were written in the 1970's and 80's, primarily to summarize military aviation research performed in the 1960's and 70's.

Boff et al 1986 [15]
Burgess 1986 [27]
Kelley 1968 [105]
McCormick and Sanders 1982 [130]
Meister 1971 [132]
Morgan 1963 [144]
Tichauer 1978 [193]
Woodson 1981 [226]

And of course, there is the classic critique of everything that's wrong with human factors design — Norman 1988 [149].

Guidelines and Methodology for Human Factors Design: The advice contained in this sampling (taken from a huge literature pool, again generated largely in the 1960's and 70's) ranges from the general to the highly specific results of detailed studies. It includes both description of ideal characteristics in a human interface for various situations, and suggested techniques for achieving good designs for particular tasks, (e.g. the use of mockups):

Bailey 1982 [7]
Bradley 1969 [17]
Brown 1957 [20]
Buchaca 1979 [21]
Burke 1965 [28]
Burrows 1965 [29]
Chapanis 1972–75 [2, 32]
Curry 1977 [41]
Jenkins 1947 [88]
Haines 1977 [73]
Konz 1966 [111]
Luxenberg and Kuehn 1968 [122]
Maurant 1980 [129]
Teeple 1961 [191]

Then we take a large jump to the present, to find the first guidelines which relate specifically to the design of haptic *robotic* (virtual or teleoperated) interfaces —

Burdea and Zhuang 1991 [23]
Hannaford 1989 [75]
Tan, Srinivasan et al 1994 [190]

Evaluation and Analysis: From the same era comes a substantial literature on testing and evaluating human interfaces. As before, this representative list combines both methodological guidelines and the results of specific studies.

Banks and Boone 1981 [8]
Beaty 1969 [11]
Childs 1980 [33]
Gibbs 1962 [63]
Harris et al 1978 [77]
Howland and Noble5 5 [84]
Meister and Rabideau 1965 and Meister 1986 [133, 134]
Stevens 1979 [186]
Wickens 1986 [219]
Wierwille 1983 [221]

Modeling of the Human Operator

Proceeding into the later 1980's and 90's, attempts have been made to go beyond empirical testing by modeling and predicting human neuromotor and cognitive processes, to varying degrees of sophistication and purposes.

Baron and Levison 1977 [10]
Chapanis 1959 [31]
Elkind et al (ed) 1989 [48]
Fasse 1992 [53]
Frost 1972 [59]
Helson 1949 [80]
Jones and Hunter 1990 [93]
Kleinman et al 1970 [108]
Rivett 1980 [162]
Sheridan and Ferrell 1974 [174]
Siegel and Wolf 1969 [177]
Watson 1989 [212]
Witkin et al 1987–90 [223, 224]

Characterization of Cognition, Workload and the Human Sensorimotor System

There is some overlap of the work listed here with the "Human Operator" section preceding, particulary with the operator modeling part. However, this list tends to focus more on quantification of aspects of human perception, motor abilities and cognition than on either designing for or modeling them.

General Psychophysics Works: These references and overviews of the field span a generation of work.

Gescheider 1985 [62]
Keele 1986 [104]
Loomis and Lederman 1986 [120]
Welford 1976 [215]

Cognition and Workload: The following represents a mix of work oriented towards psychology, psychophysics engineering.

Briggs et al 1957 [18]
Fitts 1954 [57]
Jex and Clement 1977 [91]
Lindsay and Norman 1977 [119]
Mallamad et al 1980 [125]
Moray (ed.) 1977 [143]
Navon 1979 [147]
Price 1985 [159]
Sheridan and Stassen 1977 [175]
Tobler 1976 [194]
Verplank 1977 [205]
Wickens et al 1981 [220]

Perception and Control of Haptic Phenomena: A sampling of some recent work which is specifically oriented towards characterizations useful for the application of haptic virtual environments. In general, the studies examine perception and/or motor control of single-parametered haptic phenomenon, e.g. discrimination of pure viscosity or compliance rather than of more complex environments.

MIT Touch Lab: various combinations of Beauregard, Durlach, Jandura, Karason, Pang, Srinivasan, Tan et al 1992–95 [12, 87, 99, 153, 180, 182, 188, 189]
Fasse and Hogan 1993 [54]
Jones and Hunter, 1989–93 [92, 94, 95, 96, 97, 98]
Millman and Colgate 1995 [136, 135]

Studies of Integration Between Perception in Multiple Sensory Modes: This reference list is taken from the psychophysics literature, and is concerned primarily with how people respond to conflicts in sensory modes (haptic/visual and haptic/auditory). I have not been able to locate a great deal of work relating haptic/auditory sensing and their concomitant relation to a virtual environment's performance. I believe this gap indicates an important area of new work which bridges the gap between psychophysics and engineering.

Influence of haptic/auditory sensing on virtual environment performance:
Gupta 1995 [72]
Massimino, 1992 [128]

Other work:
Kinney and Luria 1970 [107]
Lederman 1990 [118]
Marks 1978 [126]
Over 1966 [152]
Pick, Warren and Hay 1969 [155]
Ryan 1940 [168]
Rock, Victor and Harris 1963–69 [164, 163]
Tyler 1972 [199]
Warren, Welch et al 1971–86 [211, 214, 213]
Wilcott 1973 [222]

Haptic Object Recognition: One of the key findings of this research is that "intelligibility" is not the same thing as emulation "quality". Thus it is necessary to step back from the detailed picture of what people perceive in terms of single-parametered, ideal physical quantities and pay attention instead to what, out of a more complex environment, people tend to pay attention to. The object recognition literature takes this more abstract approach to the psychophysics of of haptic perception.

Lederman, Klatsky et al 1987–95 [116, 117]
Moore, Broekhoven et al [142]
Stansfield [183]
Treisman [196, 197]

Engineering Identification of Properties of the Musculoskeletal System: Distinct but overlapping with the characterization of human sensorimotor properties is the identification of the impedance of the limbs and digits in the haptic feedback loop — both passive and modulated by voluntary or involuntary motor effort.

Dandekar 1995 [42]
Hajian and Rowe 1994 [74]
Milner and Franklin 1995 [138]
Orne and Mandke 1975 [150]
Veeger et al 1991 [203]

Design Theory and Procedures

A survey of what is happening in terms of rapid prototyping and its cousins, as opposed to haptic design tools as such.

Andriole 1989 [4]: decision making
Clausing 1994 [35]: total quality development
Fisher 1990 [55]: teleoperator input device specification
Furness 1987 [61]: designing in virtual space
Hauck and Taylor 1991 [78]: derivation of design constraints
Suh 1990 [187]: axiomatic design
Ulrich and Eppinger 1994 [200]: methodologies for product design
Wall, Ulrich and Flowers 1992 [209]: evaluation of rapid prototyping strategies
Wallace and Jakiela 1991 [210]: computer-automated design
Whitney 1990 [218]: designing the design process

Robot Control and Human-Machine Interaction

General Robotic Interaction Control

Many of the problems inherent in controlling haptic virtual environments are closely related to those in robotic control, another area where end effectors come in contact with unpredictable environments. The following gives an overview of the robotic force control literature.

Anderson and Spong 1988 [3]: hybrid robotic impedance control
Goldenberg 1988–92 [69, 70]: robot force and impedance control
Hogan 1985-88 [82, 83]: impedance control
Kazerooni, Houpt and Sheridan 1988 [103]: manipulator compliant control
Lawrence and Chapel 1989-94 [113, 114, 115]: manipulator control and design
Morrell and Salisbury 1995 [145, 146]: parallel coupled actuation for dynamic range
Newman, Glosser, Zhang 1994 [68, 148]: natural admittance control
Pratt and Williamson 1995 [157]: control of series elastic actuators
Raibert and Craig 1981 [160]: hybrid position/force control
Salisbury 1980 [171]: active stiffness control
Toumi and Gutz 1989 [195]: integral force control, impact
Volpe and Khosla 1992–93 [207, 208]: comparison of explicit force control strategies
Whitney 1977 [217]: force feedback control (early work)
Wlassich 1986 [225]: nonlinear impedance control

Large scale and Complex Human-Machine Interfaces

Most of the following work was performed in the 1960–70's, an outcome of effort expended in military and aerospace research. It covers characterization of psychomotor challenges and abilities as well as more cognitive analysis of human problem solving strategies when faced with multiplicity of controls in conjunction with high-stakes / high-stress tasks. Haptic factors are not necessarily discussed or in all cases applicable; but the types of problems are highly relevant to this research.

Christensen and Mills 1967 [34]: role of operator
Fitts 1962 [58]: functions of man in complex systems
Furness 1986 [60]: supercockpit design challenges
Henneman 1988 [81]: problem solving in dynamic environments
Jennings 1977 [89]: time sharing ability
Kemeney 1979 [106]: Three Mile Island (what went wrong)
Parsons et al 1978 [154]: nuclear power plant control rooms
Reising and Emerson 1988 [161]: cockpit controls and displays
Senders 1964 [172]: human control of multi-degree-of-freedom systems
Sheridan 1991 [173]: human supervisory control, telerobotics
Singleton et al (ed.) 1967 [178]: human operator in complex systems
Thomson 1972 [192]: multi-man-machine work areas
Veldhuyzen and Stassen 1977 [204]: human control of large ships

Robotic and Teleoperated Environments

Some relevant work performed in the closely related field of robotic control and teleoperator interfaces/interaction control.

Bejczy 1980 [13]: sensors, controls and interfaces
Burdea 1991 [24]: dextrous telerobotics
Dubowsky, Durfee et al 1991-94 [45, 46]: vehicle emulation system
Durlach, Sheridan and Ellis (ed.) 1990 [47]: human-machine interfaces
Fisher 1986 [56]: specification of input devices
Hannaford et al 1989–91 [76]: six-axis force-reflecting manipulator
Jacobsen et al 1989 [86]: high DOF dextrous master/slave
Kazerooni 1990 [100]: interaction via transfer of power
Laurin-Kovitz 1991 [112]: component design for programmable passive impedance
Salcudean 1992 [169]: magnetically levitated master and wrist
Sheridan 1992 [176]: human supervisory control

Virtual Environments

Most of the work performed and literature in virtual environments focuses on graphics and to a lesser extent sound; this work is not directly applicable to that described here. The listed references are recent surveys of the field.

Barfield and Furness 1995 [9]
Burdea and Coiffet 1994 [22]

Haptic Environments

The bulk of the work listed in this section has transpired in the last five years, during which there has been an explosion in attention to the field of haptic feedback, particularly in applications apart from force reflection in telerobotics.

Design of Haptic Emulation Systems

I will differentiate here between the systems developed for academic research and those aimed for a larger and more immediate commercial market. The distinction is fine, since some of the entities now marketing devices developed their first prototypes under the auspices of academic groups listed here or in other sections.

Academic Systems:

Adelstein and Rosen 1992 [1]: manipulandum for study of tremor
Burdea et al 1992 [25]: portable dextrous master with force feedback
Buttollo and Hannaford [30]: analysis of actuation redundancy
Gillespie 1992 [64, 65]: model-based actuation of piano keys
Howe, Kontaranis, Wellman 1995 [110, 216]: vibrotactile display
Iwata 1990 [85]: desktop virtual master
Kazerooni and Her 1994 [102]: haptic interface device
MacLean and Durfee 1995 [124]: haptic emulation testbed
Millman, Stanley and Colgate 1993 [137]: six DOF virtual environment interface
Salcudean et al 1995 [170]: performance tradeoffs

Marketed Systems:

SensAble Devices Inc. — PHANToM (Massie and Salisbury 1994 [127])
EXOS Inc. — Dextrous Hand Master position sensor, SAFiRE force display, TouchMaster slip display (Marcus, Chang, Chen, Tan, Eberman 1991-94 [50, 51, 52])
Virtual Technologies Inc. — force reflecting glove (Kramer; under development)

Haptic Display Control Theory and Technique

Servo control and virtual environment modeling/implementation strategies which are aimed specifically at the problem of haptic displays.

Colgate 1992-94 [39, 40, 38, 36]: passivity analyses
Gillespie and Cutkosky 1993 [67]: study of interaction dynamics
Gotow, Friedman and Nagurka 1989 [71]: human interpretation of kinesthetic feedback
Kazerooni 1993 [101]: human-induced instability
Love and Book 1995 [121]: contact stability, virtual walls
Rosenberg and Adelstein 1993 [167]: perceptual decomposition
Rosenberg 1993 [165]: evaluation of displays
Tsai 1995 [198]: wall stability

Applications for Haptic Virtual Environments

A sample of long-range ambitions for the field.

Brooks, OuhYoung et al 1989-90 [151, 19]: molecular docking
Burdea and Langrana 1992 [26]: overview
Ellis 1991 [49]: overview
Gillespie 1992 [65]: touchback keyboard
MacLean and Durfee 1991 [123]: virtual product prototyping
McGreevy 1991 [131]: planetary exploration
Minsky 1990-95 [140, 139]: texture synthesis
Smith 1991 [179]: design of medical devices
Srinivasan 1994 [181]: overview
Stansfield 1994 [184]: situational training

Performance Indices for Haptic Displays

There is a very small list for the quantitative evaluation of the performance of haptic displays; simply determining the proper performance indices is problematic (Hayward 1995), let alone measuring them. Of the references listed, one (Jex 1988) has been the field standard for nearly a decade and while it is a good qualitative start, it is time to proceed beyond the need to defer real understanding of other systems until we have visited the laboratory and held hands with them.

Hayward and Astley 1995 [79]
Jex 1988 [90]
Rosenberg 1995 [166]

Chapter 3

Approach

In this chapter I introduce the background material which is specific and original to this project, and which clarifies the motivation for the research.

The structure of this chapter is as follows:

- **3.1** Description of the concept of **Virtual Product Prototyping**, and the implications of extending to haptic prototypes
- **3.2 Defining the problems** inherent in emulating haptic feedback
- **3.3 Objectives and specifications** for this research.

3.1 Virtual Product Prototyping

Virtual prototyping is nothing more than rapid prototyping with a prototype which exists only in "virtual" form, generally a computer model. For the case of visual prototypes, the model might be displayed to the designer through an image on a screen or in a stereoscopic viewer. This is in contrast to more traditional rapid prototyping methods where the prototype is still physically constructed in some degree of detail and accuracy — for example, a stereolithographic model which exhibits correct 3D geometry but does not share material properties or surface finish with the production version.

Needed: Haptic Virtual Product Prototyping

What designers do not yet have is a way to easily model and display the haptic feedback they might like to design into their product. Thus, in many cases the haptic feedback is not deliberately designed, rather it is left wherever the other design specifications leave it.

In the case of interfaces where haptic feedback is important, virtual prototyping may be the only way to model and display the pre-production interface because iterating on physical, feel-accurate prototypes would be prohibitively slow and expensive for most applications. However, emulation of haptic feedback in isolation will rarely be adequate. An interface's sound, for example, is often closely integrated perceptually with its feel (consider a snap-action switch, or almost any kind of detent) and overlooking this fact in the design process might be invitation to failure. Further, it is not enough to simply consider the various interface aspects at one point or another: due to perceptual sensory integration, they must be prototyped together.

All of these factors demonstrate a need to extend existing technologies of computer-aided design and concurrent engineering to simultaneously prototype *all* relevant interface aspects.

The design tool specifications which would lead to virtual product prototyping (VPP) functionality for manual interfaces include the ability to:

1. *Iterate* on haptic interface characteristics
2. *Test* haptic interfaces with users
3. *Integrate* emulations of virtual feel, sound and appearance.

Potential Realizations of Haptic Virtual Product Prototyping

There is a multitude of directions which the basic concept of haptic VPP could take. The variety occurs at many levels, including the physical form taken by the prototyping system, overall versatility of the design tool for use in different applications and optimality for use in a specific application, and its performance in a host of different indices such as emulation realism and fidelity, workspace size, output force levels, and kinematic correspondence of virtual to end product motion.

Many, if not all, of the virtual haptic interface devices already built and certainly those which are produced commercially are either oriented towards applications other than VPP or intended as a non-specific interface device. Their characteristics will not always be optimal for product prototyping. The ideal VPP haptic interface device permits high-fidelity emulation of subtle manual interface characteristics, generally fixed in space and often high-impedance.

Described below are a number of VPP concepts which lie at various points along the line between systems which could be built today and those which require substantial further development prior to realization.

"Blue Sky":

The inspiring beginning of any list of concepts should always be the Dream Machine: the blue-sky Cadillac version which we aren't going to be seeing on the market for a while. What you really would like here is a fully immersive, all-sensory virtual reality in which every aspect of every feature of the design environment is under virtual control: appearance, location, feel (in terms of both force and tactile feedback), sound and perhaps even smell. This emulator would pass the Turing Test (be indistinguishable from a similar real interface), or come close to it; and if you could tell it from the real thing, the difference would not be distracting. Most importantly, anything in the environment could be changed at the touch of a key or the pointing of a finger.

In the visual and sometimes audio realms, this sort of functionality is already being approached. The primary limitation is in environment complexity, and that ceiling goes up every day as computers get faster. For simple environments, we are already there.

With the specification of haptic feedback, however, the story changes. The challenges embodied in providing an arbitrary force display with an an arbitrary handle at an arbitrary point in the reach-space at an arbitrary moment, and then doing it again somewhere and somehow else an arbitrary moment later, are clearly substantial and I do not have a proposal for how all of these criteria might be met at once at the level of performance required.

Even aside from technical barriers, a Blue Sky invariably has a cloud in it. As is often the case, the big cloud in this one is cost; as much as they might like to have one, most commercial design interests won't want to pay for it. Another cloud is the question of whether a highly versatile and immersive system such as this is the most effective solution for all applications, or even for most of them. It is likely that custom emulators for specific families of design applications will often be the best solution; see the Steering Wheel example below.

Hand-Mounted Force Display:

The hand-mounted force display is a partial solution to the ubiquitous force display challenge stated above, versions of which are close to market by at least two commercial ventures (EXOS Inc., Burlington, MA; and Virtual Technologies Inc., Palo Alto, CA).

While details of these particular devices are not yet publicly available, the attractiveness of the concept is obvious: instead of trying to put a force display everywhere you want one, put a single display on your hand and make it display force according to where you are and what you are doing.

The technical challenges in this concept are also obvious. What does the display push against? Can it be made light enough to feel natural and not impede the user's motion? Will it be able to produce any but the smallest forces? What kind of range of motion will it have, and how will that range correspond to the user's hand mobility? What will it feel like?

It is important to keep in mind that many of these hurdles apply to the VPP application but not necessarily to others, and listing them is not in any way meant to downplay the potential value of these devices. In fact, I think they will prove enormously valuable and wish one had existed four years ago.

Ground-Mounted Force Display:

Variants on the ground-mounted force display concept are the most commonly encountered in the virtual haptic feedback field, partly because in some ways they resemble a robot and more is known about building and controlling robots than about most of the other ideas here; and partly because the concept is generally simpler and more feasible in its scope. This is the concept used for this research.

Interaction with this virtual interface is similar to that for the real one. No unusual hardware is attached to the user, and ideally the part of the display which is touched resembles in appearance and texture the real device it targets. The force display is grounded to the world frame of reference at some point, and its end effector either moves itself or is moved around by the user.

There the general part of the classification ends. The most obvious difference between different incarnations of this force display is the number of degrees of freedom, which in turn limits the sorts of interfaces which may be emulated with that system. Examples presently in existence range in DOF complexity from the Phantom [127], with three actuated/sensed and three unactuated/unsensed DOFs to the single DOF device developed for this research.

...

A low DOF device is less versatile than a higher DOF display — for example, a two DOF planar motion device could emulate both rotary and linear-acting single DOF manual controls, while a single DOF device can do one or the other. Unless incorporated serially into another mechanism for translation and/or rotation, the low DOF device will be fixed in space whereas a Phantom-like device may permit more variation in workspace.

Conversely, a single DOF device will generally do the best job of emulating single DOF interfaces, all other factors held constant, because of increased stiffness and simplified control.

Mixed Virtual/Physical Prototypes — Actuated "Steering Wheel" Example:

Design of power steering wheel and gearshift haptic behavior is an example of a useful single-purpose, single-DOF design device. The physical part conveys the correct geometric and tactile information, and all the virtual part need do is provide the right force feedback via sensing and actuation of the physical prototype.

This is an important approach for the commercial design industry. It will be most valuable in large-scale production scenarios, and/or for high-end products, such as automobiles: high-end both because haptic feedback will be most important there, and because the final product cost bracket will justify the expense of a prototyping tool built for just that aspect of the interface.

In the case of steering wheels and gear shifts, not only does the feel matter enormously in terms of both safety and sense of product "quality", limiting the iteration on haptic feedback to one or two fixed and low DOF locations in the workspace makes the location problem trivial. The cost and technical difficulty of implementing an actuated steering wheel is minor compared to other development costs, and different auto models are similar enough that the same system could be used repeatedly — just put a new steering wheel "handle" onto the actuator shaft. The primary challenge remaining in this concept is the modeling and control of the force feedback to be provided through the steering wheel; and while this may not be trivial, it is certainly feasible.

The attractiveness of this example is the ease with which it can be upgraded to 99% of the Blue Sky. Some auto designers already use various forms and degrees of immersive visual virtual prototyping environments to simulate a new automobile's interior: you sit on a regular car seat in a regular car chassis while wearing a head mounted display. The details of the interior (such as location of controls, styling, etc.) are what you see, and physical objects are what you feel.[1] The addition of an actuated steering wheel and a simulation of an open road completes the environment (Figure 1-2).

The point here is that every part of the prototype does not have to be virtual; in fact the most valuable applications will probably be like this, where virtual reality is used to augment and iterate the fine detail on the physically constructed gross structure.

Vendor-Specialized Tools: Switch Manufacturer Example:

The situation addressed is one where a design/manufacturing entity does not consider haptics a critical selling feature in product, but recognizes that it does add value and would be glad to upgrade if the cost were small. Further, they tend to purchase haptic interface components from third-party vendors rather than design them in-house. Thus, it makes more sense for the *vendor* to make the investment in haptic design tools, and amortize their cost over many customers in return for improved service.

This example has been discussed with manufacturers of medical instrumentation equipment. The scenario: consider the rubber switches used in many industrial and medical-instrumentation applications. Currently, a switch salesperson travels to the makers of the instruments with a briefcase full of switches. The customer tries them out and from the collection the salesperson has in the briefcase that day, picks the ones closest to what he/she wants.

A rubber switch's feel is defined by a few physical parameters — sidewall thickness, height, bubble curvature, and so forth. A switch could be emulated using a model constructed from these parameters. While sitting in the customer's office, the switch salesperson could iterate

---

[1] Although no references are available, work in the area of visual immersive auto interior prototyping is being performed by Nissan (through Cambridge Basic Research, Cambridge, MA) and by University of Michigan's Ocean Engineering Department in collaboration with BMW.

to exactly the switch the customer wants using his lightweight portable emulator. By the time the salesman leaves the office, the switch is ready to manufacture *because the haptics have been specified in terms of manufacturing process.* This link between emulation and physical embodiment is crucial to maintain in any type of virtual prototyping tools.

Serial Displays:

A single force display is used to prototype an interface with multiple controls, by mounting the display on a ground-mounted robot arm which then moves around in the workspace to meet the user's hand, wherever it might be.

There several variations on this concept. One is the sophisticated robot-design strategy taken by Howe [216] and Morrell [145], where a low-bandwidth and a high-bandwidth actuator are connected serially with the former proximal to ground mount and latter in contact with the user's hand; the purpose is to achieve display of both low (translational) and high (e.g. discontinuities) bandwidth features of the environment with one system.

Another version is that taken by Boeing [109] to prototype the location of controls rather than their haptic characteristics. However, a two-DOF display replacing the array of mechanical controls and used with a faster robot arm would be a feasible upgrade to that system.

Other Applications of this Research

Product prototyping is not the only field which can benefit from the characterization, mapping and general improvement in haptic feedback control which will result from the strategies described and tested in this research.

Virtual Environments:

In this application, no real device will ever be built from the virtual prototype; the actuated version is the final product because "smart" output is required. However, what that output should be still needs to optimized, and we still need to learn how to produce it. Examples are:

- Custom devices for the disabled (or very rich):
  There is a need for interfaces which meet the highly specific requirements of disabled individuals, or for luxury products for those who can afford to pay for exactly what they like. In the future, the price might not be as high as it is now.
- Flight simulators:
  the benefit of simulators is well justified elsewhere, but most of the work in that direction has been oriented towards visual displays.
- Manual tool use training:
  for example, for astronauts who must handle tools in an unfamiliar environment on a costly expedition.[2]

Teleoperated Environments:

The force-reflecting teleoperation field is at least ten years older than that of building haptic displays for virtual environments, but the two are siblings in the larger family of systems which rely on haptic feedback and information and techniques can be to some extent exchanged. The significant element by which the applications differ is in the source of the desired haptic

---

[2] Work in this area currently being performed by NASA with Bowen-Loftin of University of Houston and Colgate of Northwestern [137].

feedback: sensed by a real slave environment, versus created by a virtual model. Thus the model discussions in the following work are not relevant to teleoperation, but many of the performance and servo control issues are.

The Critical Path

Following is a discussion of some of the most crucial assumptions, limitations, technological bottlenecks and dependencies which have been made or will be faced in achieving the goal of haptic virtual prototyping tools. Some are applicable only to particular implementations of the VPP concept, others are completely general.

Model-to-Manufacture:

Perhaps the single most critical, and certainly the most general, make-or-break aspect of the VPP concept is the ease and degree of automation of the model-to-manufacture link. A prototype which nobody knows how turn into an end-product is not of great value, regardless of how nice it feels.

It is likely that the best way to handle this complication is explicitly: the emulations, regardless of the platform from which they are displayed, will have to be either specified in terms of the manufacturing process or easily translated to those terms. Cases where non-physically-based modeling and emulation strategies were employed in order to create a passive and/or high fidelity emulation will have to be accurately and quantitatively mapped back into the physical domain.

The disadvantage of the explicit manufacturing link is loss of versatility; emulation software and strategies will have to be oriented towards a specific manufacturing process rather than being designed "device independent" and then "linked" to a manufacturing translation in the same way that ANSII C code can be compiled by the ANSII C compiler on any operating system. This problem might be mitigated by the development of emulation "libraries" which would contain reusable modules for specific manufacturing processes.

Creating Emulation Models:

A product designer (in general not a haptics expert, and in general someone in a hurry) is the one who will have to create and/or modify specific emulations for his or her specific design problem. Thus creating emulations will have to be easy, but making it easy will be difficult.

This could be viewed as the inverse of the model-to-manufacture constraint, and often it will conflict: instead of requiring a correspondence between the emulation and the manufacturing process, we need a good intuitive correspondence between the emulation and the real world.

I anticipate the problem being answered in a similar manner as the model-to-manufacture case, with the gradual creation and refinement of large emulation libraries. Most emulations will be assembled out of small, pre-defined elements which can be arbitrarily interconnected and scaled to meet the geometric and force requirements of the specific application. Given such modules, all that would remain to develop is a well-designed graphical user interface.

Sensory Integration of Design Tools:

Haptic perception is closely linked to other senses; if you doubt this, try plugging your ears (really plugging them, so you can't hear anything at all) and then flicking a switch that normally makes a sound when actuated. If the switch feels exactly the same as it did when your ears were unplugged, your hearing probably wasn't very good to begin with.[3]

Sound will almost invariably be essential to the haptic prototyping process, partly because sound is part of the haptic experience in manipulating most real manual controls, and partly because I anticipate that sound will be a useful way of improving the perceived quality of marginal fidelity pure-haptic emulations. Vision will also often be important for the success of the overall design task; although in general the situations for which optimality of haptic feedback is important are those where one doesn't look at the control, and thus visual prototyping is unimportant. Therefore, the haptic VPP tools will not be complete until this sensory integration has taken place.

Multi-DOF and Multi-Control Interfaces:

The simplest and most immediately achievable versions of haptic VPP tools will apply scenarios such as the Actuated Steering Wheel, involving low DOF interfaces which are fixed in space. The extension to higher interface DOFS and/or multiple controls will vastly complicate the situation, as exploration of the VPP realization scenarios above illustrated.

3.2 Defining the Problems in Virtual Haptic Feedback

In Chapter 1 I spoke of the need to create a structure in which to think about the difficulties faced in the synthesizing of high performance haptic feedback. In its crudest form, this structure will be nothing more than a place to file information as it is accumulated; a database accessed by anyone with something to add and anyone who needs what it already contains to carry our knowledge a step further. This is essentially what already exists in the form of the technical literature and the occasional summary review.

Another Blue Sky: The Haptic Display "Matchmaker"

I would like to take the database concept beyond this point, and create the foundation for an information structure which operates on the data it contains. Imagine a black box (the "Matchmaker" illustrated in Figure 3-1) which represents a single database of information, but provides an output based on the nature of the input query.

For example, suppose you have a force feedback environment description and a force feedback task description, and would like to know how well this particular system will perform in this task. The "Matchmaker" will respond to input of system specifications and task descriptions with an evaluation of the predicted performance.

Likewise, you could use this method to generate a list of hardware and control specifications for a system you are about to build or purchase based on your definition of what, and how well, you want to do; or vice versa, to generate a list of what you can do with a proposed system and desired performance constraint. In even more sophisticated scenarios, you could specify *partial* specification lists in any of System, Task or Performance domains and have that list automatically completed for you based on the specifications in other domains.

---

[3] Alternatively, try listening while someone else flicks the switch — does it *sound* the same as it did when you were holding it? Some think that you can actually "hear" some of the high-frequency vibrations via transmission through your bones, rather than through the air to your ears. But this (the emulation of very high frequency haptic feedback) is a different problem.

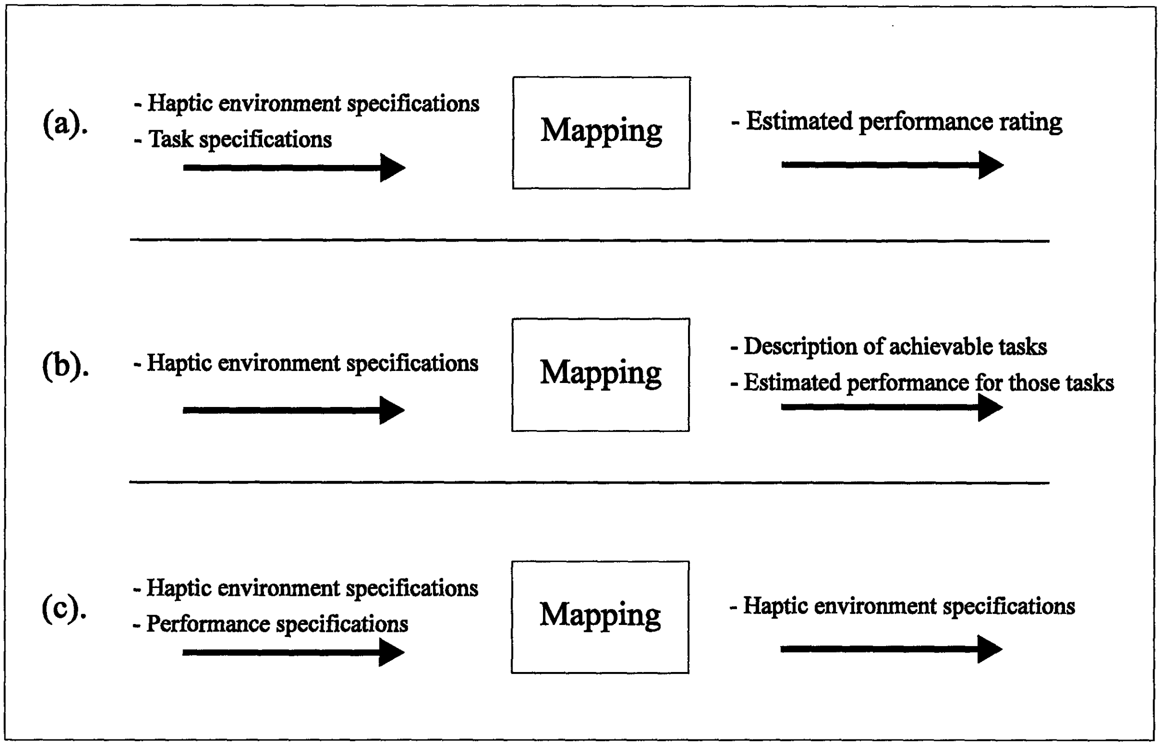


Figure 3-1: The Haptic Environment Matchmaker: a tool for virtual environment evaluation and specification.

A Beginning

Of course, a great deal of data would have to be collected and analysis performed before something like the MatchMaker could become a useful tool, and the database would change with technological advances.

Regardless of these mundane considerations, the idea of an integrated and "smart" organization of the collective information existing in the field at any time is a valuable one. This is in part because interactions of different system variables often affect performance in ways that could not be predicted based on the observation of those factors acting alone.

To this end, some definitions and organizational formats which facilitate both the logging of data and visualization of the effect of the complex influence of input on output are proposed below.

3.2.1 Performance Definitions

The performance of any virtual haptic environment is difficult to quantify. Candidates for performance characteristics abound, and which subset of all possible measures are applicable depends on the task description. The real problem with most performance measures presently used to evaluate haptic displays, however, is that they are more qualitative descriptions than metrics and thus difficult to evaluate. In general the bottom line on performance in any sort of virtual environment is a perceptual one (for the haptic case, "Does it feel right?"), and not only is this line difficult to quantify, it is likely to be in a somewhat different place for each person.

Defined here are two such fuzzy metrics which, although as hard or harder to measure than any other, will later emerge as crucial yet non-obvious characterizations of virtual haptic environment systems. I have attempted to marginally reduce the fuzziness by breaking each larger category down further into a set of benchmarks.

"Quality" Benchmarks

The *quality* of a virtual haptic interface (or of any virtual environment) is an ephemeral quantity.[4]

By illustration, consider the descriptive terms which come to mind when you think about a Cadillac. "Stately", "regal", "substantial", perhaps, rather than "heavy" or "boat-like" (the people who buy them probably do, anyway).

Now think of a fun little sports car like the MGB that my brother used to have. Driving it is quite a different experience from driving a Caddy: you can squeal around corners, let the wind blow through your hair and make a statement about your personal freedom.

Finally, imagine a ... Pinto station wagon (I used to drive one of these). It handled a lot better than a Cadillac, and its maintenance record put that of my brother's MGB to shame (in fact, it may have given the Caddy's reliability a run for the money). It carried more passengers than either, and for extra credit let me carry six bicycles on the roof. I challenge you to carry even a single bicycle on an MGB, or to find a Cadillac owner who will allow you to touch his roof. So why do people make all these jokes about Pintos? And they don't *always* blow up when rear-ended.

A Pinto is (was) a "low quality" vehicle. It is such an extreme case of low quality that no one will argue the point, even a loyal ex-owner. It isn't pretty, everything rattles, the seats are uncomfortable and you can't see the speedometer because it is blocked by the steering wheel. But you could make many of the same statements about the MGB, and in that case they were part of the charm.

The point is that the observer brings along a load of preconceptions to an evaluation of Quality. It will never be easy to measure, and the metric will be context-specific and often culturally based. Yet, it is important: people pay for Quality and they sometimes work better and feel happier under Quality conditions. So the vagueness doesn't relieve us of the necessity of trying to define and characterize it. The best we can do at this point is to break the concept down into its components and make sure we know what we are saying when we label something as of high or low quality.

For any virtual haptic environment, a partial list of the elements of a perception of "quality" are the degree to which the environment conveys a sense of:

- Realism
- Passiveness (absence of jitter, instabilities and other types of nonconservative behavior)
- Qualitative resemblance of features and textures of emulation to something the user can relate to in the real world
- Resemblance to the user's concept of what the emulation is supposed to represent
- Individual aesthetic preferences
- Cultural conditioning (what your socioeconomic background tells you is good)

Most of these elements of quality are subjective, and can only be measured through testing with interface users. Passivity comes the closest to being mechanically measurable; but almost any system attempting any but the simplest of environments will exhibit some degree of active behavior, so then the question becomes "How active can it be to still feel high quality?" and that one again requires human reactions to gauge.

---

[4] For a treatise on the subject of trying to define "quality" in general, consider Robert Pirsig in "Zen and the Art of Motorcycle Maintenance", [156].

“Functionality” Benchmarks

Functionality is a task-centered metric: a virtual environment is functional if it permits or aids the user in doing the task at hand. Sometimes qualitative resemblance to a real environment, passiveness or simply “niceness” of feel (or even how good the device looks) will improve functionality. Other times it will not; the experiment results for this research demonstrate the remarkable extent to which quality of environment can be irrelevant to task performance.

Some constituent elements of the Functionality benchmark include the following:

- Recognizability of impedances — e.g. can the emulation be recognized as a mass, and of a given magnitude, through all the defects of a low *quality* emulation?
- Range of achievable impedances
- Resolution of achievable impedances
- Fidelity to the real-world emulation target
- Resolution: How subtle a difference between two real things can a given emulation system convey?

The Functionality components listed are by nature much more easily evaluated than those for Quality; they are engineering rather than perceptual characteristics. But beware — informative and pragmatic as a Functionality-oriented design philosophy might be, exclusive reliance may well result in the haptic equivalent of the Pinto station wagon. This might be practical but probably not very satisfying.

A Gap Remains: Haptic Fidelity to the Real World

Most haptic virtual environment research to date has focused on Quality measures of environment performance, and there has been very little which has considered the specific problem of relating synthesized haptic feedback to existing real-world targets. A crucial gap thus exists in these areas:

1. Ability to create not only realistic but also ***high-fidelity*** haptic emulations.
2. Detailed, structured knowledge of how emulation variables influence emulation fidelity and realism — big-picture descriptors; more specifically, stability (passivity) and accuracy of model realization.

3.2.2 Emulation Variables

The term “emulation variable” was loosely defined in Chapter 1; we need a more rigorous definition now.

Emulation Variable : The quantitative or qualitative value (including presence/absence) of any of the arbitrarily long list of the various hardware and control elements that comprise the virtual environment.

A distinction is made here between emulation variables such as those listed in Table 3.1 and the actual magnitudes of the model parameters for the virtual environment used in a given emulation. To clarify this point, observe that while the value of “B” used to represent the dissipative characteristics of a slide switch is a model variable *parameterization*, the precision and accuracy with which this desired value for B is identified and reproduced — the emulation’s tolerance — are emulation variables. In the same way, the decision to use a dissipative model to represent a slider switch, to use viscous damping in place of or in addition to dry friction, and to compute model output based on input from a velocity sensor versus a differentiated position signal are all emulation variables.

controller update rate	actuator bandwidth
sensor noise level	force saturation
sensor set	sensor resolution
A/D resolution	controller type
controller gain set	virtual environment expression
physical damping level	accuracy of VE model parameterization
form of VE model	transmission backlash
shape of end effector handle	audible noise level of actuation

Table 3.1: Emulation variables (incomplete list)

3.3 This Research

3.3.1 Objectives of Research

The objectives for this research were designed to fill up the technological shortfall described in Chapter 2.

Identifying Components of Fidelity in Haptic Emulation: define and begin to characterize a haptic emulation variable/emulation performance mapping.

I refer here to the "smart" database described in Section 3.2, a proposal to systematize the knowledge we have regarding creation and performance of haptic virtual environments into an input/output structure with predictive capabilities. The relations in this database are not guaranteed to be simple or easily deduced and the set of emulation variables which might productively be considered is arbitrarily large. While it is far beyond the scope of this work to exhaustively identify and characterize all mappings, I here create an initial structure based on a few variables, validate it experimentally and demonstrate potential utility from this preliminary form.

Haptic Virtual Environment Testbed: design and develop a versatile haptic emulation testbed.

The hardware and control for this experimentation comprise a testbed, and must therefore be of higher performance and more versatile in capability than most application hardware used in the capacity of specific haptic emulation. I wished to span the largest possible emulation variable space and thus investigate a variety of controllers and haptic phenomenon.

Library of Haptic Emulations: create library of emulations of important haptic phenomena.

Among those who create and use haptic displays there is collective knowledge of how to reproduce a few haptic phenomenon to varying degrees of fidelity; but this data is sketchy and widely scattered. In addition, the techniques almost invariably aim to produce generic, rather than specific, characteristics: i.e. any spring, rather than that spring; any wall or detent, not a particular wall or detent. Consequently they are often dimensionally inadequate for reproducing real-world illusions — the particular real device might exhibit a nonlinearity or higher order effect not incorporated in the generic model.

This work begins to develop and categorize practical knowledge for how to emulate a variety of commonly encountered linear and nonlinear haptic phenomenon of varying degrees of complexity, to a specified accuracy and precision with respect to real-world targets, and in the language of necessary hardware and successful control strategies. The library's contents will be the object of the human subject experiments described below.

Techniques for Identifying Impedances of Real Devices: develop means of accurately identifying impedances of real devices, for the purposes of emulating them haptically.

From the experimentally driven objective of creating haptic emulations to compare with specific physical devices emerges the need to accurately and efficiently model and parameterize the target of emulation — i.e., haptic property system identification. Since the devices are often complicated combinations of many haptic phenomena, generating the system model and in particular parameterizing it can be difficult. Therefore a secondary objective to emulating real haptic environments is the converse: creating a streamlined and to the greatest extent possible, automated, methodology for generating haptic models from arbitrary real devices.

Thus I needed to take a "haptic photograph" of a real device — that is, automatically characterize those aspects which must be known in order to haptically emulate it — and then turn around and "play back" the snapshot. The modeling and parameterization step constitutes an obviously critical family of variables in the emulation space, of which an automatic procedure is one set of values; in other words, the quality of the final picture depends on the quality of both the camera and the reproduction process.

Human Subject Experimentation: design and execute human-subject experiments to quantify impact of emulation variables on emulation fidelity.

I designed human subject experiments to quantify aspects of the database referred to above. The experiment design was made challenging in part by the need to extract objective results (e.g. what is the best algorithm for creating an illusion of stiction, and what are the best variable values?) from a subjective response: an individual's perception of whether the emulation succeeds for him or her. I expected that individuals would respond differently to different stimuli (the stimuli in this case being the levels of the various emulation variables); for instance, one subject may be exceptionally sensitive to accuracy of force reproduction, while another might find discretization jitter more distracting than most people do. Experiments were designed to accommodate this effect.

3.3.2 Human Subject Experiments: Overview of Experiment Paradigm

An experiment methodology was developed and tested for the purpose of characterizing the influence of a set of emulation variables on the performance of an emulation of a haptic environment.

Comparisons Between Emulated and Real Devices

Emulations of real manual interfaces were composed employing the hardware/control testbed. For a given type of real device under study there was a set (4-6) of real devices which were visually and tactilely identical, but which had perceptually distinguishable force feedback; e.g., slider switches with the same stroke length and identical handles but which presented a different impedance to the operator. The emulation was targeted to resemble one of the real devices at a time.

A single experimental trial consisted of asking a human subject, after she had interacted with the emulation and the set of similar real devices, to state which of the real devices the emulation felt most like.

Elucidation of Emulation Fidelity

When a subject consistently chooses the real device targeted by the emulation, then the emulation tolerance in that configuration of emulation variables is shown to be within the variation inherent in the set of real devices it is being compared to; consistently erroneous choices indicate that the emulation tolerance has exceeded the real variation.

In the course of an experimental run, emulation variable settings and the targeted real device were systematically and randomly varied and the subject was queried again. Data from multiple subjects, targeted haptic phenomenon and emulation variables were compiled and analyzed to assess quantitatively how the emulation variables studied influence the emulation accuracy and/or realism.

Usefulness of Results

This knowledge can be used to determine the most cost-effective means of constructing an emulation system of a specified level of performance, where performance is defined as the subtlety of haptic distinction that a human operator can make based on the emulation — as would be useful in the case of a haptic interface being utilized as a manual interface design tool.

Specifications for Experiment Setup

This experiment design was the source of the hardware and control specifications for the experimental testbed described here. In short, its successful execution required hardware and control which permitted the critical emulation variables to be modulated over a useful range; and in its highest performance configuration, the emulation allowed a typical human subject to distinguish the degree of haptic variety present in real manual controls.

Part II

The Environments

Chapter 4

Hardware and Control

4.1 Hardware

4.1.1 Requirements

The goal of emulating single DOF haptic interfaces and the experiment design produced the following needs for the emulation system:

1. Single degree of freedom actuation and kinematics, either linear or rotational.
2. High maximum peak and sustained forces, to permit emulation of stiff environments.
3. High bandwidth in terms of force and position slew rates. This requirement is often in direct conflict with (2), since a high-strength actuator may be heavy. Force slew rates are less effected by high mass in the actuator than are position rates.
4. Stiff linkage between actuator and handle, for better control over endpoint force and/or position with a simple controller.
5. Minimal transmission effects; direct drive if possible.
6. Full sensor set, for implementation of different control and modeling strategies.
7. High power and inaudible servo amplification.

4.1.2 Realization

The haptic emulation testbed hardware pictured in Figure 4-1 and diagrammed in Figure 4-2 consists of a voice coil actuator instrumented with position, force and acceleration transducers and powered by a pulse-width-modulated servo amplifier. It is interfaced with a haptic environment controller implemented on a 60 MHz Pentium PC, permitting 5-20 kHz sampling depending on the complexity of the environment controller. A low-friction damper of controllable resistance is installed in series with the motor. Human interaction with the system occurs through a handle mounted on the motor in series with the force transducer. During experiments, the handle region was covered with a handrest plate, and the entire setup was covered with a fabric cover which both protected it from dust and prevented observation of its configuration and operation (Figure 4-3). The emulation system is designed with a hierarchy of safety features to ensure safe human interaction.

The hardware elements of the emulation testbed permit the modulation of emulation variables such as force saturation, physical damping, sensor resolution and electronic noise level. The linear

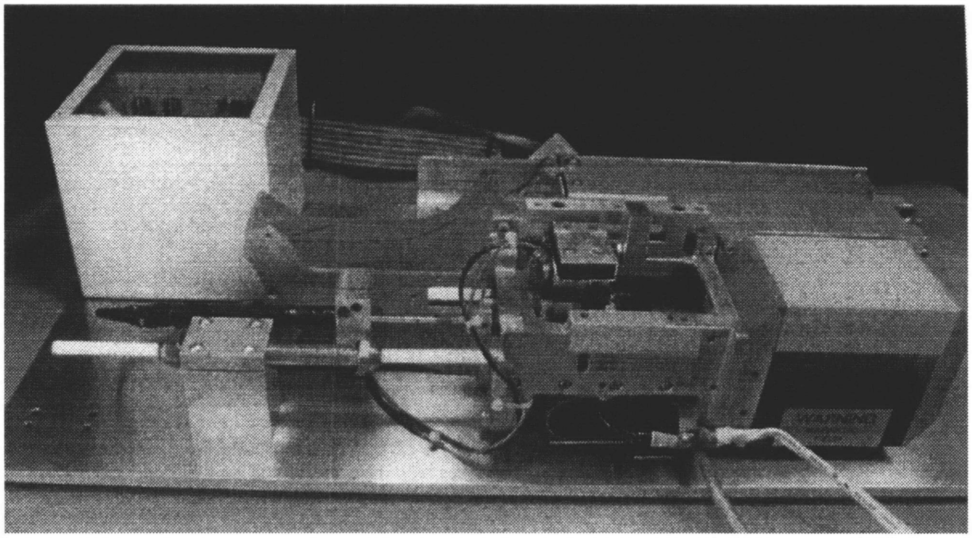


Figure 4-1: Emulation system.

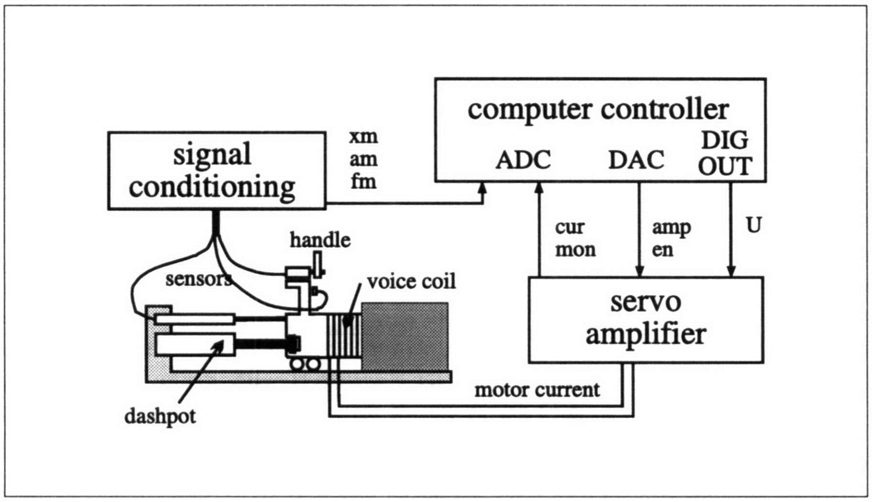


Figure 4-2: Schematic of emulator hardware.

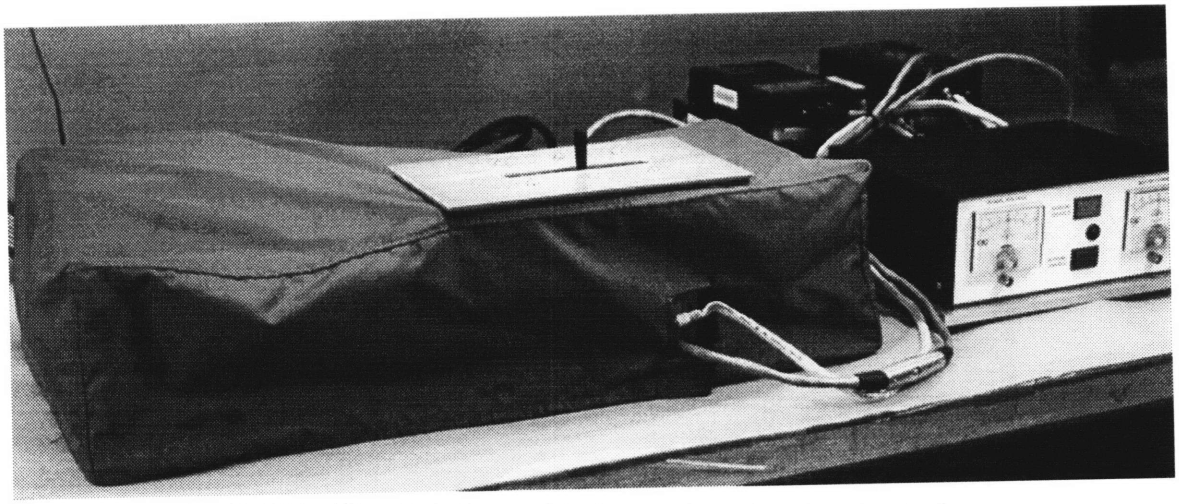


Figure 4-3: Emulation system as used in experiment, with hand plate and fabric cover.

actuation and moderate stroke length permit emulation of linear motion, one DOF devices such as slider switches, snap-action toggles and masses. With the high-speed sampling and full sensor set a variety of controllers may be tested, including impedance and admittance formulations of the haptic environment model and second-order models using both approaches.

A description of each hardware element follows, and detailed specifications and vendor information may be found in Appendix B.

Actuator

The emulation system actuation is provided by a voice coil and permanent magnet originally designed for use in a Control Data Corp. disk drive of 1970's vintage with a peak force of 60 N. The total weight of carriage, transducers and handle is approximately 0.25 kg. The 100 mm long coil, consisting of a coreless double winding of lightweight aluminum wire, is mounted on a 6-bearing, 2-rail carriage typical of older disk drive head-positioning mechanisms. The magnet-coil-carriage-rail assembly is 0.3 m fully extended, with a carriage range of motion of about 80 mm. The rail assembly is mounted on part of the original and somewhat larger cast-iron disk drive chassis.

Transducers

The emulation system is equipped with position, force and acceleration sensing, which can be seen in Figures 4-4 (force and acceleration) and 4-5 (linear potentiometer). Although direct velocity sensing is not available, sampling rates around 5 kHz for controlling behavior in the 0-10 Hz range and disturbances under 20 Hz (typical of human capacity [190]) means that even after filtering, the noise and phase characteristics of the differentiated position signal are acceptable.

Position: Position sensing is supplied by a linear potentiometer whose moving mass is negligible compared to that of the coil assembly. The potentiometer is mounted in the horizontal plane of the voice coil action, providing collocation between actuator and position sensing.[1]

---

[1] Note that collocation between position sensor and handle, as opposed to collection between sensor and actuator, is less important than for the force sensor since the reading of a position sensor located remotely from the handle will err only by the slight deflection of the relatively stiff intervening member. The output of a force sensor remote from the interaction site, however, will be distorted by potentially large accelerations of the inertia separating the force sensor and end effector; and by any dissipation present in the linkage including the linear bearings, the linear potentiometer and the dashpot.

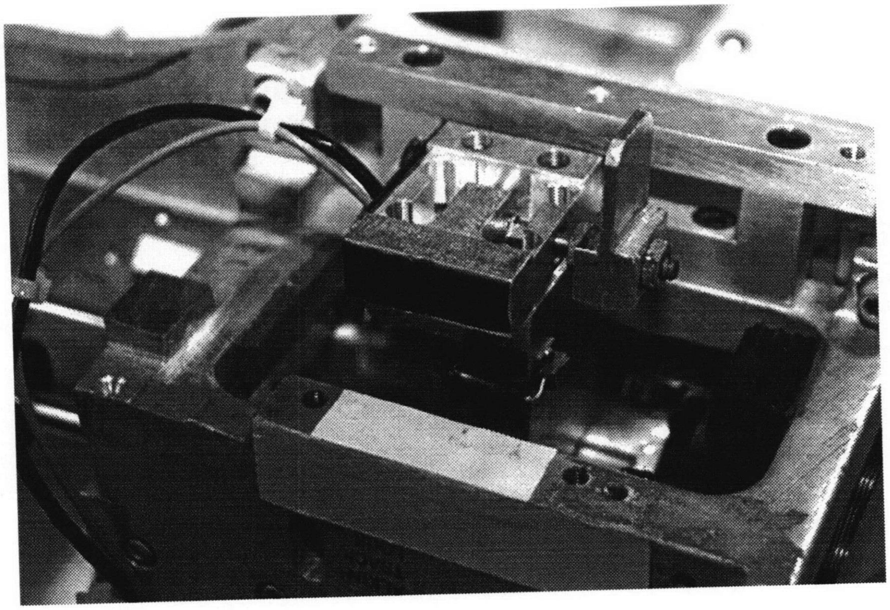


Figure 4-4: View of transducers. The force transducer is in the center of the photograph, a ~3×4×1 cm block; the handle is mounted on a beam which deflects the center beam of the transducer. The accelerometer is a pea-sized block mounted directly below the handle. Only a part of the shaft of the linear potentiometer can be seen here, at center left; see photo of physical damper for a better view.

Force: The force transducer is mounted on top of the handle support and in series between the support and the handle, in order to measure interaction force directly without error from structural dynamics occurring between points of interaction and actuation.

Acceleration: An accelerometer was needed in order to directly implement impedance-expressed inertias. A miniature piezoresistive unit was chosen for its small size and weight, bidirectional, low-G sensing range and robustness sufficient to survive trauma inflicted in hardware development. The accelerometer's output drift has been found to be manageable if operating conditions are thermally stable.

Signal Conditioning

The force sensor and accelerometer outputs are sent to commercial signal conditioners, and the potentiometer output amplified with custom-built electronics (Appendix E.2). Further digital low-pass filtering of all three signals was necessitated by the noisy environment generated by the pulse-width-modulated servo amplifier. I used a 2nd order Butterworth (chosen to minimize introduction of phase lag) with cutoffs of 80–150 Hz for position and force, and 20 Hz for acceleration and current monitor; I found the latter two signals more sensitive to PWM noise.

I experimented with analog antialiasing filters (Max280 5th order Butterworth, 200 Hz cutoff), but enough PWM amplifier noise was reintroduced between the analog filters and the ADC that digital filters were required as well. I subsequently found that removing the high-phase-lag analog filters did not result in more noise, and reduced overall filtering lag.

Servo Amplification

The motor is powered via a power supply and current-controlled pulse-width-modulated (PWM) servo amplifier, chosen for its ability to produce the required power without need for fans, which would disturb future experiments using audio displays. The PWM amplifier is equipped with current

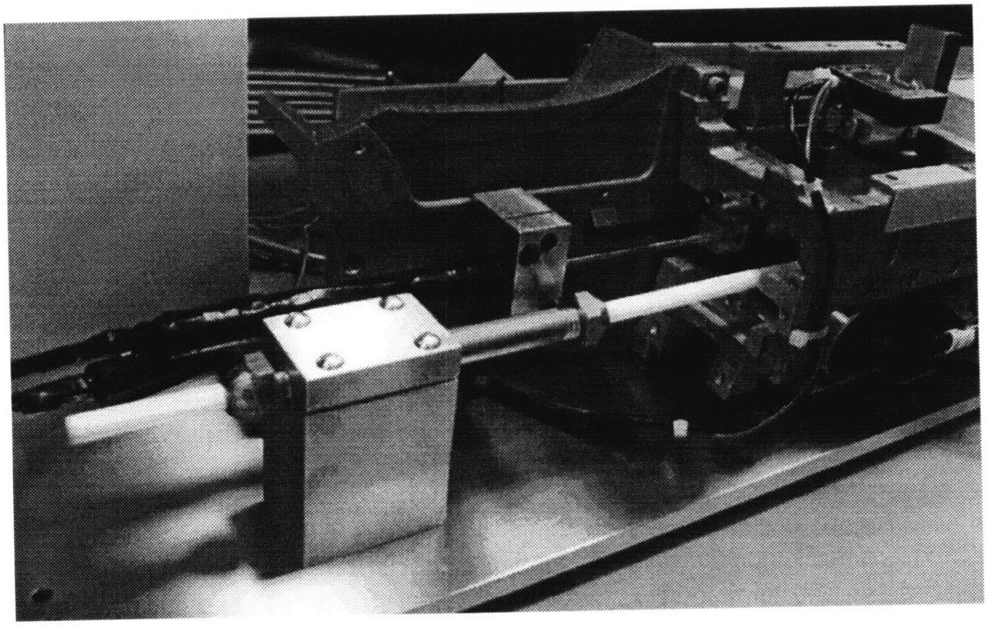


Figure 4-5: Physical damper.

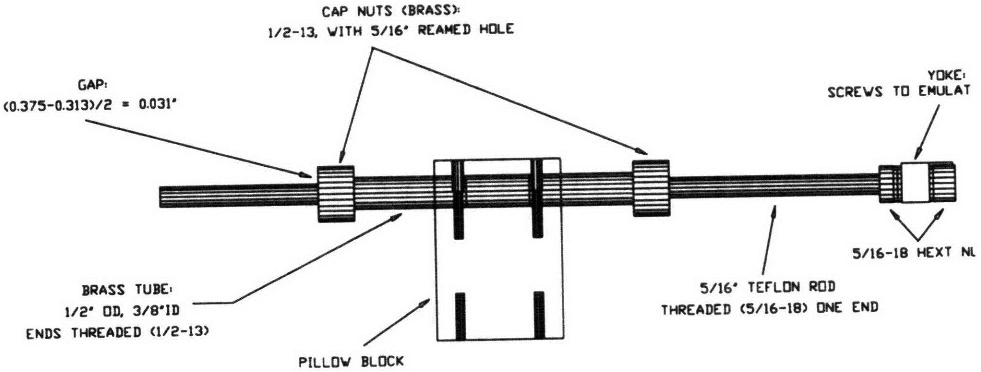


Figure 4-6: Diagram of physical damper.

monitor and external amplifier enabling features which permit the construction of external current-checking and programmable disabling circuitry, to satisfy safety requirements; schematics are shown in Appendix E.1.

The disadvantage of employing PWM amplification is the introduction of substantial electronic noise into the transducer signals and amplifier current command.

Physical Damping

It has been shown by Colgate and Brown [38] that under some control and hardware conditions, introducing physical damping increases contact stability of emulations of high-bandwidth haptic phenomenon (e.g. walls and detents). Commercially available linear-acting dampers, for the most part hydraulic cylinders, were found to exhibit unacceptably high levels of seal friction; therefore a set of low-friction dampers with a variety of resistances were constructed for installation in series with the voice coil, such that they can be rapidly detached and replaced.

The damper design was dominated by needs for bidirectional action, minimal seal friction (a source of hysteresis and resultant control challenges) and appropriate damping range. The result is pictured in Figures 4-5 and 4-6. It consists of an inner shaft connected to the motor carriage, shearing highly viscous fluid (Dow Corning; Appendix C) in a narrow gap (approximately .030") maintained by a stationary concentric outer shell; working drawings for the damper are in Appendix D.1.

The key innovation in this design is in the construction of the bearing surfaces. The damping fluid had to be largely contained in the shell, yet rubbing resistance could not be tolerated. By constructing the inner shaft of teflon rod, the inner bearing surface was given the required low friction coefficient. The outer bearing surfaces were made by threading the ends of the shell (thin-walled brass pipe), screwing acorn nuts on to either end, and drilling and reaming holes in the nuts so that the teflon shaft exactly cleared them. As a final step, the clearance holes were faced so that the actual thickness of the bearing surface was only a knife edge. The resulting damper has very low resistance when empty of fluid, and the flexible teflon shaft reduces problems with misalignment binding.

The damping level is varied by changing the viscosity of the shearing fluid; I used fluids of viscosities ranging from 1K–100K cStokes (see data sheets in Appendix C). Since the dampers are simple and inexpensive to construct, little cost is incurred by maintaining a full set of otherwise identical dampers pre-loaded with different fluids. Damping achievable with this device ranges from 0.005–0.250 N-s/mm, although system performance clearly degrades when damping was increased above about 0.080 N-s/mm and actual impedance supplied became very nonlinear and dominated by viscoelastic behavior.

Because the closed loop control was used to servo both position and force, there was no need to model the physical damping in order to compensate for it; however, non-ideal behavior in the mechanism as a result of either its design and/or fluid properties rendered even closed loop control ineffective at some point. These non-ideal effects were:

- Fluid viscoelastic behavior. This characteristic is most severe for the higher viscosity fluids and for all fluids at high shear rates, according to fluid specifications and confirmed by my observation.
- Fluid nonlinearity. Damping rate decreases at high shear rates, particularly for the highest viscosity fluids. This effect is not necessarily deleterious, as long as the decrease is gradual, since it means less actuator output is required when the environment is to move quickly. More damping is available at low velocities, where stability is most a concern (e.g., hovering in the vicinity of a discontinuity). The fluid viscosities I used in the experiments (maximum 100K cStokes), however, were relatively flat in their damping rate curve.
- Linkage compliance due to compression/stretching of teflon shaft. Although the teflon shaft served admirably in this design in many ways, its flexibility may be the cause of a serious elastic component in the overall impedance added by the damper. This elastic effect is worst at high damping levels, when there is a high axial load on the shaft.
- Fluid leakage and introduction of air bubbles. The low-friction seals are not perfect wipers, in that they allow a small amount of fluid to escape and in some cases the incorporation of air bubbles particularly when the damper is manipulated rapidly over its entire range of motion. The fluid loss and presence of air bubbles tends to reduce damping level and increase its nonlinear behavior.

Handle Interface

Human interaction with the system occurs through a handle mounted directly onto the voice coil carriage in series with the force transducer. Because the system has only one degree of freedom, no intermediary mechanical linkage was necessary and the connection from actuator to human hand is stiff.

4.1.3 Issues

PWM Servo Amplification and Transducer Noise

A pulse-width-modulated servo amplifier was selected on the basis of its high power and bandwidth performance specifications at a relatively modest price; and because it did not require a fan for cooling, the audible noise of which might interfere with the audio display that will be added to the system at a later time. The tradeoff was an unanticipated level of electronic noise introduced into the system's electronics by virtue of the amplifier's 25 kHz switching of up to 8A at ±30V (250W).

The 25 kHz noise, which enters the system through the amplifier current output, cannot be felt directly through the hand both because it is above human haptic sensory limits and because the motor's inherent inertia and damping low-pass-filter the mechanical vibratory response. However, it can introduce very large amplitude 25 kHz and subharmonics electronic noise to the sensors. Whereas a linear amplifier used with the same system produced virtually zero sensor noise and required neither analog nor digital filtering to achieve good control, PWM amplification of the system contaminated sensor output with nearly full-scale noise when sensors were unfiltered, resulting in complete uncontrollability.

The noise paths identified were

1. conductive, through the motor current lines and directly into the motor-mounted sensors;
2. radiated, from the amplifier itself, the motor current leads and the voice coil; and
3. propagated from the amplifier via the 110 V AC supply line into the signal conditioning electronics power supply.

Comprehensive noise reduction was required to reduce sensor noise levels to levels that enabled satisfactory control. The steps eventually taken which were deemed to be most effective included:

- Use of high-quality twisted-pair, shielded signal cable throughout the system, with careful attention given to grounding of cable shields.
- Use of heavily shielded and grounded motor current cable.
- Use of heavy duty (1/2" braid) grounding straps connecting servo amplifier and motor chassis; and connecting various locations on the motor chassis (contact sites found by trial and error).
- Scrupulous adherence to I/O board manufacturer's field wiring recommendations for data acquisition in severe electronic environments.
- Digital low-pass filtering of all sensor signals.
- Low-pass filtering of the current command (RC, 1 pole) with a cutoff of 500 Hz immediately prior to entry to servo amplifier.
- Housing signal conditioning boards in an electrically isolated Faraday cage.
- Independent filtering of 110V AC power to both the servo-amplifier and signal conditioning electronics power supplies via separate voltage surge suppressors.

While none of these steps resulted in great material cost and the final result did achieve the goals of high force saturation, inaudible current source, relatively low servo amp cost and noise levels allowing useful control (the first three difficult to attain with linear servo amplification), the impact on system development time was substantial and therefore expensive. Further, a price has been paid in closed-loop control bandwidth, due to the phase lag introduced by low-pass sensor filtering. It is therefore suggested that linear servo amplification should be used for haptic displays when other criteria (such as high power/cost or inaudible operation) do not preclude it.

Design for Safety

The emulation system was designed with a hierarchy of safety features to ensure accident-free interaction with human subjects. They consisted of:

1. *Safe usage of hardware:* Emulations were designed, model-parameterized and tested prior to use in an experiment so that a human user would find it difficult to destabilize the controller (note that this constraint in some cases compromised the attainable fidelity of the emulation).
2. *Mechanical design of interaction:* The user could not easily place any part of his/her body in a position where injury might occur from motor motion. When used in experiments, the haptic interface pictured in Figure 4-1 was completely enclosed; the experiment handle extended through a narrow slot in a hand rest platform mounted over the emulator's line of travel. The slot was longer than the interface's range of motion, and thus pinches were unlikely even if the user were able to squeeze a finger into the slot.
3. *Manual disabling of servo amplifier:* The servo amplifier could be disabled both manually and through computer control. Instantaneous manual disabling of the servo amplifier was possible at all times, overriding any computer enabling in effect.
4. *Automatic computer disabling of servo amplifier:* When any sensed or derived states exceeded established safe thresholds, the servo amplifier was immediately (within 0.2 milliseconds) disabled, overriding the manual enable.

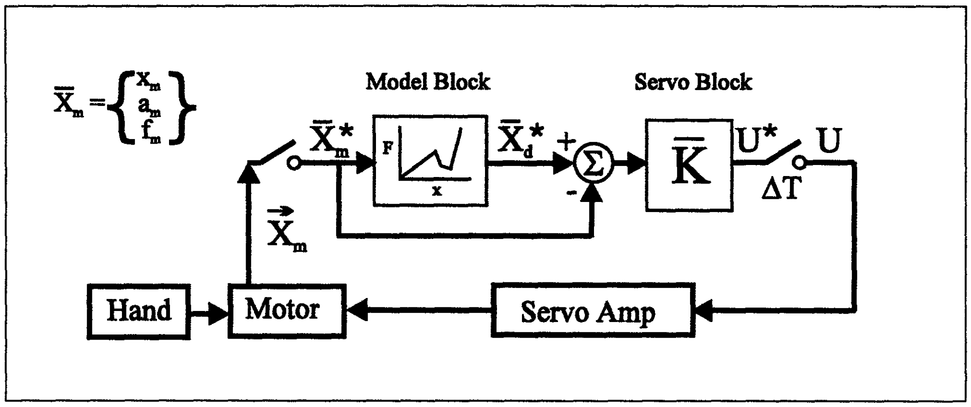


Figure 4-7: Schematic of emulation control. Subscript *d* refers to desired state and *m* to measured state.

4.2 Control

The emulation system's digital control, indicated in Figure 4-7, consists of two cascaded blocks: a parameterized model of the device being emulated, followed by a servo block that enforces the desired state computed by the model. Nomenclature for the control parameters is defined in Table 4.1.

The combination of high-speed digital sampling and access to position, force and acceleration feedback allow computation and control of 0th, 1st and 2nd order linear and nonlinear haptic models formulated in both impedance (Z) and admittance (Y) configurations. Thus the two model structures and a variety of servo controllers can be tested on the same hardware with all other variables held constant, for their relative effectiveness in reproducing a given family of haptic phenomenon.

Control Platform

The controller is implemented on a 60 MHz Pentium PC equipped with a multipurpose I/O board. Currently four channels of analog input are employed for position, acceleration, force and current monitor signals. One channel of analog output is used for the amplifier current command, and one channel of digital output for computer en/disabling of the servo amplifier.

The digital servo is regulated by polling the I/O board's real-time clock, at rates of up to 5 kHz for a 2nd order admittance (2 real-time integrations) model with complex nonlinear elements; up to 10 kHz for a moderately complex impedance model (no integrations) and up to 20 kHz for a pure servo (no model) controller.

$\vec{X}_d$ = desired state vector = $\{x_d\ v_d\ a_d\ f_d\}$
x_d = desired position
v_d = desired position
a_d = desired acceleration
f_d = desired force

$\vec{X}_m$ = measured state vector = $\{x_m\ v_m\ a_m\ f_m\}$
x_m = measured position
v_m = derived velocity
a_m = measured acceleration
f_m = measured force

u = computed control output
K_{prop} = proportional servo gain
K_{int} = integral servo gain
K_{der} = derivative servo gain

Table 4.1: Nomenclature for control parameters.

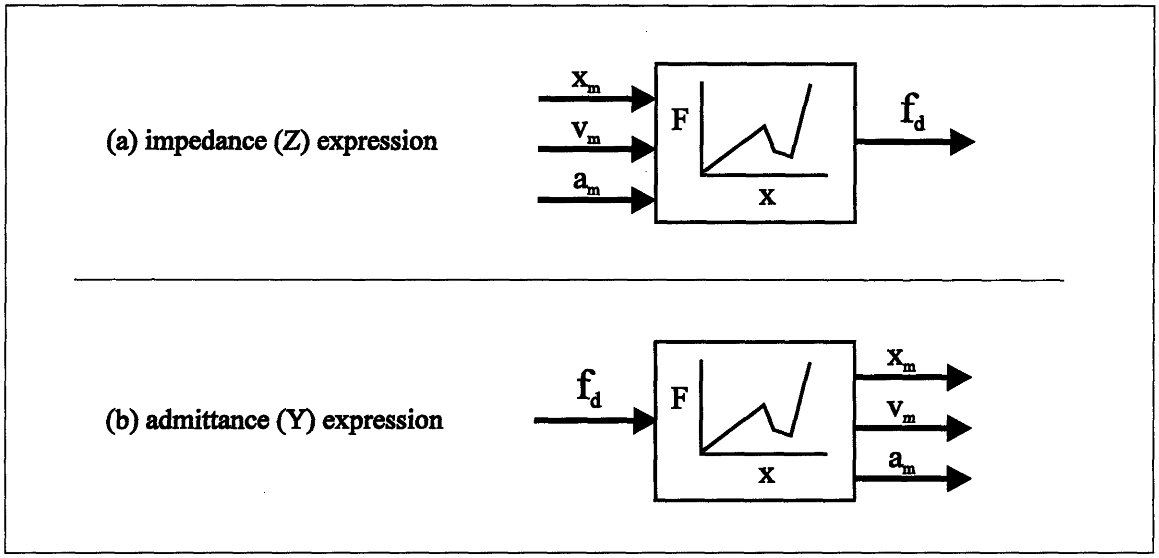


Figure 4-8: A simplified schematic of the controller model block.

4.2.1 Model Block

A detailed description for the controller model block may be found in Section 5.2.

Here, simply observe that the function of the model block is to compute a desired output state based on measured input state and a virtual environment model. As illustrated in Figure 4-8, this may be done with either impedance (Z) or admittance (Y) expressions, where impedance is measured kinematic state in and desired state output, while admittance is measured force in and desired kinematic state out.

4.2.2 Servo Control Block

Description

The second block is the servo controller, with the role of enforcing the desired state computed by the model block. The servo block computes a control command based on the difference between the desired and measured state. For Z control (Figure 4-9 (a)), the desired state is force, and the servo relation is

$$u = K_{prop}(f_d - f_m) + K_{int} \int (f_d - f_m)dt \tag{4.1}$$

for the proportional-integral (PI) force controller used here.

Closed loop force control was found to be invaluable in this implementation in that it (1) produced relative insensitivity to variations in the impedance of and disturbances provided by the human operator's input; (2) nulled any servo-amplifier nonlinearity, permitting usage of less sophisticated current amplification; (3) substantially enhanced servo stability (as well as command error) by reducing a 1 kHz open loop resonant peak found in this system; and (4) permitted additional physical damping to be counteracted without adding explicit (modeled) negative damping to the controller (particularly useful when a damping system exhibits non-ideal behavior, such as a damping rate that decreases exponentially with velocity).

For Y control (Figure 4-9 (b)), the servo relation to track position is:

$$u = K_{prop}(x_d - x_m) + K_{int} \int (x_d - x_m)dt + K_{der}(v_d - v_m) \tag{4.2}$$

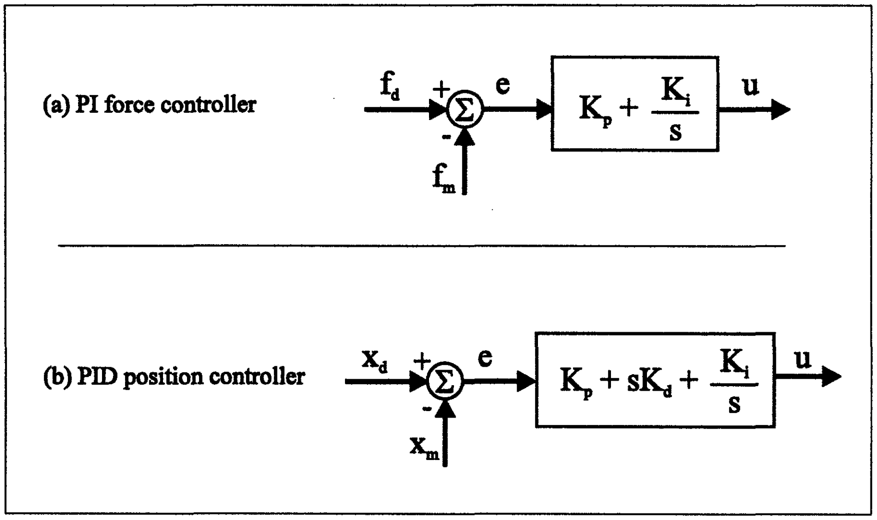


Figure 4-9: A schematic of the controller servo block.

for the case of a proportional-integral-derivative (PID) position controller. P, I and D components of the position controller are generally all required to achieve a stiff but stable servo; however, the most desirable controller pole locations depends on the nature of the impedance being emulated.

The control block of the emulation testbed permits modulation of sample rate, various aspects of the controller model block (expression, structure, order and accuracy of model parameterization) and the structure and stiffness of the servo controller.

mention antiwindup in I control for both force and position.

4.3 Performance

The purpose of this section is to give quantitative measures, insofar as possible, and for the rest qualitative assessments, of the emulator's capabilities in terms of both standard bandwidth benchmarks and "quality" of its emulation of haptic phenomena.

Bandwidth

Performance characteristics of the haptic interface under open loop and closed loop force and position control are displayed in Figures 4-10–4-11. All data were taken at a sampling rate of approximately 15 kHz with minimal physical damping, and represent response to pre-computed trajectories of desired force and position, as opposed to the output of an emulation model. Thus they convey useful information as to the capabilities of the motor and servo control; but different gain operating points must be used to maintain stability when the emulated system is added to the controller.

Force Loop

A force loop was closed using a proportional-integral (PI) force controller; the force sensor is filtered at 80 Hz with a 2nd order digital Butterworth. Figure 4-10 displays both open loop response and the closed loop response for one set of PI gains. The open loop frequency response shows a resonant peak at 100 Hz followed by 80 dB/decade rolloff, and a smaller peak at about 530 Hz; each peak produces about 180 degrees of phase lag. Although at a higher frequency than that found in position control, the force response's first resonance is still in a critical control region and thus the integral term in the PI force controller plays a key role in stabilizing it. The PI controller thus gives the

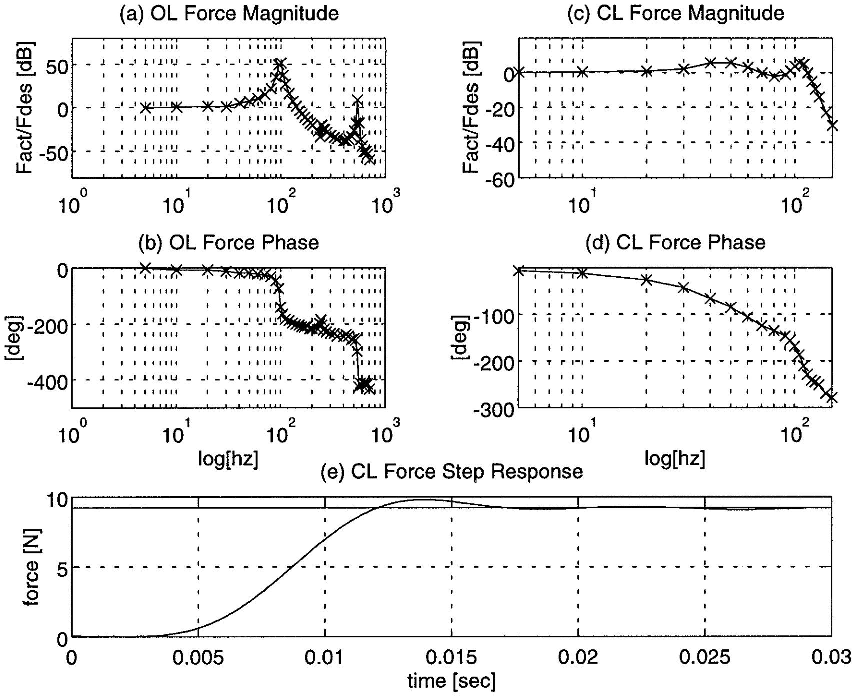


Figure 4-10: Emulator force response with handle locked. (a–b) Bode plot of open loop force response to commanded force sinusoid; (c–d) Bode plot of closed loop force response; (e) step response of closed loop force response.

capacity for stable force control through and somewhat above the 100 Hz peak and consequently increases overall system performance, although a price is paid in terms of phase lag introduced at frequencies below 100 Hz.

A closed loop force step response is shown as well, slewing force at a rate of nearly 1000 N/sec and settling a 10 N step in approximately 30 msec.

Position Loop

A position loop was closed using a proportional-integral-derivative (PID) position controller with a position signal filtered at 80 Hz with a 2-pole digital Butterworth, and backwards differenced to produce a velocity estimate used directly in the controller's D term. Closed loop position response, shown in Figures 4-11 (a-c) for one set of PID gains, exhibited a corner frequency of about 20 Hz, followed by an 80 dB/decade rolloff. The peak location and rolloff have been reproduced in simulation of a lumped-parameter fourth-order model of the voice coil, carriage and handle assembly, consisting of 2 masses (coil/carriage and handle) connected by a stiff spring (the 50mm long handle support, which contains a small but measurable compliance).

Settling time for an approximately 1 cm step in desired position is seen to be under 50 msec, with a slew of about 0.6 meters per second and about .5mm of steady state error.

Source of Bandwidth Limits

Based on the simulation validation, consistent measurement at sampling rates from 1k-15kHz, and high-bandwidth servo amplifier current slew rates, these boundaries on servo performance may be attributed to the mechanical hardware, rather than sampling rate or servo amplifier. The large

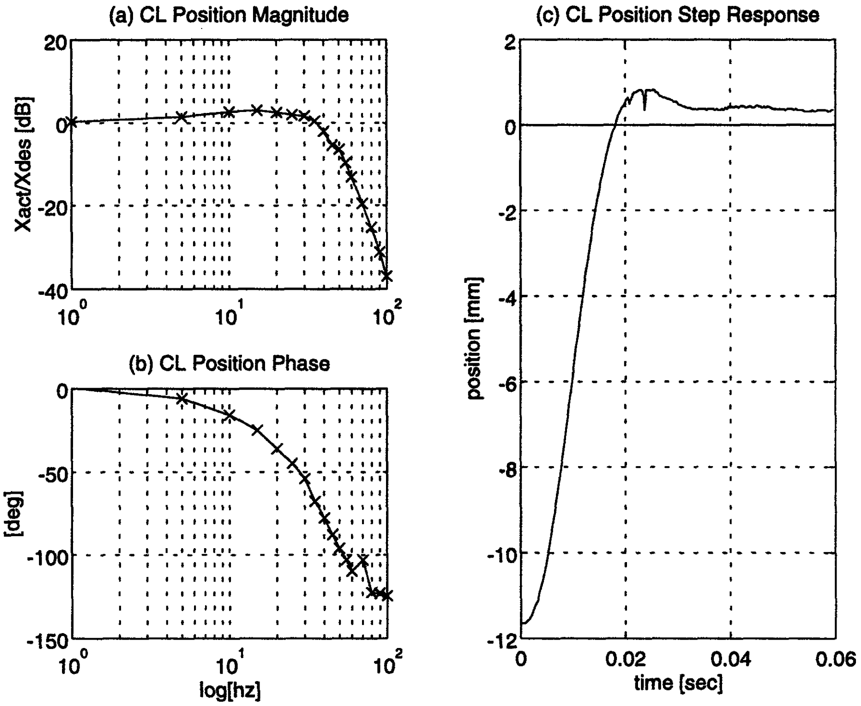


Figure 4-11: Emulator closed loop position response with handle free. (a–b) Bode plot of closed loop position response to position sinusoid; (c) step response of closed loop position to position step command.

difference observed between force and position bandwidths is typical of relatively high inertia, high force actuators.

The force bandwidth is the more critical variable in this application, since spikes in force are more important than sub-millimeter position accuracy in conveying the sense of sharp discontinuities; humans can perceive minute-amplitude force events at 1000 kHz [153], but are unable to detect small position displacements and motions. However, it would clearly be desirable to have a CL force bandwidth which more closely approaches the 1000 kHz perceptual cutoff. A lower inertia motor will serve this end, if it can be attained without overly compromising steady-state force saturation.

4.3.1 Achievable Impedances

Table 4.2 lists the impedances which this hardware and control were able to produce. Similar limits were found for Z and Y expression at adequate sampling rates and low physical damping, although in some cases extreme environments produced by the two expressions felt qualitatively different.

Quantity	Minimum	Maximum
Stiffness	0 N/mm	20 N/mm
Damping	.001 N-s/mm	4 N-s/mm
Mass	.030 kg	5 kg

Table 4.2: Basic impedances achievable with this system.

Chapter 5

Haptic Virtual Environments: Modeling and Expression

In the controller environment model (Figure 5-1), a desired state is computed from input state feedback. The purpose of this chapter is to describe how this may be done in principle and how it was implemented in reality.

Generalizability of Comments

Some of my observations and the strategies which I found most suitable will inevitably prove to be a result of the particular hardware I am using, and it will be difficult to say which are generalizable until direct comparisons are made between different hardware setups. Although this system is more versatile than many, any design has hard limitations of bandwidth, sensor resolution, stiffness and other variables which ultimately determine its performance in different metrics. Other setups have different limitations and possibly different responses to some of the modeling strategies described here.

Emulation Variables Inherent in the Virtual Environment Model Block

To put this chapter in the context of the methodology expressed in this thesis and of the experiments which will be described shortly, it is worthwhile to review the emulation variables which arise from the virtual environment model block:

Model structure: For example, a dissipative element with friction and hysteresis could be modeled as an ideal viscous damper, or in a more physically correct manner.

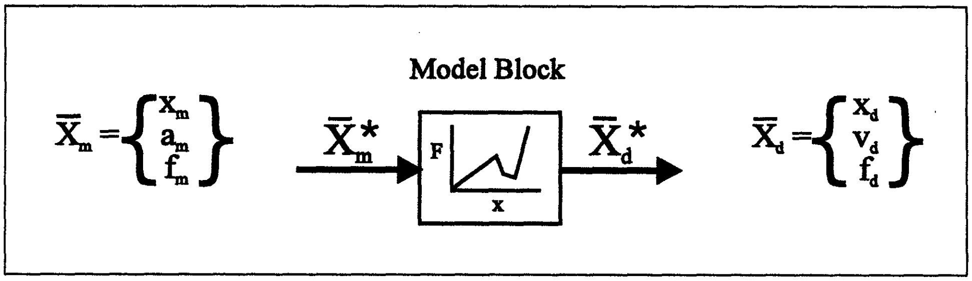


Figure 5-1: The first block of the controller is the virtual environment model.

Model expression: Is the model structure expressed in admittance or impedance formulations? And within those larger divisions, what form does the admittance/impedance expression take? One form of each expression is described here, but other forms appear in the literature.

Strategies for model realization: There are innumerable ways of computing the model within the servo loop — the more complex the model, the more implementation options. Sometimes these small refinements can make a large difference in the environment's stability.

Accuracy of model parameterization: Presuming the model is accurate, the degree to which its parameterizations corresponds to the real environment is yet another emulation variable. Accurate parameterization may require more effort than an approximation, and in some cases it will be more critical than others.

Of these possibilities, only model expression was tested as an emulation variable in the human subject experiments; the others were held constant to the extent possible. The point is that there is an infinitude of alternatives in this region of the emulation variable space alone, and since they were not all tested the ones discussed here represent only local optimums.

5.1 Creating Haptic Virtual Environments

General Features of Modeling Strategy

I will start by describing the approach to and components of the virtual environment models I used, and consider specific cases of model expression later. These comments refer to the creation of any haptic virtual environment, regardless of whether it is intended to resemble a specific real-world target.

Basic Building Block: Continuous Model Elements

The simplest model elements are those shown in Figure 5-2: they are linear, continuous and ideal. They may be combined in parallel or series and the defining equation revised accordingly, such that force is always defined as a function of kinematic state or vice versa. The Mass and Slider families were composed of such single-parametered linear continuous elements.

Emulating even these simple basic elements is not necessarily as simple as modeling them, however. For example, the model input state must be measured or derived, and the performance of the virtual environment will be influenced by the noise and resolution of the signal; while the range of achievable impedances will depend on both hardware and the servo controller.

Nonlinear Features

The next increment in complexity is to use continuous elements which are not linear or ideal. For instance, a component of stiction, static force offset, direction-dependent or v^2 damping could be employed in the same format, in some cases with little increase in modeling or computational complexity.

Piecewise Continuous Models

I used continuous elements to build piecewise continuous models of a more general and complex set of environments; I will henceforth refer to these elements as *model zones.* Figure 5-3 shows an example of such a model for a zero-order (compliance only) environment. The model parameters are stored in a lookup table which is entered in each servo loop through sensed or desired kinematic state (position, direction of motion, etc.). Linking elements in this way rather than trying to model either continuous

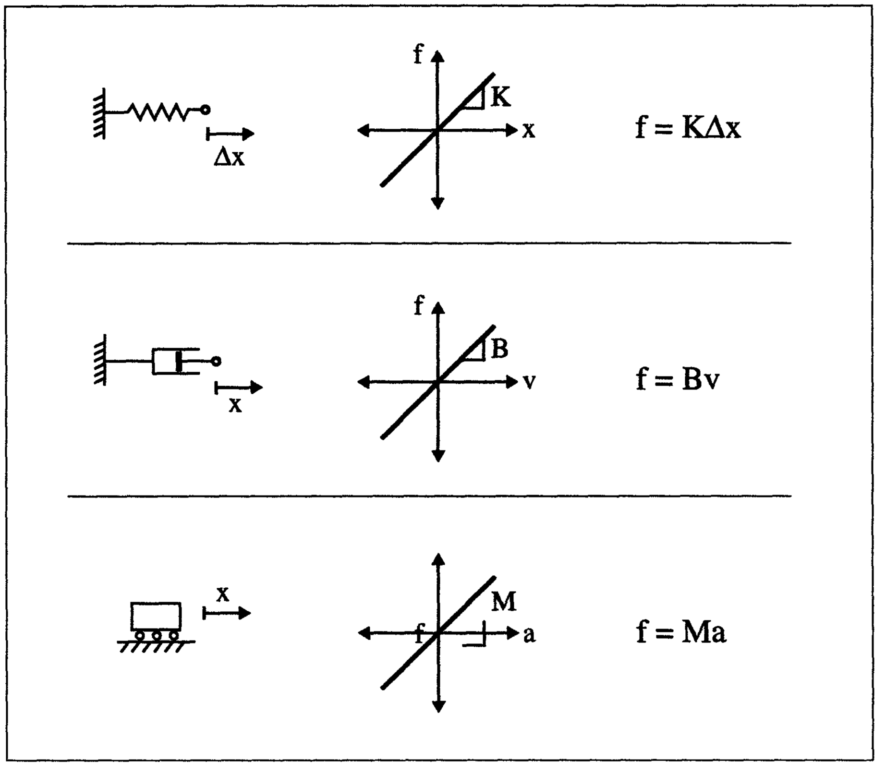


Figure 5-2: Modeling of linear continuous elements.

parameter changes or a more complex structure permits the modeling of complex environments with simple elements, providing one has a good estimate of the desired behavior either through empirical measurement or a more sophisticated model. As a result, both model parameterization and computation are more tractable. My experience suggests that increasing the number of segments beyond what is required to achieve a gross fit to the actual empirically measured $K(x)$ profile in the example shown cannot be perceived by a human subject (Section 7.4). It therefore does not merit the effort involved.

Impedance Expression of a Nonmonotonic Model :

In impedance model expression, $f_d = \mathrm{f}(\vec{X}_m)$[1] whereas in the admittance model expression, $\vec{X}_d = \mathrm{f}(f_m)$. In the admittance case the detent map of Figure 5-3 is entered through the force rather than the position axis, and the nonmonotonicity in this situation causes ambiguity. I resolved this kind of ambiguity by determining the current zone through either measured position or the desired position computed in the previous timestep.

Models Not Always Physically-Based

Sometimes it is more straightforward to model a real environment by matching its perceived or measured haptic force profile rather than modeling the mechanism which produces this profile. This is was done in Figure 5-3 and will often be the case with the piecewise continuous approach, because either the mechanical structure is complicated and/or difficult to parameterize or a mechanically accurate model is more difficult to stabilize.

---

[1] The vector notation for $\vec{X}_m$ refers to the vector of kinematic state x_m, v_m, a_m.

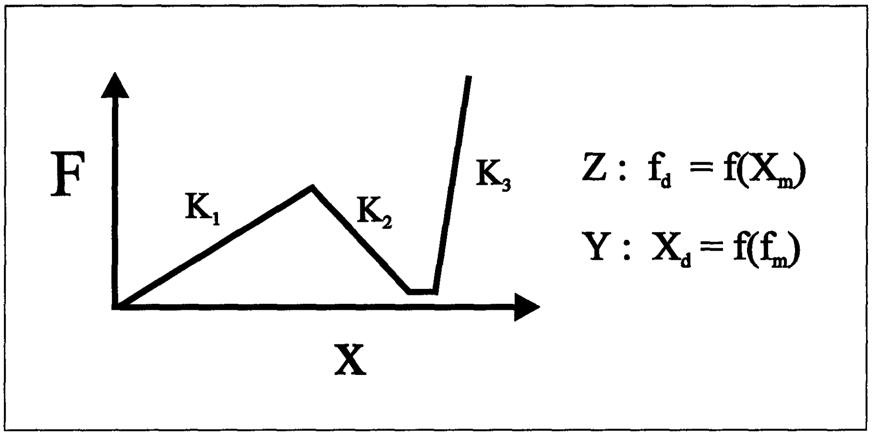


Figure 5-3: Example of a piecewise continuous virtual environment model: a typical force-position profile for a detent. Only compliance is represented.

Comments on the Emulation of Real Devices

Creating "emulations", or virtual environments which are compared directly with a real counterpart in the creation and/or usage stages, is a different and in some ways more difficult problem than that of modeling abstract environments for which no such comparison will ever take place. It is where the concepts of "fidelity" and "resolution" in addition to or instead of "realism" come into play. Model structure and parameterization cannot be arbitrary, although there may be multiple solutions which give similar haptic results; and emulation resolution (sensors and actuation) becomes critical for reasons other than control stability. Targeting a specific environment means that the controller and its gains must be adjusted more carefully to optimize accuracy.

For the application of virtual product prototyping, the requirements become even more demanding. Since the device in this case must be realized in a physical form, a virtual model which captures desired haptic features but no realizable mechanical structure is not of great use. There must be a correspondence between the model and a set of manufacturing parameters.

5.2 Impedance and Admittance Model Expression

I expressed the virtual environments developed here as both an impedance (Z) and an admittance (Y), and employed these expressions as a qualitative emulation variable in the human subject experiments. In this section, I comment on the intrinsic differences between these expressions, describe the environment expressions used here and detail the implementation of each.

Duality

Impedance control (Z) implies sensing of kinematic state and specification of a desired force [3, 36, 67, 69, 82, 127]; while admittance (Y) implies sensing of force, specification of a desired kinematic state, and nothing more [45, 68, 102, 148].

However, these descriptions encompass a broad range of specific implementations. Further, the appropriateness and performance of the model expression used in a given case is highly dependent on specific hardware (particularly bandwidth and sensing), method of desired-state enforcement (here, the servo block), and the nature of both the emulated and outside environments. It is therefore important to separate a discussion of intrinsic differences between position-in/force-out and force-in/position-out model expressions (desired state computation) from that of desired state enforcement.

The following comments are addressed at practical implementation issues related to the former question, assuming the virtual environment expressions described here.

Ideally, Z and Y controller expressions for haptic environment emulation would produce identical results for all linear environments and all invertible nonlinear environments. In practice, Z control is more commonly used. This choice is sometimes motivated at least in part by its less demanding sensing requirements: Z control is often implemented with only a position sensor in conjunction with a current-controlled servo amp, whereas Y control requires both force (for the model block) and position (for closed loop servo control, which for position is a necessity) feedback. In addition, there is the potential of simpler implementation with Z control when sensors are available: Y requires real-time integration, whereas a Z expression can be implemented without integration or solution of differential equations, as it is here.

Observations from the use of my system, one of many, suggest some practical differences between Y and Z approaches to emulation. Y appears superior to Z control in reproducing environments which are high-order (inertial) and of moderate to high impedance. High fidelity and stable emulation of inertia is a particular challenge for Z control, which can attempt it only with an accelerometer or position and/or velocity signals amenable to differentiation (a capability not present in most existing systems) or with potentially unacceptable constraints on the virtual inertia's magnitude. Conversely, it is difficult to emulate low impedance environments using Y control, largely because of the numerical complications inherent in Y models of those systems: interaction force is generally small, providing a near-singularity in the Y model operating point. Further, a Y model must contain impedance terms (e.g. M_d) in its denominator, so numerical and control instability results when these terms become too small.

Environment Expressions Used Here

Here I will describe the computation details of the impedance and admittance expressions I used for this work. The model may be of any order in either Z or Y model formulations, and may contain arbitrary linear and nonlinear elements with computational bandwidth as the primary constraint.[2] In general, the same model structure and parameterization were used for both Z and Y model expressions.[3]

The basic model for a continuous linear system is diagrammed in Figure 5-4 and generally defined by

$$f_{interaction} = M_d a + B_d v + K_d(x - x_o). \tag{5.1}$$

This expression constitutes a building block which may be combined in many configurations. More complex relations are required to compute such effects as hysteresis and direction dependence.

Impedance Expression

The Z-expressed model is computed using the following digital relation to compute desired force for the timestep k:

$$f_{d_k} = M_d(\vec{X}_m)a_{m_k} + B_d(\vec{X}_m)v_{m_k} + K_d(\vec{X}_m)(x_{m_k} - x_o), \tag{5.2}$$

---

[2] Hardware parameters (bandwidths and passive characteristics) play an important role in what model and model parameterization will be stable in implementation.

[3] The exception was in modeling some types of discontinuities, detailed later. I was generally unable to stabilize the same model of walls, stiff detents or snap-toggles at high fidelity (often requiring high gains) in the two expressions.

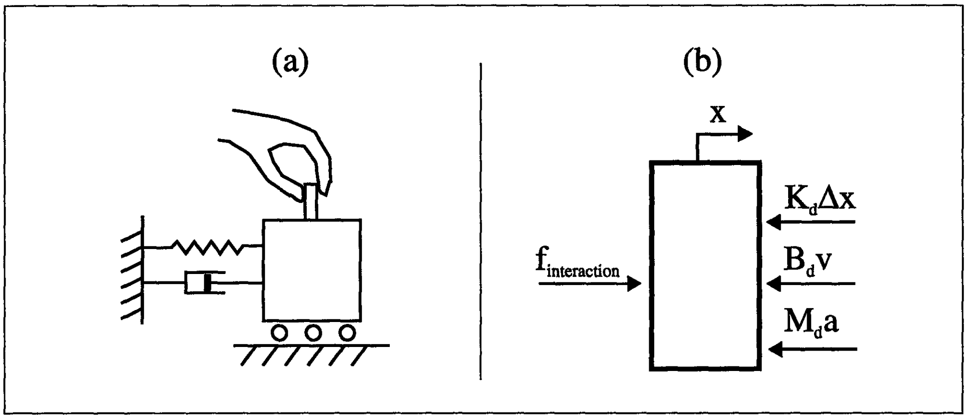


Figure 5-4: Free body diagram (FBD) of virtual mass-damper-spring system, showing direction conventions for the equations in this chapter. Only one force input (the user's hand) is shown because this is a passive virtual system and the actuator's task is to create the dynamics at the handle which would correspond to the virtual model. A different FBD could be drawn to represent the real emulation system dynamics.

where

$$\begin{aligned}\vec{X}_m &= \{x_{m_k}\ v_{m_k}\ a_{m_k}\} \\ x_{m_k} &= \text{filtered potentiometer signal} \\ v_{m_k} &= \frac{x_{m_k} - x_{m_{k-1}}}{T_k} \\ a_{m_k} &= \text{filtered accelerometer signal.}\end{aligned}$$

$\vec{X}_m$ is the kinematic state vector and M_d, B_d and K_d may be linear or nonlinear functions of sensed kinematic parameters. Systems of order lower than two are achieved by setting the higher order parameter to zero.

Admittance Expression

The admittance expression used here is the inverse of the Z model. For a second-order system,

$$\begin{aligned}a_{d_k} &= \frac{f_{m_k} - B_d v_{d_{k-1}} - K_d(x_{d_{k-1}} - x_o)}{M_d} \\ v_{d_k} &= \int a_{d_k} dt \\ x_{d_k} &= \int v_{d_k} dt\end{aligned} \tag{5.3}$$

where

$$f_m = \text{filtered force signal.}$$

and M_d, B_d and K_d may all be functions of $\vec{X}_d$ and f_m. The two integrations are performed in realtime; for a linear model, the continuous-time model specification may be converted to a discrete-time representation and computed in realtime in a period comparable to that required for the Z model. Models with nonlinear elements, the large majority of interesting systems, are evaluated using a fourth-order, fixed-step Runge-Kutta algorithm [158] at an increased but by no means prohibitive computational cost.

Lower order systems are computed in the same manner but with a revised structure. A first order system is computed by

$$\begin{aligned} v_{d_k} &= \frac{f_{m_k} - K_d(x_{d_{k-1}} - x_o)}{B_d} \\ x_{d_k} &= \int v_{d_{k-1}} dt \end{aligned} \tag{5.4}$$

and a zero order system by

$$x_{d_k} = \frac{f_{m_k}}{K_d} + x_o. \tag{5.5}$$

"Duality" of Expressions Used Here

Expressions 5.2 and 5.4 are not strictly duals but to some extent "apples and oranges": they earn their names by their respective position-in/force-out, force-in/position-out transfer functions but differ by more than input/output relations.

- Z has three inputs (x_m, v_m and a_m) and one output (f_d) and its Laplace transform is $\propto \frac{1}{s^2}$; whereas Y has one input (f_m) and three outputs (x_d, v_d and a_d) or $\propto \frac{s^2}{1}$. The latter is acausal and would be physically unrealizable in analog form, but is estimated here by using "past" zeros. Since it incorporates desired as well as measured states from the previous time step, its output is dependent on sampling rate. Below some critical sampling frequency (on the order of 300-500 Hz) this can be a noticeable effect, but generally it is possible to sample faster.

- The same effect means that the Y expression has a memory; this makes it possible to create effects such as transfer of momentum and correct modeling of losses (described later).

- In this formulation of Z control, conversely, desired state (f_d in this case) is computed directly from the kinematic state measured during the same time step. It has no memory, but also no integration delay.

Expression-Specific Implementation Details

In many cases the same model parameter values can be used for Y and Z expression, although different constraints on maximum and minimum levels for the highest-order parameter are in effect because of the different computational structure (non-unity denominator in Eq. 5.4).

However, in certain situations implementing Z and Y expressions of the same haptic phenomenon require different strategies for both environment modeling and execution. This is because the two expressions rely on servo control of force and position respectively with different inherent properties, and because specific measured signals play different roles in each (virtual model input versus controller reference). These circumstances influence the kinds of modeling strategies that may be used, for instance the components of a virtual wall which make it feel most crisp and passive. Of the environments actually tested here, only Toggles required different modeling strategies for Y and Z expression because it was the only family with discontinuities; but other environments, such as walls and detents, had similar requirements. Some general comments on these are made below, and more specific observations will be described case-by-case in the following Library section.

Admittance-Specific Details

Fine points of modeling and implementing the Y expression's integrations and manner of moving from one model zone to another are described in this section; these subtleties must be observed for any multi-zone model although they are more critical when the zone change is also a sharp change in desired force or position. The impedance expression employed here was relatively straightforward.

Changing model order:

There are at least two options for modeling a system with non-constant model order, for example in one zone a damped spring and elsewhere a mass-spring-damper: (a) always use 2nd order integration but in the 1st order region use the smallest stabilizable value for M_d; or (b) switch order of integration (structure of equation). That is, swap between Eq. 5.5 and 5.4 with appropriate resetting of integration states. I found that the latter option worked better; when a high order integration is always used, the minimal value for M_d is usually insufficiently small or else the system is a bit jittery because the small denominator causes numerical instability.

Mode of model access:

The function defined in Figure 5-3 is single-valued in position but not in force, and measured force cannot be used alone to determine which model parameters should be in effect. It is therefore most convenient to define and access the Y-expressed model in terms of either measured or desired kinematic state rather than sensed force, as for Z expression.

Zone entry/exit:

- Use of measured versus desired state to determine current model zone:

  There is a choice of whether to use measured or desired kinematic state to enter the lookup table; using x_d being more consistent with the Y model. In practice using x_d gave results which were more crisp whereas x_m tended to "round off" edges. I therefore used x_d for spring-based walls when the discontinuity was the key feature, but used x_m in emulations like the Toggles when it was undesirable to feel the edges of individual zones — they should blend smoothly together.
- Resetting of integration states on zone crossings:

  Because v_d and x_d are integrated rather than computed directly from states measured that timestep (for a 2nd order system), the integrated states must sometimes be re-initialized upon entry to a new region. When virtual model parameters change as a result of crossing zone boundaries, different effects can be achieved through the conventions followed in resetting integration states. This feature makes the Y expression more powerful in some respects than this expression of Z; since Z doesn't have any memory built into it, it does not accommodate the effect of momentum transfer, for example.

  Momentum transfer (inertial environments) :

  To conserve momentum ($M_d v_d$) at a boundary crossing, scale desired velocity by ratio of old and new masses:

$$v_{d_{new\ region}} = v_{d_{old\ region}} \frac{M_{d_{old}}}{M_{d_{new}}}$$

  This is the correct approach to use when emulating the effect of traveling through air and then hitting a stationary mass. The additional mass in that situation should have a zero initial velocity.

  To produce the effect of pushing a mass through space and picking up or dropping additional mass which is moving at the same velocity, v_d should not be reset; that is, use $v_{d_{new}} = v_{d_{old}}$.

Stiffness: When moving between zones of different stiffness, there are several options for the new desired position (x_d): the new spring equilibrium (x_o), the current measured position (x_m), the start of the region or the current desired position (i.e. don't reset it). Which convention is used, as well as the exact location of the new equilibrium x_o, influences the feel of sharp discontinuities. For example, this control over x_d can be used to ramp up contact force more quickly in a stiffness-dominated wall.
I found that leaving $x_{d_{new}} = x_{d_{old}}$ gave the most passive spring-based discontinuities for this emulation system, and since passivity was a concern in the discontinuous emulations that is the strategy I used in the experiment versions.

Switching directions:

Some types of environments undergo a change in parameters not only through translating but also at a direction reversal (see direction-dependent examples in Section 5.3). In these cases, there is a choice of choice of using measured or desired kinematic state (v_m or v_d) as a switching signal whereas in Z control, only measured (v_m) is available. The same effect was seen as for translation switching above, in that using measured velocity rather than desired gave more passive and crisp results.

Frequency of integration steps:

There is a choice as to integrating the desired states once per servo loop or more often. I integrated once per servo loop because at the sample rates I typically used (over 1000 Hz) this integration rate was sufficient for numerical stability; and the integration took a substantial fraction of the entire servo so repeating it would have compromised overall servo rate. However, there might be situations in which this is not the best choice; for instance, when input or output signals cannot be updated as quickly as the computer controller is able to carry out integration steps.

5.3 Library of Emulations of Specific Haptic Phenomena

This chapter consists of recipes for virtual models for a collection of interesting haptic environments, and anecdotal observation of the sometimes unexpected behavior which occurs in the course of creating such environments. Most of the remarks and recipes are not for emulations alone, but should apply to a haptic virtual environment for any application. Conversely, some of the observations and strategies here are likely a result of the specific hardware, servo control and virtual environment expressions I used.

What Your Mother Didn't Tell You About Virtual Environments

This section documents observations on and insights into virtual environment behavior as a function of fine detail in modeling and control. For the most part, they would have been difficult to predict; they came about only through extensive informal experimentation with the virtual environments.

"Slipthrough": Attenuation of Static Discontinuities

In virtual environments containing phenomena such as spring-based detents or other "bumps" in specified stiffness, I observed that the apparent stiffness and/or maximum force generated (according to both my perception and to sensor recordings) was dependent on speed of handle manipulation. The reason for this is finite force slew rate; the handle "gets ahead of" the motor and passes the discontinuity before the motor slews to the maximum specified force, and consequently is barely disturbed (Figure 5-5). Bumps trigger the effect rather than walls because a key feature is the

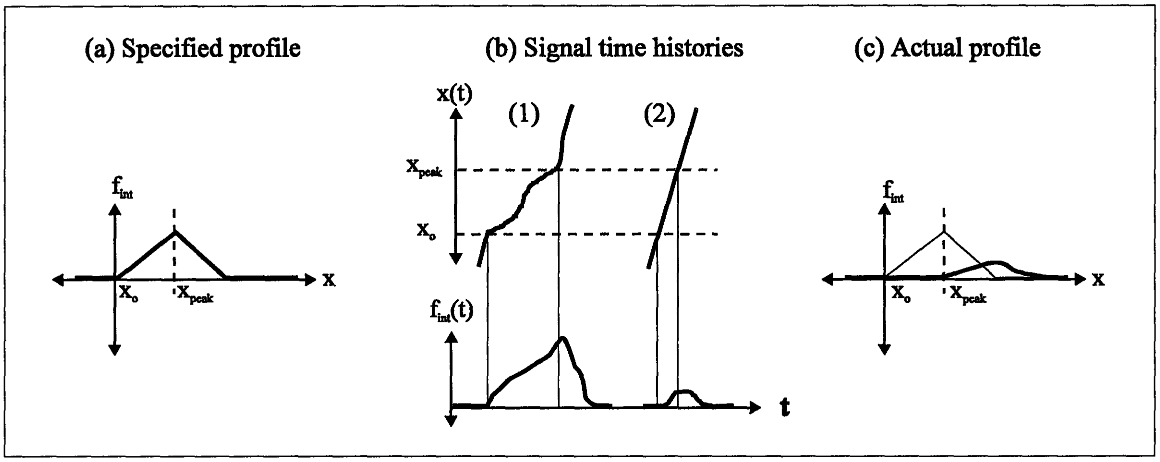


Figure 5-5: "Slipthrough": getting ahead of the motor. (a) shows a specified force-position profile for a virtual bump; which is approximately accurate for low-speed manipulation. (b) is a sketch (not real data) exemplifying the recorded handle position and interaction force as a function of time for two cases: (1) is for slow manipulation, where the bump is clearly felt. This is apparent from the slowing velocity as the peak if the bump is neared. (2) illustrates the case for rapid manipulation. The requested force is slewed too late and too little, and the handle passes through the continuity with little apparent disturbance. (c) shows the effective force-position profile for the virtual bump for high-speed manipulation, when slipthrough is in effect.

reduced specified force on the far side of the bump. Slipthrough is exaggerated by high bump aspect ratio (stiffness/width) because for the same velocity of manipulation, a shorter time period is required to pass through the bump and the motor has less time to respond.

Some quick experiments verified that for this system, critical slipthrough speed corresponded roughly to force slew bandwidth. That is, I found that slipthrough occured when I passed through an entire bump in under ~5 msec, whereas the force bandwidth for this actuator I had independently recorded at around 100 Hz, representing a time constant of ~10 msec. It is likely that the relation will hold for other systems.

I did not observe high-speed slipthrough in virtual bumps composed of changes in virtual damping; in fact I found the opposite effect of the bump becoming imperceptible as manipulation velocity approached zero.[4] For bumps in B_d, force ramps up with velocity rather than position: thus while there is still a lag as a function of change in velocity, if the manipulation is at uniform high velocity a large impedance change will be achieved. Thus, one non-physically based solution to the slipthrough problem is to combine damping and stiffness to achieve the perception of the desired "stiffness" at different speeds of manipulation. The cost is that the impedance feels slightly damped or sluggish at low-speed manipulation.

Direction-Switching Hysteresis

Several environment features developed both here and by others rely on direction-dependent effects. Examples are the non-stick virtual wall where damping is turned on while entering the wall to enhance its crispness but turned off while backing out to avoid the sticky feeling that damping otherwise causes [167], and the one-way spring regions which I used in my Toggle environments (see below).

Implementation of such effects requires knowledge of direction and direction changes. This is

---

[4] This is not a hardware anomaly but the correct physical behavior; the absolute change in impedance for a system of different levels of pure damping should feel small at low manipulation speeds.

most easily accomplished by monitoring the sign of the velocity signal. In the absence of a velocity sensor I estimated velocity by differentiating a filtered position signal, producing a result with small amounts of noise and phase lag.

Under such conditions and with no other processing, velocity sign-change transitions are blurred due to noise while the filtering causes them to be slightly late. The result is buzzing at low velocities as the velocity sign crosses zero nearly every sample and the model switches accordingly; depending on the model, the response could be more severe. The problem is worsened at high sampling rates, because if velocity is computed by

$$\begin{aligned}\hat{v}_k &= \frac{\hat{x}_k - \hat{x}_{k-1}}{T_k} \\ \hat{x}_k &= \text{Filtered position signal at time } t_k,\end{aligned}$$

as it was here, the denominator is small and the discretized velocity characterized by very large spikes unless further smoothed.

One way to improve this situation is to use a velocity sensor and get a cleaner and more timely signal; however, there will still be an indeterminate state around zero velocity where undesired model switching will occur. Therefore it is necessary to use either a hysteresis element or deadzone around zero velocity, the size of which is determined by the noise or discretization amplitude ($\sim 1/T_k$) and inside of which no switching occurs.

Difference in Z- and Y-Expressed Environment Behavior

The following observations pertain to development of the Z and Y-expressed environments, according to the computation described earlier. The same behavior will not necessarily hold for other expressions.

Energy Conservation:

> The Y environment expression used here was more successful than this Z expression in correctly reproducing model losses. For example, consider a virtual model of a damped mass-spring system with damped natural frequency and resting time determined by values of M_d, B_d and K_d. The handle is flicked with an initial disturbance and then allowed to oscillate unheld. Under these conditions, the Y-expressed model oscillates at the correct frequency and with an amplitude envelope which shrinks at the correct rate. However, the Z-expressed model feels about the same as Y model when held; but upon release, it oscillates at a higher frequency and tends to slowly increase in amplitude of oscillations with an envelope which is apparently unbounded.
>
> The key to this difference in behavior is robustness of the two expressions to small non-idealities in sensor output. The Y-expressed model operates off of measured interaction force which with the handle unheld is small, correctly reflecting the desired losses situation. Little or no energy is input, and thus the initial mass-spring potential and kinetic energy gradually dissipates.
>
> In the case of the Z-expressed model, the model causality now acts against it because of imperfect sensor readings. Recall Eq. 5.1; in a real Newtonian system obeying this law, when $f_{interaction}=0$ energy passes periodically between kinetic and potential while being steadily dissipated through the damping term. But in this system and to some degree with any system, the sensors exhibit a small degree of noise and other non-ideal effects.
>
> This causes problem at two points, the virtual model input and closing of the servo loop. Model input, a_m and x_m, include both normal sensor error and further reflect the degree to which the virtual system has already diverged from desired behavior. They describe the kinematics of

a physical system slightly different from the desired one and which is not necessarily passive, and based on those sensor readings Eq. 5.2 does not necessarily evaluate to zero.

The servo block of the controller acts closed-loop on desired and measured force. However, force sensor output is not a perfect estimator of the real $f_{interaction}$. It is affected not only by normal electronic noise and discretization error but it also registers a small-amplitude periodic signal due to acceleration of the sensor's cantilever,[5] oscillating in the same direction as the $M_d a_m$ term. f_m and f_d are thus both nonzero and in opposition, the closed loop control signal grows and the resultant behavior is nonconservative.

Different Role of Virtual Damping in Z, Y Expressions in Discontinuous Environments:

For discontinuous environments such as the Toggles, virtual damping had a different effect when expressed as Z or Y. In Z virtual damping behaves the way one would expect it to, i.e. by increasing the perception of viscosity and smoothing the sharpness of the stiffness peaks. In Y, however, the role of damping depends on the overall order of the region. If the region is order one, then Eq. 5.5 is used and B_d is in the denominator. The larger is B_d, the more numerically stable and "solid" the environment; conversely, an attempt to go from an order zero system (Eq. 5.5) with no damping to an order one system with a small level of damping means a more jittery and perhaps unstable virtual environment. Thus a substantial level of B_d has to be introduced before the increase begins to feel like a modulation of damping.

Sensitivity of Environments to Filtering/Noise:

This Z environment expression appeared more robust to signal noise than the Y expression. I found that I could run the Z emulations of all tested environments (Masses, Sliders and Toggles) with no signal filtering (position, force or acceleration) and at 1000 Hz sampling their performance was undegraded. Some Y-expressed environments became unstable in the absence of filtering (150 Hz cutoffs for force and position). This became an issue particularly in the development of low-sampling-rate emulations for experimentation; as the sampling rate dropped its Nyquist neared the digital filtering cutoff and the filtering requirement thus constrained the low-end sample rate which could be tested.

Interactions of Servo Controller Gains and Environment Performance

Gains May be Scheduled Based on $Z_{virt\ env}$:

I found that different servo controller gainsets were required to obtain comparable levels of performance (I cannot claim that they were optimized) in the emulation of different environment impedance levels. That is, different gainsets would perform better for M_d=50g, 100g and 200g; the same held true for other types of impedance (stiffness and damping) and for mixed impedances (mass-spring-damper systems). The task of finding the proper gains each time one wishes to emulate a new high-fidelity environment quickly becomes onerous. I found that I could do the job once by determining gains for 6 to 8 impedance "bins" and then computing a generalized impedance $Z_{virt\ env}$ value for the environment to enter this gain lookup table.

I did this by starting with the relation

$$Z_{virt\ env} = \frac{K_d}{\omega_{excitation}} + B_d + M_d \omega_{excitation}. \tag{5.6}$$

---

[5] The force sensor operates through a strain gauge mounted on a small cantilever; deflection of the environment system's handle deflects the beam and distorts the strain gauge. However, the cantilever and attached emulator handle together have nonzero mass and will also deflect in response to inertial loading.

An estimate of $\omega_{excitation}$, the frequency at which the virtual system is excited on average, is required to evaluate Eq. 5.6. I found that the bin boundaries I had determined already for each of the impedance types (e.g. values of $[K_dB_dM_d]_{low}$, $[K_dB_dM_d]_{med}$, $[K_dB_dM_d]_{high}$, etc.) gave a reasonable value of about 13 Hz when for each bin, $\frac{K_d}{B_d}$ and $\frac{B_d}{M_d}$ were evaluated and the entire set were averaged ($n = 6$). 13 Hz is not far from what people can achieve in voluntary motion, so this value is reasonable.

Using $\omega_{excitation}$, I could increase the number of bins and further optimize performance at each generalized impedance level without the effort of scheduling gains for each new environment. I did, however, find that in some cases better performance was still achieved by using different gains for B_d-dominated and M_d-dominated virtual impedances.

Gains, Sample Rate and Fidelity are Related:

In the course of developing and attempting to optimize the experiment emulations, I found that there was often a rough relation between the gains and the resulting emulation fidelity as defined by the closed loop servo controller error: up to a certain point, increasing gains decreased servo error and increased the recognizability of the emulation's target. Conversely, decreasing sample rate decreased the value of gains which could be used and still achieve stable emulations, and this in turn degraded emulation fidelity.

This phenomenon was problematic in terms of the experiment design, where sample rate was one of the tested emulation variables and gains were not. I wished to hold all untested emulation variables constant; but in order to use the same gainset at all sampling frequencies and achieve stable emulation at all frequencies, I had to use the lowest gainset for all frequencies and thus compromise performance at the higher sampling rates. This problem is a complication of any experiment which attempts to single out a few variables from a larger set, all of which might interact with one another.

Integral Closed-Loop Gains Increase *Stability*:

Increasing integral servo controller gain[6] tended to increase virtual environments stability. Others have also found that integral control increases stability in force control tasks [37, 195, 207], but integral control increasing stability in addition to reducing steady state error in position control is less well documented. My experience is that the chief benefit occurs in the case of low-impedance environments, when the input signal (f_m for Y expression) is generally small. The signal to noise (S/N) ratio is small, and the model output ($\vec{X}_d$ for Y) is noisy. A control signal acting on integrated closed loop error tends to be smoother and this benefit outweighs the phase lag introduced, particularly when the control signals are small. By the same argument, I found improved performance for low impedance environments by reducing or removing the derivative gain term, K_{der}.

For Z expression, increasing K_{int} improved matters in the case of high impedance emulations when kinematic signals (model input) are relatively stationary. But a large integral gain helps for low impedance environments here as well: f_m, now input to the closed loop servo controller, has the small S/N while f_d is small and possibly noisy as well. Once again, K_{int} helps smooth the noisy control signal.

Effective Gain Magnitudes Related to $\left|\frac{Z_{virt\ env}}{Z_{emul\ system}} - 1\right|$:

---

[6] Here, K_{int} for a PI (on force) or PID (on position) controller.

Less surprisingly than other observations documented here, I found that I required higher gains to achieve low CL error (high fidelity) emulation when the virtual environment was very different from the actual passive characteristics of emulation system (i.e. inherent mass and physical damping). When any component of targeted impedance was quite close to system impedance, very low gains could be used.

This meant that it was crucial to use different gainsets for significantly different physical damping levels, as during development of the experiment emulations.

High Physical Damping Mandates Large Closed Loop Gains for Stability:

Less intuitively, I found that in the presence of large amounts of physical damping, some types of virtual environments required very large servo gains for stability; nonintuitive because for this hardware and many other systems, decreasing servo gains generally increases stability. This was because without the large gains the closed loop error term became large (due to the large passive impedance) and often went out of phase. Higher gains kept the closed loop error small and in phase.

Haptic Virtual Environment Modeling Strategies

This section includes a collection of "recipes" for the modeling of interesting haptic phenomenon. None of these models were tested in the human subject experiments, although their development was part of the process of developing the most difficult experiment emulation, Toggles.

"Free" (Minimal Impedance) Environment

"Free" systems were created simply by using a Z model expression and servoing to zero closed loop force error. The result was very low impedance, markedly less than the system's passive impedance (mass and and any level of damping) and perceptually close to the reference of pushing through air. The form of Eq. 5.4 precluded achieving a similar effect in Y control. Thus whenever I expressed an environment as Y which included free regions I switched between Y and Z expression with no deleterious effect at the transition, so long as Y integration states were properly maintained and/or re-initialized at transition.

Walls

I experimented informally with a wide variety of virtual wall modeling strategies, including combining stiffness, damping and mass elements in series and parallel permutations and using directional turnoff of damping elements to avoid sticky walls as did Rosenberg and Adelstein [167]. With my hardware, I obtained the best wall performance (most realistic tradeoff of crispness and passivity) with different models for these Z and Y expressions. In both cases I found that walls emulated by ideal components in parallel worked better than did serial elements combined with directional damping turnoff, primarily because the size of the deadzone required to avoid buzzing in direction switching made the directional turnoff distracting. Use of a velocity sensor rather than differentiated position to trigger direction switching may produce better results with that strategy.

My best Z walls were parallel spring-dampers (K_d=20 N-s/mm and B_d=1.0 N-s/mm), but my best Y walls were modeled as parallel spring-masses (K_d=20 N-s/mm and M_d=2000 g).[7] In both

---

[7] I found in my Toggle models (described in Chapter 7) that the hard toggle ROM constraints were likewise most stable when Z expressions were modeled as K-B and Y as K-M, with parameterizations of about the same magnitude as isolated walls.

cases, the stiff discontinuity was approached from an adjacent free (Z expressed) region. The additional high-order element (damping and mass respectively) was necessary to ramp up force rapidly on wall contact. In the Y expression, I had great success using inertia as the primary impedance component while the stiff spring absorbed the energy which would otherwise have eventually accelerated the mass; the result was quite solid, crisp and passive. However, using virtual damping in Y walls either with or without mass produced more active contact behavior.

Using mass in Z walls had the unfortunate effect of dragging the handle into the wall because this Z expression has no mechanism for enforcing conservation of momentum, and thus on transition the perception is of suddenly picking up a large mass which is already moving at the contact velocity rather than of striking a stationary mass.

Detents

The best detents were composed as illustrated in Figure 5-3 but with the selective addition of damping and mass elements to increase contact impedance and passivity without making the result too "mushy". I tried many other strategies of arranging length and stiffness and equilibrium placement of K_d regions, and experimented extensively with negative springs; these mechanisms can be used to produce a wide variety of effects.

In general, I found that to get equivalent performance from Z and Y expressions, I could use similar model structures but often had to re-parameterize to get the same perceived impedance. This was a pervasive result in my modeling of discontinuous systems; combinations of different impedance elements (M_d, B_d and K_d) in geometrically short (0.5-10 mm) discontinuous adjacent regions produced impedance perceptions which were rarely closely related to the same generalized impedance used in a continuous region of length of a couple of centimeters or more.

Negative Springs

A virtual device for which there is no physically realizable equivalent proved valuable: spring elements with a negative stiffness which produce a reaction force that tends to further "compress" the spring away from its equilibrium rather than pulling or pushing in the direction of equilibrium. These elements elements were useful in increasing the snapping effect in an environment such as a stiff toggle switch, where a local potential energy maxima is achieved by deflecting the handle into a stiff region following which the handle must be "snapped" with as much kinetic energy as possible to the next local equilibrium. Due to the aforementioned attenuation of impedance elements in the presence of short discontinuous environment zones, it was difficult to achieve such sharp energy transitions using conventional, energy conserving elements.

Hysteresis and Backlash

In addition to the requirement of deadzones for direction-cued model switching, it was possible to explicitly emulate hysteresis and backlash effects with much the same virtual mechanism by setting deadzone size, monitoring velocity and adding a direction-dependent offset to the model output. The backlash effect was of "rattling" at direction changes.

Chapter 6

The Real Devices

The human subject experiments were based on a treatment unit in which one emulation was compared with set of real devices; the real devices were similar except for a single aspect of their haptic feedback (here, a single component of impedance, e.g. viscosity or mass) which differed in a controlled manner from one real device to the next. Three families of real devices (Masses, Sliders and Toggles) were tested at four levels each. The four devices belonging to a single family (Z_A, Z_B, Z_C and Z_D, where Z_A was of lowest and Z_D the highest impedance in the set) were similar in appearance and qualitatively in their haptic feedback; they differed in mass, viscosity and peak spring force respectively. This chapter describes the real devices, the reasoning which went into their specification and issues relating to their construction and virtual modeling:

6.1 Real devices used in this study, their justification and construction;

6.2 Virtual environment models designed to emulate them, as well as model parameterization issues;

6.3 "Switch box" mechanism designed and built to present the real devices to subjects during experiments.

6.1 Real Devices Used in Study

Criteria for Selection of Real Device Families

Three distinct types of single degree of freedom interface devices were used for the human subject experiments in order to study a variety of both impedance types and complexity. The families were chosen to span the variety typically found in real linear-acting interfaces (Sliders, Toggles) and also included inertia (Masses), considered a difficult environment to produce virtually. The families are dominated by inertia, damping or compliance, and Masses and Sliders are "simple" environments (continuous and describable by a single model parameter).

Having two experiment targets for which modeling and high quality emulation was clearly feasible, I wished to take a chance on the third by attempting a very "complex" (discontinuous and multiparametered) and high frequency environment which would be difficult to model with high fidelity (Toggles). Although success in terms of bulletproof experiment data was not assured, I felt it was time to explore the more open-ended questions embodied by difficult emulation problems. If a "perfect" model and emulation cannot be created, what is next best and how far will it take us? This cannot be easily asked of the single-parametered continuous systems, where we can come close to perfect emulation.

Once I had chosen dominant impedance characteristics and interface type, the remaining specifications for the real device families centered around modularity of design, controllability of impedance within families and maintenance of a reasonable parts and labor cost for producing eight members for each family (four in a set, duplicated once). The experiment design described in Chapter 8 requires randomization of real device presentation order from trial to trial, so the devices had to be modular and easy to switch (maximum switching time of one minute for an entire set). Four to six were to be tested together, and therefore this many devices together with the emulation had to fit in comfortable reach of a subject's single hand.

To hide the identity of the real devices from the subject while switching, I had either to make the subject move or to move the switchbox between each trial; since the experiments were already long and onerous for the subjects, the latter was the preferable alternative and this meant the devices needed to be light and untethered to any power, pneumatic or hydraulic source (ruling out the possibility of an air table for the Masses). The actual mechanism and any clue as to its function or impedance magnitude had to be well hidden from the subject. The final and in the end most difficult requirement was the devising of families whose individual member impedances could be controlled and measured.

The following three sections describe the process of meeting these specifications for each family; eight members of each family were built. Part specifications and vendor information may be found in Appendix D, along with the working drawings for parts which were designed and built here.

6.1.1 Rolling Masses

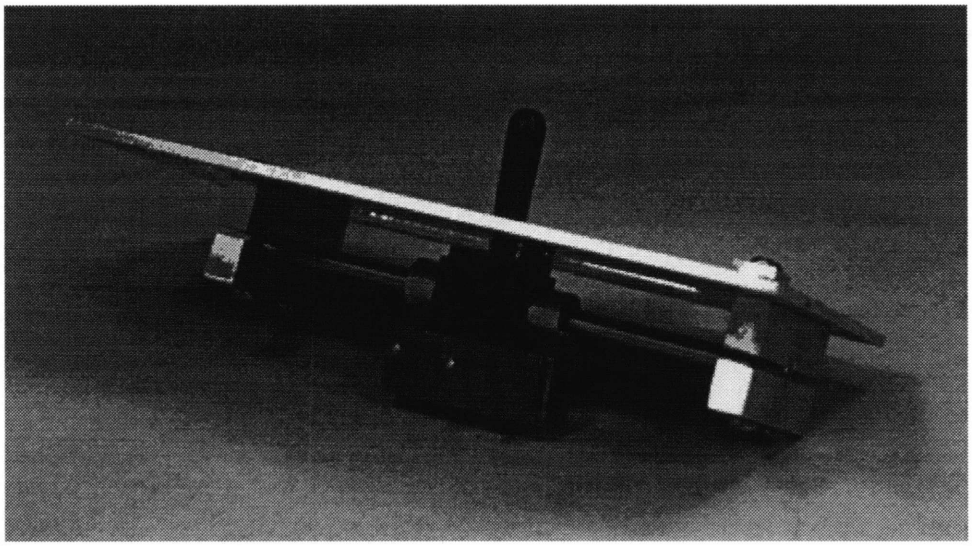


Figure 6-1: The real mass module, mounted on a plate for use in the switchbox.

As with many ideal mechanical quantities, it is trivial to write an equation of motion for a pure inertia ($F = Ma$, the "hockey puck gliding on ice" of freshman physics) but more of a challenge to build a compact, low-cost device which comes close to obeying that simple equation over an ~ 80 mm range of motion. The design which best satisfied the numerous constraints was a stack of lead slugs rolling on linear ball bearings (Figures 6-1 and 6-2). A ball-groove shaft was utilized to constrain the bearing's vertical orientation; using two shaft/bearing pairs to accomplish this would have substantially increased both the design's parts cost and its required machining precision.

The lead slugs, each weighing approximately 1/2 oz or 14 gm, can be added and removed in pairs

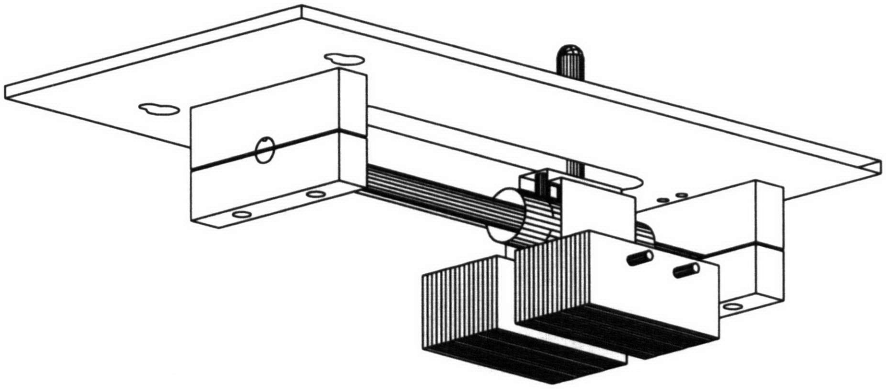


Figure 6-2: Scale diagram of the plate-mounted real mass module, viewed from underneath the plate. Lead slugs are hung on a linear ball bearing cartridge which rolls horizontally on a 1/4" steel shaft.

(for balance) to modify the total rolling mass. The moving mass of bearing cartridge, casing, handle and hangers is about 1 oz (35 gm); the wire hangers can hold up to 32 slugs (16 on each side) for a total mass range of about 35–500 grams and controllable mass resolution of 28 grams. The range of rolling motion is 3.5" or 88 mm, although only 1.5" or 38 mm ROM was used in the experiments (hard stops constrained the motion).

The finished modules exhibited low damping and friction and apparently close to pure inertial dynamics, and comfortably satisfied the requirement of easy, reproducible and verifiable control over module impedance. Design shortcomings included a faint high frequency vibration due to the recirculating ball bearings which could be felt and heard, and an unanticipated degree of lateral (rotational) bearing play. The ball groove was supposed to prevent lateral motion but such constraint required tightening the bearing (via a setscrew) until unacceptable friction resulted. The mass module handles, constructed from 1/4" aluminum rod, were coated with rubber in order to both disguise blemishes on the handle tips and to damp the transmission of bearing vibration to the user's hand. The coating succeeded in the former aim but its damping properties were insufficient for the latter.

6.1.2 Viscous Slider Switches

A member of the Slider family, shown mounted in Figure 6-3, is based on inexpensive linear slide pots such as might be used in home-built audio equipment; each has a range of motion of about 2.5" or 63 mm.

When purchased, the sliders were friction-dominated, somewhat nonuniform in impedance across their travel range but all in roughly the same impedance range. To achieve uniform viscous rather than nonuniform friction dissipation and some degree of controlled variety in damping level, I substantially modified the purchased devices. I disassembled the sliders, removed and/or modified various internal elements, reassembled them and sealed the case with epoxy. At this point, the sliders were all very low friction and nearly massless (under 3 grams moving mass).

The final and crucial modification was to incorporate high viscosity fluid (1k–100k cStokes; see Appendix C for fluid specifications) into the sealed case; the resultant damping level depended on fluid viscosity and amount of fluid. High damping levels were difficult to achieve because with very high viscosity fluid (60K–100K cStokes) the fluid was pushed to the case ends and replaced with air bubbles and thus the damping quickly became low and uneven. To get a distinguishable range

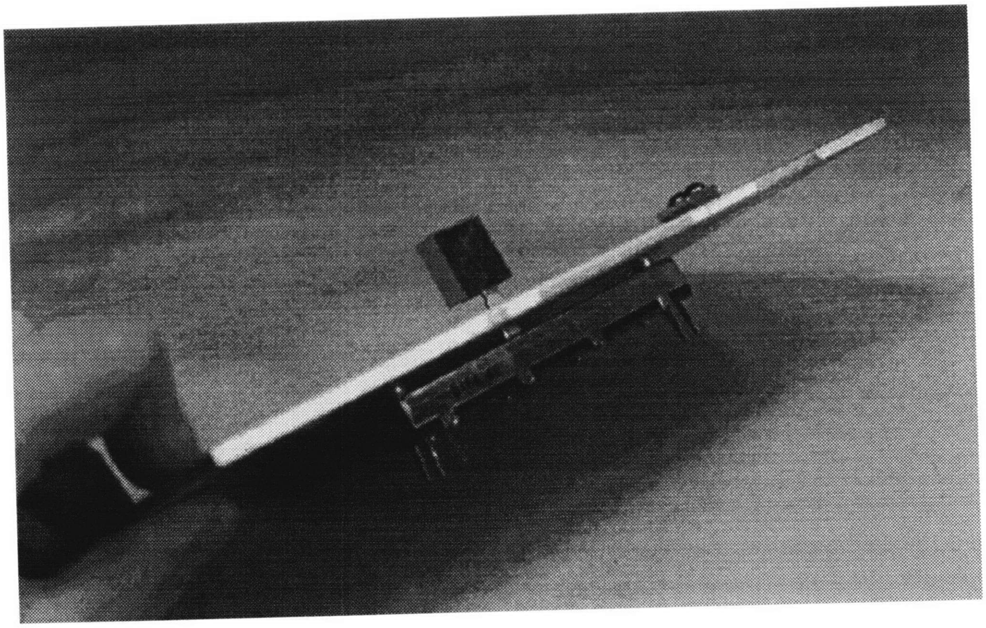


Figure 6-3: The real slider switch module.

in device impedances, I had to revert to a different assembly of slider components for the highest damping level, and this assembly proved to be higher impedance than desired, stretching out the entire range of device impedances.

The set of sliders I used in the experiments covered a reasonable range and distribution of impedances (0.001–0.032 N-s/mm), with the exception of the high value (Z_D) being a little high; and the impedance was well modeled as a pure viscosity. The drawback of this construction approach was the difficulty in controlling damping level to exact values, in reproducing damping levels (for duplicate devices) and in accurately measuring the value of B_d which characterized a given device. Unlike the Masses where a lead slug and/or the entire moving mass could be weighed, specification and parameterization was far from exact.

6.1.3 Spring Toggle Switches

Due their ubiquity in user interfaces, I had many choices in types of spring switches. The best solution was a set of large two-position toggle switches of the sort commonly found in electronic laboratory equipment. Figure 6-3 shows a plate-mounted example, and Figure 6-5 illustrates the toggle mechanism. By beginning with a set of toggles with identical handles but different number of poles (2, 4, 6 and 8) I got some range in stiffness but there were only two distinguishable levels and the snapping was so stiff and abrupt that I was unable to emulate it.

I modified the toggles by filling the cases with high viscosity fluids of different viscosities and levels and sealing them. This had the twofold effect of spreading out the impedance range as a function of viscosity and amount of fluid, and smoothing the snapping effect to the point it could be virtually reproduced although for the higher impedance devices it was still quite sharp and stiff.

After some iteration, I was able to produce toggle switches of four distinct impedance levels which could be emulated with adequate faithfulness. However, these switches suffered the same drawbacks as did the Sliders in that their impedance was difficult to control and characterize. Thus the duplicate members were not identical and the model parameterizations were crude at best. Because of the complexity of the mechanical structure, modeling itself was challenging (Section 6.2); but it was that

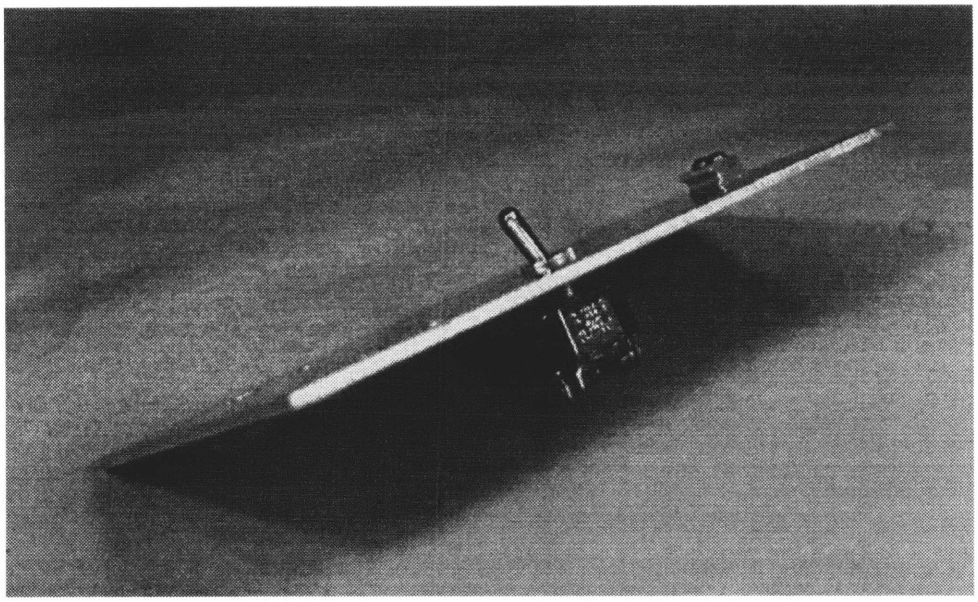


Figure 6-4: The real toggle module.

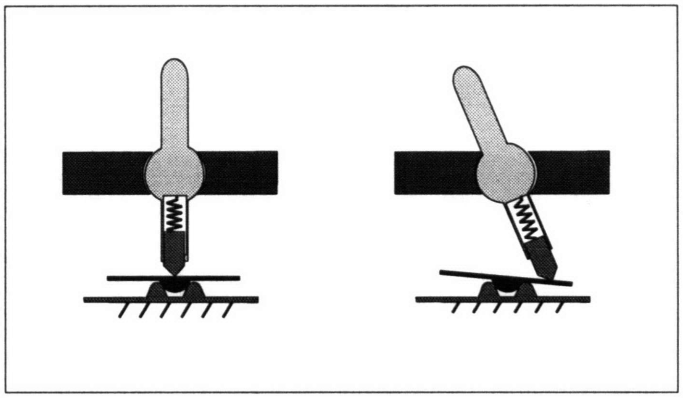


Figure 6-5: Toggle snapping mechanism.

challenge which was sought in choosing a complex device for testing.

6.1.4 Members Used for Each Family

According to arguments of assumed Weber Law perception[1] which are expounded more fully later in Section 8.2.2, the specified impedance distribution for the real devices was logarithmic (Figure 6-6). That is, the increment was to be doubled for each successive device pair given the maximum and minimum impedance in the set (Z_A and Z_B).

The method I used to approach if not fully achieve this target distribution was iterative. I began at the low end of the scale (Z_A) with the smallest impedance which I could realize in both emulation and real device; and at the high end (Z_D for a four device set) a significantly different magnitude. In between, I attempted to place Z_B close enough to the low end so that anyone would have to

---

[1] According to Weber's Law, various aspects of the human perceptual scale is logarithmic: the ratio of JND/stimulus magnitude is approximately constant [62]. Thus the difference between impedance units of "1" and "10" and between "10" and "100" should be perceived as similar.

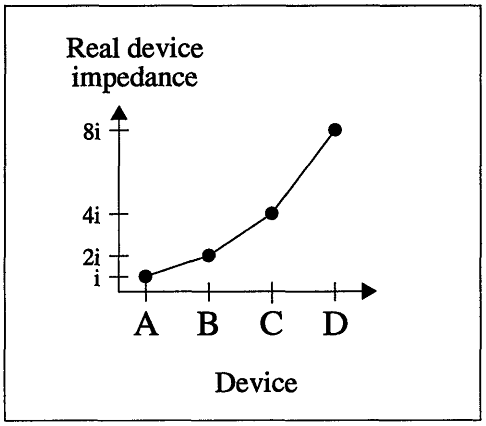


Figure 6-6: The impedance distribution of the individual real devices is specified to be approximately logarithmic.

concentrate to distinguish them, but most people could eventually make a correct identification. The last device, Z_C, I put about halfway between Z_B and Z_D.

If I found through informal experimentation with myself and other subjects that Z_A and Z_D were not different enough to allow four distinguishable values, or were so different that the C/D discrimination was going to be obvious at all times, or that the Z_B and Z_C placements did not fit the log profile, I iterated by attempting to move the scale ends in or out and/or moving Z_B and Z_C around inside.

Impedance Distributions Achieved

The actual real device distributions were somewhere between uniform and logarithmic (Figure 6-7). The first row (a-c) shows the absolute magnitudes of the four devices' impedances in real units; (d) shows the same curves normalized and translated to a scale which will be useful in later processing, an order of magnitude range [0.1–1.0]. All curves show ideal log and linear curves (dotted lines) to demonstrate the nature of the impedance spacing that was achieved, to the extent that the numbers used here reflect the true impedances of the real devices.

Comparison of Figures 6-7 (a-b) make it immediately apparent that relative impedance distribution is not the only variable at issue here. The values shown graphically are listed numerically in Table 6.1, with the ratio of the maximum and minimum impedances in the set ($Z_{\max}/Z_{\min}$) as well as the relative spacings which were actually used (Δ_{act}) next to those which would have resulted in a log spacing. Note that there is substantial variation in $Z_{\max}/Z_{\min}$, from 5 for Masses to a high of 32 for Sliders.

Variation in the ratio between $Z_{\max}$ and $Z_{\min}$ is not necessarily a bad thing, depending on how the ranges were arrived at. As long as the low end of the impedance scales for each of the families are not grossly different and the desired log spacing is attained, the result should be reasonable. There is no reason to expect either that the low end of the average human's perceptual scale will be at exactly the same impedance magnitude for different kinds of haptic environments (e.g. pure compliance vs. pure masses), or that a person's ability to discriminate different devices will have exactly the same resolution for those environments. (On the other hand, without other evidence one would expect them to be similar.)

In fact this log spacing was not always possible, and the degree to which it was departed from for specific families will render comparisons between that family and the others suspect. Features of the distributions achieved here are the following:

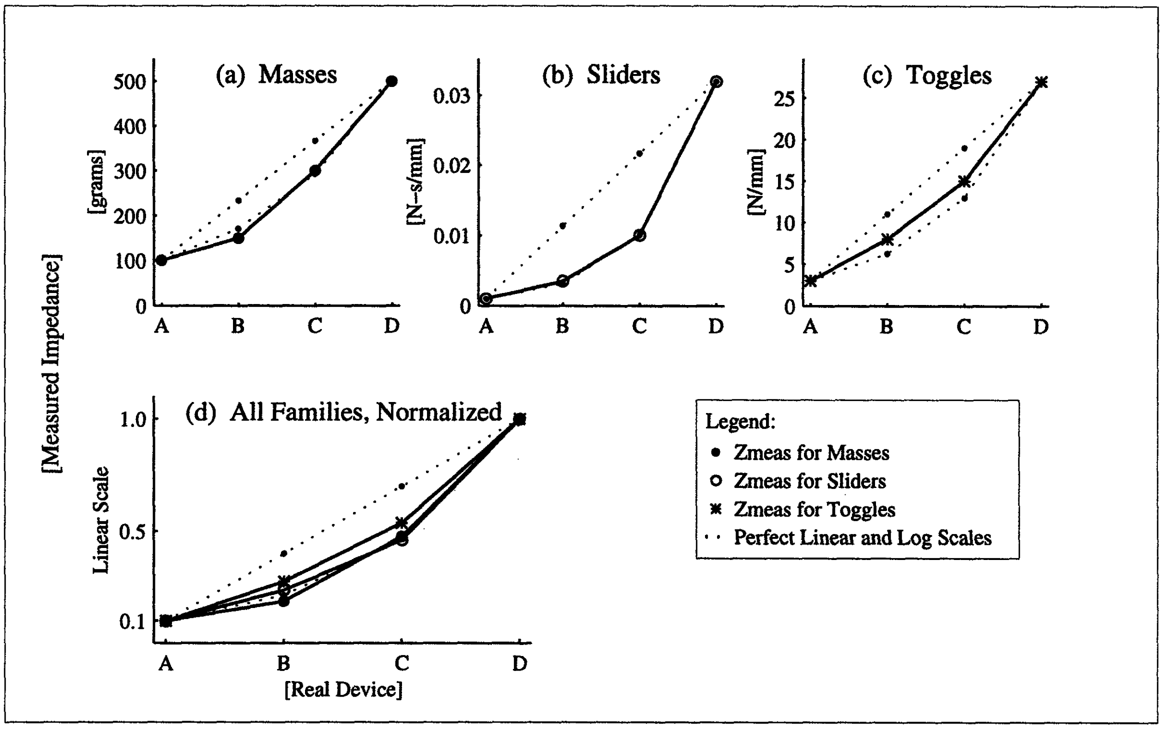


Figure 6-7: Mechanical impedance distributions achieved for real device families.

The Sliders spacing is skewed (dished) to an almost perfect logarithmic distribution. This is attributable to the fact that a slightly different mechanism had to be used to attain a Z_D which was perceptibly different from the first three, and this mechanism had a higher damping than was desired. This combined serendipitously with the fact that the differences between the first three were not well controlled (Z_C was also higher than desired) to result in a distribution that is almost perfect log, but much more widely spread than desired.

Masses and Toggles are less skewed than logarithmic, but closer to logarithmic than linear.

In some cases, *a single point determines the shape of the distribution.* e.g., with Sliders the distribution would be much less dished if it weren't for the highest-impedance value, Z_D; with Masses, Z_B is a little low.

	Masses [grams]			Sliders [N-s/mm]			Toggles [N/mm]		
	Value	Δ_{act}	Δ_{log}	Value	Δ_{act}	Δ_{log}	Value	Δ_{act}	Δ_{log}
Z_A	100	50	57	0.001	0.003	0.004	3	5	3.4
Z_B	150	150	114	0.004	0.006	0.009	8	7	6.9
Z_C	300	200	229	0.010	0.022	0.018	15	12	13.7
Z_D	500			0.032			27		
$Z_{\max}$-$Z_{\min}$	400			.031			24		
$Z_{\max}/Z_{\min}$	5			32			9		

Table 6.1: Values for mechanically measured impedances of the real device sets. Δ_{log}shows the impedance step which would have given a log distribution between the highest and lowest impedance in the set (Z_A and Z_D).

6.2 Modeling and Parameterization of Real Device Families

General Observations on Modeling

The range in device family complexity meant a range in difficulty of modeling and parameterization. Whereas the simple device families (Masses and Sliders) were relatively straightforward to target with high fidelity emulations, the Toggles required a complex model with a large set of descriptive parameters. I knew a priori that some Toggle features could not be perfectly emulated by this virtual environment system, due to inherent bandwidth limitations; the modeling challenge was to simultaneously stabilize the emulation and create a haptic illusion of the real toggle switches. The result in this sort of problem is non-physically-based model features whose efficacy may be judged only through some manner of perceptual experimentation. More will be said on this subject in Chapter 7.

The following sections describe the models and strategies used to model each of the real device families, list the model variables used in each actual emulation and comment on performance achieved in each case. Following the Masses, Sliders and Toggles modeling descriptions are accounts of some issues encountered in the modeling/parameterization process: non-correspondence of measured and perceived emulated mass, and the handling of auditory input.

Masses

Emulation Strategy

The masses were deliberately chosen to be a simple, single-parametered and continuous system so that the ability of the emulations to match inertia would not be confounded with other effects. The environment model was straightforward. For Z,

$$f_{d_k} = M_d a_{m_k}$$

and for Y,

$$\begin{aligned} a_{d_k} &= \frac{f_m}{M_d} \\ v_{d_k} &= \int a_{d_k} dt \\ x_{d_k} &= \int v_{d_k} dt. \end{aligned}$$

The only modeling decision came in the manner of representing the ends of the emulation range of motion (ROM). The real device ROM was physically constrained by the rolling mass hitting the stops at the ends of the shaft, and I initially tried to match these stops with virtual walls at the ends of the masses. While this was entirely doable, I found that the quality of the virtual walls varied more with modulation of the experiment factors than did the mass emulation itself, and thus posed a distraction (I wished this device to be a test of inertia emulation, not of walls). Therefore I replaced the virtual wall constraints with approximately zero impedance "Free" regions (see description in next section) which felt perceptibly different than the activated regions, felt about the same for all emulations but bore no qualitative resemblance to the real device constraints and invited no comparison. I also marked the activated region ROM visually on the emulator hand plate with colored tape, and instructed subjects to try to keep the emulator handle within the active region and make their assessment based on that region, not the low-impedance end regions.

The Masses were parameterized by weighing the moving mass (bearing/handle assembly and lead slugs) and calibrating the emulator sensors.

Performance Achieved

Both the subjects and I found that the emulations felt quantitatively similar to but qualitatively a little different from the real devices for high virtual impedances; the emulations were fairly close to our references of an ideal lossless mass, a "hockey puck gliding on ice", whereas the real devices had some real-world effects such as bearing vibration[2] and small amplitude lateral rocking.

The undegraded[3] emulations for Y and Z expressions were qualitatively similar to each other for high virtual impedances, but diverged slightly for low impedances. Z expressions were very smooth and almost active whereas Y expressions often had a slightly buzzy texture, apparently from the velocity sensor coming in at the PID position controller; both were stable.

Sliders

Emulation Strategy

The situation for modeling the Sliders device family was much the same as for Masses. Again, the virtual model was described by a single parameter although the Y model was in this case first instead of second order. For Z,

$$f_{d_k} = B_d v_{m_k}$$

and for Y,

$$\begin{aligned} v_{d_k} &= \frac{f_m}{B_d} \\ x_{d_k} &= \int v_{d_k} dt. \end{aligned}$$

As for Masses, I marked the ends of the activated region ROM by low impedance rather than an attempt to match the real slider end stops and by visual markers on the emulator handplate.

Performance Achieved

The results of the undegraded Z and Y emulations were both qualitatively and quantitatively similar to one another, and according to my own perception and that of experiment subjects, they were best out of all the experiment emulations in qualitative matching to the real devices.

Toggles

Emulation Strategy

The Toggle family was the only complex real device tested (containing severe, closely spaced discontinuities and described by multiple model parameters), and the effort required to model it with any degree of fidelity was much more than for the Mass and Slider families.

---

[2] I put some effort into trying to reproduce the texture of bearing vibration in order to improve the qualitative match of emulation to imperfect real device, but it proved nontrivial and I considered it a distraction from the original aim of testing emulation ability to accurately convey the sense of inertia.

[3] "Undegraded" emulation turned out to mean the highest tested sampling rate (1000 Hz) and lowest damping level (none added).

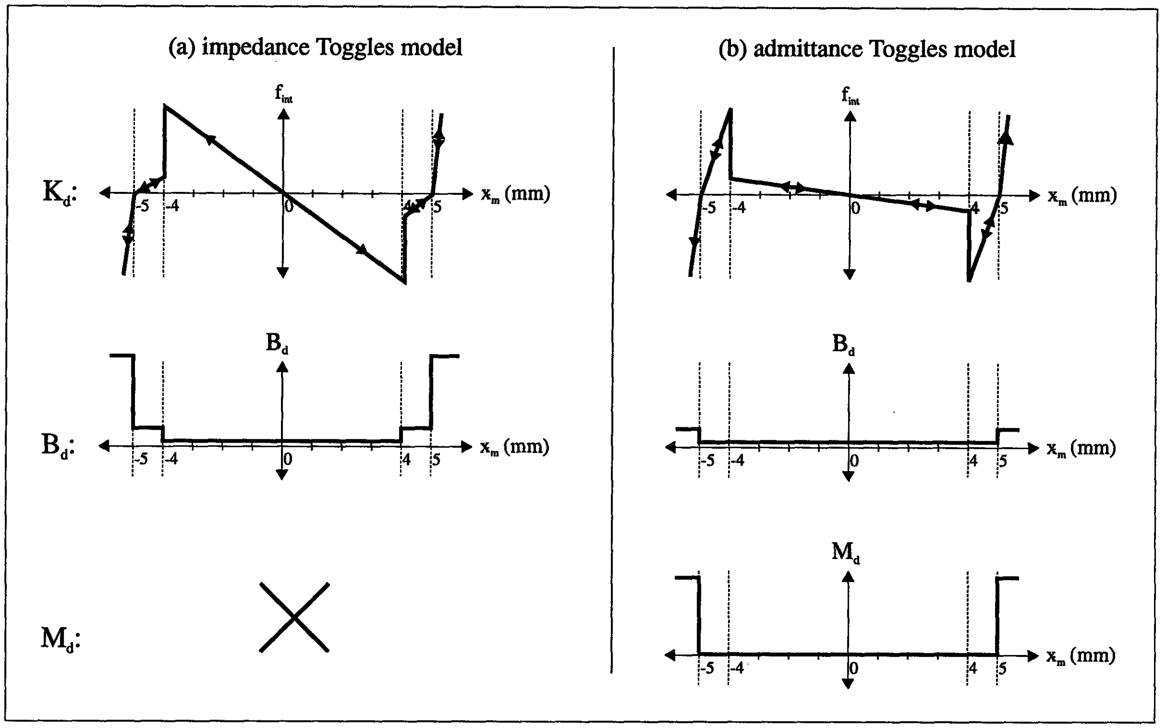


Figure 6-8: Models used to emulate the Toggle device family, for (a) Z and (b) Y expression. The first row shows stiffness components and the second row damping components; the admittance expression also uses a mass component for the toggle end-constraints. In both cases, the models are piecewise continuous but the stiffness components are sometimes active only in one direction; arrows indicate the direction of motion for which the model component is in effect.

Although many strategies were tried including a continuous model based on the actual mechanism as diagrammed in Figure 6-5, the piecewise continuous profile-matching approaches shown in Figure 6-8 eventually gave the best results. Slightly different models were used for Z and Y expressions; note the use of negative springs in the latter.

Performance Achieved

The undegraded (1 kHz sampling) results of the winning strategies for Toggle emulations bore a quite intelligible resemblance to the real devices, but they were not as high fidelity as the simple device emulations. There was some qualitative difference in feel between Z and Y expressed emulations; Y toggles felt "snappier" but this came across as a sense of increased activeness rather than increased crispness.

Disparity in Perception of Real and Emulation Impedances

Of all the real device families tested, the Masses were most straightforward to accurately parameterize. However, a significant and consistent perceptual bias was observed: both Z- and Y-expressed mass emulations felt about 50-70 grams heavier than the real devices of the mass targeted by the emulation, potentially for all levels tested between real masses of 35 and 500 grams.[4] Closed loop

---

[4] The magnitude of the bias was uncertain at the high end of the scale (Weber's Law again); it is more difficult perceptually to detect a delta of 70 grams at 500 grams than at 30 grams and my informal experiments were not thorough. However, there did appear to be a bias of some size at the 500 gram level as well as at the lower absolute

error was small in both position and force control for lower frequency sensor signals (up to perhaps 5–10 Hz manually induced oscillation), implying that the servo controller was doing its job at those frequencies. The consistency between expressions which rely on different model input sensors suggests that the obvious culprit, sensor and amplifier calibration (carefully re-checked), was not at fault.

Although I am unable at this time to explain the observation, it seems to me potentially significant. I do not know of other studies which compare perception of real and emulated physical quantities in this way. I hypothesize that the discrepancy might be due to a difference in the quality, as opposed to magnitude, of what is felt; for example, unmodeled dynamics in the real device or subtle errors in the servo control which contribute to a different perception of mass. Possible explanations include:

Activeness in the emulation:

> This argument is counterintuitive; one would guess activity would make the emulation feel lighter than modeled, a sort of "power assist". However, the unconserved energy might also appear as a resistive force; it is not modeled and we don't know exactly what it is doing. This might be a good reason to try to model it.

Unmodeled friction in real devices:

> The presence of friction in the real devices, of which there was a small amount (neglected in my model), might cause the real device to feel lighter than the emulation. If you flicked the real device and let it go, it came to a halt more quickly than the emulation; by extension, when you "tossed" the real device back and forth from finger to thumb, it took less finger-energy to stop it than for the emulation. It also took more finger energy to get it going at the same initial rate as the emulation, but this is not necessarily perceived in the same way.
>
> To test this hypothesis I implemented a simple friction model in the emulation, and iteratively matched the time-to-rest following an input energy pulse from my hand. The emulation did not feel lighter as a result of either this or a variety of virtual damping model elements, so I did not use use a dissipative element in the experiment emulations.

Energy dissipated by rolling ball bearings:

> In modeling the real masses, I approximated its linear motion as sliding rather than rolling and neglected the extra energy used to accelerate the recirculating rolling ball bearings (about 80 bearings per real device) and/or lost as heat. This energy sink might influence the perception of inertia by the same argument as above, although through a different energetic mechanism.

Closed-loop-control error at sharp discontinuities:

> Although there was low closed loop error for low frequency emulator manipulation, error was greater at points of impulse force or position inputs to the emulator — e.g. bumping the handle or tossing it back and forth. The servo controller was unable to slew force or position quickly enough to instantly bring such spikes in measured signals down to desired levels. Many individuals, including myself, rely heavily on the reaction of the device to such impulse input in evaluating its mass, in effect measuring the effort required to change the object's direction.
>
> Thus, the explanation for the perception offset might be simply mismatching of manipulator[5] and virtual impedance. The intrinsic inertial properties of the actuator are felt when its force-

---

masses.

[5] The moving mass of the actuator and handle is about 240 grams, in the middle of the range of virtual mass being discussed (35–500 grams).

slewing bandwidth is exceeded, most noticeably when the emulated mass is much smaller than the emulator's passive impedance. This does not explain the "constant" 70 gram offset, but there was uncertainty in that assessment especially at higher virtual masses.

With respect to the Mass model and parameterization used in the experiments here, I had no choice but to subtract the perceived 70 gram offset from the emulations; for example, to emulate a real mass which weighs in at 100 grams, I used an emulation M_d of 30 grams. There would have been no point in performing experimental comparisons when I knew ahead of time that subjects would perceive a bias.[6]

It is also worth observing that this problem was not encountered in modeling the other device families, because there was not an equally trustworthy means of physically parameterizing them. That is, such a perceptual offset might have existed and I would not have detected it because my primary means of parameterization was already my own haptic perception and that of pilot subjects.

6.3 Presentation of the Real Devices: Switch Box

6.3.1 Specifications

The specifications for the switch box arise from following needs:

Blind presentation: The subject must not be able to distinguish real devices within a device family based on appearance. Thus the device body must be hidden from view during use.

Quick module changeover: Devices have to be swapped between each trial. To reduce overall experiment time, swapping should take as short a time as possible; target was under 1 minute.

Rigid support: The real device modules must be supported solidly and rigidly, so there is no rattling or rocking induced by their manipulation.

Compact: All devices used in a single trial (up to 6) must fit into an area which can be comfortably reached by one of the subject's hands while seated in a single position.

Modularity of device mounting: Each device must be mounted on an interchangeable module so that different device families may be used with the same manner of presentation.

Portability: In between trials, the real devices have to be switched outside of the subject's view. Thus the presentation device must be light enough to carry a few steps across the room and cannot be tethered via electric cords, pneumatic hoses, etc.

Simple design: The design must be composed of inexpensive parts and be quick to build.

6.3.2 Design

The switchbox design is shown in Figures 6-9 and 6-10; working drawings and part information may be found in Appendix D. It accommodates up to 6 device modules, occupies a 7"×24" footprint and is about 3" tall. The chief innovation was in devising a quick-release mechanism which would permit rapid switching and locking of the device modules; thumb screws were too slow and velcro not strong enough. The solution was based on set of cam-action latches mounted on the back (distal side, furthest from the user) of the switch box. Each module plate had a small catch screwed permanently to its distal end and a pair of slotted holes at the proximal end which slipped over matching anchor

---

[6] Figure 6-7 and Table 6.1 as well as all analysis in Chapter 9 uses the real (weighed) mass values, rather than the smaller offset value used in emulation.

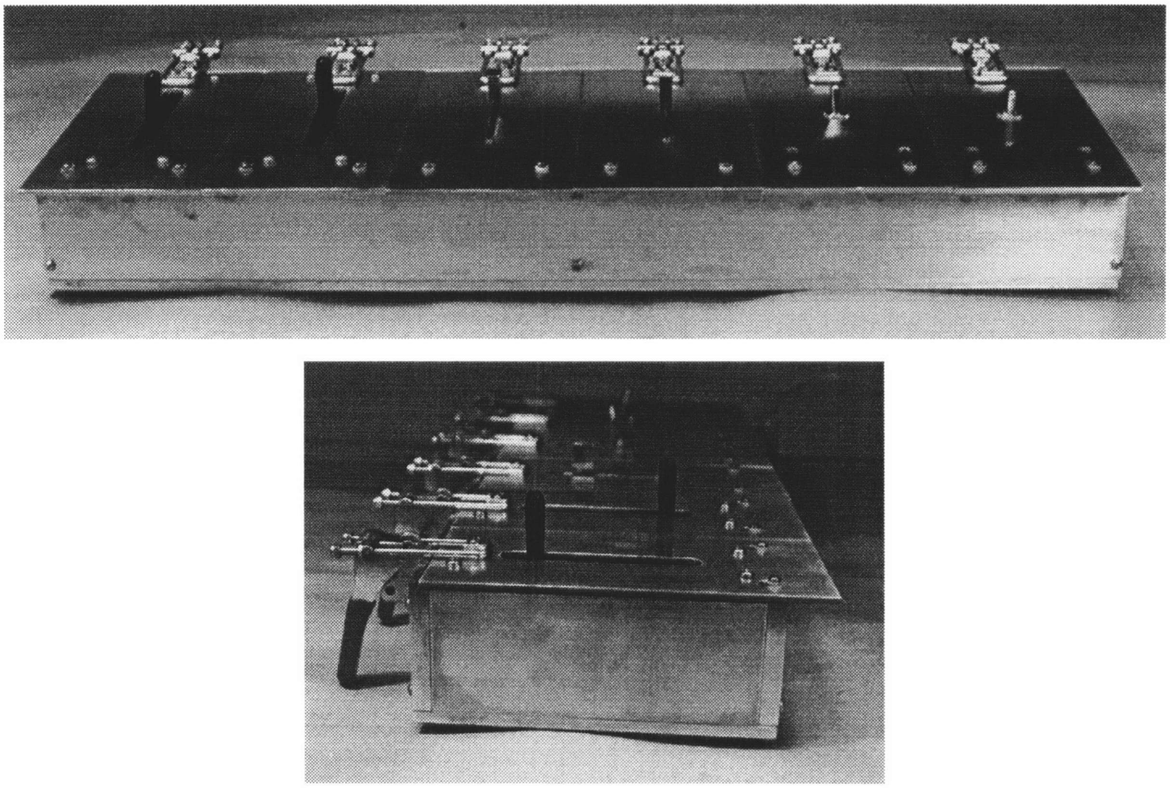


Figure 6-9: The experiment "switch box".

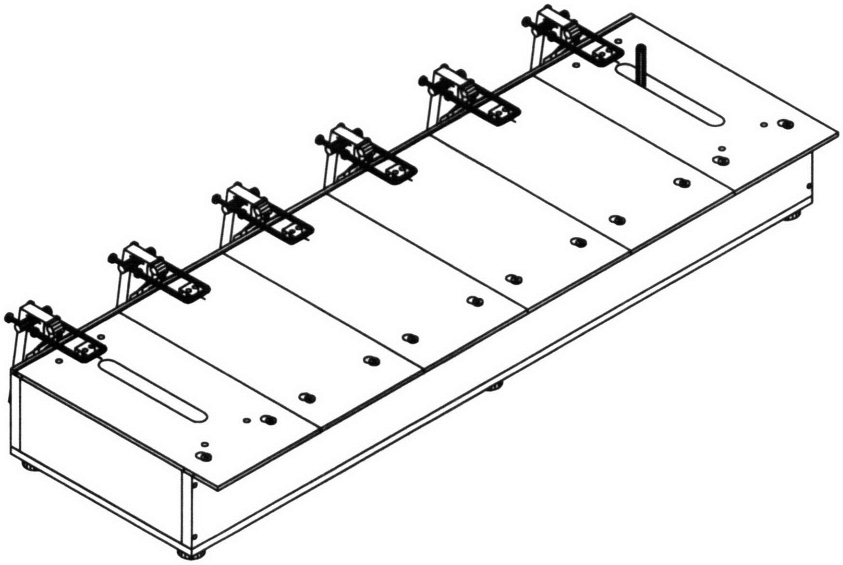


Figure 6-10: Scale diagram of the experiment switch box. This drawing shows the six spaces filled with an empty slotted mass plate on the left end, a filled mass plate (handle shows) on the right end and four blanks. In normal usage 4–5 of the spaces are filled with device modules of the appropriate family, and the remainder with blank plates as shown here.

posts on the box frame. The plate was held in tension by the closed latches between the anchor posts and the latch catch and the result was quite solid. The only component added to the high volume elements (switch plates) were the latch catches, which were trivial and very inexpensive to build; whereas only six of the more expensive components (latches) had to be purchased.

The switch box was constructed entirely of aluminum to enhance its structural solidity. Unloaded, it weighs about 5 kg and a full complement of device modules adds no more than another 2 kg; it can thus be comfortably lifted and moved repeatedly during experiments. The interior was lined with acoustical damping material because the aluminum frame turned out to be an excellent acoustical amplifier and exaggerated the snapping sound of the real toggles.

Chapter 7

Identifying Impedances of Real Devices

This chapter describes a problem and need that arises when one adds the specification of "fidelity to a particular real device" to the list of challenges in creating a haptic virtual environment, and suggests a variety of solutions.

One of the solutions presented here is an automated technique developed to identify real environment impedances for the purpose of re-playing them as a force display. Its algorithm is based on probing the real device with a sequence of tightly controlled position trajectories while measuring the forces that the device generates in response.

7.1 "Fidelity" Means Starting with a Good Model

The emulation can only be as good as the model and parameterization of the real device, but a good model means different things for different tasks.

What People Feel

When the task is to predict behavior of a mechanical system under specified excitation conditions, mechanical engineers have accumulated a broad experience base as to what constitutes an adequate model, and many tools with which to make and parameterize models. However, the task here is to recreate what people feel, or rather an input which stimulates the same sensations. Those in psychophysics are working on a model for how people feel, and others on putting it in a form which engineers can use. But at this time, we don't even have precise knowledge of what we feel, let alone how.

It is logical to start with the assumption that for the purpose of creating a haptic emulation, a human's haptic perception (what he feels, all he feels, and nothing but what he feels) is identical to what an engineer sees when he/she examines a mechanical system and attempts to capture its essence with a diagram of interconnected masses, springs, dampers and other elements. But in general, this gives an incomplete picture. Engineers are in the habit of neglecting "higher order dynamics" to make a problem tractable and because these dynamics aren't responsible for gross features of the system's behavior. But people are very good at feeling, and often those h.o.d.'s are responsible for a significant aspect of a person's haptic perception. Without them the model does not feel right and it may not even feel real.

Modeling for Feeling

Engineers neglect higher order dynamics because they are difficult to model and parameterize. In the context of using the model as a basis of a haptic emulation, they are hard to play back. High-frequency nonlinear phenomena such as wall-tapping and stiff detents demand high performance hardware and sophisticated control.

But even simple environments can be difficult to model and play back. Suppose that you have a simple viscous-dominated slide potentiometer such as those used in this research and described in Chapter 6. Assumption #1: viscous dominated. Is it really? It may be completely dissipative, but pure viscous damping is rarely encountered outside of engineering scratchpads. "Mostly" is not "completely", and chances are that the part that is viscous is not linear. And according to Murphy's Law, the part that people pay most attention to will turn out to be the part that isn't viscous *or* linear — the nice smooth linear viscous component is just a backdrop for the messy part that they do notice.

But suppose that the viscous-dominated assumption is good enough. How do you characterize "B" for a small potentiometer with a small and inconveniently shaped handle and no good way to hold onto it? You could claim to know the characteristics of the viscous fluid (if you believe the spec sheet), compose a fluid-dynamics model based on viscous shear, and hope the air bubbles don't get you. If you end up with an answer in the correct order of magnitude, you're doing quite well.

Or you could measure the impedance of the device directly. This is probably more accurate, but it is not necessarily easy either. You'll need a verifiably constant-velocity input (several, to make sure B is linear with velocity), a good force sensor, a probe to push on the handle, a way to rigidly anchor the slider while you measure it, and a realtime controller to make all of this happen. This is just to measure "B". It is, in fact, the basis of the impedance identification technique recounted later in the chapter.

Sometimes the device really is easy to model and measure; for a pure mass, all you need is an accurate scale. But you better be sure that it really is a pure mass (are you sure dissipation isn't important in how it feels?), and that the emulation playback will faithfully recreate your simple model.

7.2 Methods of Identification

Here are some comments on ways one can go about modeling a real device's impedance for the purpose of emulating it:

"Guess": Iterative Perceptual Tuning

Assume a model (not necessarily based on the actual mechanism) with parameters which seem like they should match the perceived force profile, and then rely on your own haptic perception to iteratively close the perceptual gap between the emulation and the real target.

This approach is low overhead; all you need is your hands and a rough idea of the mechanism. And in the end, this mode of measurement (haptic perception) is the only one that really counts towards emulation fidelity. If other modeling approaches give a result that feels close but not quite right, iterative guessing will inevitably have to be used to bring it in line.

Disadvantages are many. The reliance on haptic perception is good for fidelity but poor from other standpoints (product prototyping, for instance). Haptic perception is not necessarily the most accurate or high resolution measurement tool, and certainly it is not the most objective or repeatable. Because it is iterative, the method is time consuming, and if the model is at all complex and/or physically based, the connection between parameterizations and output feel might be indirect. It is entirely possible to get "stuck" with an emulation that doesn't feel right but no obvious way to get to a better one. The model constructed in this manner can probably be simulated mathematically to verify response, but if it does not closely resemble the real mechanism this might not lend much insight into how the model could be changed to improve fidelity.

The subjectivity problem can be mitigated by making the iterative comparison task one of discrimination rather than magnitude estimation. As in the experiment comparison model of Chapter 8, use multiple real devices with impedances varying impedance (or other haptic characteristics) which bracket the real device being targeted by a single emulation. The developer may then consider whether the emulation feels *more* like Z_B than like Z_A or Z_C, rather than whether or how much the emulation feels like Z_B. Having the reference difference in addition to the reference absolute greatly clarifies the evaluation process.

Divide and Conquer: Structural Modeling

Model the real device mechanism and parameterize it by mechanical measurement of individual components.

For the spring toggles used in these experiments, this would mean disassembling the toggle switch body and measuring the spring constant for the spring in the actuator, the friction coefficient of the actuator tip sliding on the seesaw and all geometric dimensions. Then put it all back together and hope the reconstructed real device feels the same as it did before. This model may also be mathematically simulated to see if the predicted behavior matches that of the real target, probably with greater benefit than in the previous method since the model can be compared with the real device for errors; e.g. unmodeled dynamics could be identified as an explanation for divergence in behavior, and added to the model.

Then, try to emulate that model. Note that correct behavior of the mathematically modeled mechanism alone is no proof that that model will be stable when it is run in a virtual environment controller and displayed on a system with impedance of its own (although the emulator plant can be added to the simulation to help with that). Moreover, the mechanical parameterization is only as good as the model, and since we don't know exactly what people are perceiving, it is hard to be confident that the model is really all there.

Active Probe

Systematically identify the impedance components of the fully assembled real device using a computer controlled force probe. One algorithm for this approach is described in more detail below.

This approach is a black box, to an extent; you supply an input to the system and measure the output. To turn the transfer function into a model of a mechanical system, however, you must make an educated guess at the structure of what is inside. This makes it more like smoked glass than absolutely opaque.

The system identification technique has the advantages of being objective, relying entirely on mechanical measures; and contextual in that it measures impedance at the same location that a person manipulating the device will feel it, at the handle. It does not try to predict output feedforward at the handle based on a model of what is downstream. In a successful implementation applied to an amenable identification problem, it is quick; the method developed here could carry out data collection and analysis for one parameter identification in about 20 seconds.

Its disadvantages are in its high overhead of hardware and program development; it is really a brute force answer to the problem. Moreover, as implemented here it does not create a model structure, it merely fills it in (parametric model). A more powerful incarnation might be a stochastic identification which makes no such assumptions about structure (a non parametric model).

All of the Above

In practice, these methods need to be used in combination. Some knowledge of the underlying structural model is helpful as a starting point for an iterative perceptual approach, particularly if it is a simple structure and/or one whose primary features are easily captured in a model. Likewise, a model-based approach will need tweaking in any but the simplest case to make it feel right, because the model will probably be imperfect and emulator dynamics will modify the model output to some extent. The same points may be made for an automated probe identification. In most cases, it will need a model structure as a starting point and iterative perceptual-loop tuning at the end.

Whatever the approach used, creativity and patience is generally required.

7.3 Methods Used Here

In modeling and parameterizing the devices used in the experiments and also those described in Section 5.3, I relied on a combination of all three techniques described above. The automatic identification technique described in the following section was used for demonstration purposes in an earlier stage of this research. To automate the identification of the one hard-to-model device used in experimentation (Toggles), further development would have been required for both the algorithm and construction of the force probe and the combination of structural modeling and iterative perceptual tuning proved more expedient, if less objective.

In the parameterization of the Masses, I used the straightforward approach of direct mechanical measurement of the single impedance component of a system which was modeled obviously and very simply.

We observed in Section 6.2 that even this presumably accurate method and simple system resulted in an emulation that didn't feel quite like the real thing, and I had to revert to iterative perceptual tuning to get a reasonable result.

It is possible that the iterative perceptual tuning approach is the most accurate method of all, rather than the method of last resort.

7.4 The Haptic Camera: An Automated Technique for Identifying Impedances of Real Environments

As I designed the emulations described in the preceding chapters, two motivations emerged for finding a means of modeling and parameterizing real environment impedance characteristics which is more objective and expeditious than iterative perceptual tuning or structural modeling.

(a) Particularly for the purpose of the proposed experiments, it was desirable to have an objective means of measuring impedance rather than relying on my own perception; in the latter case, the experiments become to some extent a measure of the degree to which the subject's perception agrees with my own.

(b) The "manual" methods were time consuming and difficult to achieve consistency in.

The automated technique described here was developed to address the need to *record and play back* the haptics of a real environment.

Algorithm

The identification algorithm is based on the partitioning of real environment impedance into components and piecewise continuous regions, given an assumption of general model structure. Model parameterizations are extracted successively from the real environment, low order to high order, by applying tightly controlled position inputs of the appropriate type and then measuring and analyzing interaction force.

The probe in this implementation was the emulator itself, with its handle replaced by a short stiff probe (Figure 7-1); its stiff single DOF design makes it ideal for this kind of task. The device being measured was mounted on a stiff plate which was in turn mounted rigidly to the emulator frame with the device handle upside down at the same height as the probe.

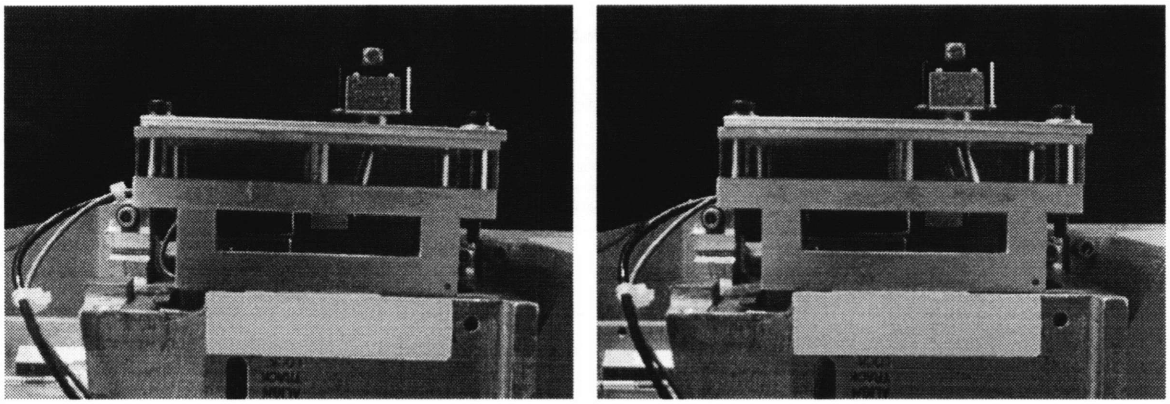


Figure 7-1: The emulator used as a force probe: data is collected by the identification program.

The data collection and analysis procedure is most easily explained using as an example an environment whose dynamics are at least locally linear, but allowing discontinuities or regions with different parameters. To identify mass, damping, stiffness and equilibrium points ($\hat{M}(x)$, $\hat{B}(x)$, $\hat{K}(x)$ and $\hat{x}_o(x)$, respectively) as a function of position or more specifically of region for such an environment through a linear acting point of contact, the following trajectories are used in succession:

Computation Step	Trajectory Type	Information Extracted
1	quasi-static	$\hat{K}(x) = \frac{\vec{f}_m}{(\vec{x}_m - \hat{x}_o(x))}$
2	constant velocity	$\hat{B}(x) = \frac{\vec{f}_m - \hat{K}(x)\Delta x}{\vec{v}_m}$
3	constant acceleration	$\hat{M}(x) = \frac{\vec{f}_m - \hat{K}(x)\Delta x - \hat{B}(x)\vec{v}_m}{\vec{a}_m}$

The actual process of obtaining the model parameters is more involved than the vector division suggested here; the most important step is in determining edges of model regions so that the piecewise continuous approximation may be made. Each trajectory can be duplicated as many times as necessary to reduce the effects of measurement noise. In the case of the velocity trajectory, it is helpful to use a variety of constant velocities in order to ascertain nonlinearity in B. The same reasoning might also apply to estimation of M; however, there is not much room over the few inches or millimeters usually available to vary acceleration, and nonlinear inertia is a less common phenomenon for simple mechanisms anyway.

Scope and Limitations

Although the example above was for an environment that could accurately be described as piecewise continuous, the same kind of algorithm could be used to parameterize nonlinear characteristics and features. For example, to characterize backlash a vibratory trajectory might be applied, and force hysteresis measured.[1] Stiction could be measured by repeatedly stopping and measuring breakaway force at different positions and breaking rates.

The operative term, however, is "parameterize". This algorithm measures parameters for an assumed model; it does not create the model structure. It can identify zero parameter values, indicating that that parameter was not a significant model component, but it cannot identify a parameter it was not told about. If the model provided does not describe the system well, this algorithm will mis-attribute impedance components (e.g., dumping all leftover impedance into its estimate of $\hat{M}(x)$ in the example above) and at best provide a statistic indicating the poor quality of the model fit.

Further limitations are inherent in the physical hardware which is used. In this iteration, the real device had to be mountable over a non-portable probe, and the probe itself could only push. This limited the types of features it could measure; hysteresis, for example, would have been problematic. There are many types of real environment ports for which a probe could be devised to both push and pull with little hysteresis or chatter, but making a probe to meet that specification for arbitrary environment ports would be a challenge.

In future development, it is possible that other more general methods of system identification may be brought to bear on this problem. For example, a frequency-rich stochastic input might be applied instead of the successive narrow-band trajectories, with less constraining assumptions about structure (and a push-pull probe). I did not begin with this approach because of complications in implementing it with the types of environments I wished to measure: their tendency to be geometrically narrow (sometimes only a few millimeters), full of even narrower discontinuous regions (where the discontinuities are of great interest) and constrained by hard stops makes a broadband input difficult to apply.

When To Use It

This method is most useful for systems which are difficult to parameterize by simpler means, and for which a reasonable estimate of model structure can be made. It excels at geometric location of subtle and higher order changes in parameters which can be felt but are difficult to decompose perceptually. It simplifies the segmenting of a model into regions by collecting and analyzing a clear multi-layered picture of system impedance.

It is less useful when there is a straightforward alternative or when the model structure is highly uncertain. Because the error is cumulative, characterizations will be less accurate when many parameters are attempted; even in the above linear example, the mass evaluation will be less trustworthy than the stiffness and damping estimates. Mass characterization is thus a good example of when not to use this technique — mass is both a high order parameter and usually can be simply and accurately weighed.

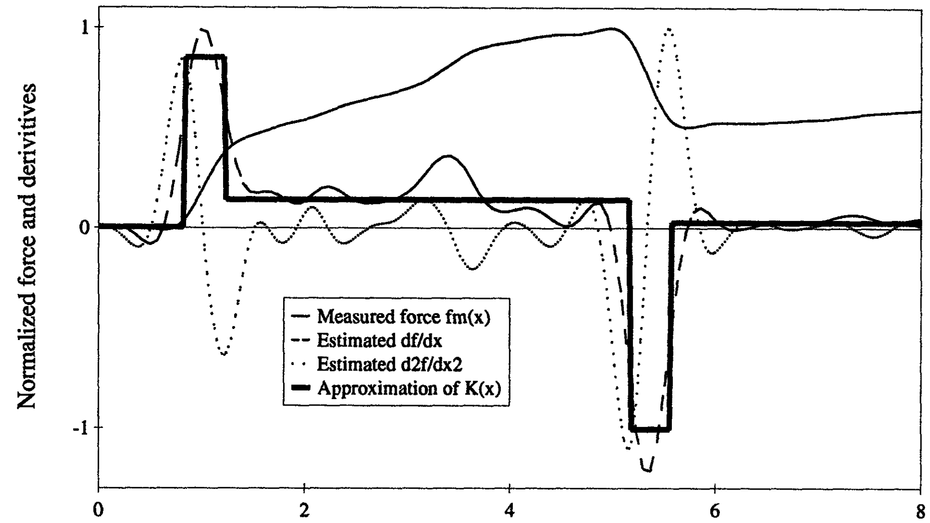
The algorithm may be strengthened in some cases by the addition of other sources of parameter information. For example in the mass-spring-damper example above, the device might be disassembled and weighed prior to the more general characterization. This value for $\hat{M}(x)$, more likely to be constant over the whole device range of motion than the other parameters, can then be substituted as a model input rather than output.

Implementation

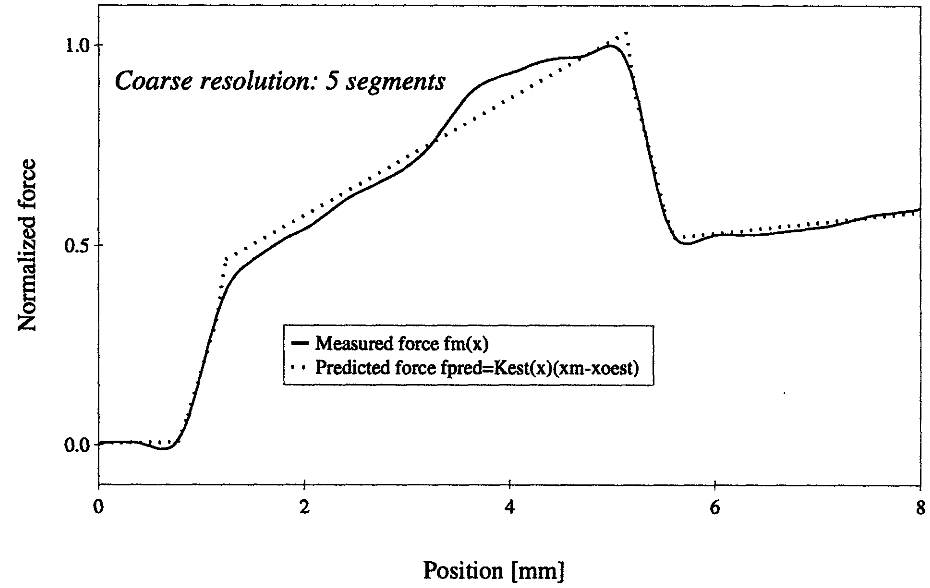
Figure 7-2 demonstrates Step 1 (characterization of $\hat{K}(x)$) of the automatic identification algorithm on a stiffness-dominated toggle switch. The switch, which can be seen in Figure 7-1, is a garden-

---

[1] For this to work, the probe, actuator linkage and probe/environment contact must have very little hysteresis itself.



(a) Measured Force and its Derivitives



(b) Estimated F(xm, Kest(xm))

Figure 7-2: The automated impedance identification algorithm on a stiffness-dominated example, set to find transitions at a coarse resolution (6 transitions found). (a) Measured force as a function of tightly controlled position ($f_m(x_m)$) and derived $\frac{d(f_m)}{dx_m}$ and $\frac{d^2(f_m)}{dx_m^2}$. The heavy segmented line are regions of constant $\hat{K}(x)$ as estimated by the algorithm. (b) Measured force ($f_m(x_m)$) and predicted output force ($\hat{f}(x)$) as a function of position computed using estimated $\hat{K}(x)$ and $\hat{x}_o(x)$.

variety "momentary" toggle; that is, when pushed it snaps to a local momentary equilibrium but springs back to its normal undeflected position when released.

The data collected for this characterization consists of the solid curve in (a), the measured force/position curve for this device ($f_m(x_m)$);[2] the normal (open) position is at the left end of the plot. Initial contact is seen in the steep positive force slope at around 1 mm; travel from there to peak force at 5 mm is gradual deflection of the spring toggle. Just after 5 mm, the handle drops "over the top" into the metastable momentary position. Data collection was halted at around 7 mm of deflection, just before the hard back stop of the switch was struck (data collected after that point is from a stationary probe).

Smoothing and acausal differencing [43] produces the dashed and dotted curves in (a), which are $\frac{d(f_m)}{dx_m}$ and $\frac{d^2(f_m)}{dx_m{}^2}$, respectively. Region transitions are located based on super-threshold local maxima in the second force derivative; and finally, an estimate of stiffness ($\hat{K}(x)$) together with corresponding equilibrium points ($\hat{x}_o(x)$, one point for each region of approximated constant stiffness) is made by finding the average slope and x-intersect of $\frac{d(f_m)}{dx_m}$, respectively, for the region.

Figure 7-2 (b) shows the functional result of this estimation procedure. The solid line repeats $f_m(x_m)$ from (a) as a reference; the piecewise constant-slope dotted line is $\hat{f}(x)$, estimated by running the x_m sequence through the estimated $\hat{K}(x)$ and $\hat{x}_o(x)$.

Control Over Region Resolution:

The characterization of Figure 7-2 produced a relatively coarse division of six transitions and five piecewise linear segments. This was not accidental; with this procedure it is easy to control the region resolution, by adjusting the threshold value of peak $\frac{d^2(f_m)}{dx_m{}^2}$ which is used to trigger a region transition (see following description of the transition detection algorithm). Figure 7-3 shows the same data set analyzed with a lower transition-detecting threshold; it produces 21 transitions and an even tighter match of $\hat{f}(x)$ to $f_m(x_m)$.

Use of the two models in emulation suggests that the lower resolution characterization is preferable, at least in this case. The small difference in output force as a function of measured position (or vice versa, for Y expression) is well below what people can detect, and in general it is desirable to minimize the number of regions in an emulation because region crossing tends to increase emulation activeness.

Auto-Transition Detection Algorithm:

The transition detection procedure is the key to the success of this identification method. It chooses transitions not based on a threshold of absolute magnitude which is somewhat arbitrary in where the cutoff occurs, but on a relative difference approach which locates and retains natural groupings.

The transition detection algorithm can be summarized as follows:

1. Begin with second derivative of measured force. We know that

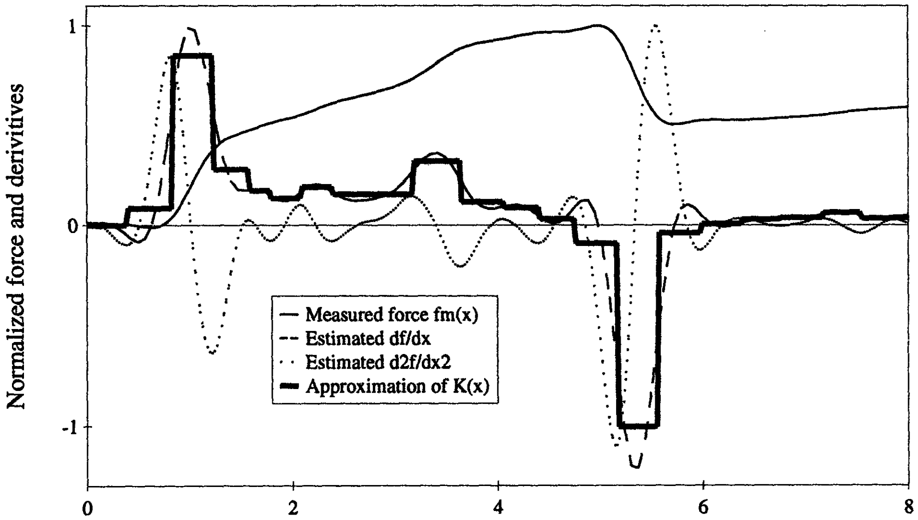
$$\frac{d^2(f_m)}{dx_m{}^2} = \frac{dK(x)}{dx_m}.$$

2. Locate all local maxima and minima in $\frac{d^2(f_m)}{dx_m{}^2}$, and sort in descending order of their amplitude's absolute value into a sequence $[\text{amp}_k, \text{amp}_{k-1}, \text{amp}_{k-2}, \ldots]$ where amp_k is the largest in the

---

[2] The data was actually collected as a time sequence, $f_m(t_k)$; if the position control had been perfect, for a constant-velocity trajectory $f_m(x_m)$ and $f_m(t_k)$would be identical. Small divergences from this ideal were handled by sorting $f_m(t_k)$ according to ascending x_{m_k}, which was collected simultaneously, and interpolating the sorted sequence to account for temporal non-uniformities.

(a) Measured Force and its Derivitives



(b) Estimated F(xm, Kest(xm))

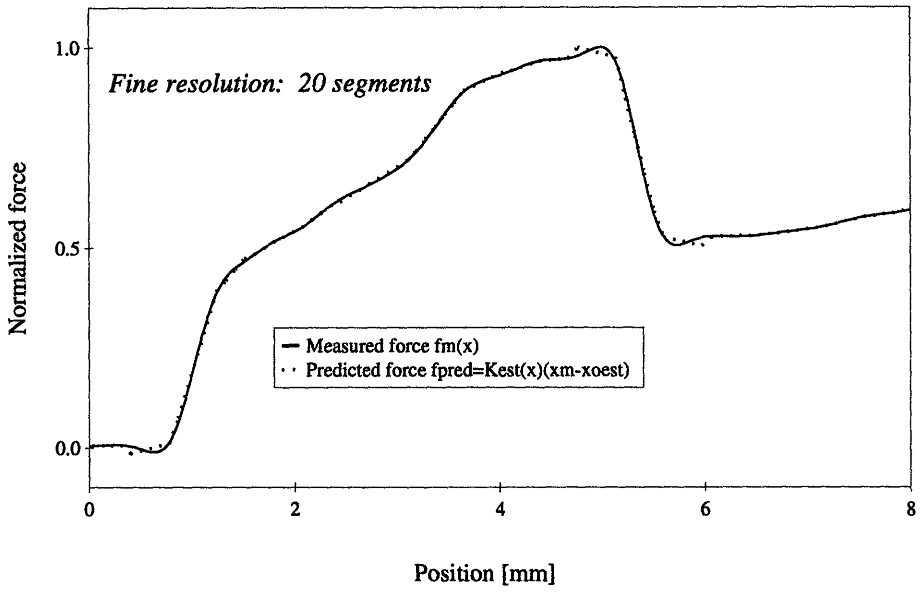


Figure 7-3: The same example and curves as in the previous example, but here set to find transitions at fine resolution (21 transitions or 20 segments found).

sequence. Region transitions correspond to these peaks.

3. However, there are too many. Select the important ones by setting a cutoff value for

$$C = \frac{\text{amp}_k}{\text{amp}_{k-1}},$$

that is, the ratio of each amplitude to the next lowest. This ratio locates transitions corresponding to divisions between natural groupings of amplitudes. A smaller value for C produces more regions.

In the example shown, the analysis which produced 6 regions used C=0.6, whereas the second analysis producing 20 regions used C=0.3.

Observations

This algorithm was tested to the degree demonstrated, i.e. characterization of stiffness only on a single real device. However, with this limited test set the results are highly promising. A completely automated assessment of $\hat{K}(x)$ and $\hat{x}_o(x)$, in playback with zero tweaking felt closer to the real device than anything I had produced previously through trial and error. With a little tuning (discontinuity handling, etc.; some of those modifications were standard treatment and could also be automated), it felt very close to the real device indeed.

The procedure also met the "expedient" criteria with ease, requiring under half a minute for data acquisition and analysis not including mount and dismount times for the real device over the probe.

Other Applications

I believe this algorithm will be tremendously useful in future efforts in the area of haptic virtual environments, and in other applications as well:

Emulation of Real Environments

As this field grows and matures, the applications where correspondence between virtual and real environments is required will expand; and in many of those cases, an accurate and objective means of identifying the real environment will be required. An example is the hands-on surgeon training system, whereby standard procedures are taught safely and haptically to novices and unusual specialized techniques demonstrated to expert surgeons. Another is NASA's need to train astronauts to handle tools in unfamiliar environments. The tool can be characterized, the strange environment modeled, and the astronaut's dexterity virtually educated.

Any Haptic Virtual Environment

Even when no direct correspondence to a real environment is required, learning how to produce those oftentimes difficult real haptic features will lend insight into creating richer virtual worlds in general.

Development of Mapping between Virtual and Real Environments

A use slightly different from the foregoing is the solution of the real/virtual mapping problem posed early in this thesis. For virtual haptic prototyping design tools to become a reality, it must be convenient to build the device which has been designed virtually. In many cases, the mapping from virtual model to physical parameters may not be direct because of modeling strategies used to overcome physical limitations of the virtual environment hardware. Here, an independent measurement technique can be used to define what an emulation output (the virtual feel which is the result of various possibly non-physically based emulation strategies) corresponds in terms of real physical characteristics. For example, an identification probe could be turned back onto the virtual device itself to characterize its output, and compare that identification with that of a passive real environment.

Production Quality Control

Returning to the haptic design application, we can find a use for this identification technique even when no virtual model is ever used in the design loop. It may be used as a tool for quality control in a manufacturing environment; for example, to make sure that all the Nissan Maxima dashboards are coming out feeling the same way and as intended. Imagine the QC inspector carrying around a handheld device and spot-checking dashboards by holding it up to individual switches: Buzz, buzz, a rating comes up on the small LCD display and is stored in memory while a beeper sounds if the interface is out of specification.

Psychophysics Research

A mechanical identification procedure may prove to be valuable in improving our understanding of haptic perception, by providing an accurate and independent measure of environments at the physical point that people feel them.

Part III

The Experiments

Chapter 8

Methods

The purpose of this chapter is to describe the experiment design. Its structure is as follows:

8.1 Overview: A summary of the experiment design, which is sufficient for the casual reader to understand its general structure and justification for critical features.

8.2 General Features: A detailed, generalized description of the experiment design and a procedure for analyzing the resulting data.

8.3 Specific Implementation: The details of implementing the generalized design described in the previous section for the present experiments.

8.1 Overview of Experiment Design and Implementation

8.1.1 Objectives in Experimentation

Motivation for performing human subject experiments of any sort was documented in Section 3.3.1. Here I discuss the more specific goals which guided the experiment design itself, relating both to the comparison model and the process of obtaining usable data from the human subjects.

Objective 1: Serve long-term goal of Virtual Product Prototyping

The long-term application for this work is haptic prototyping for human interfaces. This requires learning how to accurately reproduce essential, recognizable features of real environments such that users of the anticipated haptic design tools will be able to make subtle haptic distinctions between prototypes for real devices.

This objective of eventually creating an emulation which permits fine haptic distinctions between real devices is the philosophical basis for the following design decisions:

- "Haptic discriminability" was chosen as the primary performance index for the experiment.
- Emulation(s) were compared with real device(s) (rather than emulations with emulations or real devices to real devices).
- A [Multiple Real:Single Virtual] comparison model was used.
- Emulation variables were employed as experiment factors, in order to ascertain their influence on the chosen performance index.
- "Targets" (which real device in a set is targeted in a given trial) were designated and controlled as an additional source of variability into the experiment.

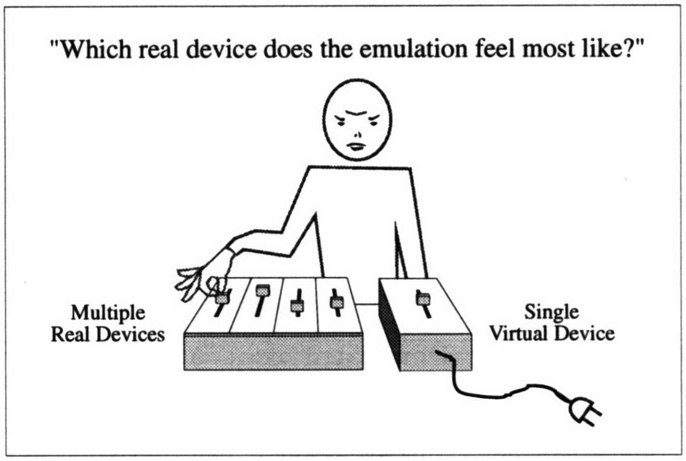


Figure 8-1: Experiments were based on comparisons between a set of real devices and a single emulated device which targets one of the real ones. In each trial, a different target and different set of emulation variable levels were used.

Objective 2: Obtain generalizable results from individual responses

- The degree of subject-to-subject variability and consistency actually present in a typical subject pool was explicitly measured and analyzed.
- Subjects were required to respond objectively in a multiple-choice task.
- As many different subjects were used as was feasible, even when not required from a standpoint of the experiment error estimate, in order to create a diverse subject pool.
- Some sessions were duplicated by a single individual to assess subject repeatability.

Objective 3: Minimize ambiguity in interpretation of responses

The first set of directives define the need and manner for obtaining subject perceptual scales:

- Subject perceptions of differences between a set of real or emulated devices is compared with the measured difference.
- Difficulty of virtual/real comparisons were measured relative to that of real/real comparisons within a set, in order to determine whether errors were due to emulation shortcomings or a resolution limit in human perception.
- The ends of magnitude-estimation style tasks were tied to known references, so that the value of the estimates would be clear.

General strategies for minimizing ambiguity:

- Subjects were encouraged to verbalize their reactions during all parts of the experiment.
- A formal place was provided in the protocol for subjective responses by using a questionnaire at the end of the session with open-answer questions.

Objective 4: Monitor quality control

- The divergence of actual design details from their specified values was documented and measured.
- Results were examined in search of patterns which might suggest such divergences when they are difficult to measure.

Emulation Variables Tested ($N_{factors}$ = 3 + Targets Factor):

Factor	Symbol	Number of Levels	Values
Virtual Environment Expression	V	$v = 2$	Z, Y
Physical Damping	D	$d = 2$	0.001, 0.070 N-s/mm
Sampling Rate	S	$s = 3$	300, 600, 1200 Hz
Target	T	$t = 4$	A, B, C, D

Device Families Tested ($N_{dev\ fam}$=3):

Name	Description	Dominated by:	A	B	C	D	Units
Masses	rolling masses	mass	100	150	300	500	[grams]
Sliders	slide potentiometers	viscosity	.001	.004	.010	.032	[N-s/mm]
Toggles	spring toggle switches	compliance	3	8	15	27	[N/mm]

Number of comparison trials per full replicate:
$N_{comp\ trials} = v \times d \times s \times t = 48$

Full replicates are used (all combinations of all factor levels tested):
$N_{trialsperreplicate} = 48$

Subjects = repetitions: $N_{reps} = N_{subjects} = 6$

Complete blocks (full replicate performed in a single session, or block):
$N_{total\ for\ device\ family} = N_{comp\ trials} \times N_{reps} = 48 \times 6 = 288$ trials per family

$N_{dev\ fam}$ device families tested:
$N_{total} = N_{dev\ fam} \times N_{total\ for\ device\ family} = 3 \times 288 = 864$ trials performed

Table 8.1: Experiment parameters used here.

8.1.2 Experiment Description

The key features of the design are summarized in Table 8.1; this table will be fully explained as the chapter progresses.

A single experiment session consisted of:

1. Signing a consent form
2. General instructions to subject
3. Practice period
4. 4 calibration trials
5. 48 comparison trials
6. Exit questionnaire

Each of these elements will be described below.

Experiment factors were a set of emulation variables plus the "Targets" factor; factor levels were varied from trial to trial. An emulation variable could be any hardware, control and modeling parameter suspected to influence the performance index being tested (in this case, intelligibility of emulation). Each experiment factor may be tested at an arbitrary number of levels (minimum two), although the experiment size increases drastically with number of levels per factor.

The "Target" factor was the source of experimental variation which arises from the emulation's targeting a different real device from trial to trial; since it is expected that error rates might be influenced by which device is targeted, this factor was varied and tested in the same manner as the emulation variables.

A single comparison trial consisted of a subject comparing a set of $N_{targets}$ real devices with a single emulation which targeted one of the real devices (Figure 8-1 and 8-2, with $N_{targets} = 4$); this is the [Multiple Real:Single Virtual] or MR:SV comparison model. The subject answered the question,

"Which real device does the emulation feel most like?"

by pointing to or giving the slot number of the chosen device.

The set of hardware/control emulation variables being tested were held at a single level for the trial; levels were changed from trial to trial, as was the targeted real device. The same set of $N_{targets}$ real devices was used in each trial, but device physical placement was reordered randomly from trial to trial.

In these experiments, an experiment *session* contained $N_{comp\ trials} = 48$ comparison trials, each lasting on average about 2 minutes including subject decision (averaging about 30 seconds) and device changeover. The entire session lasted between one and two hours, depending on device family and subject.

The haptic characteristics of the real devices varied along a preset distribution; in this experiment, the real device set members were qualitatively similar and the characteristic varied was quantitative, a component of impedance (mass, viscosity or maximum toggle-switch force). The distribution of real device impedance was approximately logarithmic, according to a presupposition of Weber Law perception on the part of the subjects (Figure 8-3).

The experiment was designed and analyzed as a multi-factor factorial, randomized block ANOVA with an arbitrary number of levels for each level; both qualitative and quantitative factors were used. Subjects were analyzed as blocks.[1]

The analysis is most simple and the results most powerful when full replicates are used on each block (subject), i.e. all combinations of all factor levels are tested in a single session. This was the case for the experiments described here.

The computed response for a trial was some measure of the difference between the device targeted by the emulation for that trial and the device chosen by the subject as feeling closest. Thus, a subject choosing the targeted device resulted in a computed response of '0'; a positive error meant that the subject chose a device of lower impedance than the targeted device, and a negative error the reverse.

---

[1] Blocking refers to the convention in the Analysis of Variance of subtracting the variation due to blocks (in this case subjects) from the overall error estimate when computing the F-test for each other source of variation. This is a good way to both measure the degree of variability due to a "nuisance" factor and prevent it from obscuring the effects of the other experiment factors.

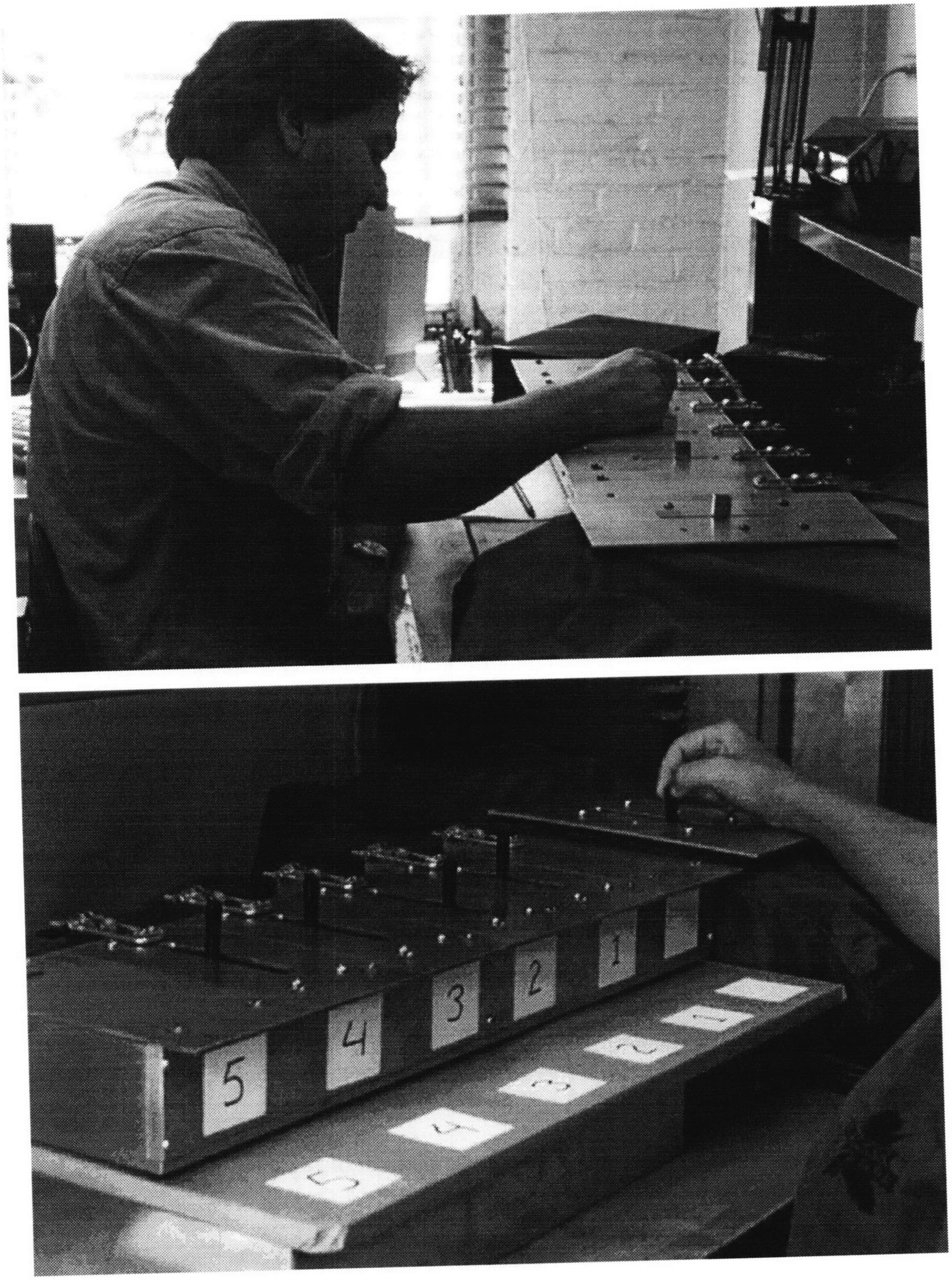


Figure 8-2: Subjects perform comparison trials.

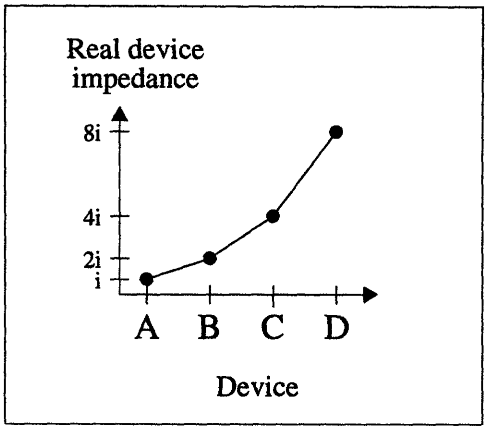


Figure 8-3: The impedances of the individual real devices were spaced approximately logarithmically in this implementation of the experiment design, based on a Weber Law assumption of subject haptic perception. In this example, $N_{targets}$=4.

The simplest response variable was computed from the relation,

$$Y_{unweighted} = \text{index}(\textit{target device}) - \text{index}(\textit{chosen device}).$$

Here, the "index" of the lowest-impedance device in the set ("A" in Figure 8-3) was 1, and went up with device ID:

Device ID	A	B	C	D
Index	1	2	3	4

Thus if the subject picked B when A was targeted, $Y_{unweighted} = -1$.

Other response variables, described in later sections, were based on weighting and transforming this unit error, given different measures of the difference between the adjacent real devices.

The number of real devices in the set ($N_{targets}$) was chosen to be and was a compromise between two conflicting considerations:

1. $N_{targets}$ should be minimized in order to shorten decision time (more time is required to select between a larger set) and thus overall experiment time/cost; and to reduce effort of experiment preparation.
2. Conversely, $N_{targets}$ should be maximized in order to increase the experimental value of each trial. More is said on this topic in Section 8.1.4. In brief, the results are more valid the more choices the subject is given, with $N_{targets}$=4 being the smallest size for which results will be meaningful.

The best compromise appeared to be (and remains) a set size of four. This resulted in a trial duration which was barely tolerable; however, the results would have been stronger if constraints on overall experiment time had permitted a larger set size. Possible solutions to this dilemma are discussed in Section 10.3.1.

"Calibration" trials were performed at the beginning of each section to determine subject perceptual scales. Subjects were asked to give magnitude estimates for the quantitative difference between the real devices which would be used in the comparison trials. No emulation was involved in this step.

Subjects were given no time limit for either calibration or comparison trials. This decision and its consequences are further discussed in Section 8.2.6.

Tradeoffs in the experiment design were made between number of experiment factors (emulation variables and Targets), experiment power (i.e., magnitude of error estimate) and experiment size. In each experiment situation with its unique set of emulation variables, number of factor levels and desired analysis power, these considerations must be examined carefully and the best compromise made for that case. This issue is discussed further in Section 8.1.4.

8.1.3 Justification of Comparison Model

"Which One?" Choices: Objective Test

The "Which one?" comparison model used here was considered most attractive of all options due to its simplicity and objectivity. It comes with a smaller degree of interpretational uncertainty than does a magnitude estimate ("*How* different is it?").

However, this very simplicity has the drawback of delivering less information than a more subjective response, such as a magnitude estimate. A multiple-choice style response restricts the subject's response to a few preset choices, whereas a magnitude estimate allows the subject to choose from a larger, in some cases continuous, array of options.

Returning to the original point, we must note that a magnitude estimation — for instance, of the difference between a single real device and single emulation — might be suspect because to the degree to which the emulation is qualitatively unlike the real device, we are asking for a numerical judgement of the difference between apples and oranges. It would thus be difficult to know what the subject meant or to make cross-subject comparisons.

[MR:SV] versus [SR:MV] Comparisons

The use of the [Multiple Real:Single Virtual] experiment comparison model versus its dual, [Single Real:Multiple Virtual] which in some respects might have been simpler to implement, is justified on both philosophical and pragmatic grounds. These arguments assume the appropriateness of the objective "Which One?" comparisons justified above.

Philosophical Justifications:

The emulation performance index used in this experimentation is the degree to which a facsimile of a given haptic environment which exists in the real world is recognizable as the targeted environment; ***intelligibility*** is the quantity which will be measured. The [MR:SV] comparison model provides an index of emulation fidelity, i.e. the degree to which a subject, on average, can use the emulation to distinguish between two real devices of a specified haptic difference.

The clearest way to present the philosophical support for this claim is by counter-example. For the purpose of ascertaining what subjects pay attention to in an emulation of a given haptic environment, it is arguable that a [Single Real:Multiple Virtual] comparison model would be more effective.[2] This is the sort of approach which was taken by Rosenberg and Adelstein [165] in their comparisons of virtual wall emulation strategies, and it worked well in that situation. Their performance criteria was influence of emulation strategy on the virtual environment's likeness to a [single] abstract "wall" ideal. The experiment's purpose was not to test influence of emulation strategy on fidelity to a particular flavor of wall.

---

[2] This would be the case if the array of real emulations varied not (or not only) in simple magnitude of key variables, but in strategy of emulation or hardware variables employed.

In the experiments performed here, it was certainly important to have an idea of what people pay attention to in a haptic environment. However, a great deal was learned on this topic during the emulation design step. Although no formal experiments were performed, a large number of emulation variable combinations were tried informally and either discarded or improved upon in the pursuit of creating the emulation which best represented a particular target. The emulations used represented the most effective strategies I developed.[3] What was left to do in the experiments was discover the power of those strategies by ascertaining their resolution; to learn the fineness of detail in real environments which could be accurately displayed.

For this, comparison with an array of real devices was necessary.

Pragmatic Justification:

It turns out to be virtually impossible to remember an emulation from one moment to the next. I had trouble with this, and subjects whom I informally tested had the same difficulty.

Therefore a comparison model which relies on a subject feeling several emulations and then choosing one of them is doomed to some kind of bias, because by the time the subject reaches the third one or maybe even the second, he will be unable to recall the first and thus make an informed decision. It is my sense that unless multiple emulation systems are set up in parallel (expensive), a [Single Real:Multiple Virtual] (SR:MV) comparison model could only work in the case of a paired comparison where a subject is asked to remember only two emulations.

8.1.4 Comments on Key Experiment Design Issues

Following is a discussion of the more important and general issues which had to be considered in creating this design. Section 10.3 continues and expands on the comments made here while incorporating observations from the experiments. A number of more specific issues are also brought up in later sections within this chapter as appropriate.

The Haptic Turing Test

To yield useful insights, the comparison model used here does not require that the emulation in any configuration pass the "Haptic Turing test." That is, human subjects need not find the emulation indistinguishable, either qualitatively or quantitatively, from the targeted real device.

While the capability of indistinguishable emulation is unquestionably a desirable eventual goal and would make these experiments even more useful, perfect emulation is not necessary to ascertain the vectors of emulation performance improvement — i.e., which variables are most influential in given emulation tasks, and how. Testing of the indistinguishable emulation would simply provide another set of non-unique data points: the improvement vector endpoints. Further, since the perfect emulation of challenging environments is not readily achievable with today's technology, knowledge of improvement vector endpoint locations does not hold the immediate practical value that an accurate identification of vector axes and direction of improvement does. In fact, the location of "ideal" performance points will be suggested by the improvement-vector directions.

However, it is important to note that results may in some situations be influenced by the degree to which the emulation qualitatively reproduces features of the real device impedance.

This results in a vulnerability in the experiment design: unexpected or hard-to-interpret results might emerge if the emulation is qualitatively dissimilar to the real devices, even though its impedance magnitude could be judged as being on some equivalent scale. This problem is the root of the Quality versus Function dichotomy which I will discuss in greater depth in succeeding chapters.

---

[3] This is not to say that these emulations represent the best strategies possible; in fact, I hope and believe they are not the best possible because the ones used are not in all cases good enough to use in a VPP application.

Implications of Fixed Effects Experiment Model

Choice of experiment model is another influence on the response variable's information content.

In these experiments, a fixed set of real devices was repeatedly compared with an emulation which was itself varied from trial to trial. The same $N_{targets}$ real devices were used in each trial; this is a *fixed effects* experiment model.

Using this model, the impedance difference between each pair of real devices was always the same from trial to trial. This means that not only was there a limit of 2–3 possible answers; the implied magnitude of each of those answers was always the same.

To illustrate this, let us say that a set of four real Mass devices were distributed (based on measurement, in this context weighing) as follows:

	Impedance [grams]			
Trial	A	B	C	D
all trials	50	100	200	400

Whenever B was targeted and C was chosen, the confusion was over 100 grams. Whenever B was targeted and A was chosen, the confusion was of 50 grams. This was true of every trial in which B was targeted and not chosen.

In the context of this experiment (multiple-choice style subject response), a *random effects* model would mean using a different set of $N_{targets}$ devices in each trial, their $N_{targets}$ impedances chosen at random. For example:[4]

	Impedance [grams]			
Trial	A	B	C	D
1	30	40	200	250
2	100	140	150	170
3	50	120	310	400
...				

Thus in Trial 1, if B were targeted and A chosen the error would represent a confusion of 10 grams; the same error in Trial 2 would represent a confusion of 40 grams, and in a different location (relative to 140 grams rather than to 40 grams). In Trial 3, this error would refer to 70 grams. Thus as long as the variation was controlled and documented, it could lead to richer and more interpretable results.

The downside of a random effects model in this context is the difficulty of producing a different set of real devices for each trial. With the real device production methods used here, a random effects model was not feasible. Such a model requires a different set of device designs which lend themselves to nearly effortless on-the-fly impedance (or other tested haptic property) modification.

Implications of Using a Real Device Set of Fixed and Finite Size

End Effects

Using a real device set of fixed size, the choices which are effectively presented to the subject depend on whether the targeted device is at the bottom, top or anywhere in the middle of the real device set (Table 8.2). When the target is in the middle of the set (B or C for $N_{targets}$=4), there are three possible answers (-1,0,1);[5] if the target is on the end of the set (A or D), there are only two possible answers, (-1,0) and (0,1), respectively.

---

[4] There are many ways in which restrictions on the randomization of real device set impedances could be employed to increase the proportion of valuable data to obvious choices, while keeping the useful property of constantly varying impedance differences between real device set members. The listed example uses no such restrictions.

[5] This assumes only nearest-neighbor errors; i.e., when B is targeted, the subject might pick A or C but never D. This did prove to be the case in the experiments performed here. However, the criticism here holds even if greater than nearest-neighbor errors occur although the distortion would not be as great.

Device Targeted	Device Chosen	Error Type
A	A	0
A	B	-
B	A	+
B	B	0
B	C	-
C	B	+
C	C	0
C	D	-
D	C	+
D	D	0

Table 8.2: Types of errors possible in the [Multiple Real:Single Virtual] comparison model.

This is a rather inefficient use of the setup — for a set of four devices, as was used here, one or two are using up the subject's time without otherwise influencing her decision. A more critical problem is the nonuniformity of near-neighbor options: once the subject has narrowed choices down to target and next-neighbors, her chance of randomly guessing the target depends on whether the device is on the end or in the middle of the set.

Pros/Cons of Using a Small Set Size

One might ask why, once we have assumed that only nearest-neighbor misses are being made, we should use more than three devices in any trial. If we use a different three-device subset out of a larger complete set of either finite (fixed effects model) or infinite (random effects model) size in each trial, we would either have to (a) target the middle device in the set and supply the adjacent high/low nearest neighbors, in which case the subject would quickly catch on that the answer was generally in the middle; or (b) target any member of a randomly chosen three-adjacent-element set, and confront the same problem as before — one neighbor for end-of-set targets, and two neighbors for set-center targets. Using a set size greater than three does avoid the bias detailed in point (a).

Conversely, we could go to the extreme and use device subsets of two members; then the target in every trial has a single neighbor. This is known as a *paired comparisons* experiment model. It would solve the problem of a nonconstant number of choices presented to the subject from trial to trial, but effectively throws the baby out with the bathwater — the solution is to further reduce information content for *all* responses whereas before we had a little more for at least some responses.

Pros/Cons of Using a Large Set Size

A way to mitigate the problems inherent in a device set of finite size, described above, would be to use as large a number of real devices as is feasible in the lineup. End effects are thus minimized since a smaller fraction of the targets will be end devices. Further, if the subject has a $1/N_{targets}$ chance of randomly guessing the targeted device correctly independent of emulation characteristics, the random-score element will be reduced by making $N_{targets}$ as large as possible — at least up to 4 or 5.

A complete set size (as opposed to subset of larger set, as described immediately above) of three would by the Fixed Effects model criticism stated above yield highly distorted results, because only a single set member would have two neighbors. A complete set size of five would be an improvement on the four-member set used here by this argument, but introduces new problems in terms of exploding experiment size and the cost involved in producing larger real device sets.

Increasing the set size past five is probably unwise for perceptual reasons as well, since it is well established that at around that point, short-term memory becomes cognitively overloaded [62, 143].

Resolution of Real Device Set

An important factor upon which the usefulness of experiment results hinges is that the haptic resolution of a set of real devices being tested be chosen such that it is comparable to the emulation system's best-case resolution.

Satisfaction of this requirement will ensure that the entire region of emulation variable space spanned by the emulation system is in fact studied: if the real device set had a resolution more coarse than that of the best-case emulation, no observation could be made of the performance resulting from best-case emulation variable settings. If on the other hand the real device set resolution was much finer than the emulation, subjects would be forced to guess randomly at all times and variable degradation would reveal no change in performance.

Experiment Size

Once the length of a single experimental trial and a reasonable limit on entire session time were determined, the design size was determined within those constraints by choosing:

$$\begin{aligned} N_{factors} &= \text{number of factors} \\ N_{levels} &= \text{number of levels for each factor} \\ N_{reps} &= \text{number of repetitions (full or partial)} \\ N_{subjects} &= \text{number of subjects} \end{aligned}$$

This process was and generally will be iterative, resulting in a compromise between desired strategy (full repetitions per subject session) and desired number of factors and levels.

For efficiency and to study emulation variable interactions, it is desirable to test many emulation variables at once. However, this increases the number of trials which must be run for a single replicate.

One would like to run replicates on blocks (subjects) in order to best study this important variability source. However, this mandates replicates which can be performed in approximately an hour and no more than two hours, after which subject concentration and patience dwindle; this in turn limits the number of trials which can be performed in a single session / replicate.

One would also like to run either complete replicates (all combinations of all factor levels), or incomplete replicates in which only high-order (3-way and higher) interactions are confounded[6] with main effects and 2-way interactions in order to maximize the strength and consistency of the analysis. However, complete replicates of a multifactor, multilevel factorial design grow rapidly in size with increase in number and levels of factors. Incomplete replicates with desirable confounding patterns can be tricky if not impossible to devise when mixed numbers of factor levels are used.

Ways to increase number of emulation parameters studied while only slightly increasing experiment size, if at all, are:

Use only two levels for every factor except Targets.

The levels of the Targets experiment factor must correspond to the real device set members if a Fixed Effects model is used. However, the other factors may be tested at only two levels, and this reduces the number of trials required for a complete replicate.[7]

---

[6] "Confounding" refers to what happens when multiple factors are varied simultaneously, and a result any measured variation cannot be properly attributed. Simultaneous factor variation is often necessary when the experiment is too big to vary each factor independently, but a good design will control confounding patterns by varying important and likely-to-be-significant factors only with unimportant and unlikely-to-be-significant factors (such as high order factor interactions).

[7] For example, in this experiment I tested one emulation variable (Sampling Rate) at three levels because I expected the response to be nonlinear — fairly flat at high sampling rates, with a break point and rapid dropoff at an unknown

Use incomplete replicates.

This refers to the standard practice of confounding higher-order factor interactions [141]. Thus, every combination of factor levels (as laid out in Table 8.4) is not tested; instead a subset of trials is used based on the chosen confounding pattern. The cost of the reduced experiment size is loss of information about the confounded interaction. Thus it is crucial to choose a confounding pattern which confounds an unimportant effect (one expected to be insignificant) with an important one.[8]

Modeling and Parameterization of Real Devices

The goal of these experiments was to measure how well a subject perceives a haptic emulation as matching the real device it targets. Ideally, the degree to which the emulation is truly representative of the physical model and parameters that are embodied by a given real device is known and testable.

In fact, it is often difficult both to model and parameterize the real device and to measure the impedance of the emulation. For example, among the device families used here Masses were trivial to measure but Toggles were quite difficult to model and parameterize in a manner which an emulation could stabilize. Even when using a haptic system-identification procedure like the one discussed here, there is no guarantee that the features measured are the ones that a subject will key on perceptually; or that when measured, certain high-frequency features will be faithfully reproduced by the limited-bandwidth emulation just because the computer model asks for them, or that the emulation will be stable.

Thus, the assumption that the emulation's target is the "right" answer to the comparison problem posed to the subject in every trial is on occasion stretched severely, even when the emulation variables have not been degraded. In fact what is being measured is the extent to which the subject agrees with whatever means of measurement was used to compose the emulation model, whether it be a direct mechanical measure, a parameterized mechanical model or the investigator's own perception.

---

location. In such a case, it is useful to use more factor levels in order to more closely pinpoint the location of the dropoff.

Here, the tradeoff between experiment size and number of Sample Rate factor levels proved to be a good one, because not only was the Sample Rate response nonlinear, it interacted strongly with other emulation variables. Thus it was, in the end, important not only to test Sample Rate at more than two levels, this needed to be done in conjunction with varying other emulation variable levels.

[8] When, as here, a mixed number of factor levels are used (two factors had two levels, one had three and one had four), safe confounding becomes a challenge. A valuable reference on this subject is Binet [14] which lists confounding patterns for a wide variety of mixed level designs.

8.2 General Features of Experiment Design

8.2.1 Structure

Description of Experiment

To systematically study the effect of each emulation variable on a subject's ability to discriminate between real devices using the emulation, a multifactor, multilevel factorial design is used. That is, $N_{factors}$ factors are considered, each at an arbitrary number of levels. In addition, the variability due to using a different target with each trial results in another element of variability ("Targets") which is most effectively treated as an additional experimental factor, T, with $N_{targets} = t$ levels. The virtual environment emulation for a given comparison trial is then defined by the levels of the set of emulation variables being studied, all other emulation variables being held constant to the extent possible; and by the emulation target.

For illustrative purposes, the arbitrarily designated number of factors and levels for each factors shown in Table 8.3 will be used. They result in $N_{comp\ trials}$=24 trials for a single full repetition. When full replicates are used, the number of trials per experiment session is also $N_{comp\ trials}$, and the total number of trials available for analysis is $N_{comp\ trials} \times N_{subjects}$.

Standard Order

An example of how the two-factor, multi-level study defined in Table 8.3 would be laid out is shown in Table 8.4. This convention of deriving trial components by systematically listing all combinations of all factor levels for all factors is called "standard order". In an actual experiment, the order would be randomized to prevent subject bias. The case shown is for a full replicate; that is, all combinations are tested in the 24 trials.

Partial replicates are created by testing some subset (systematically arrived at) of all possible factorial combinations.

Blocking on Subjects

An important experimental issue in any human subject study is the degree of variation between different subjects, and the repeatability and/or consistency of single subjects. It is reasonable to expect a priori that individuals will vary in perceptual acuity, attentiveness and strategy. These experiments were designed to simplify both subject repeatability and subject variation issues, by

- Using a randomized block design with blocking on subjects, where a block is an experiment session.[9] Thus if N_{reps} replicates are performed for the entire experiment,

$$N_{reps} = N_{blocks} = N_{subjects}.$$

---

[9] Blocking on subjects means performing each replicate (whether full or partial) of all the factor level combinations being tested on a different subject, and then computing variation within the block (subject) before pooling variation to estimate population variation. In this manner, variation between subjects does not obscure within-subject patterns.

Factor	Number of Levels
A	a=3
B	b=2
$N_{targets}$	t=4

$$\begin{array}{rcl} N_{factors} & = & 2 \text{ (not including Targets)} \\ & & \\ \text{Factorial Type} & = & t \times a \times b = 4 \times 3 \times 2 \\ & = & \text{Number of trials for a full replicate} \\ & = & N_{comp\ trials} = 24 \end{array}$$

Table 8.3: Factor Levels: general case and an example of a hypothetical implementation. The example is smaller in scope than the experiment described in this thesis.

Trial	Target (4 levels)	Factor A (3 levels)	Factor B (2 levels)
1	A	low	low
2	A	low	high
3	A	mid	low
4	A	mid	high
5	A	high	low
6	A	high	high
7	B	low	low
8	B	low	high
9	B	mid	low
10	B	mid	high
11	B	high	low
12	B	high	high
13	C	low	low
14	C	low	high
15	C	mid	low
16	C	mid	high
17	C	high	low
18	C	high	high
19	D	low	low
20	D	low	high
21	D	mid	low
22	D	mid	high
23	D	high	low
24	D	high	high

Table 8.4: Example of how emulation variable levels are varied between comparison trials. The pattern shown here is "standard order" for a two-factor experiment, with one factor at 3 levels and the other at 2 levels. In experiment execution, the trials would be applied in a random order.

- Including "repeat" sessions within each real device family wherein a single subject is tested twice with the same experimental variables, as well as single subjects tested across device families to look for consistency in cross-device performance.

- Using multiple subjects (five different individuals for each device family) both because the experimental time required is too great to inflict on a single individual and because we are generally interested in the degree of variability found in the population at large.

In this case, subjects are not really a nuisance variable because we are in fact interested in the characteristics of individual variation, or lack of it. However, blocking lets us both focus more clearly on experiment factor variability by showing trends which are common to all subjects, *and* obtain an explicit significance value for the effect on response of subjects themselves.

The requirement for subjects blocking has implications for overall experiment size and layout. As performed here, comparison trials (described fully in Section 8.2.5) take about 2 minutes apiece. Thus, to keep an experiment session (including preliminaries and calibration trials) down to 1–1.5 hours per subject, 50 comparison trials is an upper limit on session size and it would be desirable to use fewer. Since a replicate must be performed in a single session,[10] a replicate (as well as a session) may contain no more than, and preferably fewer than, 50 trials.

The 50 trial budget is quickly expended with 2–3 factors (excluding Targets) of 2–3 levels each. To consider more factors or more factor levels, either incomplete blocks[11] or incomplete replicates must be used. Either modification will reduce the experiment power and increase complexity of design and analysis. The power reduction is affordable when a large number of factors is used because the error estimate in that case has a very large number of degrees of freedom, and for some confounding patterns the analysis complexity may not be extreme [141].

The factors and their levels chosen for this experiment are detailed in Section 8.3.

---

[10] A block must be limited not only to a single individual, but also to a finite period of time to keep track of day-to-day variation in a single individual and minimize unmodeled variability.

[11] Incomplete blocks are full replicates spread across multiple subjects, meaning either that a full dataset is not obtained for each subject or that each subject performs multiple sittings.

Use of Targets as an Experiment Factor

I treated Targets as an experimental factor — meaning that each real device was targeted and tested with all levels of all experimental factors — because I expected there to be some interdependence between the impedance of the targeted device and the other emulation variables being studied in their effect on emulation performance.

For example, I observed prior to the experiments that an emulation of a small mass was more sensitive to sampling rate than that of a large one; a stiff toggle more sensitive to the level of physical damping than a very compliant one. This would have been likely to introduce unmodeled noise into the experiment results if the target was not modeled explicitly.

Further, I was explicitly interested in this dependence; using targets as an experimental factor allowed me to study it.

Choice of Level Number and Values for Each Experiment Factor

It is generally advisable to rely on 2-level factorial designs because of their high efficiency. Increasing the number of levels in a factor makes the size of the experiment rise rapidly, particularly in an experiment with many factors, such as this one; and when linear relations are expected, the higher number of factors contributes little additional information.

In this investigation, however, nonlinearity of response to factor is often expected for at least some of the factors. Without a third factor level, it is not possible to test for the presence of a quadratic relation, far less model it; and this expectation must be traded off against available experimentation resources — most compellingly, a limit on the reasonable length of a single experimental session.

There are more "efficient" means (in the sense of total number of trials performed at the end of the day, and the experimental value of each data point) of modeling high-order relations (e.g. response surface methods, [141]). However, these methods are generally sequential, requiring an iterative process of gathering and analyzing data. More sophisticated techniques, they are not so efficient in terms of time spent in designing. Such an approach will be interesting to study in future efforts.

8.2.2 Spacing of Real Device Impedances

The real devices were spaced as nearly as possibly at uniform intervals along a logarithmic scale of their measured impedance[12] (Figure 8-3).

The considerations justifying this logarithmic spacing of real device impedances were:

1. According to Weber's Law, various aspects of the human perceptual scale are logarithmic: the ratio of Just Noticeable Difference (JND)/stimulus magnitude is approximately constant [62]. Thus the difference between impedance units of "1" and "10" and between "10" and "100" should be perceived as similar.

2. A log spacing permits more efficient coverage of a given experimental space than a linear spacing — given Weber's Law. We would like more devices concentrated in regions which will be perceived at higher resolution, and more sparse coverage in regions perceived at low resolution.

The goal of controlled spacing of real device impedances is not always an easy one to attain. For the device families used in these experiments, it was met more easily for Masses than for Toggles and

---

[12] A distinction is made throughout between the measured impedance and the subject's internal scale for perceiving real device impedance.

Sliders, because there was more direct and fairly fine control over the dominant impedance variable in the former case.

To a limited extent, a divergence from specified distributions can be counteracted by use of post-processing (see section describing response variables, 8.2.7); however, it is important to come as close as possible to specification for reasons described there.

8.2.3 Subjects

Number of Subjects

Three considerations enter into choosing the number of subjects. The outcome will be somewhat influenced by other determinators of overall experiment size (number of factors and factor levels, including real device set size); and whether full replicates can be performed in a single session or must be split across sessions or not completed.

The considerations are:

1. Need a strong ANOVA error estimate through replication of trials;
2. Want to study variability among different subjects;
3. Want to minimize overall experiment size, i.e. number of sessions.

These considerations are likely to conflict. A small number of replications might produce a sufficiently strong error estimate, particularly when many experiment factors are used (may pool high order interaction effects into error estimate, reducing requirements for replication). However, a respectable number of individuals must be tested to examine individual variability — in a sense, individuals become yet another qualitative experiment factor which must be tested at a representative number of levels.

In retrospect, six subjects (used here) were more than necessary to obtain an adequate ANOVA error estimate; four would have been sufficient. On the other hand, four subjects might have been enough to give an idea of individual variability but six subjects was better, and also allowed space to repeat a subject.

Type of Subjects

Since the eventual goal of this project is to provide design tools, we would like to test a representative variety of potential manual interface users. These may be a wide spectrum of individuals, ranging from fighter pilots to secretaries to occasional VCR users. To begin with, however, the products targeted by these tools will probably be relatively high performance and high cost, used by professionals of one type or another.

Therefore it seems reasonable at this point to use professionals (as opposed to, for example, construction workers who might have a degraded haptic perception from using a jack hammer all day) as a subject pool.

An additional motivation for using professional or otherwise well-educated subjects is that it is important that subjects comprehend, be motivated and willing to perform the somewhat obscure tasks which are asked of them.

This still leaves a good variety. It is desirable to represent a range of gender, profession (a mix of technically and nontechnically trained) and age. All subjects should be naive as to the details of the experiment, and most should be unfamiliar with the explicit field of research.

It is not necessary to vary handedness at this point, and this variation might in fact may introduce an unhelpful corruption into the results. The setup used here is inherently one-handed, and requiring some subjects to use a dominant hand while others use their non-dominant hand will likely affect results without telling us anything about the influence of emulation variables on emulation performance.

Instructions to Subjects

Subjects must be instructed precisely in their assigned tasks in order to achieve uniformity in instructed behavior and limit response variation to more interesting factors such as perceptual acuity and strategies for haptic discernment. Instructions should be read from a script.

Subject Compensation

It is helpful to compensate subjects, in the interests both of recruitment and fair compensation for time proffered, and to minimize unmodeled error in subject response. A token compensation is generally sufficient to achieve desired results.

Compensation for Participation To help recruit subjects and encourage in them a positive and conscientious attitude, compensate subjects simply for participating in the experiment.

Performance Incentive

In pilot experiments, I observed that the responses of some subjects varied in an apparently inconsistent manner. I hypothesized that these subjects were being inattentive — not surprisingly, as most people would find the sessions lengthy and boring, a typical feature of psychophysical experimentation. Therefore in the real data collection effort, I tried using a performance incentive with good effect. I promised subjects a small monetary compensation according to their performance, e.g. a quarter for each “correct” answer; but did not inform them as to their performance until the end of the session, in order to minimize undesired training effects. Following the subject’s answering the questions in the wrapup questionnaire, in which they were required to guess their score, I informed the subjects of their score and paid them immediately in cash.

Although I did not do a controlled study to ascertain whether the compensation actually improved subject concentration, my impression from observation and post-experiment discussion with the subjects is that it did introduce both a financial incentive and a mood of competition which significantly increased attention.

8.2.4 Calibration Trials

The subject calibration trials were used to obtain the subject’s personal perceptual scale, based wholely on a consideration of the real devices (no emulation was involved).

Purpose

Interpretation of subject response requires both (a) knowledge of subjects’ internal impedance scale and where they locate members of the real device set being tested on that internal scale, and (b) knowledge of those devices’ locations on a measured impedance scale. The value of this knowledge is fourfold:

1. It allows comparison of subject impedance scales with a measured scale, thus providing an indication of how different the subject’s impedance scale is from the measured scale.
2. It indicates whether the emulation targeting is varying what the subject uses perceptually to differentiate between the real devices.
3. It determines the threshold of subject impedance perception. This is important: it determines the degree of fidelity the emulation should achieve. If the emulation allows the subject to

discriminate between devices of an impedance difference x, and the subject can discriminate between real devices of an impedance difference y, then it is only necessary to achieve $x \leq y$.

4. It facilitates dissemination of results: the resolution of the set of real devices needs to be measured or otherwise described/defined for the emulation comparison results to have physical meaning in a written (non-touchable) report.

The subject's perception of the ***relative*** differences between device family members is also a useful quantity to know, in that it may be inconsistent with their perception of absolute magnitudes of member impedance. This would lend insight into successful emulation strategies.

Description

Subjects were asked to give magnitude estimates for the quantitative differences between the real devices being tested in that session.

Specifically, subjects performed the following two tasks on a set of real devices arranged in random order:

1. *Order* the devices by "hardness", from lowest to highest without physically rearranging them. The actual instruction read to the subject was:

   > "Please rank the devices in order of how "hard" they are. By ***hardness***, I mean the physical force you must use to make the device move. The hardest device will be the one which is stiffest, or heaviest, or otherwise presents the greatest resistance to motion.
   >
   > If two devices feel the same after a reasonable attempt to distinguish them, pick their relative order arbitrarily. You may re-visit the devices to make your choice. You may make notes on a piece of paper if that will help you organize your thoughts."

2. Assign a ***magnitude*** to the "hardness" of each device, in the order already defined by the subject, on a scale of [1:10]. The actual instruction read to the subject was:

   > "Now, tell me how hard you think each of those devices is, by assigning it a number between 1 and 10. '1' means it's the easiest in the set; '10' means it's the hardest in the set. Thus for each trial, you're always going to give me a '1' and a '10' ranking, and the others should be distributed depending on how hard you think they feel.
   >
   > If you can't tell the difference between two or more devices, give them the same number. This goes for the '1' and '10' levels; there can be more than one of them as well.
   >
   > Be aware that you are not necessarily getting the same set of devices in each trial."

One such presentation and pair of responses constituted a *calibration trial.*

Number of Calibration Trials

$N_{calib\ trials} = N_{targets}$ calibration trials were used so that each family member could be duplicated once (see below for what duplication means). Thus for a set of $N_{targets} = 4$ real devices, four calibration trials were performed at the beginning of each session and the responses analyzed to estimate the distribution in impedances of the real device set perceived by the subject for that session.

This resulted in $N_{targets} + 1$ estimates of each subject's perception of each device's impedance; the "1" was due to device duplication.

Device Duplication in Calibration Trials

One device family member was duplicated in each calibration trial in order to test the repeatability and resolution with which subjects perceived the real device impedance distributions, and to compare the subject's ability to detect "identical" devices versus those the investigator independently measured as non-identical. Thus a total of $N_{targets}+1$ devices were presented to the subject in each calibration trial. The order of duplication was randomized.

If a subject was highly consistent and repeatable and could reliably tell the difference between all $N_{targets}$ members in the set but could not distinguish the duplicate pair, he/she would

1. always sort the devices into the same order, except for the duplicated device pair which would be arbitrarily ordered;
2. give the same magnitude estimate for a particular device every trial regardless of which set member was duplicated that trial.

Divergences from this "ideal" behavior was expected to (and does) lend insight into what the subjects actually perceived both individually and as a group.

As an example, for Subject Q the devices might be presented in the device sequence shown in Table 8.5; both presentation arrangement and duplication order are randomized.

	Switchbox Slot				
Calibration Trial	1	2	3	4	5
1	B	A	C	C	D
2	A	D	A	B	C
3	B	A	D	D	C
4	A	D	C	B	B

Table 8.5: A representative shuffled trial and presentation sequence in which $N_{targets}+1$ (one device duplicated in each trial) real devices are presented to the user in the course of the $N_{calib\ trials}=N_{targets}$ calibration trials. All $N_{targets}+1$ devices in a single calibration trial are arranged in the switch box so the subject can explore them with his own strategy.

The portion of subject instructions which exacted the desired subject action given the presence of duplicated devices (extracted from those printed above) were:

> "If two devices feel the same after a reasonable attempt to distinguish them, pick their relative order arbitrarily. Be aware that you are not necessarily getting the same set of devices in each trial."

Tying of Ends of Perceived Impedance Scale

Subjects were instructed to assign real-device impedance magnitudes on a scale of [1:10], where '1' was associated with the lowest-impedance device in the set for that trial, '10' the hardest and the remainder spaced in between according to the subjects' judgement.

The strategy of tying both ends of the perceived impedance scales to a reference point which is contained in the real device sets has both advantages and shortcomings: (a) it maximizes information as to the subject's idea of relative device magnitude, which is what we care most about; although (b) this is at the cost of information as to perceived absolute device magnitude and range, which would be interesting to know but is not crucial to this analysis. For the purpose of these experiments, end-tying seems to be the best solution.

The implications of scale end-tying to a reference within the device set are that:

1. The subject is given a bounded scale for each trial. This is necessary to make sure we have some idea of the reference the subject is using each time.

2. However, it introduces the complication that we are not only forcing subject's magnitude range by fixing it at both ends to arbitrary values (why not use [11:20] instead?), we are also forcing an implied offset. A scale of [1:10] suggests that the low value is very close to 0, with an order of magnitude (OOM) of absolute difference between lowest and highest. That is, if $N_{targets} = 4$ then

   $$|\text{Device D}| = 10\times|\text{Device A}|$$

   An exact single-OOM top-to-bottom difference and a very low end is difficult to achieve precisely, and it is not clear that such a large difference is a good thing to strive for in future iterations of this experiment design. It is possible that the subject is clever enough to not be biased by the offset of magnitude range permitted him/her; however, this would require another study.

3. An alternative is to bound the perceptual range with "absolute" (0 or infinite) impedances, or other explicit references which are not contained in the real device set. The problem then becomes, as with many magnitude estimation problems, uncertainty as to what the subject is using as a reference particularly when the reference is even slightly qualitatively dissimilar to the test impedance — for example, an abstract "zero" impedance reference for the low end, and/or a "brick wall" for the upper bound. This leaves too much room for uncertainty in interpretation.

8.2.5 Comparison Trials

Description

The comparison trials comprised the major part of an experiment session. They were based on comparisons between a set of multiple real devices and a single emulation which targeted one of them. A full replicate of trials (factor level combinations) was performed in a single subject-session. In each trial the subject answered the question, "Which real device does the emulation feel most like?" There were $N_{targets}$ real devices to choose from, and a different (randomly chosen) real device was targeted by the emulation. The [visually identical] real devices in the set were arranged according to a new preset, randomized order so that for each new emulation the subject started with a "clean slate." He/she had no knowledge of what each device in the trial set would feel like, thus minimizing learning effects.

Non-Duplication of Targets in Comparison Trials

Device duplication was not employed in the comparison trials. Once the subject's ability to discriminate duplicates or close matches as well as his/her susceptibility to other biases (e.g. the real device's position in the switch box), was investigated in the calibration trials, there was little more specific information to be obtained by duplicating devices in the virtual/real comparison trials.

It would not reveal, for example, whether the subject could discriminate between two real devices which were intended by the investigator to be haptically distinct. This is learned, rather, from using two real devices that the investigator considers distinct but very close haptically.

The price of identifying bias such as switch-box location — a legitimate concern, since arm reach and physical distance of real device from emulation could quite possibly be error factors — would be an experiment where not only is the emulation's target treated as an experimental factor, but the duplicate must be considered an experiment factor as well. Worse, this factor must be considered in two cases: duplication of the target, and duplication of a device that is not targeted. The experiment size immediately gets out of control, and the benefit does not justify it.

8.2.6 Time Limits for Calibration and Comparison Trials

Subjects were not given time limits, but encouraged to make a quick "gut" judgements rather than attempt to haptically dissect the emulation. The instruction by which subjects were directed to pace their decisions for both calibration and comparison trials was:

> "Take the time you feel you really need, but don't dawdle: go with your gut reaction."

Justification for Unconstrained Decision Time:

1. Other work performed in this lab and elsewhere (Fasse [53] and Tan, personal communication) suggests that when comfortable time limits are set in the constrained case, there is little difference in results between time-constrained and unconstrained decisions. The observation was that subjects generally settle into a consistent decision period which is comfortable for them and not inordinately long. Thus, imposing time limits makes no difference so one might as well let the subject do what is most comfortable).
2. I wished to measure the time a subject required to make comfortable decision, on the hypothesis that it might be useful as a measure of decision difficulty and/or subject performance.

I expected that as the experiment progressed, the subject would learn to make relatively quick decisions out of tedium and eagerness to finish, and that the remaining degree of variation in decision time (trial-to-trial, subject-to-subject and family-to-family) might yield useful information if recorded.

In the experiments performed here, this expectation was indeed fulfilled in some cases; less so for other subjects who were extremely conscientious and continued to be very careful in their decisions even to the end of the final session.

Justification for Encouragement of Quick Decisions:

1. In the project's long-term objective (haptic design tool), it is the quick impressions that will matter most.
2. A brisk pace keeps trial length to a reasonable length, and avoids overtaxing subject patience and attentiveness.

Recording of Decision Times:

Subject decision time was recorded for each comparison trial, to test the hypotheses that decision time might act as indicators of:

1. the subject's ability to discriminate between the real devices based on the degraded emulation: do shorter decision times correlate with better performance (quick judgments win); or the reverse, implying that the "right" answer is generally available if enough care is taken in the choice?
2. the emulation "quality", which is generally very difficult to either define or measure — when the emulation is "high quality", is the choice more obvious and the decision quicker?

8.2.7 Response Variables

The basic response employed,

$$Y_{unweighted} = \text{index}(\textit{target device}) - \text{index}(\textit{chosen device})$$

implies a unit difference in magnitude between adjacent pairs of real devices in any set, because the indices of adjacent devices differ by one unit. A variety of *response variables* (the input to the Analysis of Variance) may be computed by weighting and transforming $Y_{unweighted}$ with the actual differences between adjacent real devices, based on either measured or subject-perceived scales.

The different response variables are worthy of investigation because the ANOVA will compute significance of the analyzed effects based on relative error magnitudes. Improperly weighted responses could result in distorted and misleading significance tests.

Limited Value of Response Weighting

Note that while this approach is in theory an improvement on an implicit assumption of equal spacing of the real devices, it still is limited in "grain" by the discrete nature of the responses themselves and what they refer to. That is, when a subject makes an error for a given targeted emulation, there are only two possible errors (sometimes only one) he/she can make, and only two possible magnitudes regardless of what those magnitudes are (Table 8.2). Thus, weighting an error between widely spaced devices as more egregious than a confusion of near devices improves the analysis once the confusion has happened; but the size of the relative gaps have a great deal to do with which confusions are made in the first place. For example, gaps which are very small or very large relative to human perceptual limits will cause some discriminations to be either impossible or obvious.

The message here is:

Weighting cannot make up for a poorly specified or poorly implemented impedance distribution.

That statement is better illustrated with an extreme example: suppose a device set were spaced on some scale as $\vec{Z}$=[1 2 3 10], and we were able to measure that scale accurately. If a subject ever confused Z_C=3 and Z_D=10, it would obviously help the analysis to weight that error with a '7' and the adjacent error, confusing Z_C with Z_B=2, with a '1'. However, it is probable that unless something was quite wrong with the emulation the subject would never make the Z_C/Z_D confusion, and this would show up as a very low, perhaps zero, error *rate* for D. The error weighting would therefore not make much difference.

The problem that this illustrates is that by choosing a Z_D which is likely never to be confused, we've lost a third of the information possible from this experiment; there are six errors possible with this paradigm used on a set of four devices, and this eliminates two of them. We might as well have used just three device members.

Definition and Expected Impact of Response Variables

Here the five response variables used in the analysis for this experiment are defined and the rationale and expected behavior described. The expected behaviors are based, when stated, on subsets of the following assumptions:

1. Subject perception follows Weber's Law at least approximately.[13]

---

[13] Recall that Weber's Law states that the subjects' perceived scales are logarithmic relative to a measured scale — that is, devices spaced logarithmically on a measured scale are perceived as uniformly spaced.

2. Subject perception does not follow Weber's Law, but is better approximated by the measured scale.
3. Device impedances are distributed on a measured scale which is accurately logarithmic.

These quantities are defined for reference by the following formulas:[14]

Z_{mech}	=	Impedance values, measured scale
Z_{perc}	=	Impedance values, subject-perceived scale
Z_{mech_T}	=	Z_{mech} for the device targeted by the emulation
Z_{mech_C}	=	Z_{mech} for the device chosen by the subject
Z_{perc_T}	=	Z_{perc} for the device targeted by the emulation
Z_{perc_C}	=	Z_{perc} for the device chosen by the subject

1. $Y_{unweighted}$ = index(*target device*) – index(*chosen device*)

   Weighting/Transformation Performed:
   All errors are weighted equally, regardless of the measured or perceived impedance difference between the targeted and chosen devices.

   Assumptions for prediction: (1), (3)

   Predicted effect:
   The ANOVA based on $Y_{unweighted}$ should show no target effect, if Weber Law holds (Assumption 1). That is, subjects should find high-impedance devices (more widely spaced on a measured scale) just as hard to discriminate as low-impedance devices.

2. $Y_{mech} = Z_{mech_T} - Z_{mech_C}$

   Weighting/Transformation Performed:
   Errors are weighted by the measured impedance distribution.

   Assumptions for prediction: (2)

   Predicted effect:
   This weighting makes confusions between widely spaced devices count more heavily in the ANOVA significance prediction than confusions between closely spaced devices. The implicit assumption is that subjects are perceiving the measured distribution, and this justifies using that distribution to determine the importance of each error. Assumption 3 is not required, because the compensation accommodates an arbitrary spacing of real device impedances.

   Thus, when Weber's Law doesn't hold (Assumption 2), Y_{mech} should make the errors appear more uniformly distributed in the face of a nonuniform real device impedance spacing, and reduce the ANOVA Target effect.

   When Weber Law does hold, however, the ANOVA results based on Y_{mech} will be skewed towards a greater emphasis on more widely spaced devices. The Target effect will be exaggerated because this weighting places undue emphasis on mistakes made between adjacent members at high ends of the scale.

3. $Y_{perc} = Z_{perc_T} - Z_{perc_C}$

---

[14] When Z_{perc} is used in these formulas, it always refers to the average perceived scale for a single session rather than the average for all sessions for that device family. The session data is then weighted by its own average perceived scale.

Weighting/Transformation Performed:
Errors are weighted by the impedance distributions perceived by each subject.

Assumptions for prediction: (1)

Predicted effect:

The ANOVA significance predictions should theoretically show the smallest Target effect, i.e. the most uniform perceived distribution.

The difference between Z_{perc} and Z_{mech} results will depend on how different the perceived scale is from the measured scale. If Weber's Law does hold (Assumption 1), then there should be a marked difference between these response variables; otherwise they should be similar.

4. $Y_{\ln(mech)} = \ln(Z_{mech_T}) - \ln(Z_{mech_C})$[15]

Weighting/Transformation Performed:
A log transformation is performed on Z_{mech}.

Assumptions for prediction: (1)

Predicted effect:

If Weber's Law holds, then ANOVA significance based on $Y_{\ln(mech)}$ should resemble that based on Y_{mech}; however, it should show a smaller Target effect than the nontransformed version.

If perception is linear on a measured scale (non-Weber) *and* Assumption 2 (log spacing of real devices) holds, however, this transformation should give ANOVA results which most resemble $Y_{unweighted}$ (unweighted, untransformed). This is because the log transformation of the log-spaced measured distribution results in a transformed distribution in which adjacent impedances are equidistant, i.e. linearly distributed. Thus, errors are weighted by the same amount as for $Y_{unweighted}$.

The matching will diverge to the extent that the real measured distribution is not actually log spaced. Thus, this transformation is a potentially a way to achieve an effectively linear distribution which reflects the non-specified nature of the real device impedance distribution.

5. $Y_{\ln(perc)} = \ln(Z_{perc_T}) - \ln(Z_{perc_C})$

Weighting/Transformation Performed:
A log transformation is performed on Z_{perc}.

Assumptions for prediction: (2)

Predicted effect:

This transformation is included more for completeness than usefulness. It probably does not reveal useful new information, regardless of nature of perception.

If Weber Law perception holds (Assumption 1), then using $Y_{\ln(perc)}$ tends to skew ANOVA significance results which have been rendered uniform by Y_{perc} and inflate the Target effect

---

[15] In computing the log transformation, it is necessary to use

$$\ln(Z_{mech\,T}) - \ln(Z_{mech\,C})$$

rather than

$$\ln(Z_{mech\,T} - Z_{mech\,C})$$

in order to handle case of zero error ($Z_{mech\,T} = Z_{mech\,C}$), because the log of a nonpositive number is undefined.

Factor	Number of Levels
A	$a = 2$
B	$b = 3$
T	$t = 4$

N_{reps}=$N_{subjects}$ = 3

Trials per repetition:
$N_{comp\ trials}$= $a \times b \times t = 24$

Total trials:
N_{total}= $N_{reps} \times N_{comp\ trials} = 72$

Source of Variability	DOFs	
Blocks (subjects=repetitions)		2
Main Effects		6
A	1	
B	2	
T = Target	3	
1st-Order Interactions		11
AB	2	
AT	3	
BT	6	
2nd-Order Interactions		6
ABT	6	
Error		46
Total = $a \times b \times t \times N_{reps}$ - 1		

Table 8.6: Analysis of Variance, general case. Arbitrary numbers of levels are used for illustration.

If perception is on a measured linear (non-Weber) scale, then $Y_{\ln(perc)}$=$Y_{\ln(mech)}$ and $Y_{\ln(mech)}$ is then preferred because it is more easily arrived at.

If perception is neither linear nor log but something else, probably in between, then a transformation is needed but this is not the correct one. Choosing the right one will require a model for the scale which *is* being used.

8.2.8 Method of Analysis

Analysis of Variance

Use of the factorial design makes it possible to analyze the data with a standard analysis of variance (ANOVA).[16] For each device family, variability contributions can be computed for all factor main effects and interactions as well as for subjects.

A simple example of the type of analysis used here is shown in Table 8.6. High level interactions (3-way and more) may be pooled with error to increase the error degrees of freedom. However, due to the large number of factors and factor levels there is already a very strong error estimate even with a small number of subjects and the high-order pooling is not necessary.

Study of Linear and Quadratic Effects

All factors with two or more DOFs may be partitioned into linear and quadratic effects to study their high-order effects; likewise interactions.

Modeling of Data

The factorial design permits the use of regression techniques to model the variability in the data, so that subject response may be predicted with a known degree of certainty from emulation variable settings which were not tested [16, 141]. The modeling analysis was not carried here because the data was not of sufficient quality to justify the effort, but in later iterations it may prove a useful tool.

---

[16] For an in-depth review of the analysis of variance, see any statistics text; for example [16, 141].

When factors of three levels or higher are used, curvature in the response may be estimated as well; this is the best reason to test more than two factor levels in a factorial design.

8.2.9 Randomization

To obtain completely randomized experiments, randomization and/or shuffling must happen at three levels:

1. Random shuffling of order of presentation of real devices from trial to trial
2. Random shuffling of order of trials from the standard order in which they were generated
3. Random alternation between members of any duplicate sets produced for the experiment. For example, there were two Device A's for Toggles, and which was used in a given trial was determined randomly.

In addition, other randomization procedures are required when there is a restriction on randomization. For example, in this experiment altering the damping level took too long to permit changes between each trial (Section 8.3.3).

8.2.10 "Double Blind" Protocol

Experimental sessions need to be run double-blind; i.e. it is important that the investigator administering the session not be aware of a critical subset of the emulation target, the setting of the emulation variables or the identity of the specific real devices being presented to the subject for that trial. By "critical subset" I mean that it is not necessary that the investigator be unaware of all of these variables, but only enough of them such that she will be incapable of conveying a cue of expected result to the subject, even if she has any expectation. For example, it would be sufficient that she be blind to the target of the emulation (most important) and most, preferably all, emulation variables settings, and at least uncertain about identity of real devices in the lineup; or vice versa about the target and lineup. Even knowing some emulation variables, she will not know which member of the lineup is targeted, or whether she expects the subject to guess correctly or incorrectly that trial.

8.3 Specific Implementation of Design

This section contains details of and comments on the general design and protocol described in Section 8.2.

8.3.1 Elements of a Single Experimental Session

A single experimental session consisted of the following elements:

1. Reading and signing of a consent form
2. Delivery of instructions to subject
3. Practice Period
4. Calibration Trials ($N_{targets}$=4)
5. Comparison Trials ($N_{comp\ trials}$ = 4 sets of 12)
6. Debriefing (Questionnaire and Payment)

An experiment lasted between one and two hours, depending both on the device family being tested (some proved more cognitively challenging than others) and on the subject. All experimental sessions used in the data analysis followed the same protocol. They differed by real device family and by randomization of trials and device presentation.

Signing of Consent Form

The informed consent form used here, approved by MIT's Committee on the Use of Humans as Experimental Subjects, appears in Appendix F. Signed forms were kept on file. Consent was obtained at the beginning of the first session performed by a subject.

Instructions

Subjects were asked to read, while simultaneously having read and demonstrated to them, a set of instructions which are included in Appendix F.

Practice Period

Three or four typical emulations of the environment being tested in that session (e.g., Masses or Toggles) were played in succession so that the subject could become familiar with the range of emulation behavior to be expected in that session. The subject could have the practice emulations replayed as many times as he/she liked; however, no one ever requested that they be played more than once. Generally the practice session took 3–4 minutes.

Calibration Trials

A single calibration trial consisted of the following:

1. Presenting the subject with a set of 5 real devices ($N_{targets}$+1, or the full targeted set plus one randomly chosen duplicate), in pre-randomized order, a different order for each trial.
2. Allowing subject to interact with devices for as long as he/she felt she needed to make the ranking.

3. Asking subject to rank real devices in order of increasing "hardness" (without removing them from case and physically re-ordering). The subject was asked successively,
   "What is the slot number of the least hard device?";
   "What is the slot number of the 2nd-least hardest device?", etc.

   I recorded the response in a response file via the computer program and manually on the investigator script as a backup.

4. Asking subject to give a hardness rating on a scale of 1-10 of the devices he/she had ranked, in the order of the ranking (recall that the least and most hard devices were pegged to magnitudes of 1 and 10 respectively). i.e., I asked the subject to: "Give a ranking (1-10) of the device in slot X [which you just said was second-least hard]";
   "Give a ranking (1-10) of the device in slot Y [which you just said was third-least hard]"; etc..

   The computer program kept track of all ordering.

Comparison Trials

There were 4 sets of 12 comparison trials in an experiment session, for a total of 48 trials. The sets were separated by 2–3 minute breaks during which the subject left the room while I changed the physical damper. The subject was asked to leave whether or not the damping level actually had to be changed for the next set or remained at the same value.

In a single comparison trial, the subject compared a single emulation with a set of 4 ($N_{targets}$) real devices, of which one was targeted by the emulation and none were duplicates. A single experiment trial consisted of the following:

1. Presenting the subject with a set of 4 devices (a full set), in pre-randomized order, a randomly different order for each trial.
2. Execution of the emulation for that trial and allowing the subject to interact with the emulation and the real devices for as long as he/she felt she needed to in order to make the choice.
3. Asking the question, "Which real device (slot number) is the emulation most like?"
4. Recording the response via the computer program in a response file (the computer automatically added all trial information, including factor levels) and manually on the investigator's script.

Debriefing

At the end of the session, I asked subjects a set of questions in order to document their subjective impressions to the emulations and the specific comparison tasks they had performed. The questionnaire is included in Appendix F.

8.3.2 Experiment Size

This section describes the outcome of the $N_{factors}$ / N_{reps} / experiment power compromise process for the set of experiments performed here. Table 8.1[17] summarizes the resultant experiment size and scope.

Full replicates were used largely for simplicity of design and analysis, particularly in light of the 3- and 4-level factors which were used.[18] However, this meant that only three emulation variables could be studied at once, whereas I was equipped for and would have liked to look at up to twice that many.

---

[17] See Chapter 6 for more complete description and discussion of the real device sets used here.

[18] Mixed-factor designs are more difficult to break into partial replicates of convenient size.

Number of Experiment Factors:

Three emulation variables in addition to Targets were used in order to achieve a manageable experiment size while proving the concept of the experiment methodology.

Which Factors:

The primary criteria for choosing emulation variables to use as experiment factors was that (a) they could be conveniently modified in the emulation system built for these experiments, and (b) I knew through prior experience with the hardware that modifying the variable influenced some aspect of emulation performance.

The following variables were chosen:

Virtual Environment Expression: Few virtual environment setups have the sensing capability to implement both impedance and admittance environment expressions, particularly with a closed loop servo (helpful in producing high-fidelity emulations) and in the variety of environments tested here. Thus the two expressions have not been well compared with each other, although it is logical to expect that each would perform best in a different set of situations.

Physical Damping: Colgate [38] demonstrated that adding physical damping to a haptic virtual environment system could have a dramatic influence on emulation performance, primarily through improved passivity characteristics. I chose to test physical damping to (a) verify whether the same kind of results would hold for a different hardware system, and (b) document this variable's behavior in interaction with other emulation variables. For example, will it have the same effect in impedance control as in admittance control?

Sampling Rate: Servo update rate was an obvious choice for testing, because (a) it is obviously crucial to emulation quality. Below some cutoff, any virtual environment will degrade. (b) The location of the cutoff is not known, and almost certainly depends on other emulation variables such as actuator bandwidth and the type of environment being modeled; and (c) it is trivial to modify sampling rate in the region of interest.

Number of Levels for each Factor:

VE Expression: 2 levels (Z and Y)

The virtual environment may be expressed as either Z (impedance) or Y (admittance). It is a qualitative factor with only two choices.

Physical Damping: 2 levels (0.001 and 0.070 N-s/mm)

Damping was restricted to two levels (intrinsic emulator damping and one level of added damping) because preliminary observation suggested that it was not as important in interactions with other factors and increasing damping consistently reduced the emulation "quality" (high damping emulations were more difficult to stabilize and tended to be less passive). I thus felt that the results of testing damping might only reinforce what was available to observation.

Sampling Rate: 3 levels (300, 600 and 1200 Hz)

Through prior experience working with the emulator, I expected to find that sampling rate exerted a nonlinear influence on emulation quality (quality approximately constant and high above some critical sampling rate cutoff, but dropping off sharply as sample rate is decreased below the corner frequency). By testing three levels of this factor, I allowed both the possibility of estimating curvature in subject response and, more informally, establishing the general location of a corner-point in the anticipated performance plateau.

Targets: 4 levels ($N_{targets}$)

Using $\geq$ 4 targets reduces the probability of the subjects' randomly guessing the target; but there was incentive to minimize number of targets in order to shorten trials and stay within a reasonable subject cognitive load.

8.3.3 Restrictions on Randomization

The primary exception to the double-blind protocol for both the comparison and calibration trials — and also the only restriction on randomization in the design — was the administration of the d levels of physical damping, D. I could not change out the physical damper without knowing its level — hence the absence of investigator "blindness"; and further, I couldn't change its level for every trial because the change took too long (2-4 minutes) and introduced undesirable levels of wear and tear on the device — hence the randomization restriction. Thus, I always knew the damping level, and it was necessary to execute trials for a given damping level in sets rather than making each application arbitrary.

In order to minimize the restriction on randomization, these sets were made as small as possible — i.e., multiple visitations of each damping level — and a different order of damping levels was used for different subjects. In these experiments, with d=2 levels of damping and $N_{comp\ trials}$=48, it was reasonable to break the 24 trials for each damping level into 2 sets of 12 trials each. Thus, one subject might get D = [1001], another [1010], another [0110] and a fourth, [0101].

8.3.4 Factor Level Combinations Applied in a Single Session

Table 8.7 shows the factor-level combinations which were used in a single session, listed in standard order. This display convention is different from the sequence in which the trials are actually presented to the subject, but makes it easier to see how the factors are varied.

In practice, the applications were randomized and thus differed from session to session. Table 8.8 shows an example of one random application; it is the sequence of trials used for subject

Patty (see subject description below) for a Mass session. Note that the two levels of damping were applied in four constant-damping sets, while other factors were varied randomly within those sets.

8.3.5 Subjects

Number and Identification of Subjects

The experiment was designed to accommodate a total of six subjects, each performing in three sessions. Multiple sessions per subject were employed both to minimize recruitment and training effort and to facilitate comparisons of single-subject performance across devices.

Three subjects performed one session in each device family, while three performed multiple sessions for a single device family.

Subjects are referred to throughout this document by the following pseudonyms:

Subject Number	1	2	3	4	5	6
Pseudonym	Josephine	Patty	Mike	Sam	Lorna	Hank

I use pseudonyms to minimize confusion of subjects with session numbers, and to establish an identity when it is useful to do so. These are not the subjects' real names, nor do they necessarily represent the real subjects' genders — gender was not explicitly studied in this study, and the gender scrambling helps to disguise real identities.

Trial	Target	Sampling	Damping	VE Exp
1	A	300 Hz	0.001 N-s/mm	Y
2	A	300 Hz	0.001	Z
3	A	300 Hz	0.070	Y
4	A	300 Hz	0.070	Z
5	A	600 Hz	0.001	Y
6	A	600 Hz	0.001	Z
7	A	600 Hz	0.070	Y
8	A	600 Hz	0.070	Z
9	A	1200 Hz	0.001	Y
10	A	1200 Hz	0.001	Z
11	A	1200 Hz	0.070	Y
12	A	1200 Hz	0.070	Z
13	B	300 Hz	0.001	Y
14	B	300 Hz	0.001	Z
15	B	300 Hz	0.070	Y
16	B	300 Hz	0.070	Z
17	B	600 Hz	0.001	Y
18	B	600 Hz	0.001	Z
19	B	600 Hz	0.070	Y
20	B	600 Hz	0.070	Z
21	B	1200 Hz	0.001	Y
22	B	1200 Hz	0.001	Z
23	B	1200 Hz	0.070	Y
24	B	1200 Hz	0.070	Z
25	C	300 Hz	0.001 N-s/mm	Y
26	C	300 Hz	0.001	Z
27	C	300 Hz	0.070	Y
28	C	300 Hz	0.070	Z
29	C	600 Hz	0.001	Y
30	C	600 Hz	0.001	Z
31	C	600 Hz	0.070	Y
32	C	600 Hz	0.070	Z
33	C	1200 Hz	0.001	Y
34	C	1200 Hz	0.070	Z
35	C	1200 Hz	0.070	Y
36	C	1200 Hz	0.070	Z
37	D	300 Hz	0.001	Y
38	D	300 Hz	0.001	Z
39	D	300 Hz	0.070	Y
40	D	300 Hz	0.070	Z
41	D	600 Hz	0.001	Y
42	D	600 Hz	0.001	Z
43	D	600 Hz	0.070	Y
44	D	600 Hz	0.070	Z
45	D	1200 Hz	0.001	Y
46	D	1200 Hz	0.001	Z
47	D	1200 Hz	0.070	Y
48	D	1200 Hz	0.070	Z

Table 8.7: All trials performed in a given session are listed here, in standard order (in the actual session, the trials would be randomized). The convention is to update the variable with the least number of factor levels the most often.

Trial	Target	Sampling	Damping	VE Exp
1	D	600 Hz	0.070 N-s/mm	Z
2	B	600 Hz	0.070	Z
3	A	1200 Hz	0.070	Z
4	B	300 Hz	0.070	Y
5	C	1200 Hz	0.070	Z
6	D	300 Hz	0.070	Z
7	A	1200 Hz	0.070	Y
8	A	600 Hz	0.070	Z
9	C	1200 Hz	0.070	Y
10	C	600 Hz	0.070	Y
11	D	1200 Hz	0.070	Z
12	D	1200 Hz	0.070	Y
13	D	600 Hz	0.001	Y
14	C	300 Hz	0.001	Y
15	D	600 Hz	0.001	Z
16	A	600 Hz	0.001	Z
17	D	1200 Hz	0.001	Z
18	B	600 Hz	0.001	Z
19	D	300 Hz	0.001	Y
20	A	600 Hz	0.001	Y
21	D	300 Hz	0.001	Z
22	C	1200 Hz	0.001	Y
23	A	300 Hz	0.001	Z
24	A	1200 Hz	0.001	Y
25	D	1200 Hz	0.001 N-s/mm	Y
26	C	300 Hz	0.001	Y
27	D	1200 Hz	0.001	Z
28	A	600 Hz	0.001	Z
29	D	1200 Hz	0.001	Z
30	B	1200 Hz	0.001	Z
31	D	300 Hz	0.001	Y
32	A	600 Hz	0.001	Y
33	D	300 Hz	0.001	Z
34	C	300 Hz	0.001	Z
35	A	1200 Hz	0.001	Y
36	A	600 Hz	0.001	Y
37	B	600 Hz	0.070 N-s/mm	Y
38	B	1200 Hz	0.070	Z
39	B	1200 Hz	0.070	Y
40	D	300 Hz	0.070	Y
41	B	300 Hz	0.070	Z
42	C	300 Hz	0.070	Y
43	A	600 Hz	0.070	Y
44	A	300 Hz	0.070	Z
45	C	600 Hz	0.070	Z
46	C	300 Hz	0.070	Z
47	A	300 Hz	0.070	Y
48	D	600 Hz	0.070	Y

Table 8.8: Shows an example of a randomized experimental sequence of the trials listed previously in standard order. This sequence was used for Patty (pseudonym) in a Mass session.

Subject Profiles

The following summarizes the demographics of the subject set used; for reasons of confidentiality[19] more specific information cannot be revealed.

Seven subjects were used in the main experiments described here. One of the pseudonyms actually refers two different individuals; one individual was unable to complete the full three sessions for reasons unrelated to the project and was replaced for the remaining sessions by a seventh individual. The pseudonym of this dual-identity subject cannot be revealed, again for confidentiality. The performance characteristics of the sessions performed by each of these two individuals were, however, fairly typical of other subjects in terms of difference from and similarities to each other, and thus lumping them as one subject does not alter the cross-family comparisons performed in the latter part of Chapter 9.

Handedness: All subjects were right-handed.

Gender: Three female, four male.

Age: All in 25-35 age group.

Backgrounds:

2 MIT engineering graduate students in a haptics-related field

4 MIT engineering graduate students in a non-human factors field

1 non-technical professional.

The demographic profile was satisfactorily balanced in all respects except for background (see Section 8.2.3 for criteria in choosing subjects); a subject pool of 50/50 technical to non-technical background would have been preferable.[20] However, it proved quite difficult to recruit outside subjects, given the time-consuming nature of the experiments.

Repetition of Subject Sessions

I felt it important to study not only ***variability*** within a representative subject pool, but also ***repeatability*** of a single individual. That is, to what degree is an individual's performance a characteristic of the individual — haptic acuity, motivation, etc. — and to what degree random noise? To some extent this question can be answered by comparing the performance of single individuals across real device families; but the performance changes are then confounded with another substantial source of variability (type of environment).

To solve this problem, I required three subjects to perform "repeat" sessions in a real device family, one subject for each family; the resulting assignment of subjects to sessions is shown in Table 8.9. Using the repeated sessions, an N=2 repeatability analysis is performed for each device family in addition to the general N=6 analyses carried out in Chapter 9.

---

[19] Refer to Appendix F for documentation of this project's application to MIT's Committee on the Use of Humans as Experimental Subjects.

[20] In addition to the subjects used in the primary experiments described here, several subjects were employed in preliminary pilot experiments. One of these was of non-technical background; my primary observation in comparing this subject's performance with that of others in both pilot and primary experiments was a much higher error rate. However, this may have been due to problems with the experiment protocol which was still being debugged, and at best is an observation of n=1.

	Device Family		
Session	**Masses**	**Sliders**	**Toggles**
1	Patty	Josephine	Josephine(1)
2	Mike	Patty	Josephine(2)
3	Sam	Mike	Patty
4	Lorna(1)	Sam	Mike
5	Lorna(2)	Hank(1)	Sam
6	Hank	Hank(2)	Lorna

Table 8.9: Assignment of subjects to experiment sessions, by family. Note that three subjects performed a repeat session in one family, and skipped another family (Lorna, Hank and Josephine in Masses, Sliders and Toggles, respectively); this is to permit an subject-repeatability analysis. Note that "subject", as used here, is not synonymous with "session". For example, Subject 2 (Patty) performs Sessions 1, 2 and 3 of Masses, Sliders and Toggles, respectively.

Compensation of Subjects

Subjects were compensated according to the guidelines outlined in Section 8.2.3.

Participation: Subjects were promised a $10 gift certificate at a popular local ice cream store for contribution of three 1-to-2 hour sessions. The payment was delivered after all sessions were completed.

Performance Incentive: At the beginning of each session, subjects were promised a quarter-dollar for each "correct" answer; their score was computed and the subject paid at the end of the session. The maximum compensation which could be earned in a session was thus 48 trials × $.25, or $12; the average session compensation was about $9.

Refreshments: Due to the long and tedious nature of the experiments, drinks and food were made available to the subjects at all times.

8.3.6 Control of Auditory Input

Sound was a crucial component in the perception of some of the real device families (most notably, Toggles) and all of the real devices made some sort of a sound; see Chapter 6. However, I neither emulated sound[21] nor found a satisfactory method for blocking the frequencies of sound present in the real devices. Thus auditory input was not blocked.

8.3.7 Masking of Auditory Input

One of the challenges in conveying the haptic sense of sharp discontinuities such as was present in the Toggle families is either recreating or masking the high frequency auditory component of the real devices. Since it was beyond the scope of this work to emulate sound in the virtual devices (although that will be a crucial element of future work), I considered masking the sound of the real devices so the subjects could make comparisons based on haptic input alone.

I recorded the sound of the real toggles and found the frequency content of the snap of the unmodified toggle to be in the 1.3–1.5 kHz range, i.e. far beyond what I could hope to achieve with this emulation system. Following modification, they did not snap quite as loudly or sharply but it

---

[21] Augmentation of the haptic emulations with sound are a key element in the future work section.

still produced a very different sound, both qualitatively and in magnitude, than did the emulations. Thus auditory blocking seemed desirable.

I experimented with various means of auditory blocking, and found that toggle snapping was very difficult to mask; aviation-type earplugs and industrial ear protection are designed to mask lower frequency sound. The only effective solution was a combination of blocking-type stereo headphones (the kind which are worn deep in the ear canal and not only transmit amplified sound to the inner ear but also block external noise sources) with loud, jangly music played through those earphones to camouflage the toggle clicks; and augmented with another layer of sound-blocking ear protection outside the headphones. Such sound blocking unquestionably made a difference in perception; the emulations "felt" much more like the real devices when you could not hear either. However, the paraphernalia was uncomfortable, disorienting, annoying if the subject disliked the music and potentially a source of ear damage after 1–2 hours of use. I therefore chose not to use it on my subjects and performed the experiments without any auditory masking.

8.3.8 Interaction Constraints

It turns out that perception of the haptic environments tested, particularly of the emulations, is influenced by subtle details of the hand grip and orientation of the arm and hand. In order to ensure maximum consistency in haptic perception of the emulations and real devices from trial to trial and from session to session, I constrained the manner in which subjects interacted with both the real devices and the emulations. These constraints were expressed verbally, and enforced through my visual monitoring and correction if a subject failed to observe the rules.

The constraints, the same for emulator and real devices, were:

- The same (dominant) hand must be used to interact with both the emulation and the real devices.
- The heel of the hand must always rest on panel surface while holding switch handle.
- The handle must be held with the thumb and one other finger, the same each time.
- The handle must be held lightly but firmly with the finger and thumb at all times (no flicking).
- Subject must always remain seated while performing a trial, and the seat must remain in the same place relative to switch box and emulator throughout the session (a choice of different-height chairs was offered to subjects so that all could reach all handles comfortably and from approximately the same angle).

Result of Unconstrained Comparison Trial Decision Times

Subjects required an average of 31 seconds to make a choice over all trials and all device families. The subject all-family averages ranged from 19.7 to 48.8 seconds, with minimum decision time of 6 seconds (Sliders) and a maximum of 127 (Toggles) recorded.

8.3.9 Handling of Missing Data Points Due to Unstabilizable Emulations

In a few cases where emulation variables were substantially degraded,[22] the emulation was difficult to stabilize. Making the emulation safe for subjects would have necessitated reducing servo gains so much that the performance of other emulations using the same gains would have been compromised.

---

[22] The specific factor combinations were for the Toggle device family: Sampling Rate = 300 Hz; Target = C or D; VE Expression = Admittance; Damping = 0.001 or 0.070 N-s/mm.

This situation was dealt with by simply skipping the four unstabilizable trials in each Toggle session, and counting those trials as response errors in the data analysis.[23]

8.3.10 Automated Protocol Generation and Data Collection

I developed computer-automated procedures in order to implement the double-blind protocol requirements as well as the necessary randomizations. Appendix F holds sample computer-generated protocols and the randomized trial treatments used for each session; this section outlines the procedures used.

Computer Output

A set of protocol generation and execution programs were created which:

1. generated batch files that called the actual virtual-environment computer control program, with command-line arguments reflecting the emulation variable settings for that trial. This VE program was responsible for, in a given trial,
   (a) running the emulation with appropriate factor settings;
   (b) accepting/recording subject responses in a response file, along with all relevant experiment information; without revealing experiment factor setting information on the screen
   (c) running in "invisible" mode such that there was no computer-screen output indicating details about the emulation.
2. generated a script file, corresponding to the batch file, for my own use which was printed out and kept on the "changing table" (the bench where I carried the switch box and swapped device modules between each trial without the subject's watching). The script file contained directions for the switch-box order of the device modules for each trial, but no factor-setting or target information; the latter information was not available to me during experiments.

   The devices, otherwise visually indistinguishable to a casual observer, were color-coded with color tape in a location not visible to subjects.

Randomization Procedure

In creating a protocol for a given experimental session, I followed this procedure for randomization:

1. for (d = 1 : N_{damp}):
   (a) Schedule out all factor/target combinations, excluding N_{damp} (randomization restriction). That is,

   $$N_{VE\ express} \times N_{sample} \times N_{targets} = 2 \times 3 \times 4 = 24 \text{ combinations for this experiment.}$$

   (b) For each trial, randomize order of presentation of the $N_{targets}$ devices.
   (c) Randomize order of trials.
2. Shuffle order of D (physical damping) presentation by choosing sequentially by session number from a list of all possible combinations, to ensure there is no repetition in order.

---

[23] In terms of computing subject score for determine subject compensation, the trials were counted as "hits" since the subject never had a chance to answer.

3. Associate color codes with all devices in all trials.
4. Record randomized sequences in a batch file and a script file (described above).

A similar procedure was followed in generating protocol for calibration trials.

The randomization algorithm used here is based on the ANSII rand() function. Properly seeded, it produced different randomized sequences for all randomization tasks described above, which appear visually to be uncorrelated uniform deviates. If there was more correlation than was visible to my eye, it was not critical in this context.

Chapter 9

Experiment Results

Overview of Chapter

The human-subject experiments were designed such that a set of analyses of variance could be performed on the data to ascertain the significance of the emulation variables studied here:

Virtual environment expression	[Admittance, Impedance]
Added physical damping	[0.001, 0.07 N-s/mm]
Sampling rate	[300, 600, 1200Hz]
Real Device targets	[A, B, C, D]

The bulk of this chapter is devoted to a detailed and in-depth elucidation of the insights offered by the experiment design; the reader who is short on time should merely skim this chapter and proceed to the next (Discussion) for summaries. Here, I invite the reader to be as attentive to the methodology and possibilities offered by this design as to the results themselves. At times I have carried the analysis further than this first-run data set merits, for the sake of demonstrating the richness of the design.

In presenting the experimental results, I have taken a general approach of starting from a moderate level of detail, then both generalizing to a larger picture and focusing in to view the fine-grain detail: the idea being to arrive at a fairly complete picture with a minimum of misleading simplifications as early as possible, and then to substantiate that picture once its outline is in place.

The chapter is organized as follows:

9.1 The Short Answer: A brief summary of the rest of this chapter, and an overview of the main experimental results.

9.2 A Road Map: Description of an iconic representation of the components of the dataset and analysis, and its use throughout this chapter to orient the reader.

9.3 Perceived Impedances of Real Devices: Presentation of results of the calibration trials taken at the beginning of each experimental session to determine the perceived impedance distribution of the real devices used during that session.

9.4–9.6 Pre-Analyzed Data: An initial analysis of the comparison trials, in which subjects matched a real device from a randomized set with a single emulation of given emulation variables settings and real device target; focusing on the analyzed effects of each of the primary experimental factors (VE expression, physical damping, sampling rate and targeted real device).

9.4 *Data as it was Collected*: Here we look at comparison trial data in the form it was taken, to give the reader a sense of where the more elaborately analyzed results originated.

9.5 *Analyses of Variance*: Shows the results of a series of analyses of variance conducted to look at experiment factor effects, based on the most simple response variable.

9.6 *Effects Decomposed*: Goes beyond the ANOVA to tell the "behind the scenes" story of the comparison trial data, demonstrating subtleties of the data when it has been reduced to effects averages but not to a single significance value.

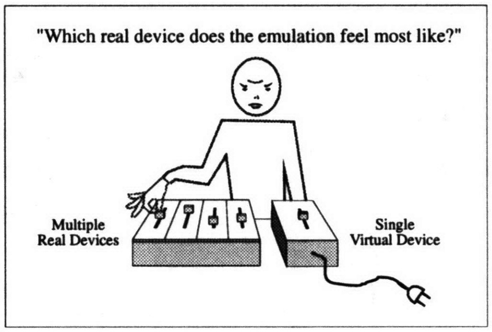
9.7 Response Variables: Detailed examination of the impact of using different response variables prior to effects averaging. The primary difference between the response variables is in how they account for the impedance distribution of the real device sets, and the differences between the measured and subject-perceived impedance distributions.

9.8 Subjects: Repeatability and Variation: Analysis of the "repeat" subject sessions for single-subject repeatability, and non-repeat sessions for subject-to-subject variation.

9.9 Decision Times as Indicator of Performance and/or Emulation Quality: Testing of the hypotheses "Does shorter $T_{Decision}$ = better performance?" and "Is $T_{Decision}$ an indicator of emulation quality?"

9.10 Subject Debriefing: Analysis of subjective post-session questionnaire responses.

9.1 The Short Answer



Controlled Variables	**Performance Indices**
• Real device targeted by emulation • Tested Emulation Variables: 1. Sampling rate (300, 600, 1200 Hz) 2. Physical damping (0.001, 0.07 N-s/mm) 3. Virtual environment expression (impedance, admittance)	• Score (% correct) • Factor significance (ANOVA) • Factor effect (effects analysis) • Decision time

Figure 9-1: Experiments were based on comparisons between a set of real devices and a single emulated device which targeted one of the real ones.

In a single trial, the subject compared a single emulated device to a set of real ones (Figure 9-1). One of the real devices was targeted by the emulation, and from trial to trial both the target and the emulation variable levels were varied. The response is the difference between the chosen device's index [1–4] and the targeted device's index, also [1–4]. Thus the simplest response variable is

$$Y_{unweighted} = \text{index}(target\ device) - \text{index}(chosen\ device).$$

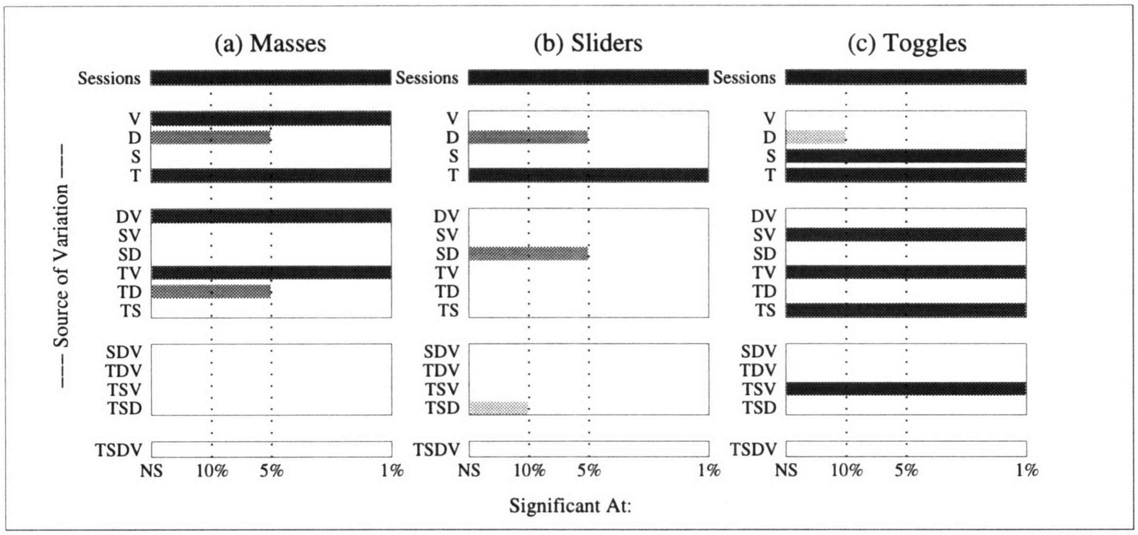
Throughout this chapter, I will present the results of computing the various performance indices listed in Figure 9-1.

Figure 9-2 summarizes the ANOVA significance results for each device family and a single response variable, $Y = \text{index}(target\ device) - \text{index}(chosen\ device)$.[1] Device families are shown in adjacent columns; the sources of variation are grouped into main effects and 2-way, 3-way and 4-way interactions and displayed in successive rows. The letter(s) at the left of each bar signifies the experimental factor(s) which are included in that significance test for that main or interaction effect.

---

[1] The Analysis of Variance (ANOVA):
For the reader unfamiliar with interpretation of the analysis of variance, note that the significance percent represents the probability that the hypothesis in question may be rejected. Thus, a small percent means that the effect is likely to be significant. The rule of thumb is to regard with skepticism significance results below 10%; significance at 1% is a strong indicator of that factor's influence on response variable variation. The significance of high-order interactions (greater than 2-way, i.e. the SDV interaction and below) is not greatly meaningful.

For more background on the experimental methods used here, see any reliable statistics text; recommended is [141].



V = Virtual environment expression	D = Physical damping	S = Sampling rate	T = Targets

Figure 9-2: Overview of ANOVA Effects Significance: all families and all effects, $Y = Y_{unweighted}$. Rows correspond to device families; columns to source of variation as computed by the ANOVA. The variation sources are grouped, from upper to lower, by Subjects; main effects; 2-way interactions; 3-way interactions and a single 4-way interaction. Shading and length of bar are redundant cues to significance; longer/darker means the source's variation is more significant.

Families

There is substantial variation in the magnitude and patterns of significance for the devices tested, which differed in type of impedance and complexity. Sliders had low error rates in general and few factors were significant. Toggles were characterized by very high error rates (average "scores" — % of correct identifications in a session) were the lowest for Toggles, see Section 9.8, and the various emulation variables tested greatly influenced the distribution of those errors.

Subjects

It is clear that subjects[2] were highly significant for all device families, and in fact they were the most significant source of variation with the exception of targeted real device. This is an expected result: individuals differ in their strategies of choice, the haptic features they notice, their tactile sensitivity and the degree of attention they gave the task. Every effort was made in the design and setup to control and/or minimize these sources of variation, and thus what shows up here is a good indicator of intrinsic variability. As such it should be regarded as an experimental factor of interest in and of itself rather than an obstruction to meaningful interpretation of the other factors, or a flaw in the experiment design. To this end, subjects are examined in detail in Section 9.8.

Targeted Real Devices

Varying the member of the real device set which was targeted in a given calibration trial greatly influenced the response variable for all device families. The effects analysis later in this chapter will

---

[2] Subjects correspond to experiment sessions, which were generally performed by different individuals and were analyzed as experiment "blocks".

reveal that that lower performance was usually correlated with the lower-impedance devices being targeted.

This is in part because Weber's Law in subject perception did not hold strong sway. The real device impedances were deliberately composed to approximate a log spacing on a mechanically measured scale, in the expectation that subject perception would be linear on this scale. Instead, the subjects appeared to have perceived real device impedances as approximately log-spaced. That is, the subjects' perceptual scales were closely correlated with the mechanically measured scales. The result was that the low-impedance devices were perceived as more closely spaced than the high-impedance devices, and the discriminations were consequently more difficult.

In addition, the low-impedance emulations were more often more difficult to stabilize for control reasons. The cost of stabilization in some cases was reduced servo gains, and consequent lower fidelity of emulation.

Damping

Variation of physical damping had a consistent but not strongly significant influence on average subject performance. We will see later that higher levels of the damping used here resulted in degraded emulation quality and this did impact on subject performance.

Sample Rate

Sampling rate was a significant main effect only under Toggles; this despite the fact that reducing sample rate to its lowest level, 300 Hz, did result in a perceptible degradation of emulation *quality* in some cases. The qualitative degradation apparently did not hinder the subjects' ability to make *functional* distinctions, except for the extreme case of Toggles. We will find shortly that much of the sampling rate significance in Toggles came from a set of four 300 Hz, Admittance expressed trials whose degradation was so severe they were unstabilizable; I did not present them to the subjects, but counted them automatically as errors. In most other cases, subjects made correct distinctions in spite of degradation.

This is an important example of the Quality versus Function phenomenon which will be discussed in greater detail below.

Virtual Environment Expression

The influence of admittance vs. impedance environment expression had a complex impact on subject performance; the degree and nature of the influence differed depending on the device family.

Masses: VE expression was strongly influential for the Mass device family. Looking at the effects will show that in fact the amplitude of average subject error was similar for each of the controllers; however, the polarity was flipped. Y=[Target-Chosen] tended to be positive for Y-expressed environments (subjects picked a real device of lower impedance than the targeted one); but Y was negative on average for Z-expressed masses (subjects picked real devices of higher impedance than targeted). This strong bi-polar effect implies that with a Y expression, subjects on average made positive errors, while with a Z expression they averaged negative errors.

Sliders: Varying VE expression did not alter average subject performance. This turns out to be a doubly-disguised result, which highlights the importance of looking at the data at both the big-picture level and in fine detail. The effects analysis will show this result to be flat but negative; i.e., both Y and Z expressions resulted in fairly large negative average error, implying some sort of systematic bias which acts for both expressions.

However, when the responses are broken down into subjects, it will be seen that the negative average was due almost entirely to a single outlier session characterized by a very high error rate, and is not really indicative of the "average" performance. The flat small-average-error response (yielding a non-significant result in the ANOVA) is, in the end, probably the most accurate description of the situation. The keeping/discarding of this single session will be discussed later, but the point here is that the big picture alone can be misleading.

Toggles: VE expression did not appear as a significant main effect with Toggles, but it was strongly influential as an interaction with sampling rate. This is because as sampling rate was drastically reduced (300Hz), the admittance-expressed environment became quite difficult to stabilize, whereas the quality and fidelity of the impedance-expressed toggle environment was almost insensitive to sampling rate even as low as 300Hz. At higher sampling rates, the two environment expressions resulted in similar performance rates.

Interactions

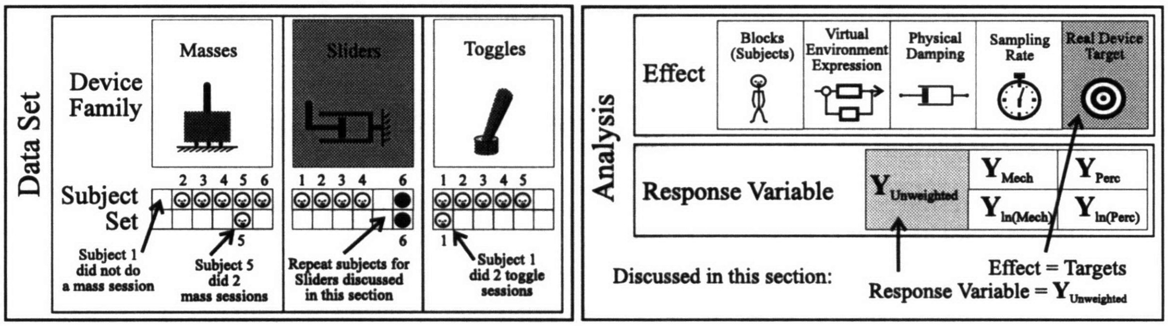
Most of the notable interactions have already been mentioned in conjunction with interpreting a main effect. Some features of Figure 9-2 remain to be noted; one is simply the number of significant interactions.

The Toggle device family has the most complex set of interactions; three main effects, three 2-way interactions and even a 3-way interaction are significant. In fact, every factor is significant at 1% in a main effect and/or a 2-way interaction. The Slider column, on the other hand, is notably lacking in interactions.

These results are consistent with experiment observation and with my experience in developing the emulations: the toggles were the most difficult devices to model and stabilize, and the resulting emulation fidelity and sense of mechanical passiveness were lower than for the other device family targets. Sliders were relatively trouble-free, simple to model and stabilize except at very small values of desired damping.

9.2 Navigating the Data: A Road Map

A large amount of data and a variety of analyses are presented here. In order to make the presentation more coherent, the graphical mnemonic shown in Figure 9-3 will reappear throughout the chapter to help orient the reader as to which subset of the data and which method of analysis is being examined in a given section. The left half describes the data set used in that section, while the right half identifies the analysis being presented.



Subject:	1	2	3	4	5	6
Pseudonym:	Josephine	Patty	Mike	Sam	Lorna	Hank

Figure 9-3: Map icon introduced. The left half identifies the data set used in a section of this chapter, while the right half summarizes the analysis that was performed on that data. A map region is turned "on" when darkened; the highlighted parameters in this example are for a hypothetical example

Data Set

The data collected during these experiments can be characterized and related to the data portion of the map icon as follows. The data portion of the map icon is the left half, identified as "Data Set".

- Three **real device families** were used. Each piece of data was obtained with respect to either Masses, Sliders or Toggles. The current device family is indicated by the highlighted Mass, Slider or Toggle icon at the top of each device family box.
- A single **session** consists of a set of 4 calibration trials and a set of 48 comparison trials performed by *one subject* on *one device family* — e.g. masses. Six sessions were performed for each device family, for a total of 18 sessions in the experiment.
- Six different **subjects** participated in these experiments. Each subject performed a total of three sessions: three subjects performed sessions for three different device families, and the other three subjects performed sessions for two different device families, with repeat sessions on one device family (see Methods, Section 8.3.5). Thus, there were five *different* subjects for each device family.

  The subjects that participated in a given device family set of sessions are indicated by the boxes in the Subject Set regions of the map icon that have people-faces in them (e.g., subjects 2, 3, 4, 5 and 6 participated in the Mass data set).

  The subject(s) whose data is "current" are indicated by darkened people-faces.
- A single subject is **repeated** for each device family; i.e., that subject performs two sessions for that family, and only one for another family. Conversely, for each device family, there is one

Reference Number	Response Variable		Computing Formula
1	$Y_{unweighted}$	$=$	index(*target device*) – index(*chosen device*)
2	Y_{mech}	$=$	$Z_{mech_T} - Z_{mech_C}$
3	Y_{perc}	$=$	$Z_{perc_T} - Z_{perc_C}$
4	$Y_{\ln(mech)}$	$=$	$\ln(Z_{mech_T}) - \ln(Z_{mech_C})$
5	$Y_{\ln(perc)}$	$=$	$\ln(Z_{perc_T}) - \ln(Z_{perc_C})$

Table 9.1: Computation of response variables.

subject who does not perform a session for that family (because that subject doubled in another family).

Subjects are referred to elsewhere in this chapter by pseudonym, according to the table beneath the map icon in Figure 9-3; this is to minimize confusion of subjects with session numbers, and to establish an identity when it is useful to do so. These are not the subjects' real names, nor do they necessarily represent the real subjects' genders. Gender was not explicitly examined in this study, and the gender scrambling helps to disguise real identities.

The "repeat" subject is indicated in the map icon by a person-face in the second row of the Subject Set region. The "omitted" subject is indicated by a blank first and second row for that subject number. e.g., subject 5 (Lorna) is the repeat subject for Masses. Lorna also performed a single Toggle session, but was the omitted subject for Sliders.

Analysis

The analysis parameters currently in effect are indicated in the "Analysis" section of the map icon. The scope of the analysis can be characterized by just two parameters: the effect(s) under examination and the response variable(s).

- Experimental **effect** is the emulation variable or emulation target whose variation is being examined.

  The effect being analyzed in the current section is indicated in the map icon by highlighting of that effects box.

- The **response variable** is the math formula used to take the raw subject response data to the input of the ANOVA and effects analyses. The response variables were used as defined in Table 9.1. A complete description of these relations can be found in Methods, Section 8.2.7. The current response variable is indicated in the map icon by highlighting of that response variable's box.

9.3 Calibration Trials: Perceived Real Device Impedance Distributions

Here I present and analyze the data collected in the calibration trials which were conducted at the beginning of each experiment session. The purpose of the calibration sequence was twofold:

1. To gain insight into how subjects perceive the set of real devices used in these experiments — including such questions as impedance distribution, confusion as to ranking of near neighbors, and repeatability of both impedance magnitude judgments and ranking. Such information will be particularly useful in future iterations on this design.

2. As the basis of response variables used to analyze the comparison trial data. Thus, the calibration data comprises part of the total data set used in the section of this chapter which looks at the effect of response variables, Section 9.7.

Some terms need to be reviewed before proceeding:

Measured Impedance Distribution: The mechanically measured impedance of the real device set.

In general the procedure departed from the ideal in two ways:

1. An assumption and/or simplification was made as to the dominant impedance characteristic of the real device; e.g., mass for the Mass family (a very good assumption) and compliance for the spring toggles (less good, as those devices had a complex impedance).

2. Due to limited control over the impedances of the real device set, the set's impedance distribution diverged to some extent from the intended true log scale.

Although these observations are detailed earlier in Section 6.1, I repeat them here in order to emphasize that contrary to the ideal scenario, the emulations were to some extent modeled subjectively and imprecise in both measurement of their impedance magnitudes and in adherence to their specified impedance distribution. This renders suspect some of the language used here — references to "hits", "misses" and "scores" which imply that there is in every case a "right" and "wrong" answer, when in fact the situation is not so clearcut.

However, it does not undermine the basic experimental question and premise, which are (a) to ask how a specific set of emulation variables influence the emulation of a specific set of haptic environments; and (b) to employ emulation variable modification in conjunction with randomized variation of emulation target to determine those influences. The strongest evidence that the emulations were of reasonably high fidelity under optimum conditions — whatever means were used to attain that fidelity — is that the majority of the time subjects did choose the targeted real device. The errors then served to ascertain what are and aren't optimum conditions.

Subject Perceived Impedance Distributions: the relative difference which subjects perceive between adjacent real device family members (here, four real device members were used).

Elucidation of subject perceived distributions are the point of the calibration data. The data are always relative rather than absolute because of the ranking method enforced, which pegged the low-impedance end of any given device set to a ranking of "1" for every trial and the top to "10".

Since four calibration trials were taken for every session, the subject perceived distributions shown here are averages from those four trials. Trial-by-trial data is shown at the end of this section.

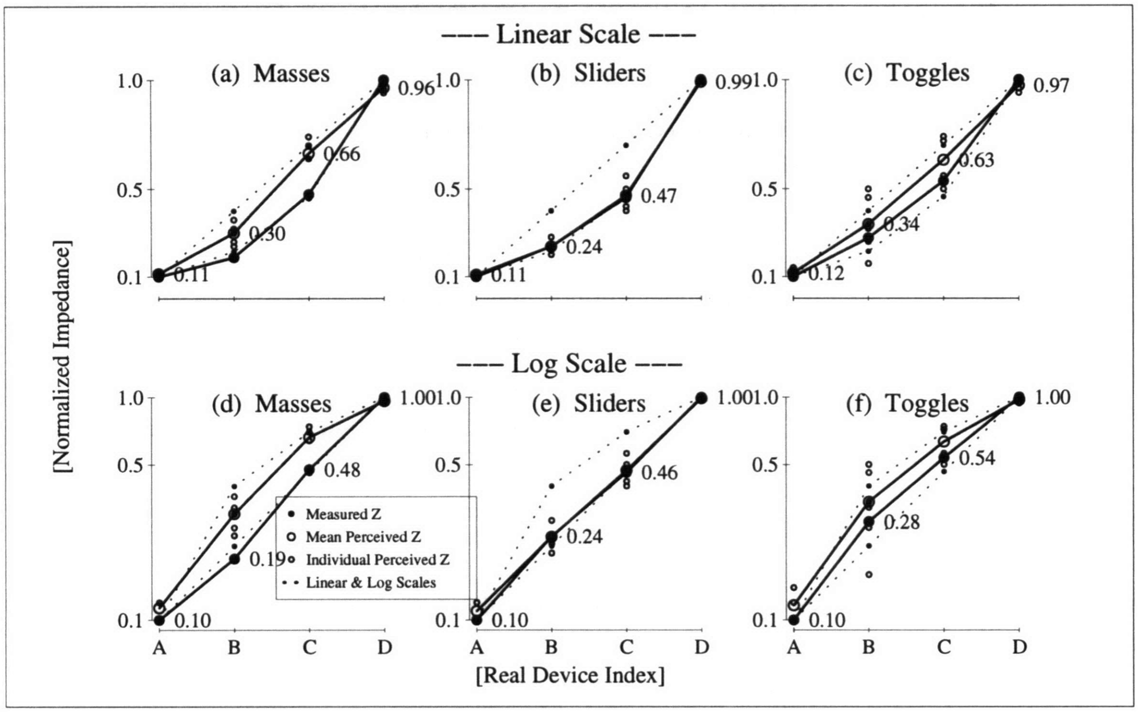
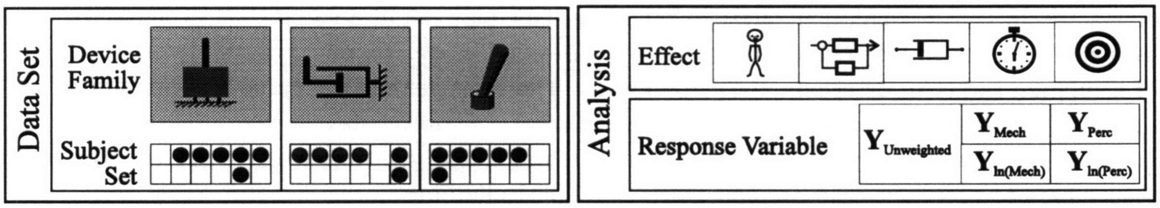


Figure 9-4: Measured and average perceived real device impedance distributions.

Average Subject Perceived Impedance Distribution: averages of all subject perceived impedance distributions.

In some instances during the analysis it was useful to generalize from the individual case and see how the subjects' perceived impedances compare, as a whole, with measured device impedances. For this purpose, the perceived impedance distributions obtained from all the sessions for a given device family are averaged together to come up with a single four-point distribution.

Comparison of Average Subject Perceived Impedance Distribution with Measured Distributions



The average subject-perceived impedances for the real devices used in this experiment are shown in Figure 9-4 for each family, along with the measured impedances. All plots show ideal log and linear curves (dotted lines) to demonstrate the nature of the impedance spacing that was achieved. The first row (a–c) have a linear Y-axis; the second row (d–f) shows the same data with logarithmically a scaled Y-axis. Thus curves which demonstrate close to a log distribution in (a–c) appear as a straight line in the second row.

Each perceived curve is the average of results of the calibration trials for each of the six sessions taken for that device family; and each of those session curves (whose values are indicated by open circles) is the averaged response from the four calibration trials taken during each session.

Subject Ordering of Real Devices

All subjects produced a correct ordering every time, although there was the expected confusion in ordering the duplicate devices which indicating that the subjects could not reliably distinguish them. There is therefore no need to account for out-of-order data in what follows.

Anchoring of Impedance Scales

Subjects supplied impedance magnitudes on a scale between 1 and 10, with the bottom and the top of the scale pegged to their choice of the lowest and highest impedance device in the set of five devices that they used in a given trial.

Each value shown in Figure 9-4 is normalized to the maximum in its set (always 10 for the subject distributions), pinning the top of both the measured and perceived scales to 1.0. In addition, the low end of the measured distributions are tied to 0.1, in order to make cross-family and measured/perceived comparisons possible.

This might lead to the following incorrect conclusions:

1. That the low end of each real and perceived device-family impedance range is the same absolute percentage of maximum impedance, i.e. 10 percent.

   In fact, the absolute value of any of the impedances, in perceptual terms, is difficult to know. It is much easier to obtain the subject's perception relative to other impedances (this was discussed further in Methods, Section 8.2.2).

2. That the same range, in proportion, is covered in each real device family's impedance range.

   In fact, it is a ratio of 5:1 for Masses, 32:1 for Sliders and 9:1 for Toggles (see Real Devices, Table 6.1), extending well into both sides of a single order of magnitude.

Artifact of Calibration-Trial Device Duplication

Since the each device was duplicated in one of the 4 trials for each subject, most subjects did not consistently rank duplicate devices as equal; that is, when Device D (Z_{max}) was duplicated, they did not supply 2 '10' rankings, but generally gave them 9,10 or 8,10 rankings.

This inconsistency (a predicted tendency to not recognize two devices as identical when not told to look for a duplicate pair) explains why the individual session marks and the mean perceived curve do not always fall exactly on '1.0' and '0.1' for the highest and lowest-impedance devices, respectively (Z_D and Z_A); although the measured curve does because it was explicitly forced to those values.

Correspondence of Subject-Perceived to Measured Impedance Distribution

In all cases, the distribution of average subject impedance perception lies between the measured distribution and a linear distribution; i.e., it tends to be more linear and less log-spaced than the measured distribution. However, there is a notable difference in the degree to which the two bold curves in each plot in Figure 9-4 agree, and in the manner in which they differ.

The subject-perceived distribution for Masses is nearly linear, compared to a measured distribution that comes reasonably close to log spacing; this family is nearly a perfect illustration of Weber's Law perception.[3]

---

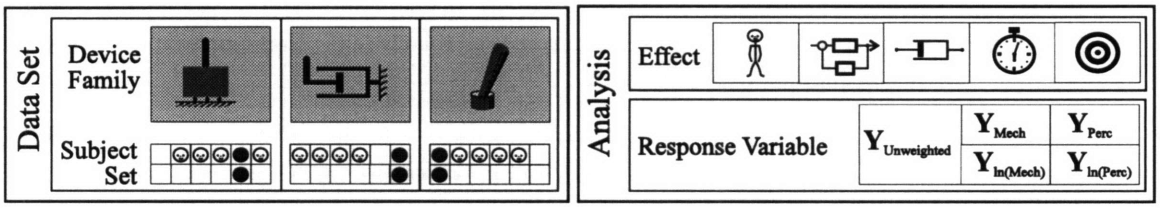
[3] According to Weber's Law, various aspects of the human perceptual scale is logarithmic: the ratio of JND/stimulus magnitude is approximately constant [62]. Thus the difference between impedance units of "1" and "10" and between "10" and "100" should be perceived as similar.

For Sliders, the measured and subject perceived distributions are virtually identical in shape, implying that the subject perception of the log measured distribution is linear; i.e., it does not follow Weber's Law at all.

Toggles are somewhere in between; but note that the spread here (visible in the degree to which the small open-circle markers, representing points from the six sessions for that family) is much larger than for Masses and Sliders. In fact, the variability in this case envelopes both of the reference distributions (linear and logarithmic), not to mention the measured distribution which is somewhere in between. In this case, the only safe conclusion is that the subject variability was the dominant feature of this distribution. It is possible that for this device family, by far the most complex tested, subjects were paying attention to different features in making their impedance rankings. It could also be hypothesized that they were internally inconsistent — generally confused about what they were measuring — but a look at trial data in the next figure does not support that this was more of a problem than for the other families.

In summary, (a) there is consistently a divergence between measured and perceived spacings; and (b) the divergence is consistently in the direction of Weber's Law, but not by a an overwhelming amount. Weber's Law is neither strongly upheld nor discounted; but linear perception is probably a better description of what these subjects did than logarithmically transformed perception.

Subject Perceived Impedance Distributions



The "average" picture is a useful overview, but it disguises some important detail. This section presents the data shown in Figure 9-4, individual data points may be viewed by session, subject, device and trial.

Figure 9-5 shows all of the calibration data taken during the 18 sessions, organized in a sessions × families grid (each of the 18 plots represents a single session). The subject who performed each session is identified, and each plot holds three kinds of curves:

1. A heavy solid line indicates the mean (based on response from all trials in the session) impedance distribution. Each solid line is the average of the four trials taken in a session; each trial consisted of a ranking / magnitude assignment of five real devices (including 1 duplicate). Thus each point in the heavy solid line represents an average from five data points.
2. Two light solid lines show the mean for all sessions for that device family and for the measured distribution for that family, respectively.
3. Ghost lines indicate ideal log/linear distributions.
4. In addition, open-circle markers (large for the subject, small for the device family average) signify the subject perceptual data, as opposed to solid-dot markers for the measured distributions.

Influence of Audible Cues

Several of my subjects offered the comment that members of the Masses and Sliders families did not sound the same, and that they found this either a giveaway as to which real device they were

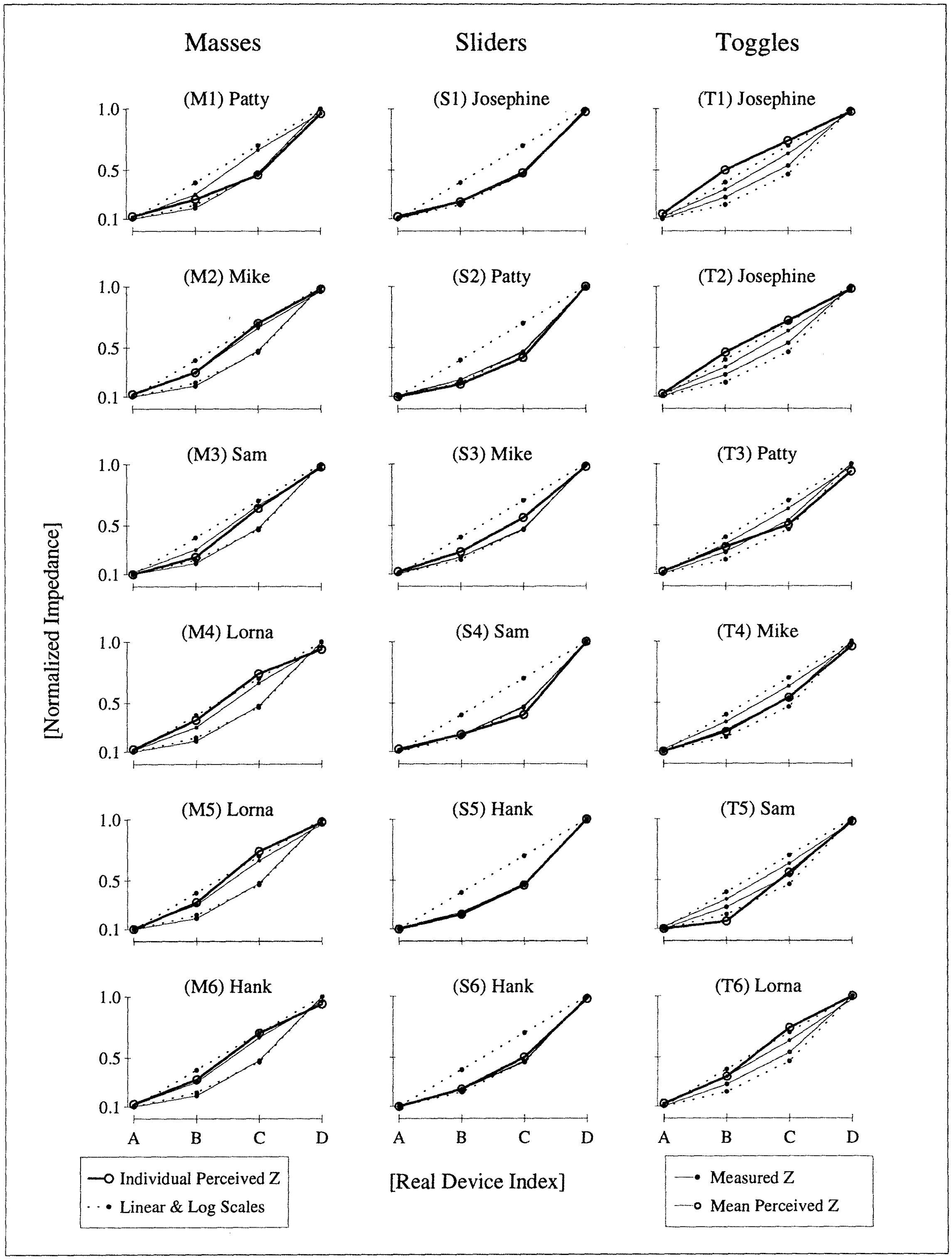


Figure 9-5: Perceived real device impedance distributions for each subject.

handling or simply distracting. The Masses, for instance, made a subtle "whir" from the rolling bearings which was dependent on the device's mass, and was audibly different in pitch between the lowest and highest in the same family; and all of the real devices sounded qualitatively different from the emulations.

This is evidence that sound may be important in the emulation both of environments containing high-frequency components (e.g. walls and clicky detents) and for some continuous environments.

Log/Linear Character of Subject Impedance Distributions

Distributions for Masses and Sliders are remarkably consistent between subjects, with a few exceptions. M1 (Patty) is much lower than everyone else in her estimation of Z_C, and much closer to the measured distribution. M3 (Sam) is low for subjects on Z_B, but typical for Z_C. For sliders, S3 (Mike) is the only session which is clearly distinguishable from the mean.

Toggles are a different story. The spread suggested by the individual point markers in the previous figure is clearly seen here. Note, however, that most sessions are consistent in shape of the distribution; i.e., few cross over the mean. Most are always either above (a high estimate for both Z_B and Z_C) or below (a low estimate for both Z_B and Z_C).

This is the correct interpretation of an above-mean or below-mean (or above/below-measured) curve, since the top and bottom of the scale was fixed to Z_A and Z_D for each trial; a '0.9' magnitude for Z_C in Figure 9-5 means that in that session, the subject's response averaged 9 on a scale of [1–10] where Z_A=1 and Z_D=10. If either end of the subject's impedance scale had been allowed to float, the magnitude estimations would have been relative to something unknown rather than relative to the devices in the set, and the meaning of a '0.9' would have likewise meant something unknown and probably different for each subject and possibly each trial in a session.

Specific Toggle observations include Josephine, the repeat subject for this device family (T1 and T2), who is notable in that her two sessions are quite different from all the other sessions, and nearly identical to each other in high overestimation of the difference between Z_A and $Z_B$and extreme departure from the measured distribution. Also, much of the variation in the Toggle sessions can be accounted for by a single device; T3, T5 and T6 (Patty, Sam and Lorna) more or less agree with the mean except for Z_C, Z_B and Z_C, respectively (not in the same direction).

Patterns Among Subjects

The single-session display of Figure 9-5 allows us to track patterns in subjects. Note that sessions are not the same as subjects; thus tracking across rows does not facilitate subject comparisons — names must be checked explicitly. For example, Patty appears in M1, S2 and T3 while Josephine appears in S1, T1 and T2.

Specifically, Patty and Sam are consistently (for all families and every point) slightly below mean. Lorna is always slightly above mean, particularly for Z_C (both Masses and Toggles; she is a repeat on Masses). Mike refuses to be characterized: he is right on the mean for Masses, above for Sliders and below for Toggles.

The presence of such within-subject consistency and consistent across-subject differences suggests that subjects might have been keying on different things or have slightly different internal scales — or that some have internal scales, and some were inventing their scales on the fly.

Consistency in Identifying Duplicate Real Devices: A Trial-by-Trial View

Can subjects reliably tell the difference between all the devices, or do they consistently confuse some of them? This is particularly of concern at the small end of impedance scale, since we have now

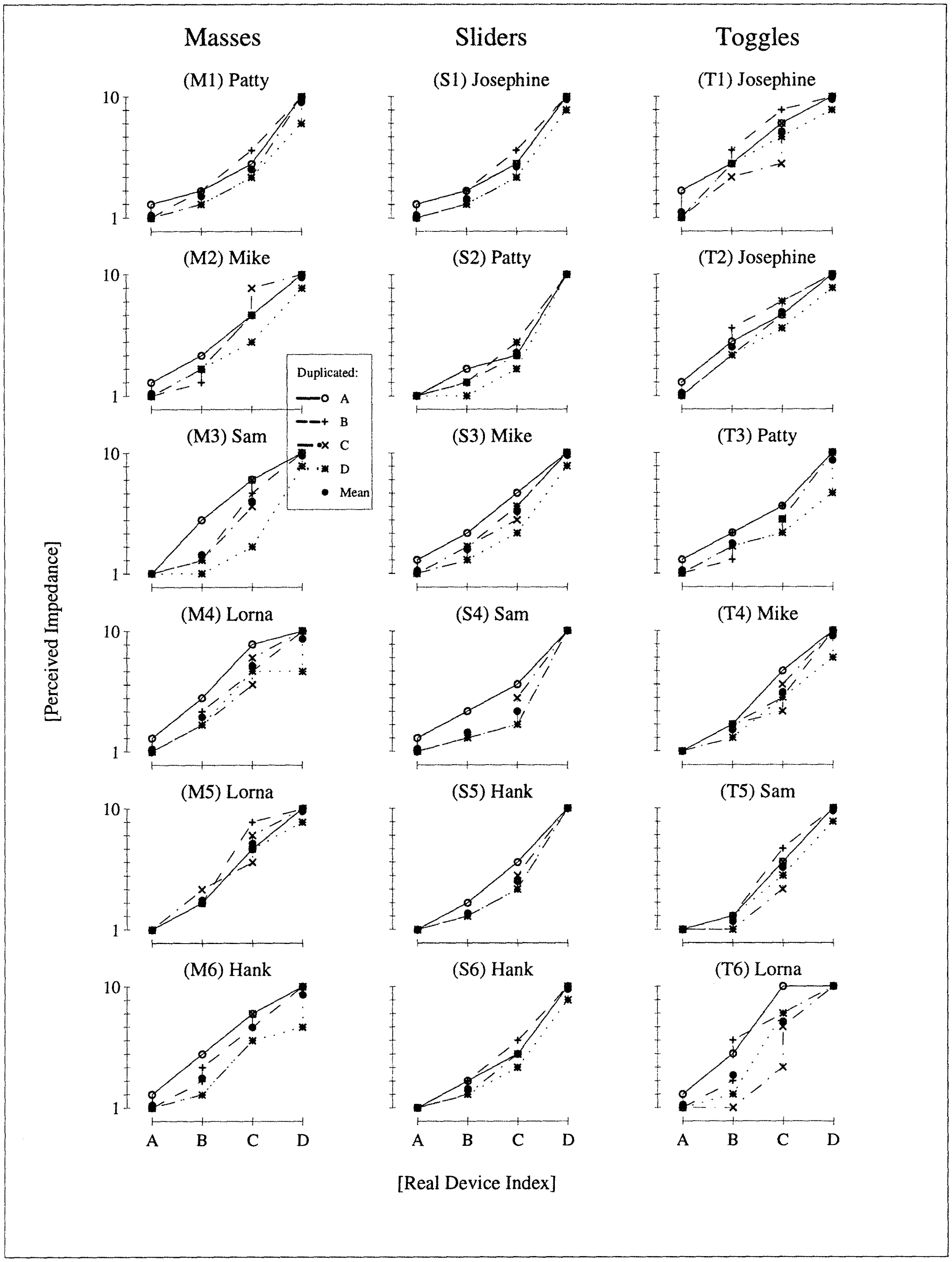


Figure 9-6: Subject perceived impedance arranged by trial.

seen that most subjects are viewing the spacing of these real-device sets as non-uniform, closer at the small end.

Turning to Figure 9-6, a substantially noisier picture emerges than before. Here we see the averaged session data of the previous figure shown in its full disordered detail of every data point for every trial. The usefulness of this representation is that it lets us see what a subject did trial by trial, and in particular it lets us see what subjects did when presented with duplicates.

The lines connecting magnitude markers for each real device member are coded by trial; thus the solid line is always Trial 1 and the dotted line Trial 4. Each line makes five stops across the four devices, because five devices were used in each trial (one duplicate). The length of the vertical segment of each line joining the duplicate pair shows the difference in magnitude assignment that a subject made for a duplicate pair that was presented together.

Note that during the trials, the devices were presented in random order, and the order in which the duplicate member was presented was shuffled as well. Thus the A–D ordering suggested by this display has no bearing on the experiment presentation order. Likewise, the trial ID shown here (1–4) indicates the order of duplication, not the order of presentation. Device A was always duplicated in Trial 1, B in Trial 2, etc.

Once we understand how to navigate through this figure, some observations may be made:

- Toggles are not obviously more messy/inconsistent than the other families in general correspondence of within-session curves, regardless of what the variation seen in session means in the previous figure (9-5 suggested. Similar-appearing amounts of fine detail variation is averaging out to different amounts of mid-level variation.

- There are some major mismatches in duplicate magnitude estimation. Toggles do tend to have more of these than the other families; every session, in fact. Lorna (T6) is the worst. See Section 9.3 for more on this.

- Masses and Toggles have about the same frequency (counting down columns) of misses on duplications of Devices A and D. Sliders did better in that respect at about 50% of sessions calling the A and D duplicates correctly.

- There is some degree of subject trend in trial consistency between device families. Patty (M1, S2 and T3) and Josephine (S1, T1 and T2) are quite trial-consistent in all their sessions, particularly in the matter of duplicate identification (Josephine makes a bad duplicate mid-ID in T1, but so does just everyone when they are confronted with Toggles). Sam (M3, S4 and T5) is the least trial-consistent subject for Masses and Sliders, although he does relatively well with Toggles. Lorna (M4, M5 and T6) appears to have more trouble than most with identifying duplicates as identical.

Consistency in Magnitude Evaluations: A Device-by-Device View

Now, we will focus on the devices themselves, looking at how much variation there was in magnitude estimations cumulative over the four calibration trials for each device (Figure 9-7). Each plot represents a single session; the x-axis is real device family members of increasing impedance, and the y-axis is subject magnitude assignment.

A subject has 20 magnitude rankings to "spend" in a session (five per trial), and each real device ultimately gets five of them over the course of a single session's four calibration trials. There are five dots arranged above each real device in each plot in this figure; the vertical positioning of the dots above the device compose a miniature histogram that demonstrates the distribution of discrete magnitude rankings assigned that device.

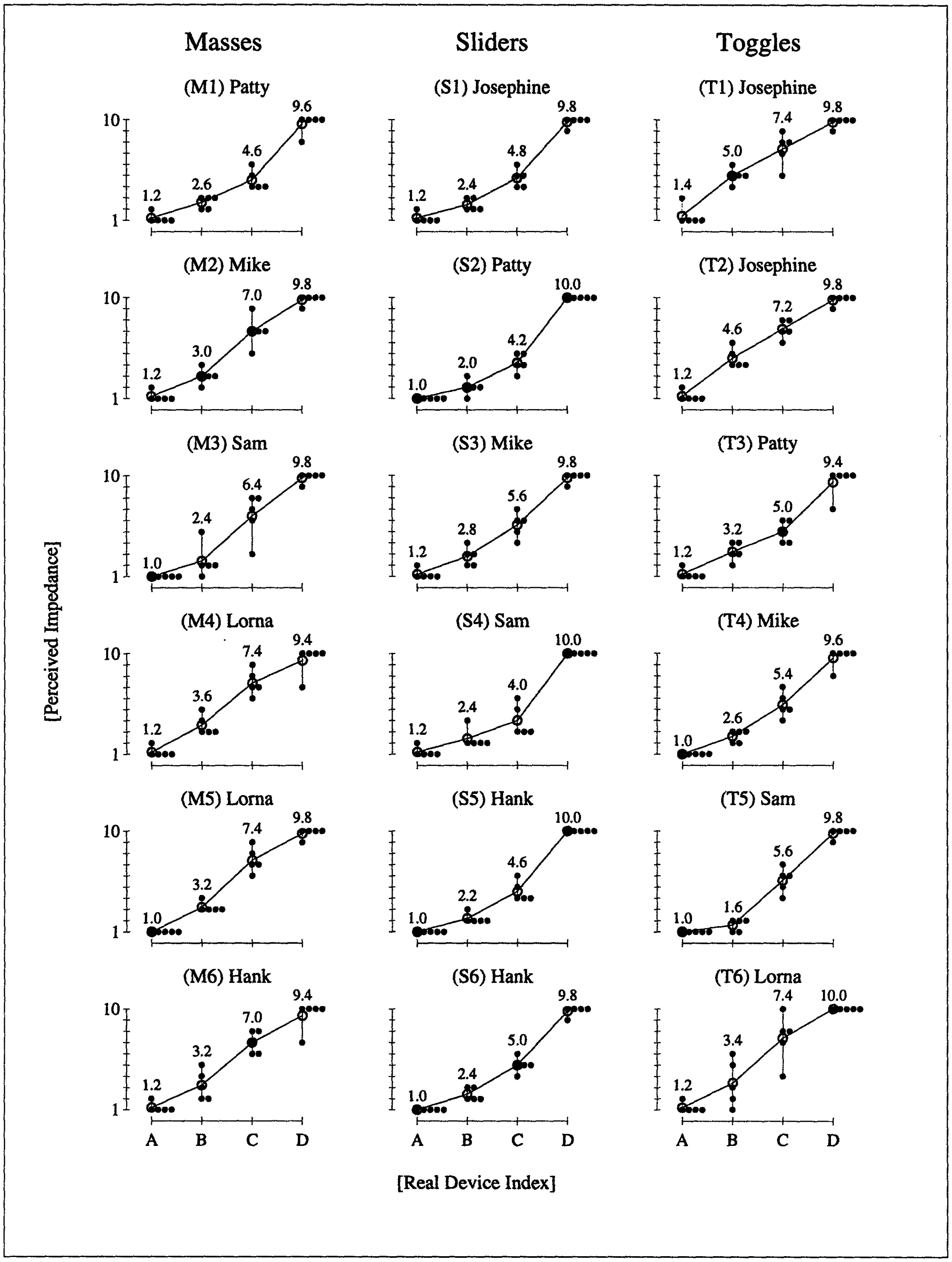


Figure 9-7: Subject perceived impedance, with distribution histograms arranged by device.

While there is no information about which trial that ranking came from, its relation to other rankings in the same trial and to a duplicate device, it is now easier to see what is happening for a single device in terms of range, outliers and adjacent-member confusions.

We might have predicted a roughly Gaussian spread at each level, and in fact this is approximated several times under Sliders: Josephine, Patty and Hank(s) (S1, 2, 5 and 6) have nicely clustered groupings with range not exceeding two or three magnitude rankings and a peak in the middle, if there is more than a two-ranking spread.

In general, however, the situation is not so pretty. Outliers are a regular sight (M2, M3, M6, T1, T3 and T6) — but we do know from the previous figure that sometimes these are associated with an offset trial due to duplicating. e.g., the over-estimated outlier for Device B in M3 is in Trial A, and the under-estimated outlier for Device C in the same plot is with Trial D.

Perception of Duplicated Devices

Does the duplication data reveal whether each "duplicate" pair of devices really felt the same?

One real device (A–D) was duplicated in each of the four calibration trials (review Table 8.5 in Methods). Therefore, in a single session each real device receives five rankings, three of which correspond to non-duplicate trials and two to the trial in which it was duplicated. Figure 9-8 shows the rankings given to the primary and duplicate member of each duplicate pair during the duplicate trials for each device:

Primary: The first member of the duplicate pair; this one is used exclusively in all non-duplicate trials. This quartile, however, contains only points from the trials (one per session) in which this device was duplicated. 1 point per session × 6 sessions = 6 points in this quartile.

Duplicate: The second member of the duplicate pair is used only in the trial in which this device was duplicated. There are 6 points in this quartile as well.

All: All magnitudes given for this device from all trials, all sessions for this family. Thus there are $5 \times 6 = 30$ points in the quartile, of which 12 are data from trials in which this device was duplicated, 18 from trials in which it was not duplicated. 24/30 points are for the primary device, and 6/30 for the duplicate pair.

Measurements were reasonably consistent for most members of Masses and Sliders; Device D for Masses shows a little more range than we would like. Toggles were less consistent, particularly for Devices B and C. The duplicate Device C is almost as low as the the duplicate Device B; thus we expect to see confusion in the comparison trials for this device.

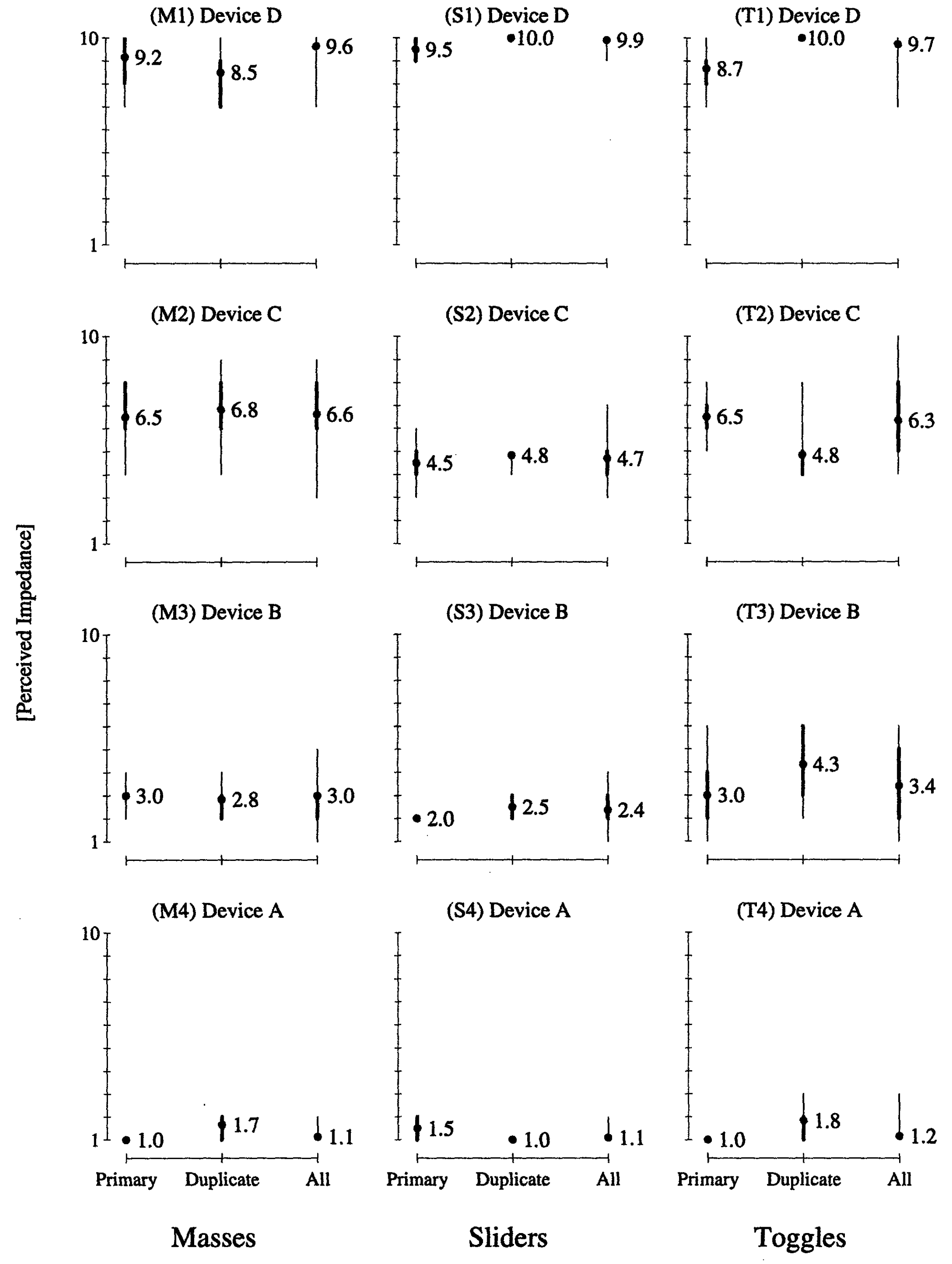
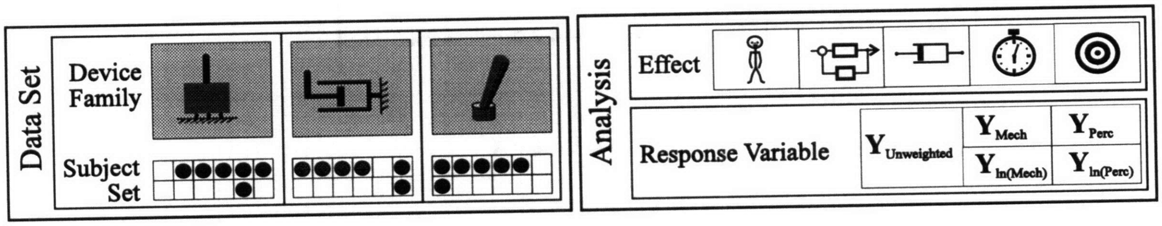


Figure 9-8: Consistency in perception of duplicated devices: Data for all sessions are shown. For a single device, the three categories are "Primary" (data from one member of the duplicate pair during the duplicate trial), "Duplicate" (from the other member for the duplicate trial) and "All" (the sum of the first two plus data from all non-duplicate trials for that device).

Trial	Target	Sampling [Hz]	Damping [N-s/mm]	VE Exp
1	A	300	0.001	Y
2	A	300	0.001	Z
3	A	300	0.070	Y
4	A	300	0.070	Z
5	A	600	0.001	Y
6	A	600	0.001	Z
7	A	600	0.070	Y
8	A	600	0.070	Z
9	A	1200	0.001	Y
10	A	1200	0.001	Z
11	A	1200	0.070	Y
12	A	1200	0.070	Z
13	B	300	0.001	Y
14	B	300	0.001	Z
15	B	300	0.070	Y
16	B	300	0.070	Z
17	B	600	0.001	Y
18	B	600	0.001	Z
19	B	600	0.070	Y
20	B	600	0.070	Z
21	B	1200	0.001	Y
22	B	1200	0.001	Z
23	B	1200	0.070	Y
24	B	1200	0.070	Z
25	C	300	0.001	Y
26	C	300	0.001	Z
27	C	300	0.070	Y
28	C	300	0.070	Z
29	C	600	0.001	Y
30	C	600	0.001	Z
31	C	600	0.070	Y
32	C	600	0.070	Z
33	C	1200	0.001	Y
34	C	1200	0.070	Z
35	C	1200	0.070	Y
36	C	1200	0.070	Z
37	D	300	0.001	Y
38	D	300	0.001	Z
39	D	300	0.070	Y
40	D	300	0.070	Z
41	D	600	0.001	Y
42	D	600	0.001	Z
43	D	600	0.070	Y
44	D	600	0.070	Z
45	D	1200	0.001	Y
46	D	1200	0.001	Z
47	D	1200	0.070	Y
48	D	1200	0.070	Z

Table 9.2: Standard order convention.

9.4 Comparison Trials: "Raw" Data



With this session we begin analysis of the emulation/real device comparison trials. The data is shown here in a minimally processed form to clarify the origins of the more sophisticated ANOVA and effects discussions, covered in Sections 9.5 and 9.6 respectively. The data shown here is not quite "raw" in that it has undergone a first level of processing (differencing and sometimes averaging), but only here is it displayed point by point, by trial and subject and in a manner suggestive of the way it was collected.

Recall that a full factorial design was employed: all 48 combinations of experiment factor levels were presented exactly once in every block (session) (Table 9.2; for a review of standard order and the randomization process used here, see Section 8.3.4). The pre-processed data is presented by device family. For each set of device family data, we will look for (a) indications of non-randomness in subjects for any of the families, and in whole-family averages; (b) learning effects, particularly through differences between Mass averages (first session performed for most subjects) and the other two families; and (c) systematic patterns in the standard-order data presentation for each family. Any such patterns should appear later as statistically significant effects.

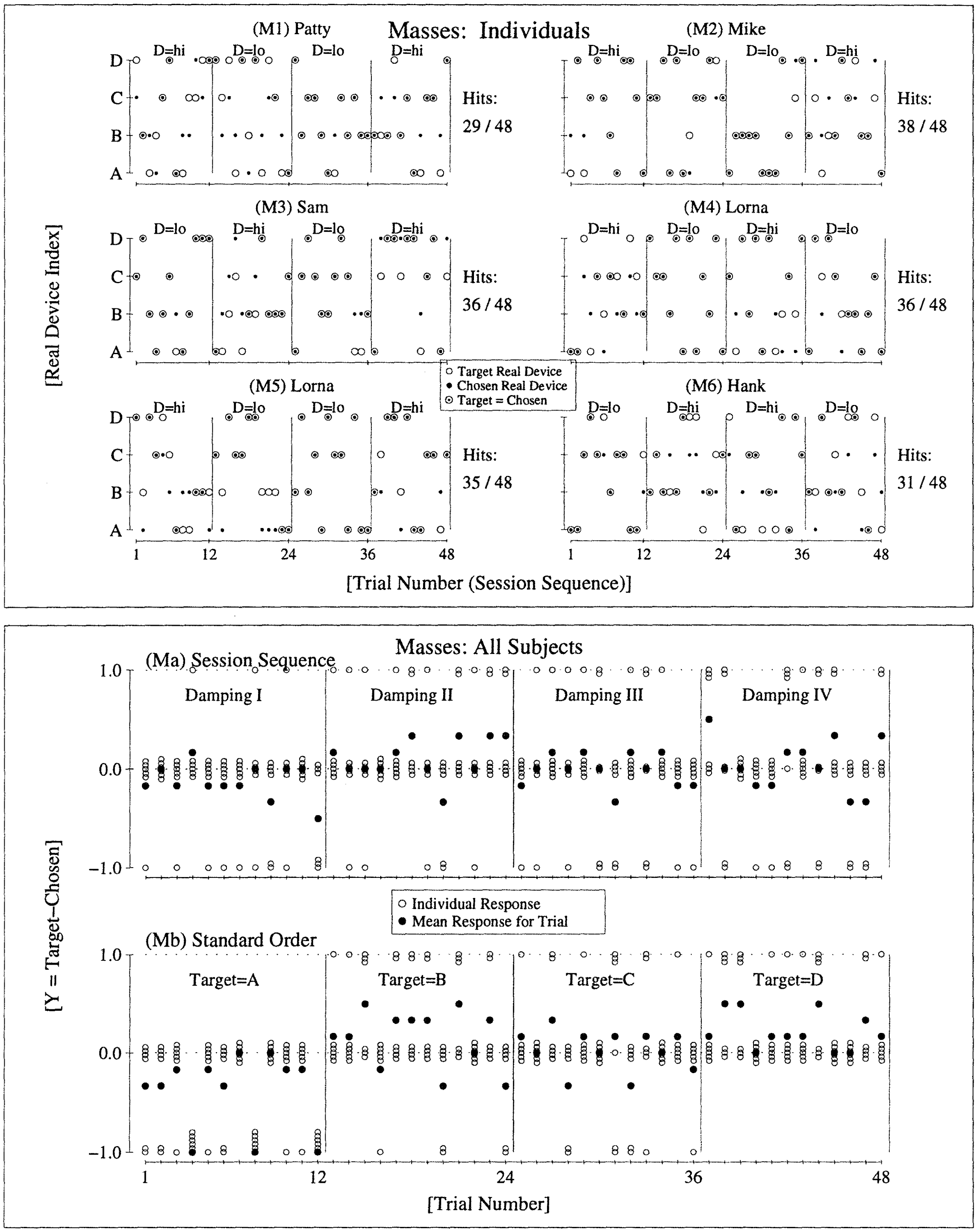


Figure 9-9: Minimally analyzed data by trial: Masses.

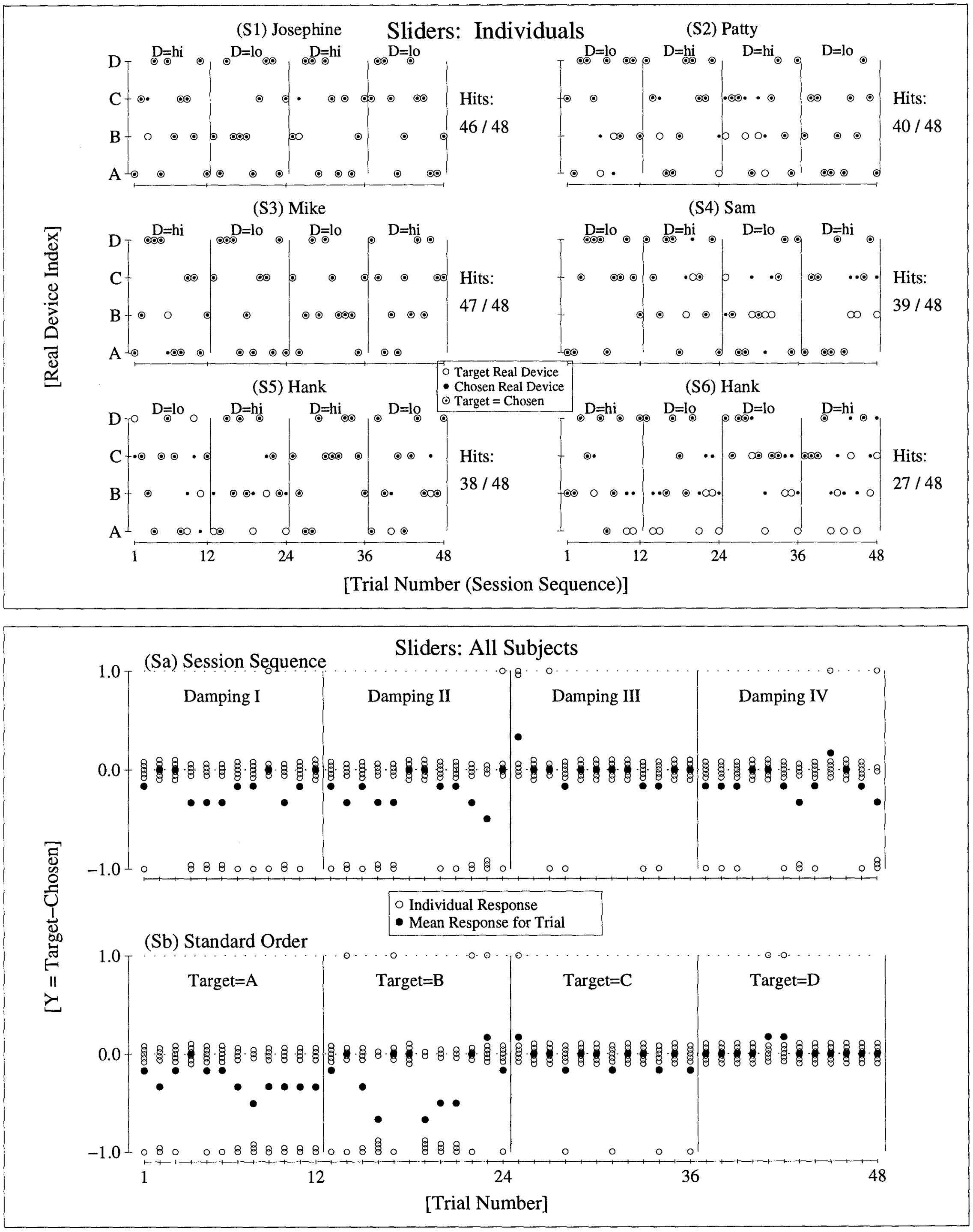


Figure 9-10: Minimally analyzed data by trial: Sliders.

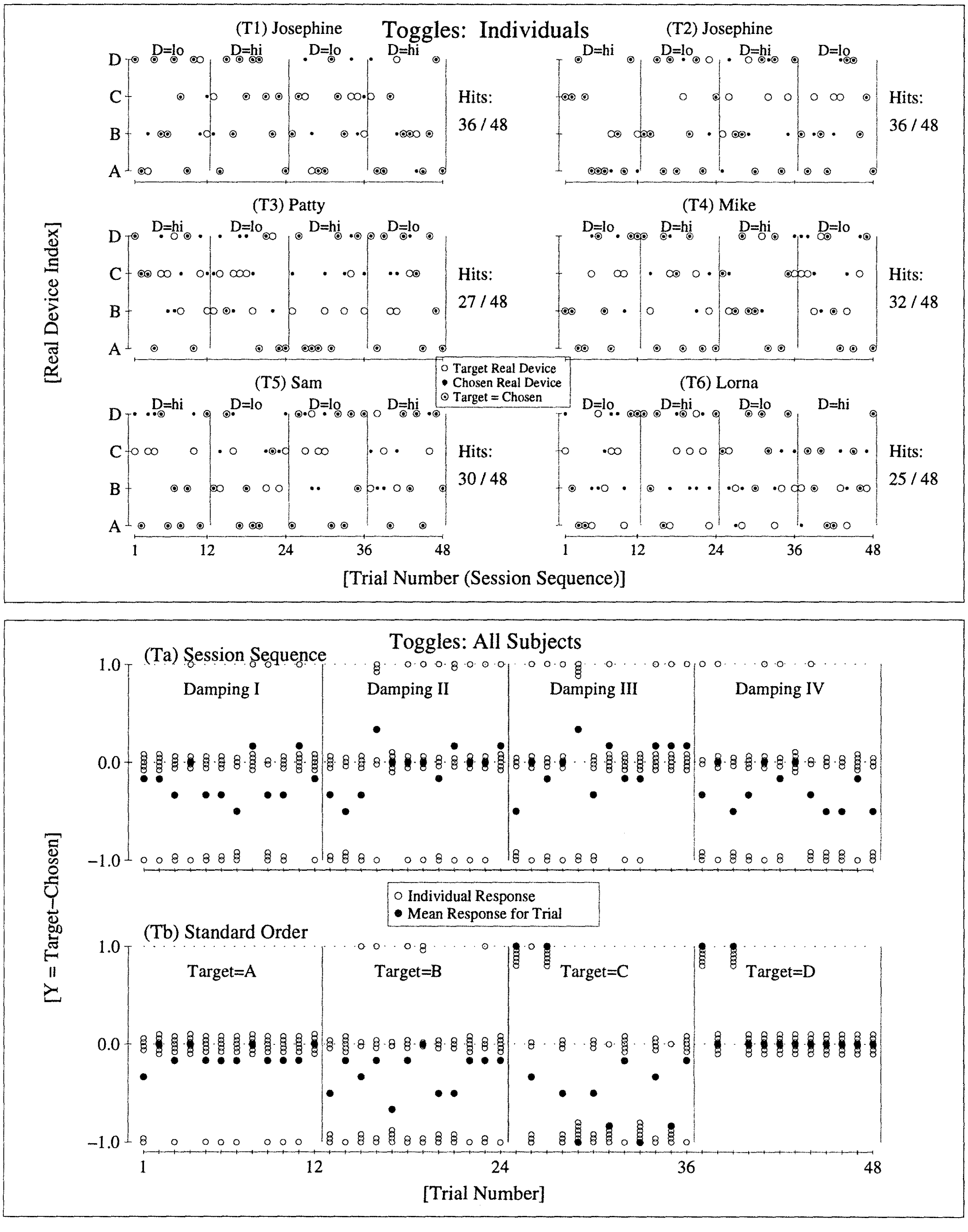


Figure 9-11: Minimally analyzed data by trial: Toggles.

Individual Sessions

The upper half of Figures 9-9, 9-10 and 9-11 each shows the response given by every subject for every comparison trial, for the indicated device family. Figure 9-9 covers responses taken for the Masses family; we will start there to explain how this data display works, then look at the other device families.

Plots M1–M6 are for individual sessions, indicating which real device (Y axis) was targeted that trial (x-axis) by an open circle; and the real device chosen by a small filled dot. A "hit" is indicated by the target and chosen marks being overlaid: a dot inside a circle. The "score" (number of trials in which the subject chose the targeted device) for each session is listed to the right of each bulls-eye plot. The trials are listed in session sequence; thus the dotted lines indicate regions of constant damping.

While the randomized order of trials is represented by the order of target/chosen responses in the x direction, this plot does not say anything about the order of device presentation. The real devices are listed vertically on the Y axis in the same order for each trial display; during the experiments the order was shuffled for each trial.

Subject Performance Patterns and Rates

The overall rate and pattern of hits and misses is not notably different by subject, nor does it appear to be skewed in the direction of any real device.

Session Sequence Trends

The trials are arranged in session sequence, so we can see on an individual basis whether there are trends across the session which may be attributed to learning and/or fatigue and boredom. We can also look at the quarter-session sets (demarcated by the tick marks at Trials 12, 24, 36 and 48) for signs of change in overall performance due to damping, which was altered at those points.

As before, the sequences appear to be uniformly distributed. The most notable pattern is Sam's (M3) fourth quarter, where he makes a large number of D identifications, about half of them positive. This is more a function of the randomization algorithm than his response.

Comparison across Device Families

The Subjects plots for Sliders and Toggles show different average error rates and, for Toggles, more variability among subjects in terms of score. Hank under Sliders (S6) stands out, with a large number of negative errors (Chosen > Target) for low impedance devices. Toggles shows a higher error rate than the other families. None of the families show any obvious temporal trend in response.

All Subjects Shown Together

The bottom half of each of these three raw data figures (9-9–9-11) plots the same data as their upper halves, but with all subjects shown together. Again, we begin with Masses to see how these plots work.

Trends in Mean by Session Sequence

In Plot (Ma) the data from all sessions in that device family are plotted with the same x-axis convention as in the upper half of the page (M1–M6); trials appear in session sequence, but with the y-axis now showing the difference between the index of the targeted device and the index of the chosen device in a given trial. When only near-neighbor misses are made (the case here), there are three possible values for each data point: [-1.0, 0.0, 1.0]. Session data points are represented by a small open circle; the mean for all the trials which occurred at that trial sequence number by a filled circle. Thus any point in the first Subjects plot (M1) which appeared there as a bull's-eye (dot and circle overlaid) appear here in the same horizontal position but as a single open circle located vertically at 0. Positive and negative errors are stacked up at 1 and -1, respectively.

This display makes it possible to see time-trends; that is, change in error rate either by absolute value or polarity as the session progresses. If found, such trends are likely to be attributable to learning and fatigue or boredom. There are no visible temporal trends in this data.

Because of the restriction on randomization for damping, each quarter-session of the session-sequence chart represents a change in damping level. However, the value of damping for a given 12-trial set varied from subject to subject and thus no conclusion can be made about general effects as a function of damping from this figure.

Trends in Mean by Standard Order

(Mb) uses the same conventions as (Ma), but in this case the trials on the x-axis are arranged by standard order and have no relation to the sequence in which the data was actually obtained. The advantage of this display is that now every point on the x-axis corresponds to a single setting of levels for all four experimental factors; thus every data point (represented by small open circles) may be compared with the other five in its column knowing that they were taken under the same experimental (factor level) conditions, although generally at a different point temporally in the session. The type of observation made most clearly from the second format are trends in response to varying the Target factor level. Target varies by quarter-session when the trials are listed in standard order.

Error polarity swaps between Device A and Device D: for A, errors are always negative, and for D they are positive. This is because the plotted Y=index(*target device*) – index(*chosen device*), and thus a negative error for Device A means if a subject didn't pick A when A was targeted, he/she picked B. When Device D was targeted, the response was either 0 or positive. When stated that way, the reason for the pattern becomes obvious: because A and D are at the end of the real device set, they have only one neighbor each and thus only one type of error may be made.

Another such skewing of response as a function of Target occurs because subjects did not always perceive the devices used here as being distributed on a linear scale. Thus we expect to see a higher error rate for low impedance targets in cases where the subject-perceived distributions had close to the same shape as the measured distributions, as these should have been more difficult discriminations. For Masses, this effect is not obvious. When we look back at the subject-perception plots (Figure 9-4, however, we also recall that on average,subjects were much closer to a Weber Law perceptual scale than for the other device families, so we would expect to see more small-to-large skewing for, say, Sliders (measured and subject-perceived distributions almost identical) and to a lesser extent Toggles. A look at Figures 9-10 and 9-11 bear this statement out.

Observations for Specific Device Families:

Masses: The only obvious trend is a cluster of negative errors at Trials 4, 8 and 12; all subjects got this one wrong. The every-four-trial repetition in that region of (Mb) corresponds to a confluence of Target=A, Sampling rate=anything, Damping=High and VE expression=Impedance.

Figure 9-2 (the bar chart showing ANOVA significance of all the effects studied) shows no TDV interaction. However, there are strong TD, TV and DV interactions; other data features counteract this observed interaction to mitigate the 3-way effect while maintaining significance of the respective 2-way interactions.

Sliders: Sliders have negative-dominated errors with a trend of more negative error for lower impedance devices; that is, subjects most often chose sliders of higher impedance than was targeted.

Toggles: Trials 25, 27, 37 and 39 in (Tb) correspond to VE Expression=Admittance, Sampling rate=300 Hz and Target=C or D: these are the four conditions for which the emulation could not be stabilized, and no data was taken. The all-positive-error values shown are artificial errors introduced to indicate that these emulations were worst-possible (unusable), penalize them and let the analysis proceed (see Section 8.3.9 for a more thorough discussion/justification of this situation). In fact, the presence of positive "errors" for Device=D is a tipoff that this is not normal data.

Otherwise, the overwhelming features of the Toggles plots are the zero-error region in (Tb) corresponding to Device=D, and the series of odd-numbered trials (27–39) in which all or most subjects made either a positive or a negative error.

These observations turn out to be the result of a single very strong effect (VE Expression = Admittance in the case of high impedance emulations). The over-estimation error in the C region is insensitive to any other factor level: it is made regardless of sample rate (switches to a new level every four trials) and damping level (toggles every two trials); and it is true for every appearance of Target=C and VE expression=Admittance. Further, it is not clear from this data whether the same type of error (overestimation of target) would have been made for Device D, since there is no upper neighbor to choose from in this case. As it happens, verbal comments confirm that most subjects were not content with the choices they were given when D was targeted with an admittance-expressed model. The emulation felt much harder, they said, than any of the real device choices, but in that situation they were forced to pick the hardest real device available.

This highlights a problem with this approach mentioned earlier: the discrete set of target choices sometimes results in the paradigm getting fooled. The data shows a very low-error segment for V=Y and Target=D even though those conditions produced a terrible emulation.

The consistency of poor performance and/or poor emulations for Devices C and D, V=Y suggest that at least part of the problem is that these emulations are mis-targeted — all subjects agree that they feel harder than the real device under all other conditions. In fact, subject comments and investigator experience demonstrate that the problem is worsened with low sampling rate: the particular form of degradation in that case is for the emulation to feel much harder than it does at higher sampling rates. However, since the emulation appears already to have been over-targeted, this extra divergence did not show up in the data.

9.5 Analyses of Variance

The Analysis of Variance comprises a succinct and efficient summary of the influence of changing experimental factor levels on subjects' ability to make a "correct" discrimination; it is good at identifying which factors of those tested are important. In this section we look carefully at the results of the ANOVA, by device family and effect but for only a single response variable (the simplest one, $Y_{unweighted}$=[Target−Chosen]). The impact of using different response variables will be discussed later, in Section 9.7.

The Analysis of Variance: Tabular Results

Tables containing the ANOVAs themselves have been relegated to Appendix G due to their volume. Those from which results are shown for this section may be found in Tables G.1 (a), G.2 (a), and G.3 (a); results from other response variables fill out the remainder of those tables. These analyses are conducted for the full N=6 session data set.

Description of the ANOVA Used Here: In the following I assume that readers are generally familiar with the ANOVA table; if not, please refer to any basis statistics textbook (e.g. [141]). The ANOVAs presented here follow the standard format. Key features are summarized as follows (see Methods, Section 8.2.8 for more detail):

- Blocks correspond to sessions, of which there were 6 for each device family, and hence the 5 blocks degrees of freedom (DOFs). Interactions between blocks and the other experimental effects were not analyzed.

- Aside from blocks, the following sources of variation are analyzed:

Type of Effect	Number	Total DOFs
Main	4	7
2-way interactions	6	11
3-way interactions	4	23
4-way interactions	1	6
Error	1	235

- Degrees of freedom for each main effect range from 1 to 3, depending on the number of levels used for that factor. DOFs for interactions are the product of DOFs for the constituent main effects.

- Error degrees of freedom are related to the number of effect degrees of freedom, including those of interactions. It is typical of designs with large number of factor levels to have many error DOFs; and such is the case here, with 235 for the full N=6 analysis.

- Because of the large number of error DOFs and the use of full replicates (no effects were confounded with blocks), it was not necessary to pool high-order interaction DOFs with error in order to get a strong error estimate.

- A single number for each source of variation, that effect's significance level, is the output of the ANOVA table.

Further Analysis Possible: More extensive analysis is possible using the ANOVA than is conducted here. Using the sums of squares used in generating the ANOVA table, the data may be modeled in order to proceed from a purely descriptive analysis to a predictive one — the usefulness of this step in the context of this study should be obvious. In the presence of high-DOF effects such as Sampling Rate and Targets, the variance can be partitioned into quadratic and linear components for an even more subtle model. These procedures are further discussed in Methods, Section 8.2.8.

The "Star" Significance Plot

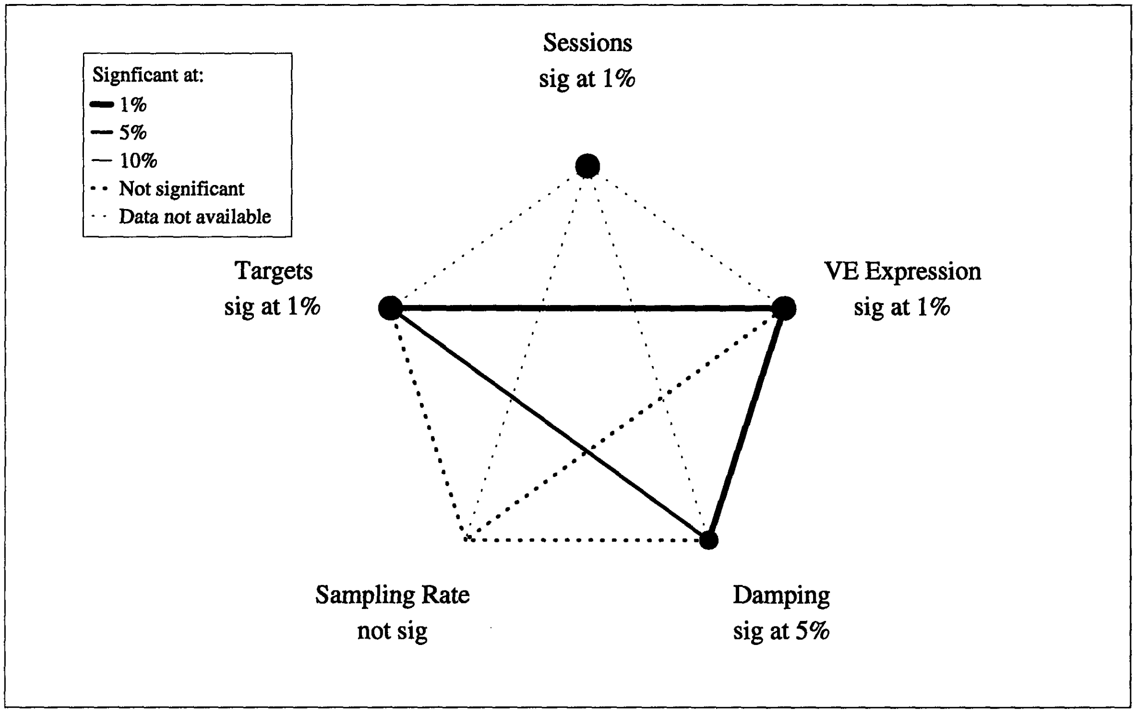


Figure 9-12: Effects significance for Masses: all subjects averaged, $Y_{unweighted}$.

Because of the large number of analyses that were conducted here, a graphic was developed for the purpose of efficiently displaying ANOVA results, focusing on main effects and interactions. This graphic is described in Figure 9-12. Its usefulness is in clustering all the main and 2-way interaction effects for one family in a compact and hopefully intuitive manner, by visually relating the interactions to the main effects. The value of a succinct format will become clear in later sections when as many as 18 ANOVAs need to be compared at once.

Vertices of the polygon represent main effects, including Blocks; lines connecting the vertices are 2-way interactions between the respective effects. All the interactions and vertices are explicitly labeled in this verbose version of the graphic, but will not be in future occurrences.

Interactions between Blocks (sessions) and other experimental factors are not customarily computed, and they were not computed here (a way to get at this information, if it were desired, would be to perform the ANOVA on sets of N=1, i.e. single sessions and look at significance in that way). Thus the interaction lines connecting Blocks and the other effects are ghosted, to signify that no significance data is available in those cases.

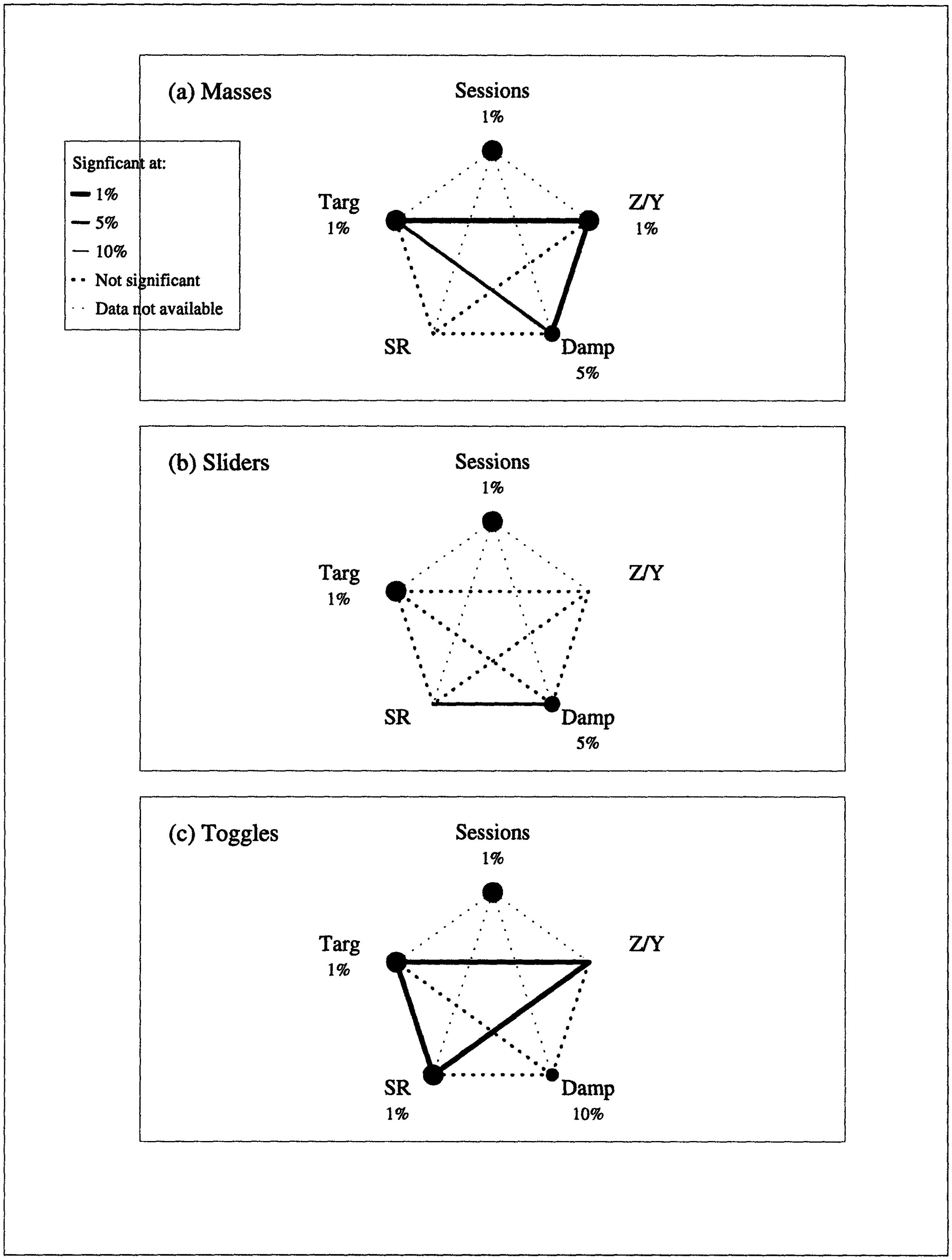
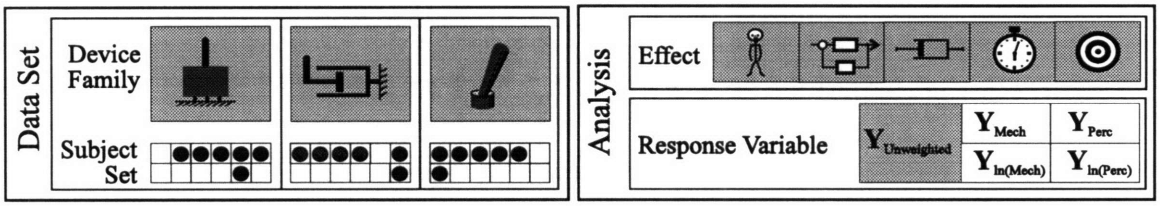


Figure 9-13: Significance of main effects and 2-way interactions by family: all subjects averaged, $Y_{unweighted}$.

Main Effects and Interactions Compared By Family



The star significance graphic is used in Figure 9-13 to illustrate the ANOVA results for each device family and the single response variable used here, $Y_{unweighted}$. It is apparent that each device family has its own "fingerprint", a pattern of influential effects which differs from family to family. Targets are always significant, and damping is always moderately significant, but each family is otherwise unique.

Most of these patterns have already been discussed and are presented again here for completeness and to give reader familiarity with this tool, which will be used gain later. However, note that in this format, a surprising anomaly such as is seen for Toggles, VE expression effect is instantly visible: a main effect is not significant when two of its interactions are.

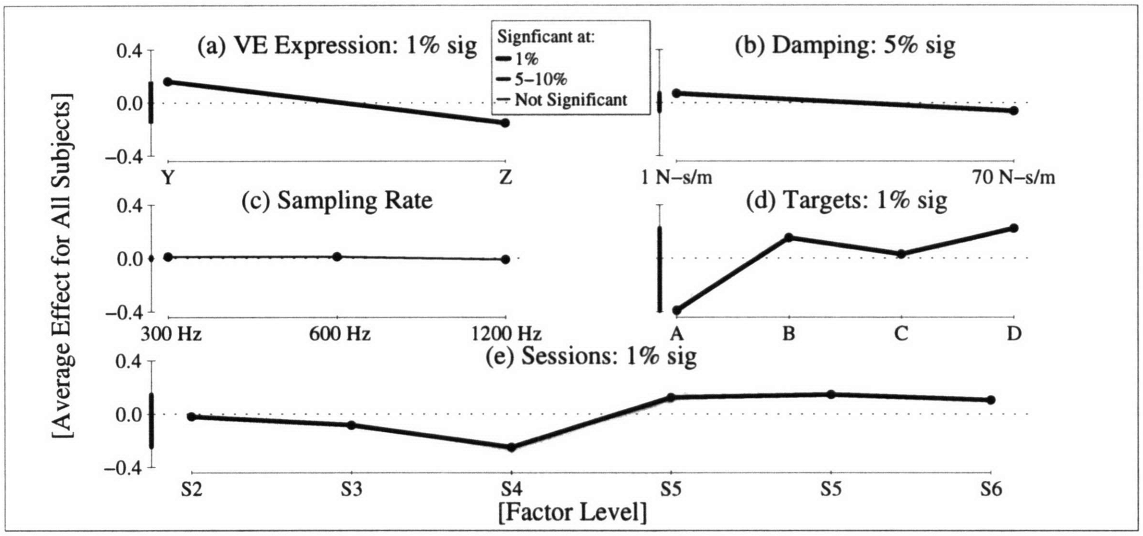


Figure 9-14: Main effects for Masses: $Y_{unweighted}$.

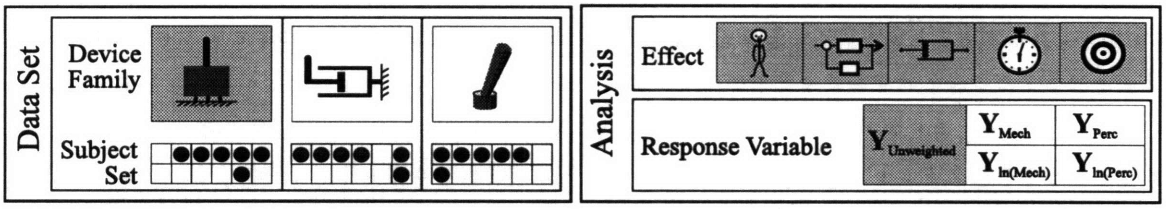
9.6 A Closer Look at the Effects

The goal of this section is to display what went into the Analysis of Variance shown in the last section; there is much more information than the single number describing significance of an effect, but effort is required to extract it. Each ANOVA table referred to in the preceding section is a distillation of 288 responses taken over the six sessions for a single device family, into a single number for each effect and interaction; 15 in total for a data reduction ratio of about 20:1.

Lost in this reduction is information about polarity of the effect — which direction, if any, is better; and it doesn't give us much feel for the morphology of the data, which is important when trying to map out a territory as we are doing here.

A single response variable ($Y_{unweighted}$) will be used throughout this section, to keep it uncluttered and because response variables are the focus of a later section. The difference in effects by response variables are not large enough to make this a problem.

Description of a Main Effects Plot



I will start the examination of the experiment effects by explaining how an "effects plot" works. Figure 9-14 shows the five effects, including Blocks (Sessions) on the bottom, while device family and response variable are held constant — here, Masses and Y=$Y_{unweighted}$.

Plot (b) shows the average effect on the response variable for the two levels of the quantitative Physical Damping factor. The point over the "0.001 N-s/mm" level of damping equals the average response for

$$Y_{unweighted} = \text{index}(\textit{target device}) - \text{index}(\textit{chosen device})$$

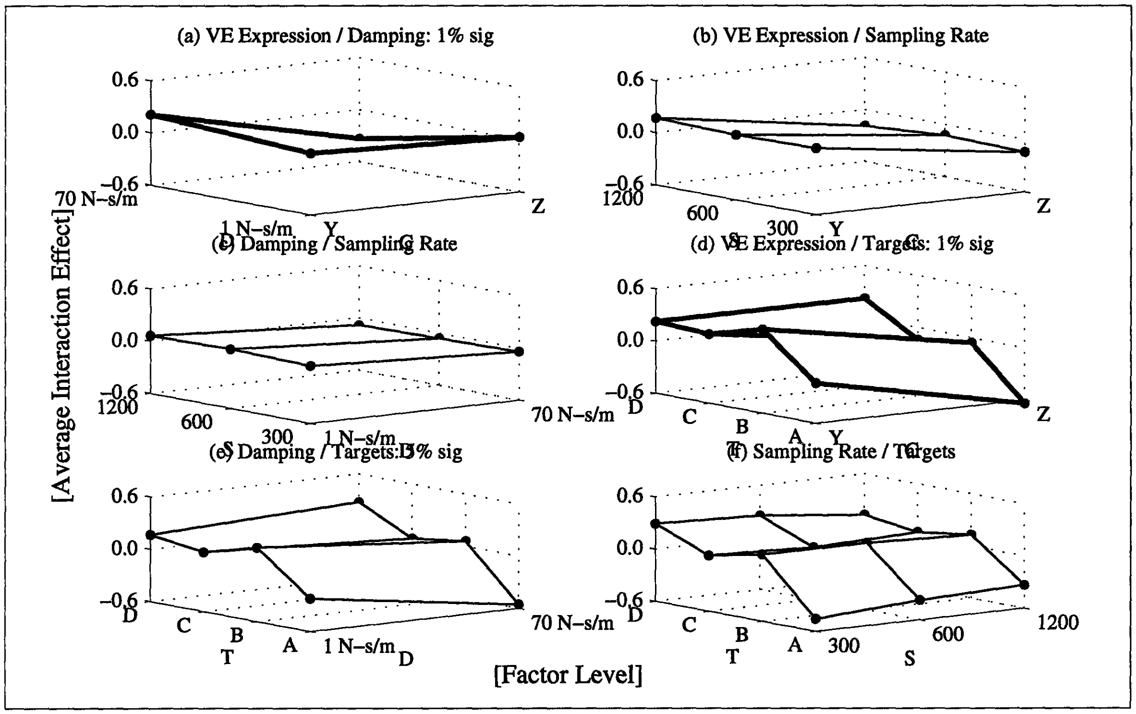


Figure 9-15: First-order interactions for Masses: $Y_{unweighted}$.

of the 288/2 (all data for the six Mass sessions, divided by the number of levels for this factor) subject responses in which Damping was set at the 0.001 N-s/mm level. Likewise, the point over the "0.070" level represents the average response for the other 144 data points. The heavy line segment superposed on the Y axis is simply a range axis representing the Y-axis divergence of the points plotted on those axes. Thus, we see that when damping was 0.001 N-s/mm, subjects in the Mass sessions on average produced a small positive error, i.e. chose a real device of lower impedance than the targeted one. This polarity of response was reversed (on average) when damping was set to 0.007 N-s/mm.

The significance of the effect, as determined by the analysis of variance for this data set, is indicated by the boldness of the line, and reiterated in the caption of the effect: Damping is significant at 5%. VE expression, on the other hand, produces a larger variation in response and this is reflected in the significance (1%). The sampling rate effect plot (c), virtually flat relative to the others, is not significant at all.

The Targets and Blocks effects plot demonstrate a situation possible for higher DOF factors. Targets are significant at 1%; however, most of the variation is due to a single real device — there is a disproportionate rate of negative errors when Device A was targeted. This is because, as we saw in Section 9.3, subjects perceived Device A as somewhat closer to B than they did D to C, and consequently confused the lower pair with greater frequency.

What we can't tell from this plot is whether similar rates of confusion occur for B and C; if there was a high rate of both positive and negative errors (indicating that the subject couldn't reliably use the emulation to distinguish those devices) it would be averaged and canceled here. That is, this data is still too highly processed to tell the whole story. To overcome that problem, we will look at more detailed data later.

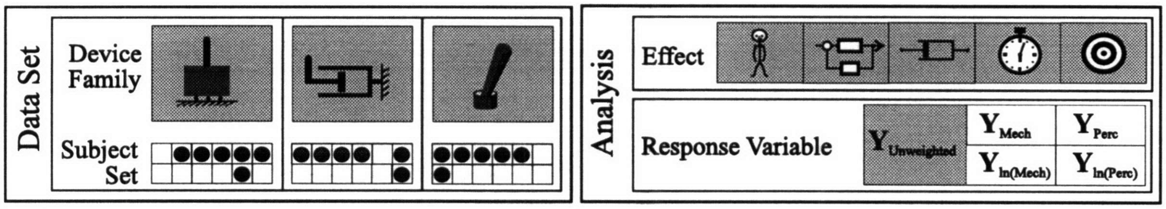
Description of a 2-Way Interactions Plot

The same approach is taken to plot 2-way interactions, shown for Masses and response variable Y=$Y_{unweighted}$ in Figure 9-15. Since two variables are being varied simultaneously, the interactions must be represented as a surface, the number of facets of which increase multiplicatively with the number of factor levels involved on the two horizontal axes.

Factor levels are specified in shorthand: "Lo" is whatever appeared at the left in the previous Main Effects display (Figure 9-14), and "Hi" is what appeared at the right. In the case of the qualitative factor VE expression, this should be interpreted as Lo=Admittance and Hi=Impedance.

Only three 2-way interactions are significant: VE expression/Damping, VE expression/Targets and Damping/Targets — the threesome which we noted earlier in Section 9.1. As was true of the main effects plots, the bold lines indicating ANOVA significance correspond to the surfaces which stray furthest from the zero-plane, demonstrating positive or negative average identification errors. However, significance of an interaction additionally requires the surface to be skewed, changing as both factor levels change. This is the case for (a) and (d) but not for (f), which does diverge substantially but varies only with Targets (T) and not with sample rate (S).

Average Effects by Family



Moving up in complexity, we will now take a look at Figure 9-16. Here, Main Effects for all families are shown in a matrix array to facilitate cross-family comparisons by effect. The Mass effects which we just examined are repeated here for the purpose of comparisons.

Blocks

Blocks (sessions) dominate the page along with Targets; the degree of variation is similar from family to family. All families have some sessions that were quite low in average error (although Toggles did not have many), and some that were very high. The record is held by session 6 of Sliders.

VE Expression

Masses have an average VE expression effect footprint substantially different from Sliders and Toggles; it is the only family for which the polarity of average error for admittance ("Lo") and impedance expressions obviously changes.

We will see in the next section, when the average effects are broken down into sessions and examined by subject, that the VE expression effect for Sliders would be approximately 0–0 (small average error for both factor levels) if it weren't for a single session (6), and in that case the apparent substantial offset may be anomalous.

However, there isn't that sort of explanation for the bias in Toggles where there really is a fairly strong negative bias. This may be due to targeting problems in the emulation; at least one of the emulations did not correspond well to the actual target, and consistently resulted in subjects picking a higher real device than was targeted. We can also eliminate Device D from the list of devices which would produce this result, since only positive errors could be made when it is targeted — this

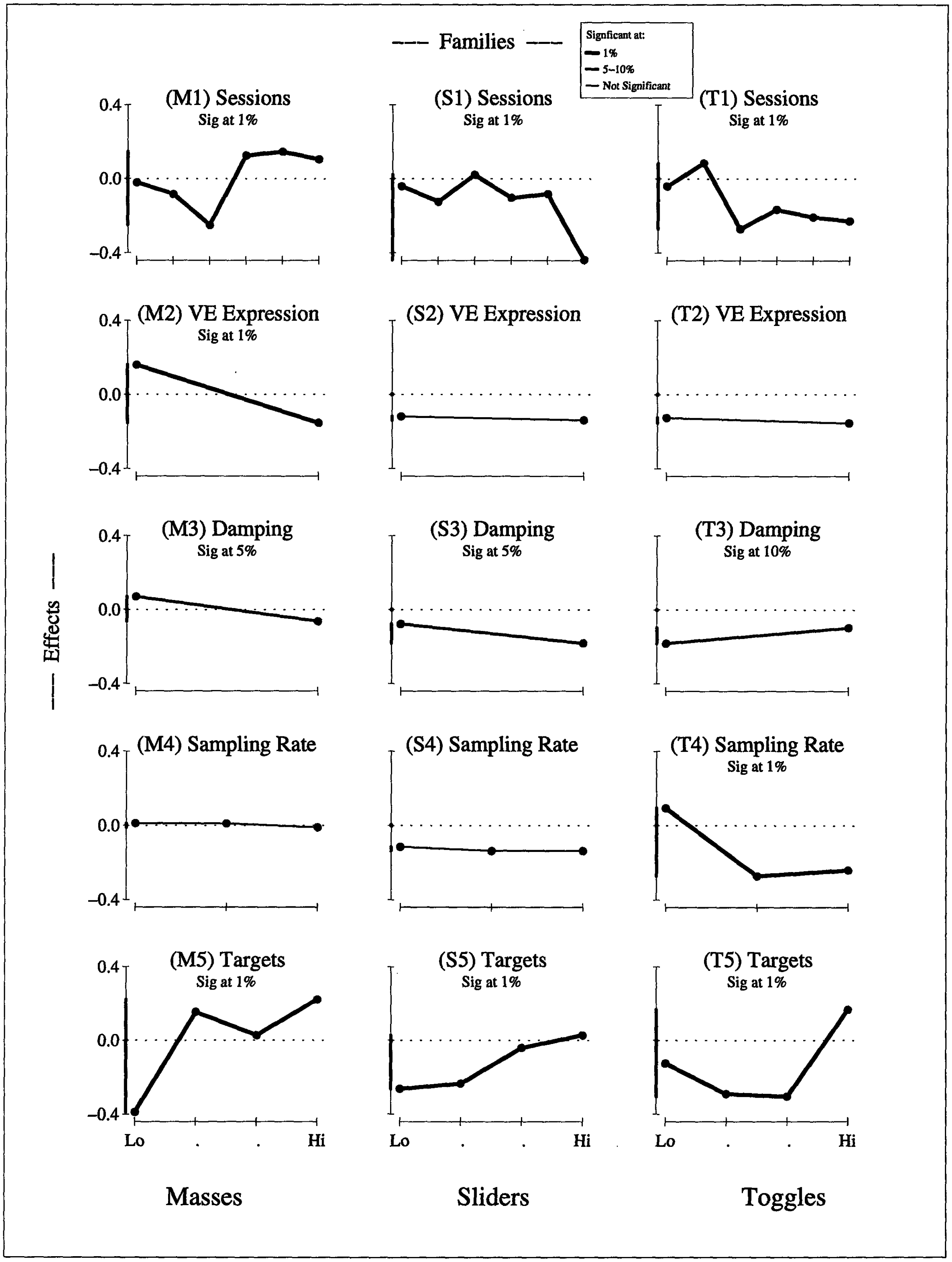


Figure 9-16: Main effects for all families: all subjects averaged, $Y_{unweighted}$.

does not mean that D wasn't mis-targeted, just that such mis-targeting could not have resulted in negative average errors here.

Damping

Damping did not play a major role in the rate of errors for any of the device families; although it was mildly significant at 5% for Masses and Toggles. It appears that despite the lower quality and active nature of the emulations at higher damping rates, subjects were generally able to make the intended identification.

Sampling Rate

Sampling rate is not significant for Masses and Sliders. This was an unexpected result because the emulations decreased in quality at the lowest sampling rate, 300Hz. Again, subjects proved able to make discriminations even while commenting on the relatively poor feel of the emulation.

For Toggles, average error increases at higher sampling rates, making this a 1% significant effect. The larger positive error corresponds to the four unstabilizable 300 Hz emulations, where the error was arbitrarily chosen positive (Section 8.3.9); thus the change in polarity and magnitude between 300 and 600 Hz is also arbitrary.

Targets

Targets are consistent in high significance: error rate depended strongly on which device was targeted. The effect is always negative for Device A targeted and positive for D because only one type of error could be made in these cases.

For Toggles, the four unstabilizable trials (associated with Devices C and D) again influence the results. The two artificial points associated with Device D in every session account wholly for the fact that the average error in Target effect is not zero for Device D.

It appears from looking at the Target effect for B and C that subjects generally made negative errors, picking a device of higher impedance than the one targeted. Then they got to D, and now we know that the error there was zero except for the unstabilizable trials. That is, they couldn't make a negative error in that case because they'd run out of real devices; and thus the D emulation seems to have been produced with too high an impedance as well as the B and C Toggle emulations.

Effects by Subjects

"Composition" Plots

In order to properly interpret the Effects plots shown and discussed in the preceding section, we need to take a look at what went into the averages. For this purpose, I'll use another type of plot which can be seen in Figure 9-17. The composition plot shows the same dataset three times at different levels of averaging. This makes it easy to quickly eyeball the composition of the effect, tracing the source of the variability.

Each quadrant of Figure 9-17 represents one of the main effects; this means that the 288-element data set is shown four times on the page, but partitioned in a different way to illustrate the average effect of varying each factor. The top row, with a single set of axes, is the average of the entire three-family, eighteen-session, 864 response value data set divided into, for (a), two halves (432 responses for each level of the VE Expression factor).

The middle row shows the next level of detail; it subdivides the all-family data of row 1 into data for individual families. Thus, each curve in row 1 represents six sessions or 288 responses; for this

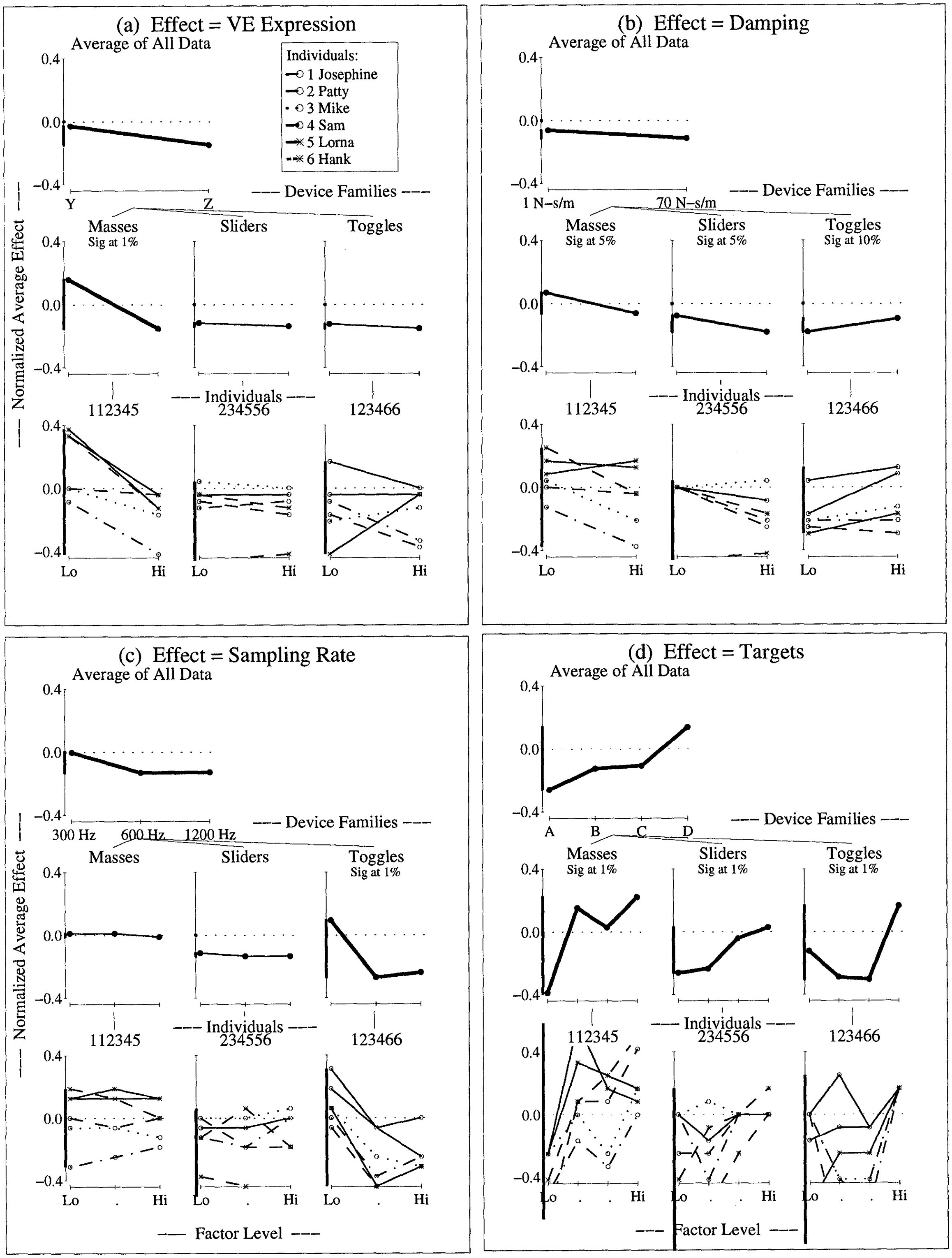


Figure 9-17: Composition of effects: [Subjects : Families : All Data] ($Y_{unweighted}$).

effect, again divided into two for the two levels of the VE Expression factor. Note that these curves are the same ones that appeared in Figure 9-16.

Finally, the bottom row in each quadrant breaks the family curve down into individual sessions; thus there are six curves in each plot in row 3. The numerical subject identifiers for the subjects who performed sessions for that device family are listed above each plot; thus it can be seen that Subject 1 (Josephine, from the legend above), didn't do a Mass session but Subject 5 (Lorna) did two of them. Don't worry about picking to pick out and compare the repeat subject performances here; we'll be looking at that explicitly later on.

Effects for Subjects, Arranged by Family

Now let's look at Figure 9-17 again with the intent of seeing what new insights it yields. Key features to note are:

- Single outlier curves which distort the behavior of the average so it doesn't reflect the average of the others.
- General "noise" — how much variation/consistency among sessions is there?;
- Do subjects generally agree in their relative changes in over-/under-estimation (i.e., do the session curves have similar slopes, or of at least the same sign); or are the differences simply in rate of error (are some subjects always close to zero error rates, while others have both large positive and negative error rates for certain factor level settings)?

Outlier Subject and Handling of Data

When the effects are plotted at the session level (bottom row of each quadrant), we finally see the identity of the single outlier session which has been flavoring Slider data since the beginning of this chapter. So far out of range that it is almost off the bottom of the page, it is Hank, Subject 6. Hank was the repeat subject for Sliders, and his other session (look for another dash-star line on the same axes) is perfectly typical of everyone else's. (You can't tell from this plot, but the outlier is the second of Hank's two sessions).

It would thus appear that either Hank was having a bad day for his second session or the investigator was. Examination of patterns in the error make it appear likely that Hank was not concentrating as well as he did the first time, and consequently made a lot of mistakes on the discriminations which required concentration.

(a) VE Expression

The VE Expression (Plot 9-17(a)) shows good consistency in direction. Most subjects underestimate the target (choose a device of lower impedance than the target, resulting in positive $Y_{unweighted}$=[Target−Chosen]) for an admittance-expressed environment, and achieve a much smaller error rate, close to zero but negative if anything, for the impedance-expressed environment. One curve (4, or Sam) shows some individuality, with low error for admittance and large negative error for impedance. This one curve is informative, because it (a) is largely responsible for the quite large negative average error for impedance seen in the middle-row effects plot; and (b) suggests a new interpretation of all the other curves in this plot. Without Sam's data, a reasonable interpretation of the remaining data would be that admittance was simply a bad situation and impedance was a much better one. With Sam's repetition of the slope of the other curves, but addition of a substantial negative offset, we must ask whether the errors are all relative: the admittance emulation always feels lighter than the impedance one, and one subject just happened to have a different perceptual offset in the difference between his perception of real devices and this type of emulation. Sam thought both emulations (admittance- and impedance-expressed) felt heavier than the real devices, but impedance was always heavier than admittance.

In the case of Masses, where the model was simple and easy to parameterize, this is an interesting observation. Recall that in the Mass modeling process, I found that subjects in the pilot experiments generally thought the emulations felt about 70 grams heavier than the real devices, a perception that I shared. Therefore, in emulating a real mass of 500 grams, I used an emulation mass of 430 grams for both the impedance and admittance expressions. It now appears that this was about right in most subjects' opinion for the impedance-expressed emulation, but it made the admittance emulation too light; whereas in the pilot experiments, subjects seemed to prefer the same offset. Whatever else this means, it reopens the question of whether sensor calibrations were at the bottom of the 'perceptual offset' problem since the two expressions, which employed different sensor sets, gave different results.

Compare the Toggle third-row plot with that of Masses. While both exhibit about the same amplitude of variation, it is qualitatively different: the relative changes in error rate between admittance and impedance is different in magnitude and direction for nearly every session. That is, while Masses show some kind of trend, Toggles are well distributed. Negative error rates (subject chooses a device of higher impedance than the target) are more prevalent, but as the average (not significant) above shows, there is no clear improvement in performance for either environment expression. The most that can be said for Toggles is that subjects had a hard time with them, but on average tended to think the emulations were stiffer than the target.

(b) Damping

For Damping, all three families are significant at only 5 or 10%.

Sliders are the most interesting case here; with the exception of Hank's outlier session, there is an extremely low error rate for no extra damping, and variable but sometimes quite high error for the higher damping level. This result is consistent with my experience in developing emulations, but since the Sliders were the most stable and high-quality of all the emulations, I would not have expected them to be where higher-damping-is-worse result came out strong.

(c) Sampling Rate

The overall picture in Plot 9-17(c) is not as extreme as might have been expected, particularly at the low end of the Sample Rate scale (300 Hz).

Masses seem remarkably insensitive to sampling rate on the average (middle row) and the real story is seen in the subjects plot in the bottom row. There are quite high error rates for the different subjects, but they are distributed almost symmetrically around zero and thus cancel each other out on average, resulting in the Not-Significant ANOVA result. This is a case where if what you cared about was the average effect, the ANOVA would be right. But if what you care about is knowing what any given individual is going to do, and if they are different, how; then the ANOVA and even the average effects plot before it is broken down to subjects is quite misleading.

Sliders are uneventful, showing no particular trend for Sampling Rate decreases. Toggles, however, present a picture which is at first surprising. Performance at the higher sampling rates (600 and 1200 Hz) is fairly nonuniform, although characterized by negative error on the average. The nature of the degradation at 300 Hz, however, is a consistent and large jump to positive error. Before trying to draw a conclusion from this, recall that the four Toggle data points per session which had to be artificially created were for the lowest sampling rate and Targets one of C or D; and that the data introduced was arbitrarily a positive error value. We'll look for the Target anomaly in a moment, but can see here that this artificial data could easily explain the change in polarity. The same data introduced as negative errors would have produced a jump in the opposite direction, resulting in a steady average decrease in performance as sample rate dropped.

(d) Targets

All families but particularly Masses show a great deal of variability. Some of the subjects have extremely large error rates, both positive and negative, for the lower impedance devices; and the fact that the variability is off the axes scale here and now where else is an indication that indeed, the targets were responsible for the lion's share of the variability in the experiment.

Sliders are dominated by Hank's second session, which we are finally able to identify as being fairly normal for Devices C and D, but off the scale for B and particularly A (follow the off-axis range axis down to see the extent of the maximum error).

The anomalous up-turn in the curve for Toggles, Device D, is all the more striking because every single subject "responds" with the same error rate for that device. In fact, we know that the same artificial data points were given each session for that point, and otherwise every subject got every trial for Device D correct — because the emulations felt heavier than the hardest real device and therefore the choice, although not satisfying, was obvious.

Effects for Subjects, Arranged by Subject

We've seen what goes into the data for a single family effect, but it's equally informative to see what subjects are doing across the different families. We'll be examining this in more detail later on, but let's look quickly at it here: Figure 9-18 is the same format as the previous Composition figure, but instead of breaking the "All Data" curve down by families, it does it by subjects. Thus, the same 18 curves appear in the bottom row because this is the same data set, but they are sorted into a different set of combinations. Note that not all subjects performed the same families — repeat subjects did two of one type and skipped another. Also note that there is no significance information for the subject sums (middle row of this figure).

The key features worth noting in these plots are consistencies and variations of subjects across families. For example, in plot (a), VE Expression, look at Lorna, who is the repeat subject for Masses. Lorna is highly consistent for her two Mass sessions, but has a different response for her one Toggle session. In both families, Lorna is perhaps the most extreme of all the subjects in her responses, with low error for the impedance expression and quite high positive or negative average errors for the admittance expression. On the other hand, Sam one session each on Masses and Toggles, and the sessions for the different families look very much alike. Indeed, if we quickly scan over Sam's reaction to the other effects, we see that he is often a little more consistent than the other subjects across families (look at the length of his range-axis bar relative to the others).

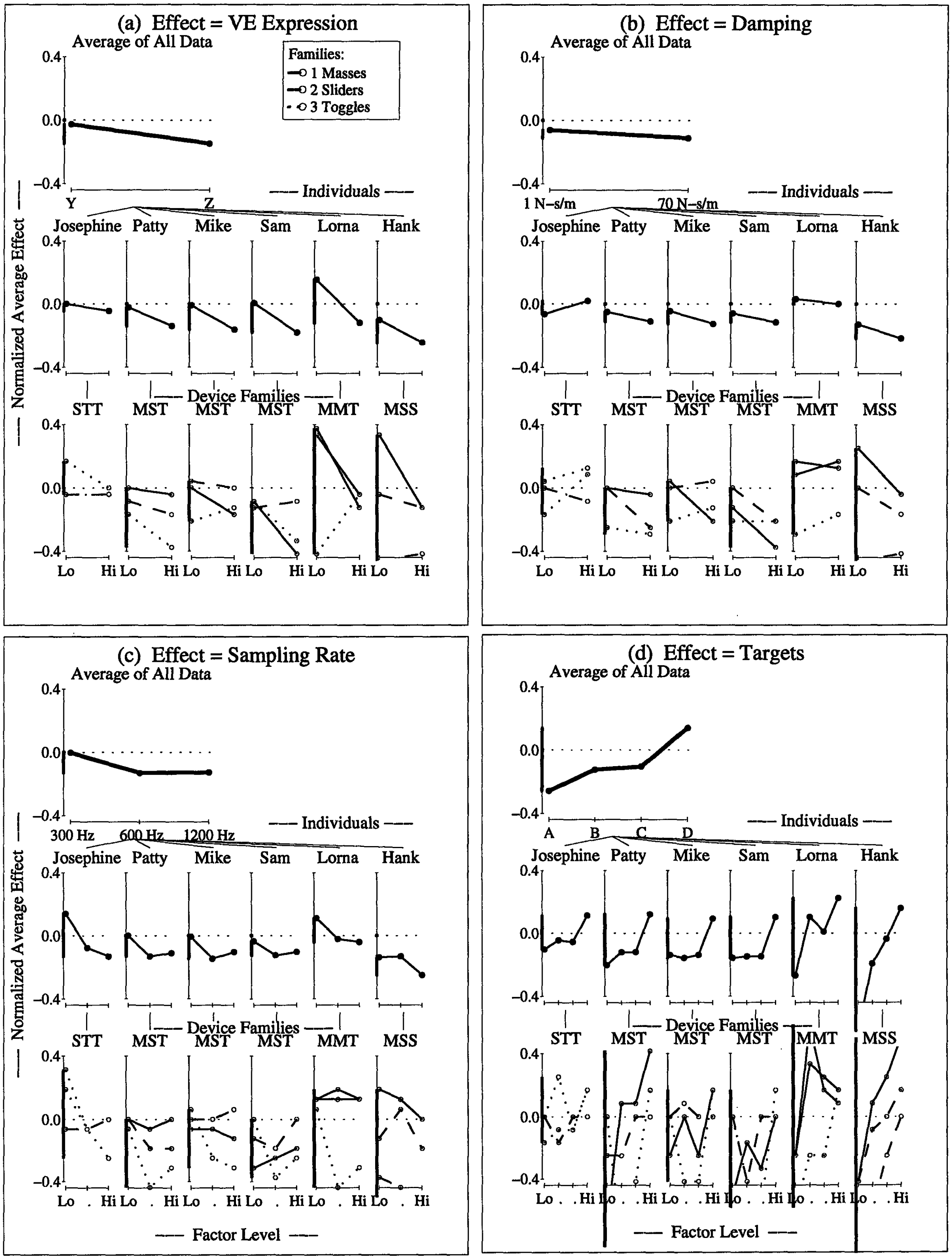
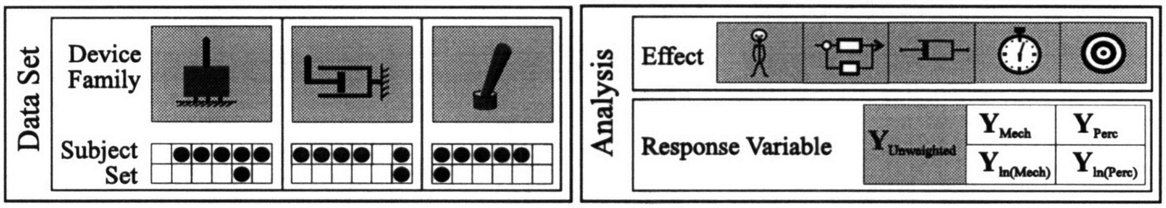


Figure 9-18: Composition of effects: [Families : Subjects : All Data] ($Y_{unweighted}$).

9.7 Influence of Response Variables



A computation of response based on the real device's impedance index (1 for Z_A, 2 for Z_B, etc) institutes a constant unit spacing between each adjacent pair of devices. That is, a subject's confusion of two devices which are closely spaced, either on the subject's perceptual scale or a measured one, is given the same weight as a confusion between two devices which are in reality more widely spaced. This is the case with the response variable which has been used up to this point, $Y_{unweighted}$.

In Section 9.3 we saw that peoples' perception of the impedance distribution of the real device sets was in general not identical to the measured distribution. In Section 9.5 and 9.6 we used a simple impedance-index-based response variable to analyze subjects' ability to identify the real device which was targeted by a given emulation. This section explores the result of combining the information we've collected about subject-perceived and measured impedance distributions with the comparison trial responses, by weighting the errors according to the two distributions (see definition of response variables in Section 8.2.7 and Table 9.1). Specifically, in this section we will (9.7) compare ANOVA significance results across response variables; (9.7) interpret the influence of error weighting on analyzed effects; and (9.7) study detail in subject responses.

Subject Perceptual Scales Used

When the subject perceptual scale is used in weighting the response errors, it is on an individual basis. That is, each session's comparison trial responses are weighted by the perceptual impedance distribution obtained from the calibration trials for the same session. Average perceptual scales are never used in these weightings.

Using Relative-Impedance Weighting in Response Variables

No analysis weighting scheme will alter the fundamental response patterns which are a function themselves of the target set impedance spacing, both relative and absolute gap sizes. Because of the discrete nature of the choices (multiple choice) given the subject each trial, the spacing influences the frequency of each type of error (Table 8.2); weighting the errors by relative spacing only modifies the response variable once the errors have already been made. Thus, weighting is of most value and is likely to make the most difference when the impedance spacing was wisely chosen to begin with, to prompt an informative pattern of errors and correct responses.

As we saw in the analysis of the calibration trials (Section 9.3), it is not clear that the real device impedance spacing used here was always ideal in terms of extracting a maximum of useful information. I expected subject perception to follow Weber's Law more closely than it sometimes did, and spaced the real devices non-uniformly. When Weber's Law was not followed, the results was subjects perceiving the higher impedance devices as more widely spaced than the others, and consequently making fewer errors. Since these identifications turned out to be obvious, that data was not as informative as that of the lower impedance devices which were harder discriminations. This is not a flaw in the design, but a procedural detail which would have been difficult to predict ahead of time; future iterations will benefit from this knowledge.

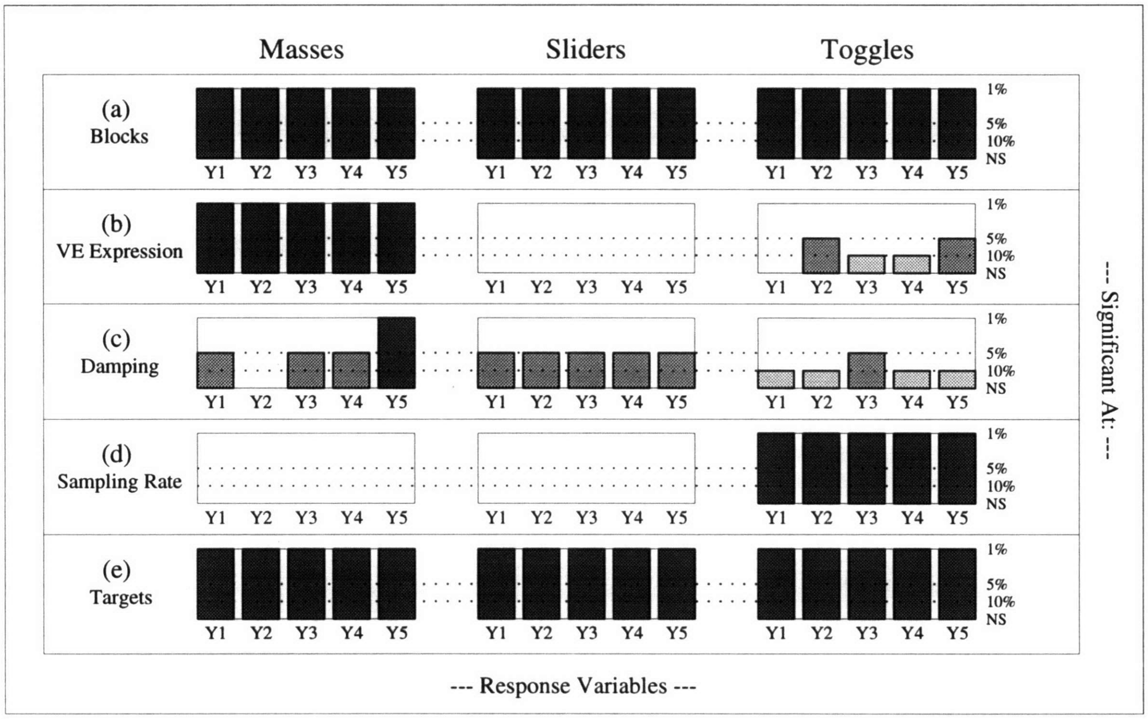
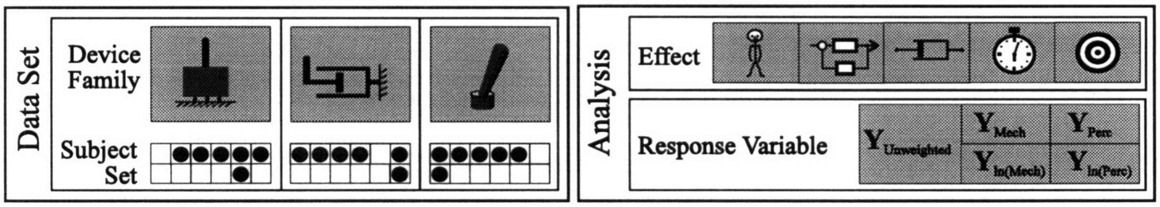


Figure 9-19: Effects significance by effect, all response variables: all subjects averaged. Bar shading and bar length are redundant cues as to effect significance for that response variable; darker and longer mean more significant.

ANOVA as a Function of Response Variable



In this section we will summarize the significance of all effects, by family and response variable. Refer again to Appendix G, Tables G.1 (Masses), G.2 (Sliders), and G.3 (Toggles) for the actual analysis.

ANOVA = *func*(Response Variable): by Effect

The bar plot of Figure 9-19 summarizes the significance results of the ANOVA for all main effects, response variables and families. The effects for which response variable makes any difference are clearly Damping (Masses), VE Expression (Toggles) and Damping (Toggles).

The most marked change as a function of response variable occurs for Damping in Masses, where analysis based on a response normalized by measured impedance indicates that Damping (Y2=Y_{mech}) is not significant at all. Conversely, using the log transformation of the perceived-scale weighting (Y5=$Y_{\ln(perc)}$) results in a Damping effect significant at 1%.

Damping under Toggles also exhibits more variation as a function of response variable than do most of the effects, but not as much as for Damping under Masses. Here, only one response variable

($Y3=Y_{perc}$) stands out at all; and it is not the same variable as for Damping under Masses. The fact that Y_{perc} is most significant for one family whereas $Y_{\ln(perc)}$ is most significant for another family suggests that we should look at the shapes of the subject perceived impedance distributions relative to the measured distributions for those families; it appears that the log transformation in response variable did something important in one case (Masses), but not in the other (Toggles).

Finally, the significance of VE Expression under Toggles appears to be noticeably influenced by response variable; but the polarity is reversed from what we saw with damping (compare rows (b) and (c) under Toggles for Y-bar height). Whereas the analysis seemed most sensitive to Y_{perc} and $Y_{\ln(perc)}$ for Damping, Y_{mech} (as well as $Y_{\ln(perc)}$) bring out the strongest results for VE Expression. Most importantly, significance based on all response variables is stronger than that for $Y_{unweighted}$, which has been used in all the preceding discussions.

ANOVA = *func*(Response Variable): by Family and Response Variable

Figure 9-20 uses the previously introduced "star" graphic to illustrate the tabular ANOVA significance findings for all main and 2-way effects, all families and all response variables, averaging the influence of each subject. The primary value of this mode of presentation is that you can quickly pick up changes in the effect and interactions pattern by scanning differences in the location and boldness of lines and vertices as you look down rows (response variables) and across columns (families). In addition, interactions are physically related to the main effects of which they are composed.

Thus, as we look down the Toggles column, we see that there is a characteristic triangular interaction pattern (TV, VD, and TD, indicating significance for all interactions involving Targets, VE Expression and Damping) for all the response variables. This is a different pattern than for Sliders, where only one interaction is ever significant at all (SD) and there is no variation according to response variable.

Toggles has a different fingerprint based on 3 2-way interactions between Targets, VE Expression and Sampling Rate, as opposed to Damping. Damping is, however, always significant at at least 10% for Toggles, even though Damping isn't involved in any interactions.

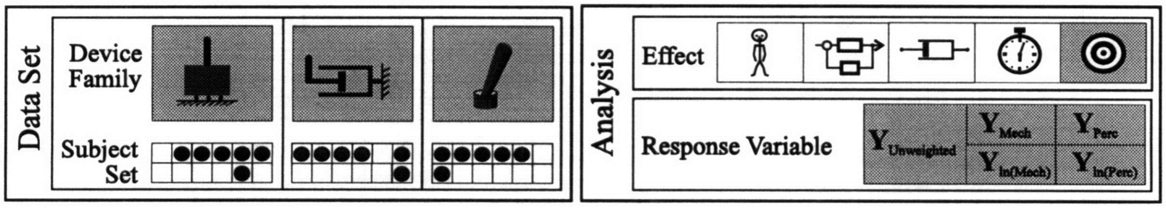
The interactions which vary as a function of response variable are:

Masses: there is some variation in the TV and TD interactions, the former running the whole range from not-significant to 1% significant; the latter only varying between 5% and 10%.

Sliders: interactions are all the same, and generally not significant. The lack of variation is not surprising, since the calibration experiments found that $Z_{mech} \simeq Z_{perc}$ for Sliders.

Toggles: interactions are all the same, and when they are significant, they are strongly so (1%).

Target Effect as a Function of Response Variable



Once again, we need to go to a visualization of the effects themselves in order to properly interpret the ANOVA significance results. Figure 9-21 was constructed to facilitate comparisons among selections of response variables and their impact on a single effect (Targets). The combinations shown in successive rows (columns are device family, as usual) are listed in Table 9.3 and their line-type

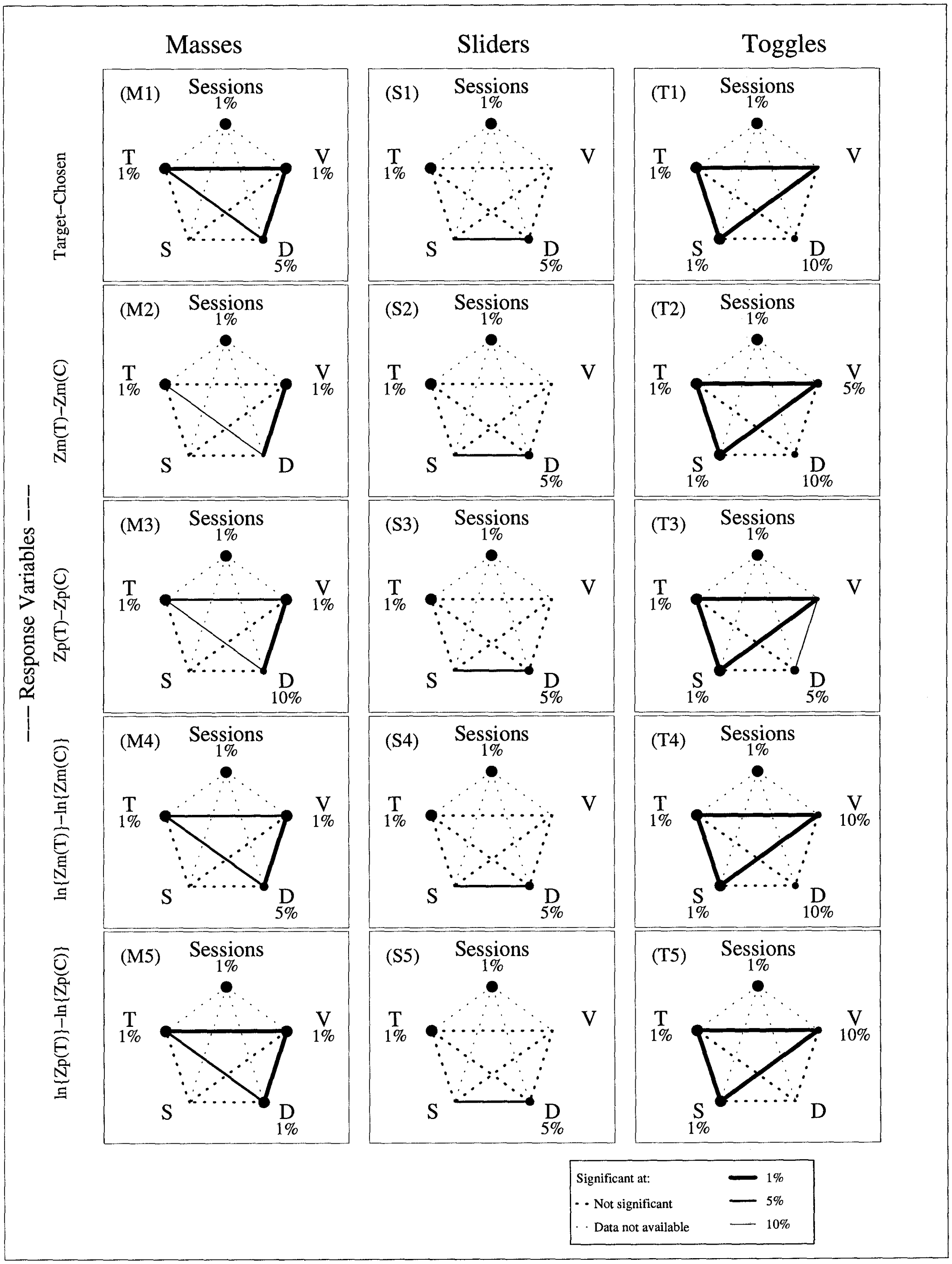


Figure 9-20: Effects significance by family and response variable: all subjects averaged.

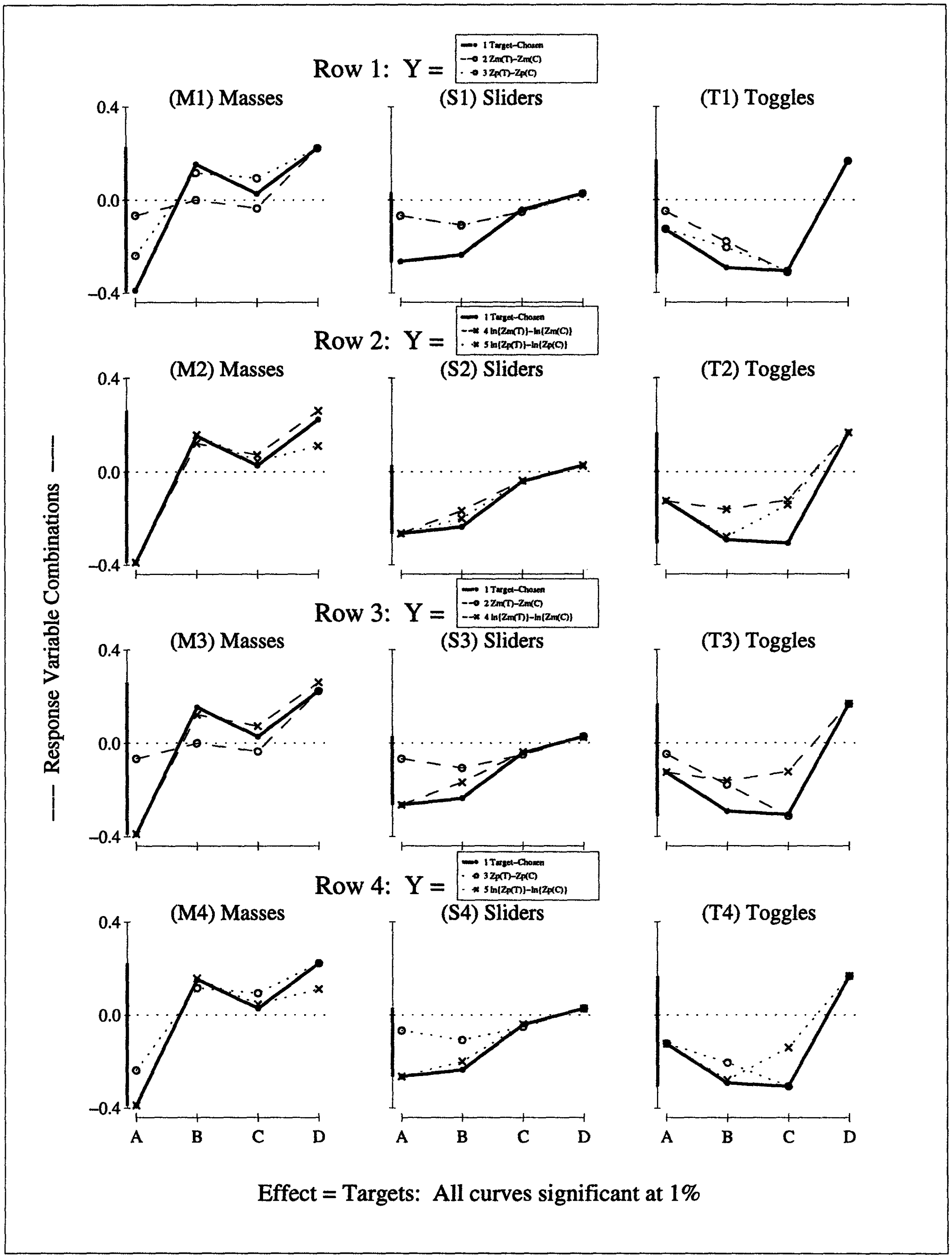


Figure 9-21: Target effect as a function of response variable and family.

Row	Curves Shown			Features Common to This Row
1	1 $Y_{unweighted}$	2 Y_{mech}	3 Y_{perc}	Weighted only
2	1 $Y_{unweighted}$	4 $Y_{\ln(mech)}$	5 $Y_{\ln(perc)}$	Log-transformed
3	1 $Y_{unweighted}$	2 Y_{mech}	4 $Y_{\ln(mech)}$	Weighted by measured impedance scale
4	1 $Y_{unweighted}$	3 Y_{perc}	5 $Y_{\ln(perc)}$	Weighted by perceived impedance scale

Table 9.3: Combinations of response variables used in Figures 9-21–9-25.

identifications listed in the legends posted for each row. $Y_{unweighted}$ (solid bold line) was used in every column in order to give a uniform reference of comparison, and because one of the issues of interest is to see how the more complex transformations appear relative to this simple initial response computation. The measured scale based response variables, Y_{mech} and $Y_{\ln(mech)}$, always appear as dashed lines; the perceptual-scale based variables, Y_{perc} and $Y_{\ln(perc)}$, as dotted lines. The weighted response variables, Y_{mech} and Y_{perc}, have open-circle markers while the weighted and log-transformed variables, $Y_{\ln(mech)}$ and $Y_{\ln(perc)}$, use '×' markers.

Analysis Limited to Targets Effect :

Because the differences among the response variables used here is due to variation in the real-device impedance distributions used to estimate the error implied by a subject's choice of target device, the influence of the various response variables on the analysis is best seen by study of the Targets Effect as a function of the response variable used to compute it.

Convergence of Curves for High Impedance Devices :

The curves based on different response variables tend to converge for Devices D and to a lesser extent C, an artifact of the normalization process. The important feature to note is the relative gaps between the curves which open up at the lower impedance levels.

(Row 1) Weighted Response Variables

In general, we see that $Y_{unweighted} > Y_{perc} > Y_{mech}$. This is because

1. $Y_{unweighted}$ weights errors for low-impedance devices disproportionately, therefore it generally has largest average error for Devices A and B. Note that the greatest discrepancy comes for Sliders (plot S1), and in this case we saw that both the measured and perceptual impedance distributions were the furthest from linear for any family; the linear skew is thus most extreme here.

2. The relative magnitudes of the Y_{mech} and Y_{perc} curves are understood by looking at the respective calibration (impedance distribution) curves in Figure 9-4: for Sliders (S1), these curves are nearly overlaid, and so are the effects based on their corresponding response variables. For Masses (M1), Y_{perc} is closer to the linear-distribution-based variable, $Y_{unweighted}$, than is Y_{mech}. For Toggles (T1), the case is more complicated — the Y_{mech} and Y_{perc} curves are close together and actually cross at Device B. We will see in the next section, when we look at subject effects, that this is due to the fact that there was a wide range in the subject perceptual distributions for Toggles, with the variation putting some subject curves as more dished (logarithmically spaced) than the measured curve. The average effects plot is the sum of these individually normalized components, and reflects its ambiguity.

(Row 2) Weighted and Transformed Response Variables

Row 2 compares the second set of response variables, those which have undergone a log transformation. Making downward comparisons between Row 1 and Row 2, we immediately see that there is generally much closer correspondence between the weighted and log-transformed variables and $Y_{unweighted}$ than there was between the merely weighted variables. This is because the log transformation effectively takes the distribution back to a linear case, by amplifying the low-impedance-device errors which had been minimized with the weighting.

The transformation has a different effect for Toggles; the transformed average effects are flattened and smaller than $Y_{unweighted}$, although they are both tied to $Y_{unweighted}$ for Device A. Again, relative differences are most important: the normalization of the log-transformed variables changed the shape of the Target effect curves, such that Device A got the biggest, or close to it, average error; whereas for $Y_{unweighted}$, Device C did. Without the normalization, the log-transformed curves would have been larger average error magnitude than $Y_{unweighted}$ for all real device targets.

(Row 3) Measured Scale Based Response Variables

Here we see the relative effect of the weighting-only and weighting plus log transformation, for the measured based response variables. Interestingly, for Masses the curve which $Y_{\ln(mech)}$ most resembles is neither of the other curves in M3 but the Y_{perc} curve in plot M3: log-transforming the nearly log-spaced Y_{mech} distribution resulted in something similar to the in-between-log-and-linear distribution used by Y_{perc}, by placing more emphasis on confusions between Devices A and B.

Relative to Y_{mech}, $Y_{\ln(mech)}$ generally inflates low-impedance confusions; the most notable instance of this is for Sliders (S3). An interesting case occurs for Masses (M3), Device B: Y_{mech} gives a zero average error, while $Y_{\ln(mech)}$ gives a substantial positive average error. This is due to the different weightings being given positive and negative errors on an individual basis; coincidentally, the weightings result in cancellation for Y_{mech}, but there are more B–A (positive) errors which are weighted more heavily for $Y_{\ln(mech)}$, resulting in net positive error.

(Row 4) Perceptual Scale Based Response Variables

Essentially the same relative trends show up in Row 4 (Y_{perc} compared with $Y_{\ln(perc)}$) as for Row 3, although more so for Masses (M4), where Y_{perc} was closer to $Y_{unweighted}$ than was Y_{mech} to begin with.

Target Effect as a Function of Response Variable and Subject

Figures 9-22–9-25 illustrate the same four combinations of response variables, but decomposed to show effects for single sessions.

Essentially the same trends are seen here as in Figure 9-21, the last one discussed, but with substantially more variability — particularly with toggles, where there was a great deal of variation both in the perceptual curves and in the error patterns. Note the Toggles column (T1–T6) of Figure 9-25, which shows the first group of response variables ($Y_{unweighted}$, Y_{mech} and Y_{perc}): the polarity of average error changes from session to session. This is true of $Y_{unweighted}$ alone (and was noted earlier); but you can see that in addition, the relative amplitudes of Y_{mech} and Y_{perc} are not always the same. In T3 (Patty), Y_{perc} is smallest amplitude; for Sam and Lorna (T5 and T6, respectively), the Y_{mech}-weighted error is smallest. These kinds of inconsistencies are generally due to differences in the number of negative versus positive errors made in each case, and the relative weighting of each.

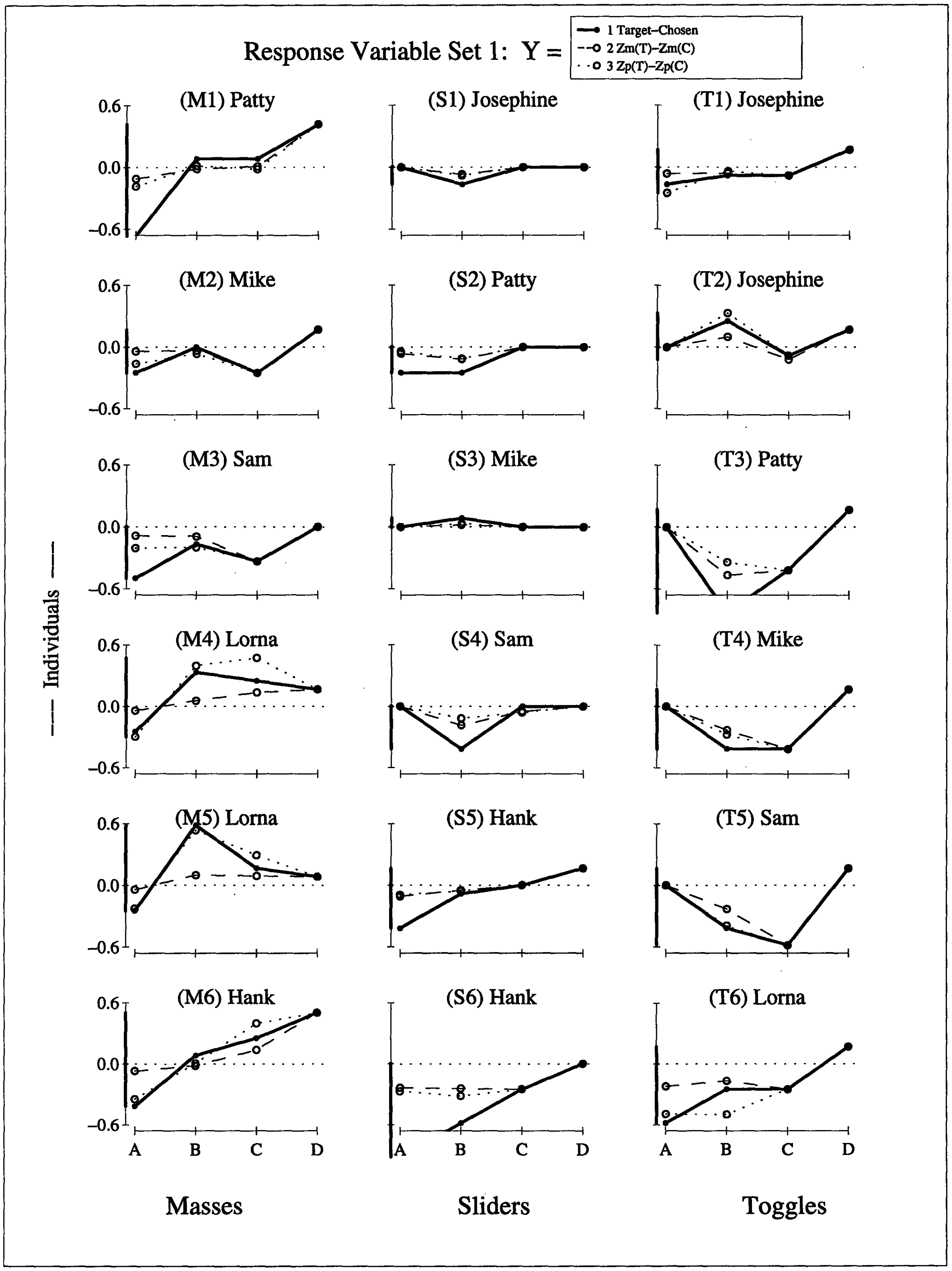


Figure 9-22: Target effect by subject: Y=[$Y_{unweighted}$, Y_{mech}, Y_{perc}].

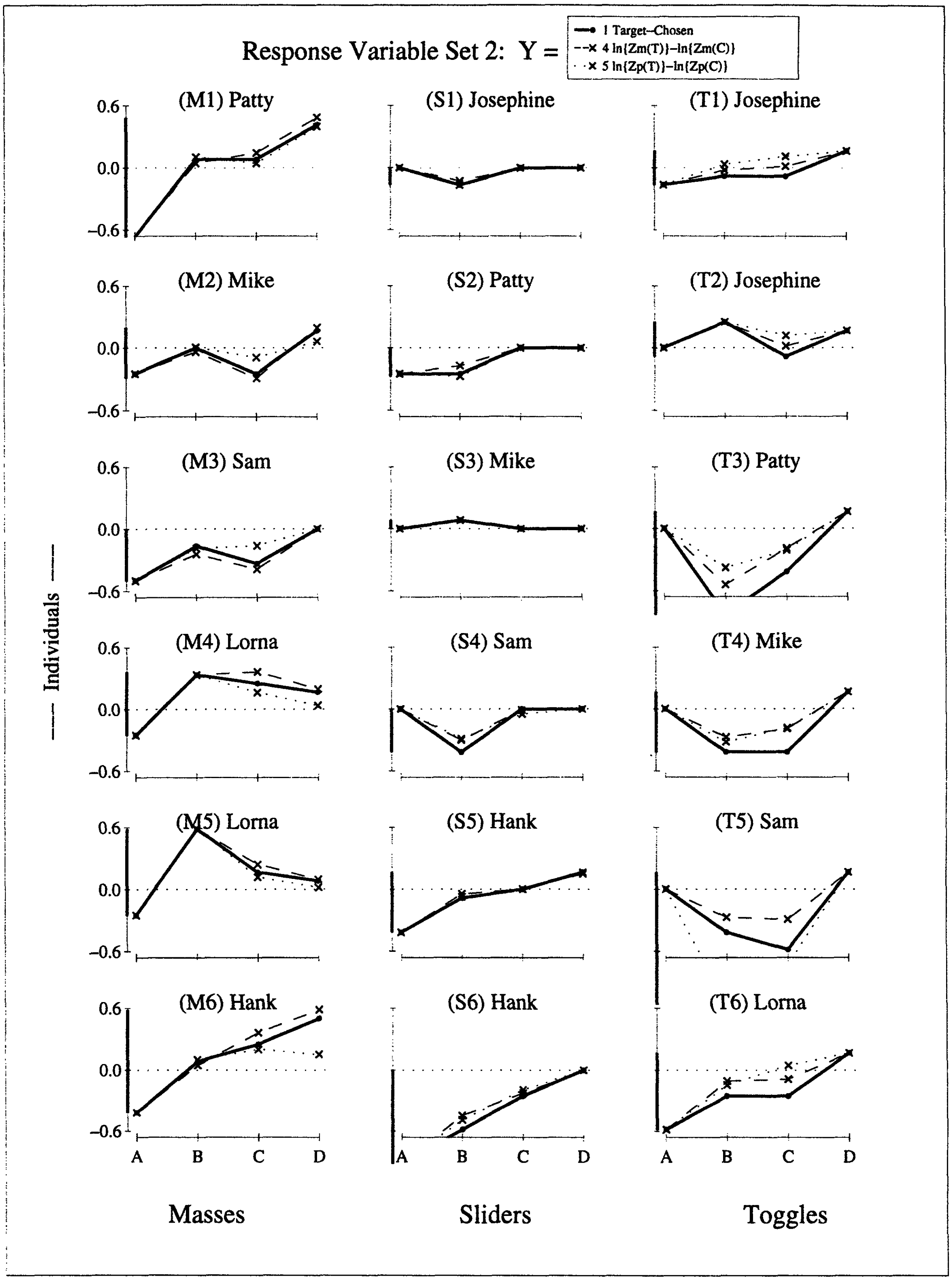


Figure 9-23: Target effect by subject: Y=[$Y_{unweighted}$, $Y_{\ln(mech)}$, $Y_{\ln(perc)}$].

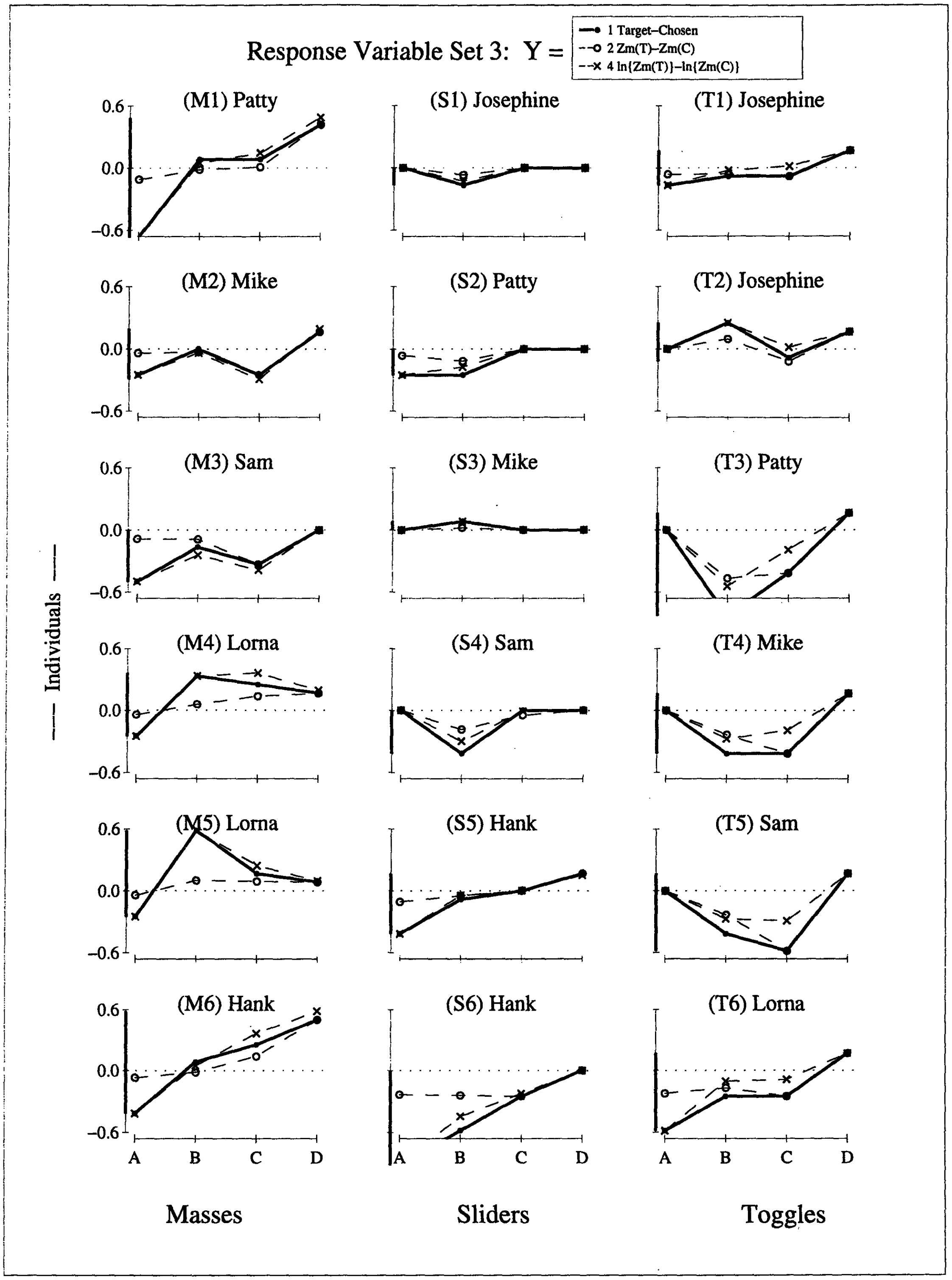


Figure 9-24: Target effect by subject: Y=[$Y_{unweighted}$, Y_{mech}, $Y_{\ln(mech)}$].

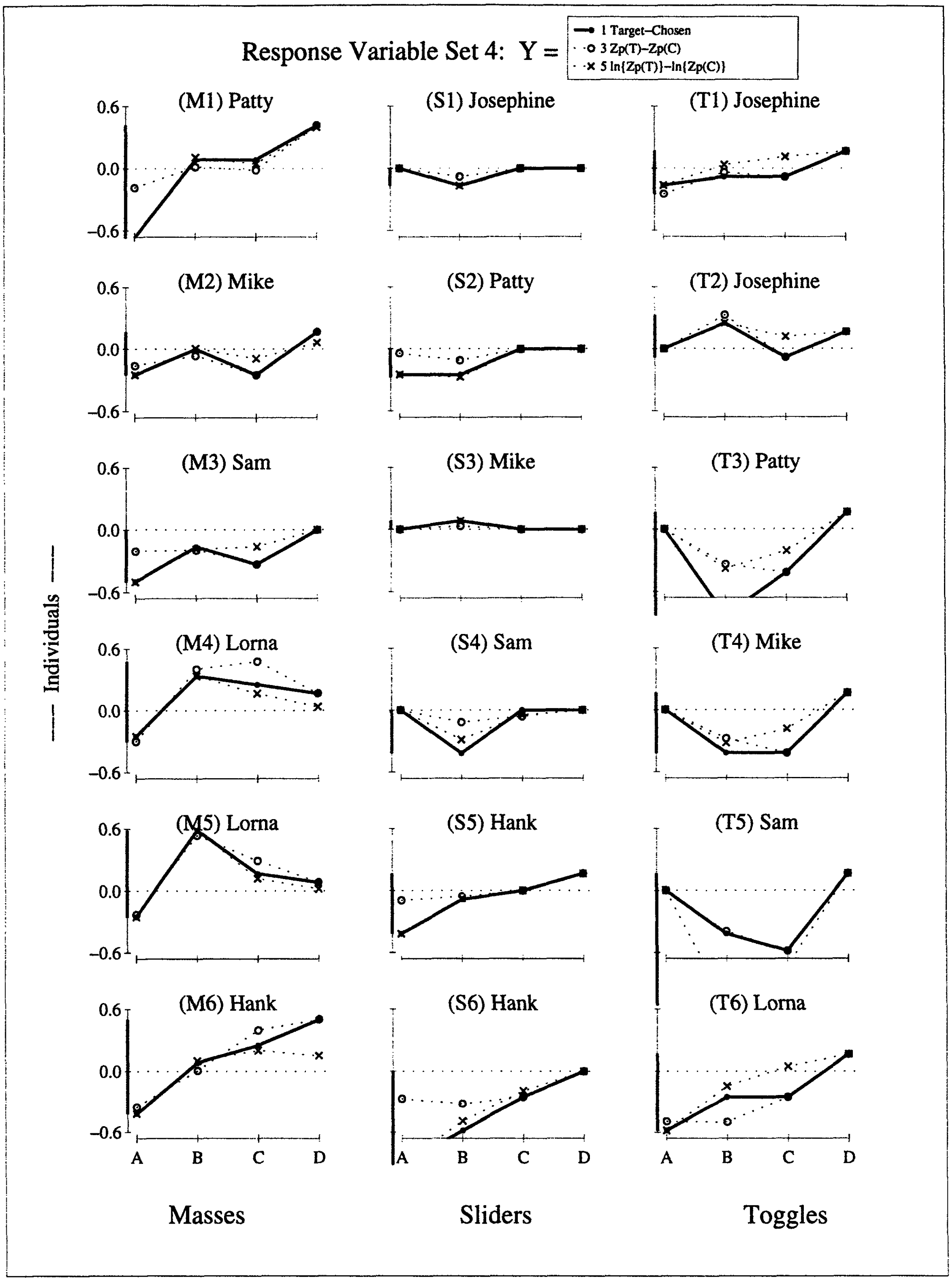


Figure 9-25: Target effect by subject: Y=[$Y_{unweighted}$, Y_{perc}, $Y_{\ln(perc)}$].

Subject	Real Device Family Masses	Sliders	Toggles	Subject Mean
Josephine	—	95.8	75.0, 75.0	81.9
Patty	60.4	83.3	56.3	66.7
Mike	79.2	97.9	66.7	81.3
Sam	75.0	81.3	62.5	72.9
Lorna	75.0, 72.9	—	52.1	66.7
Hank	64.6	79.2, 56.3	—	66.7
N=6 Mean	71.2	82.3	64.6	72.7
N=6 Std Dev	7.1	15.0	9.5	12.8
N=2 Mean	74.0	67.7	75.0	72.2

Table 9.4: Subject scores: individual and averages.

9.8 Subject Response: Repeatability and Variation

The data shown here is designed to test the hypothesis:

Variability within in an individual is less than variability between individuals.

This section will cover not only the explicit N=2 repeats analysis, but is also an opportunity to examine on the performances of individuals in general. Comments will be directed towards assessing characteristics and trends for the individuals in and of themselves. In some cases, previously presented data will be referred to.

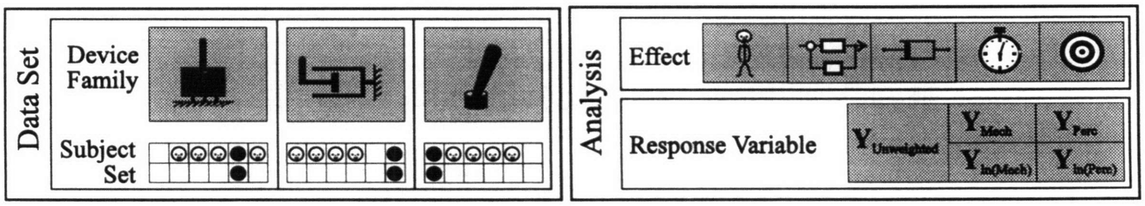
Scores for All Subjects

Table 9.4 and Figure 9-26 display the "scores" achieved in each session, by subject and family and in average for both N=6 (the full complement of sessions for each device family) and N=2 (repeat sessions only). A score is the percentage of trials, out of the total 48 in a session, where the subject chose the real device which was targeted. The four unstabilizable Toggle emulations in each Toggle session were automatically counted as 'misses' to signify an inadequate emulation, even though the subject did not have an opportunity to choose in those cases.

There were three repeat subjects: Lorna (Masses), Hank (Sliders) and Josephine (Toggles). Hank's second session with a score of 56.3 was far out of line with all the other Slider sessions, but the other repeat subjects' two scores for their respective families were well within an N=6 standard deviation of each other. In fact, they are as close to each other as any other pair of sessions for the family (Josephineachieved identical scores). This suggests that aside from an occasional anomalous performance, subject behavior is more internally consistent than cross-subject behavior.

Conversely, there is evidence subjects are also somewhat distinctive in their behavior: except for Masses (Lorna), the N=2 mean hugs the outside of the first N=6 standard deviation.

Subjects Repeated in a Device family



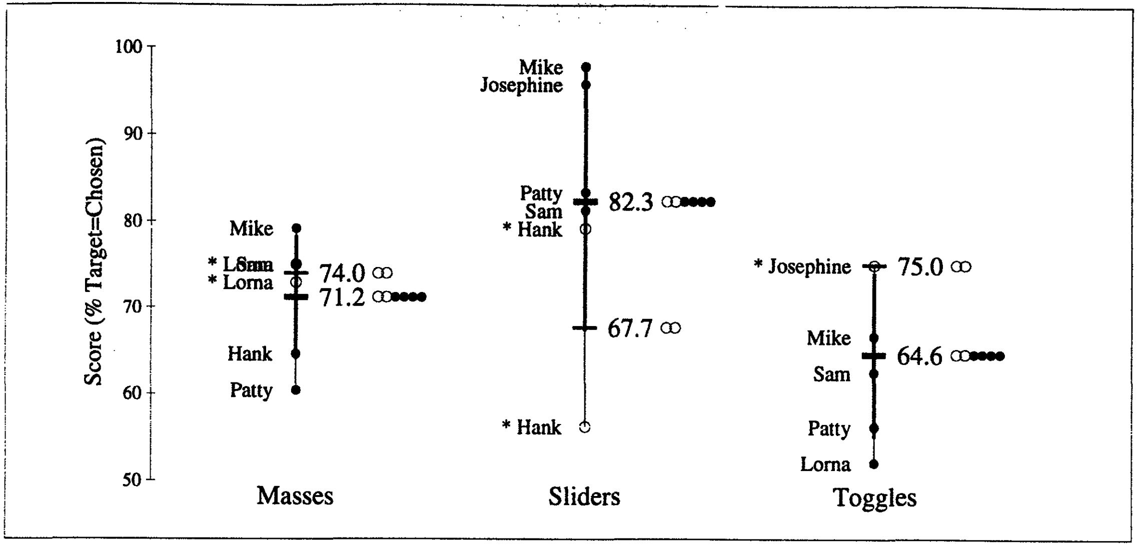


Figure 9-26: Score distributions by device family, all subjects. The thin vertical line for each family is range of scores, while the heavy vertical line indicates the N=6 standard deviation. Repeat subject names are marked by asterisks. The heavy horizontal bar marks the N=6 score mean, the light bar marks the N=2 (repeats) mean; the open (repeats) and filled (non-repeats) circles to the right of each mean bar signify the type of data contained in the mean.

This section will analyze the N=2 repeat sessions, which were designed to study the repeatability of single subjects. In some cases I will refer to the data of non-repeat subjects as well, in order to provide a basis of comparison for the degree of variability and repeatability found in the repeat subjects.

ANOVA Results by Family and Response Variable, N=2

A separate analysis of variance for each family was performed on the two repeat subjects, to see how significance patterns changed for the subject subset. The details of this analysis for all response variables and N=2 are shown in the Appendix, Tables G.4, G.5 and G.6).

The significance results for the N=2 ANOVAs are shown in Figure 9-27, by family and response variable. This figure should be compared with Figure 9-20, which has the same format but represents the N=6 analyses for each family and response variable.

The overall patterns of main effect and interactions significance is similar; there is no effect or interaction that shows up as significant in the N=6 case but and not in the N=2 case, with the exception of Blocks in Masses and for some response variables in Toggles. Already, this tells us there is more to the story than what was suggested by the scores: Josephine, the Toggles repeat subject, got identical (75% = high) scores for her two Toggle sessions, and yet the N=2 analysis shows Blocks mildly significant. The 25% (12 trials) which she missed in each session must have been different trials. We will look for evidence of this in the Effects analysis.

On the other hand, the fact that Blocks are strongly significant in the Sliders N=2 analysis is not surprising at all, since the repeat subject in that case (Hank) is responsible for most of the block variation in the family.

The N=2 significance patterns sometimes show interactions as significant which are not so marked in the N=6 analysis; for instance, the V–D (VE Expression–Damping) and T–D (Target–Damping) interactions for Toggles. Apparently Josephine's performance was more affected by these factor settings than were the rest of the subjects

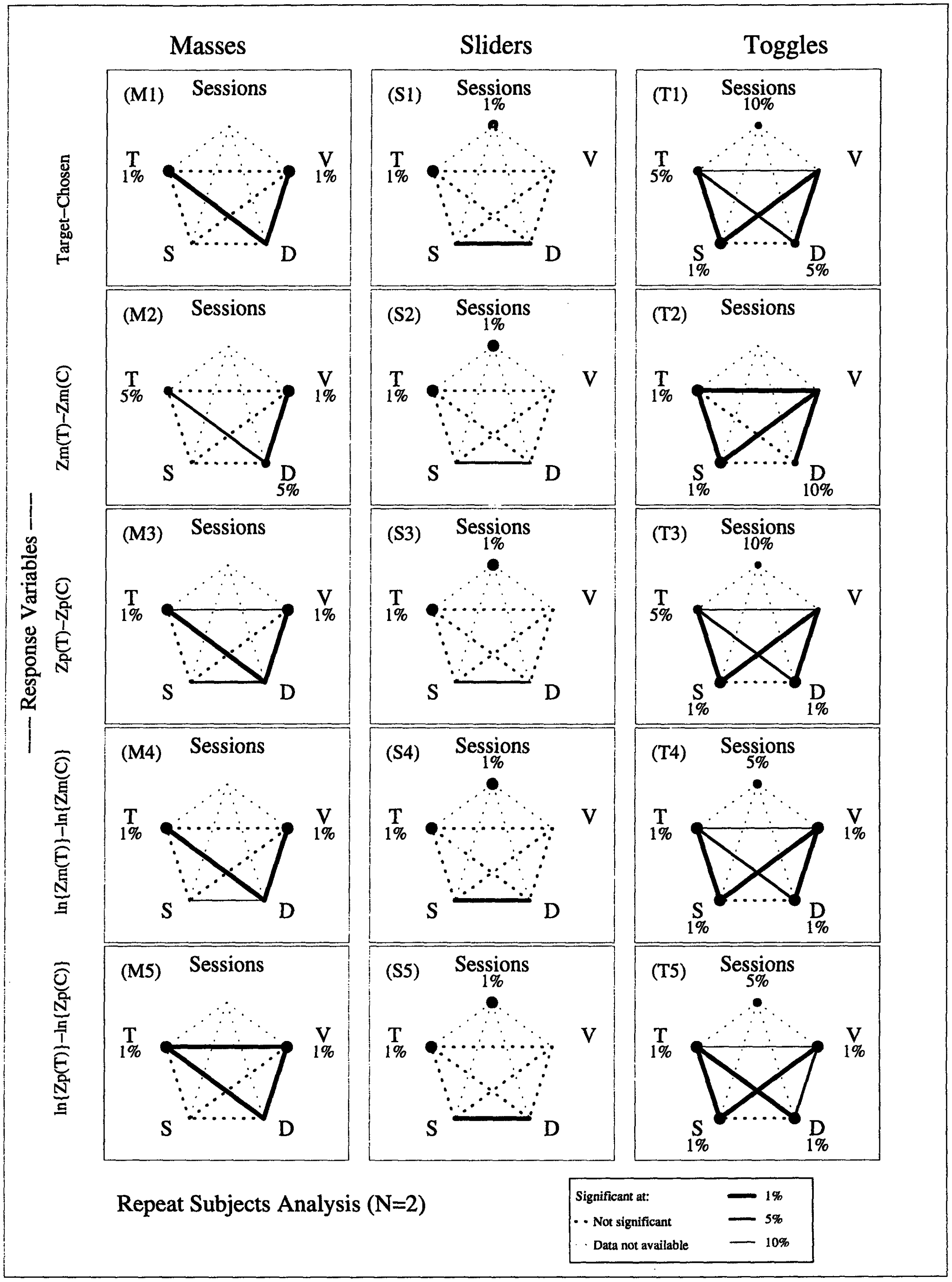


Figure 9-27: Effects significance by family and response variable: repeat subjects averaged.

Effects for Repeats Analysis

The subject effects for the N=2 analysis are shown in Figure 9-28 for all families and a single response variable ($Y=Y_{unweighted}$). The two curves shown on each subplot represent the two repeat sessions performed for that family, both by the same subject. Blocks are not shown in this figure, because blocks do not have meaning when looking at single sessions. Likewise, significances are not shown here because a significance analysis cannot be performed on individual blocks without changing the method of analysis. This set of effects may be compared with Figure 9-16, which showed the same effects in the same format but for the full N=6 analysis, with averages for all six sessions rather than individual session curves.

While there are similarities between the two sets of plots (N=2 and N=6), the differences are evident:

Masses (Lorna) : The repeat subject for Masses, Lorna, appeared by her scores (very close to the N=6 mean) to be the most typical of the repeat subjects; and this is exemplified by her effects plots, which follow the N=6 effects in shape and are not far off in amplitude. The N=6 curves are generally a little smaller in amplitude, reflecting the averaging effect of the other four sessions.

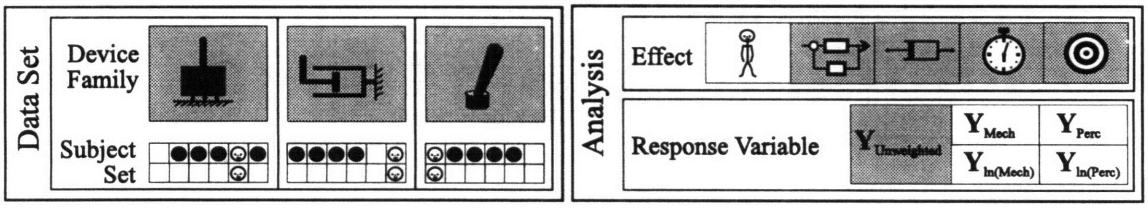
Sliders (Hank) : Hank of the anomalous second session produces one N=2 effects curves that is in line with the N=6 average curve and a second which is far lower than the N=6 average. In fact, the N=6 curves have been pulled down from zero largely by this single large-negative-amplitude session.

Toggles (Josephine) : The star Toggles subject's effects session curves are each of smaller amplitude than the N=6 average curve; however, we know that Blocks were significant for these two sessions. Closer examination reveals that there is more variation between each curve than for the Masses N=2 curves (blocks not significant) than less than between the Sliders curves (blocks significant). Although it isn't clear from this plot whether one session was consistently higher error than the other, we can assume from the fact that each session produced the same overall score that the error was balanced between them.

...

Figure 9-29 shows the same results again, but in the "Composition" format that was used earlier; compare with Figure 9-17, where the same format and scope was used in the N=6 case. Now the average for each session pair is shown in the middle level with significance coded, and the average for all repeat data across families (and different repeat subjects) for each effect can also be compared with those obtained for the N=6 case. There are no surprises here, or evidence of patterns which haven't already been noted and discussed.

Single Subjects Across Device Families



We just examined the repeatability of the three subjects who performed two sessions for a single device family; those repeat subjects were given the identical situation (except for randomization differences) twice and tested for consistency of performance.

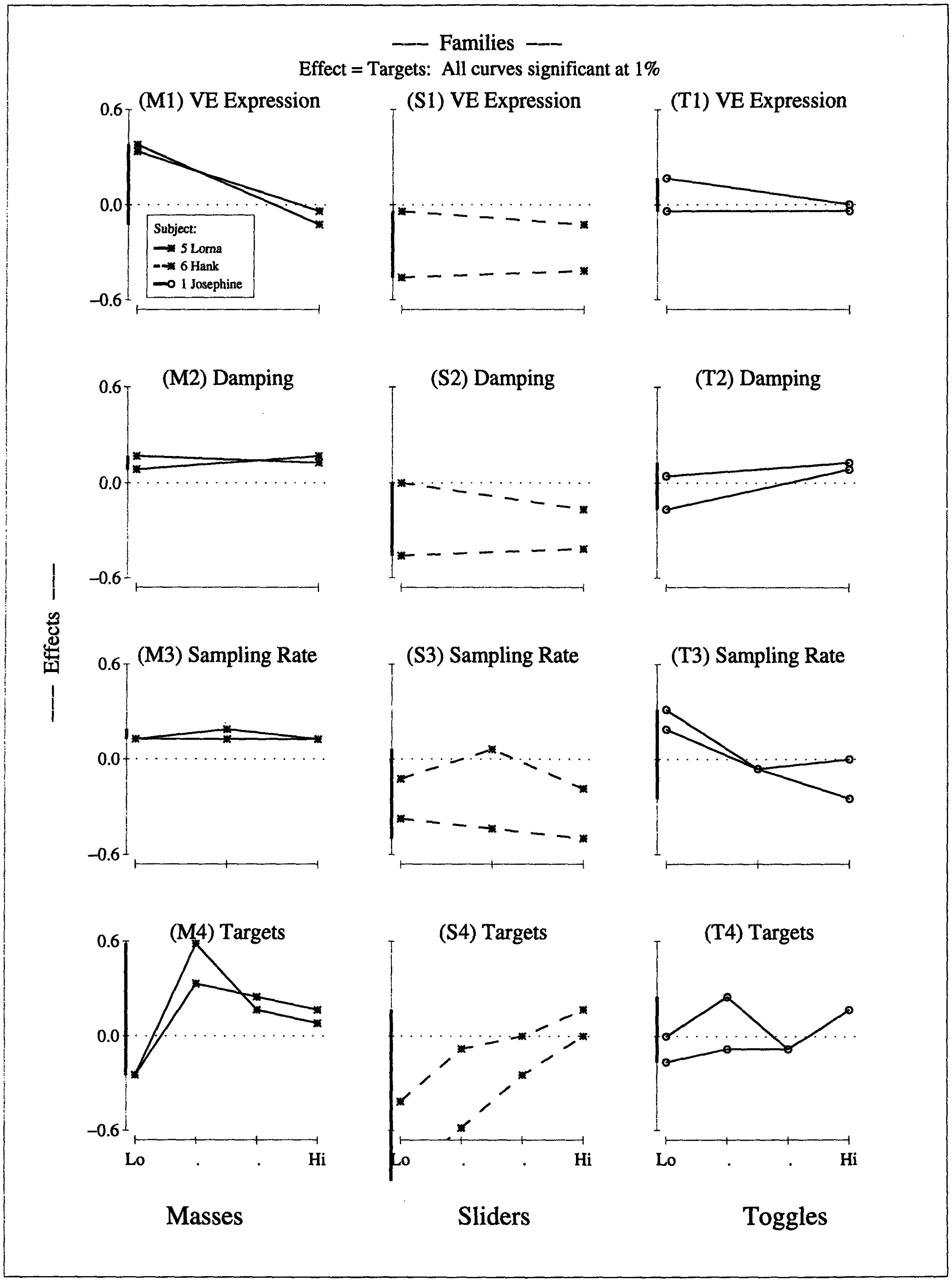


Figure 9-28: Main effects for all families: repeat subjects, $Y_{unweighted}$.

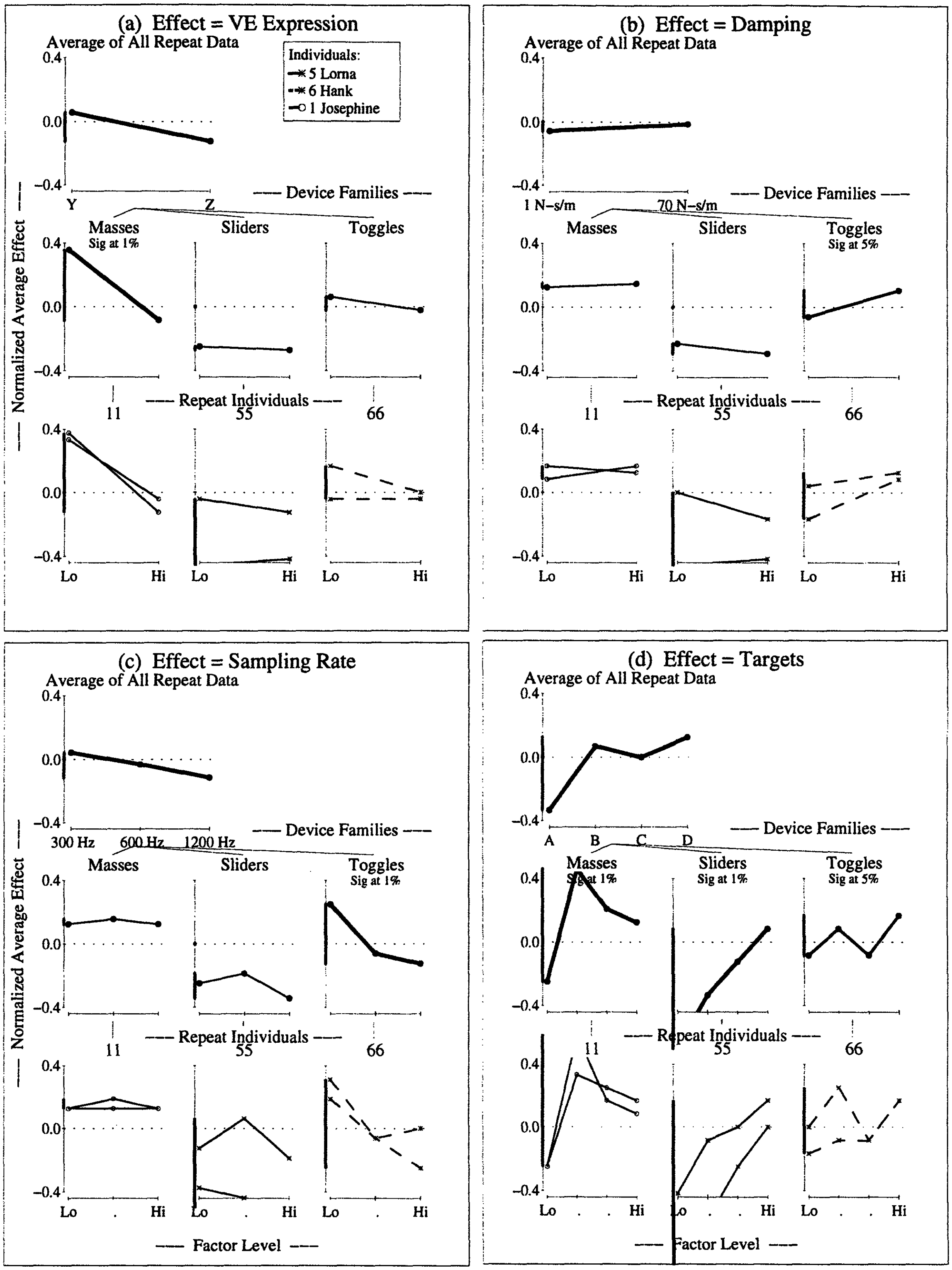


Figure 9-29: Composition of effects: [Repeat Subjects : Families : All Repeat Data] ($Y_{unweighted}$).

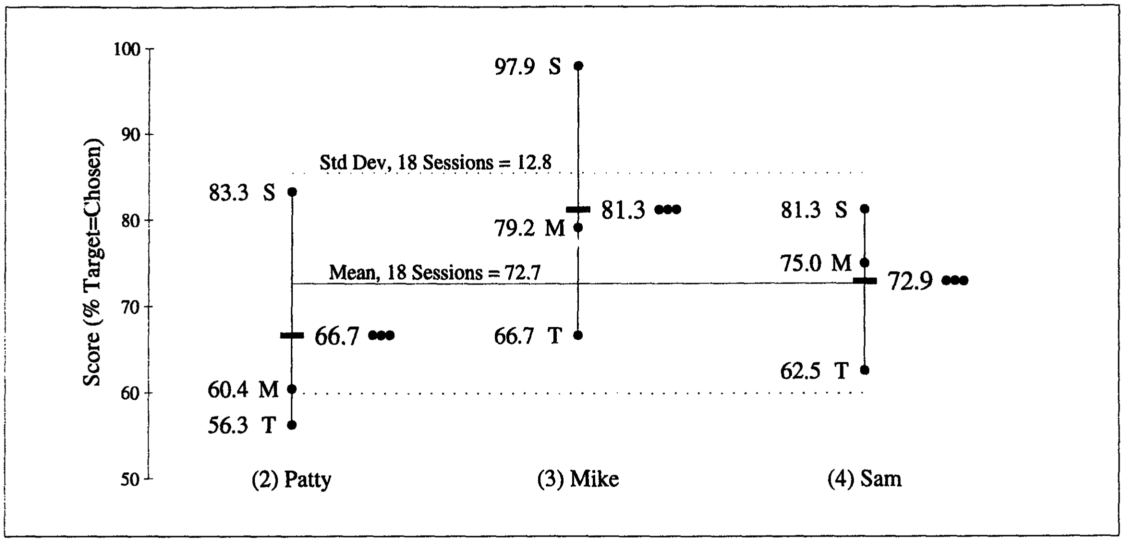


Figure 9-30: Consistency in scores across families for non-repeat subjects.

Now, we will do a similar type of investigation of the remaining three subjects who performed a single session on all device families: these are the non-repeat subjects, and a different set of insights is available from looking at how different or similar their cross-family performances are from each other.

In the following presentation, refer again when necessary to Table 9.4 , which listed both repeat and non-repeat score statistics. The non-repeat subjects are Patty, Mike and Sam.

Scores for Single Subjects

Here we will look at the scores of individual non-repeat subjects across the families they tested in. The question to be answered here is whether the variability in the scores for a single subject is less than than that for the group as a whole; e.g., do some individuals tend to score higher, regardless of device family?

Consider Figure 9-30: the format here is different from that of the last scores figure. Rather than families, each non-repeat subject is listed across the horizontal axis and the scores for each of their three different-family sessions listed above them. Standard deviation cannot be computed reliably for a three-member set, but the range and location for each non-repeat subject can be compared with the 18-session mean and score standard deviation.

A few features are immediately clear.

- The non-repeat subjects vary in range; Sam is more consistent across families than Patty and Mike.
- The groupings for each subject are located at different horizontal points on the page: Mike always scored the highest for each family, and Patty usually scored the lowest (except for nosing out Sam on Sliders).
- Each non-repeat subject scored in the same order for the respective families: highest on Sliders, lowest on Toggles. Even the distribution is similar for Mike and Sam; Patty's Masses score was very close to her low Toggles score.

These patterns confirm that even across families (i.e. given different tasks) individual performances are internally consistent, and well-differentiated from that of other individuals.

Effects for Single Subjects

The primary question to be considered in this section is whether a single individual consistently has problems with the same emulation variable degradations, when other individuals seem relatively insensitive to them. Figure 9-31 shows, in "Composition" format, the variability in average session error exhibited by each of the non-repeat subjects in their performance for three different family sessions. This data was seen previously in Figure 9-18, which shows all sessions; but here only the non-repeats are displayed and the comparisons are easier. As before, linetype in the third row (most detailed level) indicates family: solid lines are Masses, dashed Sliders and dotted are Toggles. We can now look through these plots by effect (quadrant of the figure) and subject (columns within quadrant) and see what is different or the same.

- Perhaps the most noticeable feature of this set of plots is the dominance of negative session-average errors; referring back to the all-subject plot, Figure 9-18, we see that the averages are slightly lower here than for the N=6 averages. This is probably coincidental, rather than an artifact of the repeat analyses; the same families are represented at the same frequency (1/3) in each set.
- Mike's session curves usually exhibit the lowest session-average error, consistently with the scores results.
- The polarity of average effects for a given family is not the same for all the subjects; for instance, look at what the dotted line (Toggles) does in (a) VE Expression, and the dashed line (Sliders) in (b) Damping. Sampling Rate and Targets are more consistent across subjects in their shape.
- Toggles (dotted line) most often shows highest average session error, independent of subject; Masses usually has the lowest.
- It is hard to say that subjects are more internally consistent in the shape of the various family-session curves than across subjects. For example, in (a) VE Expression, there is only one subject whose various session curves all have the same direction (Patty) and only one, a different subject, who has at least two curves with both magnitude and direction closely correlated (Sam). This lack of correspondence is even more apparent in (b) Damping.

Thus, the result of both the repeat and non-repeat subject studies is that there is better correspondence between subject and performance (how often the subject makes the intended identification, as evinced by score) than between subject and repeatable pattern of errors (which would imply different strategies and/or perceptual abilities and sensitivities).

Combined with my own observation that the subjects who consistently turned in the highest performances also tended to be the most conscientious (although not always the slowest), as opposed to the most confident or the most haptically "experienced", it seems likely that the individual differences are best explained simply by attention.

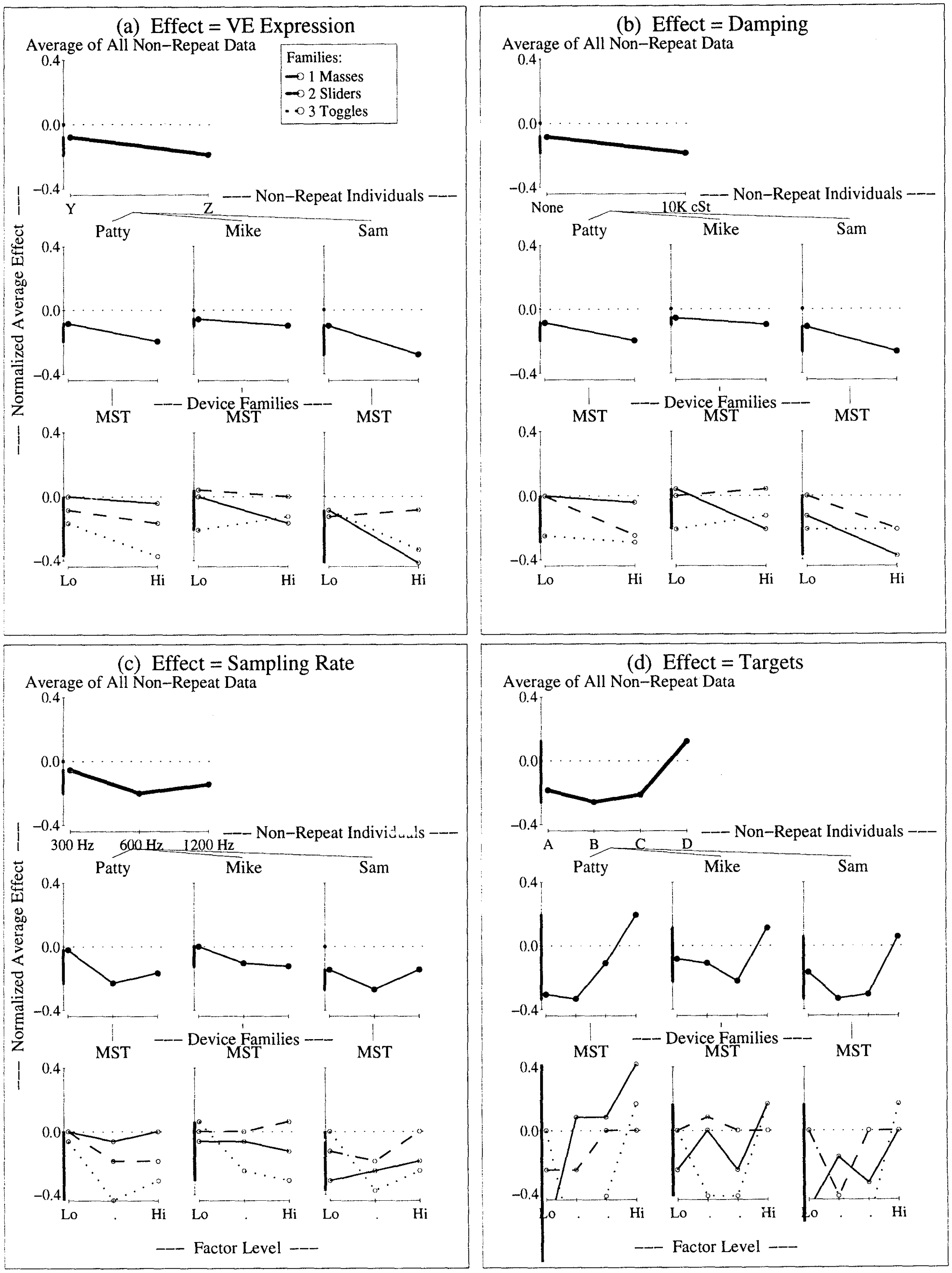


Figure 9-31: Composition of effects: [Families : Non-Repeat Subjects : All Non-Repeat Data].

9.9 Decision Times

Subjects were not given a time limit in making their comparison-trial decisions. It became obvious during the pilot experiments and the first real experiment session that there was a great deal of variation in decision time, both on a trial-by-trial basis within a session, and between subjects and device families. I felt that it would be informative to record these decision times and use the data to test two hypotheses:

1. Decision time is an indicator of subject performance (based on comparisons over all subjects of (Subject i score) / (Subject i average decision time).
2. Decision time is an indirect indicator of emulation quality, based on comparisons of variation in decision time within individual subjects. Reasoning: lower quality emulations are harder to choose from, thus lower quality emulations will result in longer decision times than high quality emulations even when the subject eventually makes the correct response.[4]

The latter hypothesis is of particular interest in that the primary limitation of this experimental methodology is that while it measures emulation functionality in a specific set of tasks, it does not measure the more vaguely defined variable(s) of emulation "quality", which in many applications will be as important, if not more so, as functionality. Thus if the metric of decision time could be used to estimate "quality" of the emulation, this would be useful indeed.[5]

A further value of recording and studying the $T_{Decision}$ data is to validate the decision to let subjects choose the time they spend on a discrimination; it is possible that constraining $T_{Decision}$ would have resulted in different strategies. Given that the open-ended time rule was used here, a definitive answer can't be found from this data, but we may learn something.

Data Estimated for One Subject: Decision times were recorded explicitly for all sessions except the first one performed. For this session (M1, Mike:Masses), the decision times had to be estimated from the timestamp which was recorded for each trial, and knowledge of the fairly uniform time required to execute the rest of the trial. long I take to change out. This data is not always reliable because other time-factors, such as a subject stopping to make a comment, could not be reconstituted. However, it is a reasonable guess and will show, in most cases, a good indication of the time the subject required for a given trial, as opposed to average time for all trials.

Decision Time Data

The minimally-processed decision time ($T_{Decision}$) data is presented in Figure 9-32, in much the same format as was used for the pre-processed comparison trial response data in Figures 9-9–9-11. Refer to Section 9.4 for a review of the randomization procedure and standard order used here.

Figure 9-32 shows two plots for each device family (M–T): (a) is decision times for all sessions for that family in the order the trials were performed, and (b) is the same data resorted into standard order. The first difference between this data presentation and that of the comparison trial response is the discretization level. Whereas the comparison data could only be [-1 – 0 – 1], $T_{Decision}$ ranges from 6 to 127 seconds, with a resolution of 1 second.

---

[4] Although I have a similar degree of confidence in the data shown in this section as in the rest of the chapter, these observations and results are of a conjectural nature because they are supported only indirectly by the data. No direct quality measurements were made here.

[5] Note that in trying to relate trial-by-trial decision times to emulation quality, we are to some extent confusing results with another factor which strongly influences decision time: the low-impedance targets (real devices and emulations) are harder to distinguish than the high impedance ones, because they are more closely spaced. i.e., not only is the emulation quality changing, so is the difficulty of the discrimination task independent of emulation variable changes. Thus the we expect to see a strong influence on decision time of both general emulation variable degradation and Target effect (which is not a degradation, but a change in task).

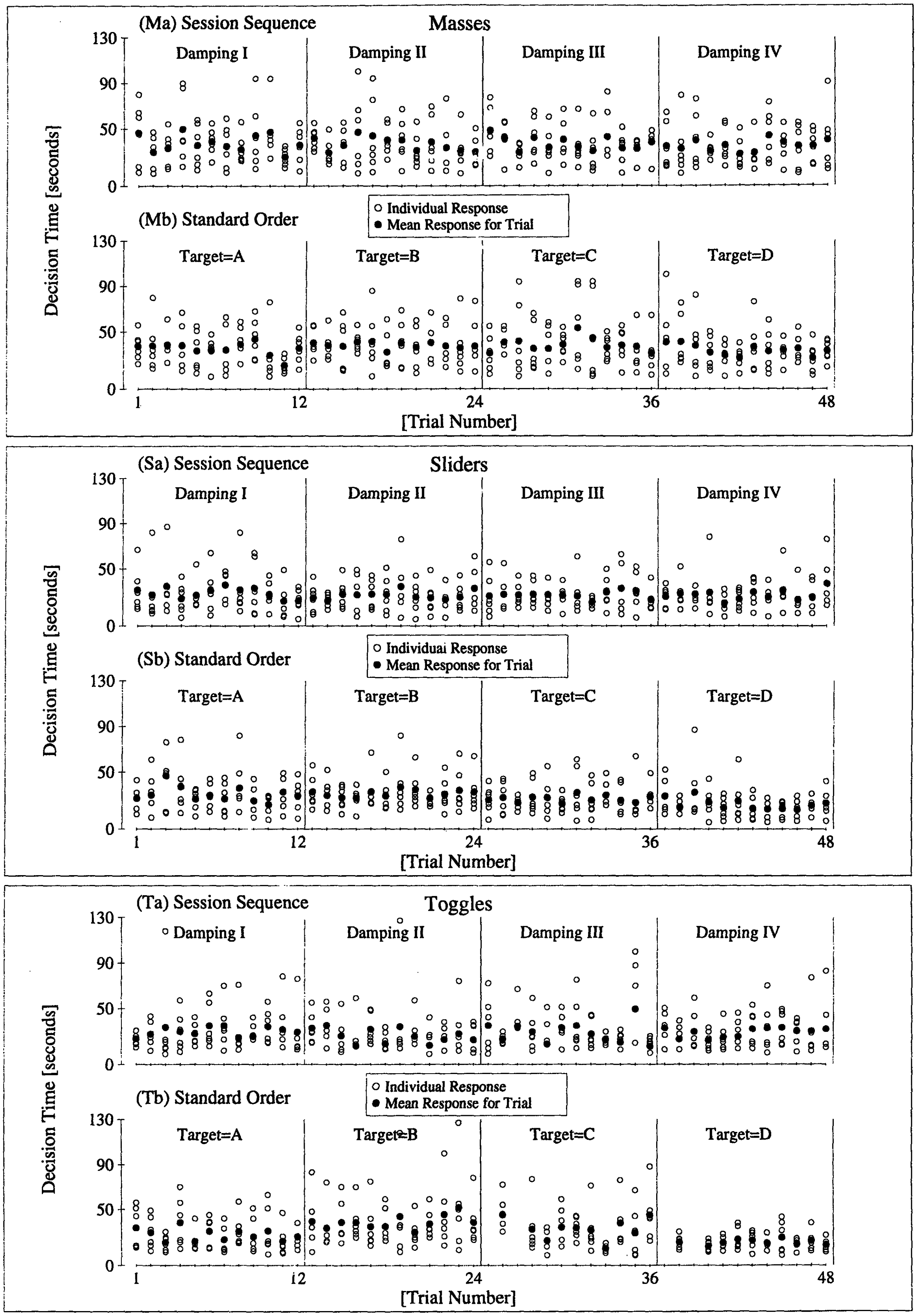


Figure 9-32: Minimally analyzed decision-time data by trial.

As with the comparison trial data, we should look for temporal trends in the session sequence plot for each family, and for repetitive patterns in the standard order plot.

(a) Temporal Trends: Surprisingly, there is no visible evidence of temporal decrease in $T_{Decision}$ for any of the families (Ma, Sa, Ta); subjects on average took the same amount of time to make a decision in the first few trials as for the last few trials in a session, even for Masses (the first session for most subjects, and thus the one most likely to exhibit learning effects). There is an apparent difference between families in all-session means, with Masses looking a little higher — this will be considered shortly.

(b) Standard-Order Patterns: There are no obvious patterns in the standard-order plot for Masses (Mb). For Sliders (Sb), there seems to be a reduction in both mean decision time and occasional large-magnitude values in the last few trials, 43–48. This set of trials corresponds to the highest-impedance target and the highest sampling rate. We know already from Section 9.3 most subjects considered Slider Device D very different from the lower-impedance targets, resulting in the more highly dished log distribution for Sliders (both perceptual and measured), so the Target dependency is not surprising. Thus, the lower decision times for that area does reinforce the concept that subjects took less time to choose for an easier discrimination. It also appears that there is some influence from sampling rate, however, and we should look for that effect later on.

In addition, there seems to be a reduction in mean decision time and decision time variability for the first quadrant of (Tb), which corresponds to the lowest-impedance target, A. We will examine this effect later to see if there is indeed any interesting correlation.

Hypothesis 1: Subject $T_{D_{avg}}$ as Indicator of Performance

The first hypotheses is whether average session decision time ($T_{D_{avg}}$) constitutes a useful indicator of session (subject) performance. We will look at gross whole-session trends in decision time to see if it tells us something specific about that subject's strategy and a possible relation between fast/slow strategies and high performance in general. We are not looking at trial-by-trial decision times for information about which emulations were hard to discriminate, etc.; that will happen in the next section.

Several of the forthcoming figures (9-33, 9-34 and in a later section, 9-38) plot average scores vs $T_{D_{avg}}$ (or predicted score, for Figure 9-38) in a variety of formats. Some comments are in order with regards to this plotting method; then the results themselves will be discussed.

Use of Diagonal Mean-Reference Line: The diagonal dotted line present in each plot, which starts at the origin and proceeds with a slope determined by average score (rise) and $T_{D_{avg}}$ (run) for the data in the plot, is drawn with the assumption of a 50% score as baseline (in the same way that a decision time of 0 seconds is baseline on the horizontal axis). That is, the line's slope is really

$$\frac{(\text{Score} - 50\%)}{T_{D_{avg}}}.$$

That is based partly on the observation that no subject scored below 50% ever, although three session scores (16% of all sessions) were in the 50–60% range; but more strongly on the consideration of 50% as a statistical breakpoint. A session score of less than 50% is unlikely except in the presence of strong misleading emulation biases, implying as it does wrong answers of higher frequency that correct answers. Assuming nearest-neighbor misses only, which assumption was justified by the results, in any given trial a subject has either two or three possible choices (target + 1 (A,D) or (B,C) neighbor devices). In the two-choice case, which happened 50% of the time, the subject has a 50% chance of randomly choosing the right answer; 33% chance in the three-choice case. Thus

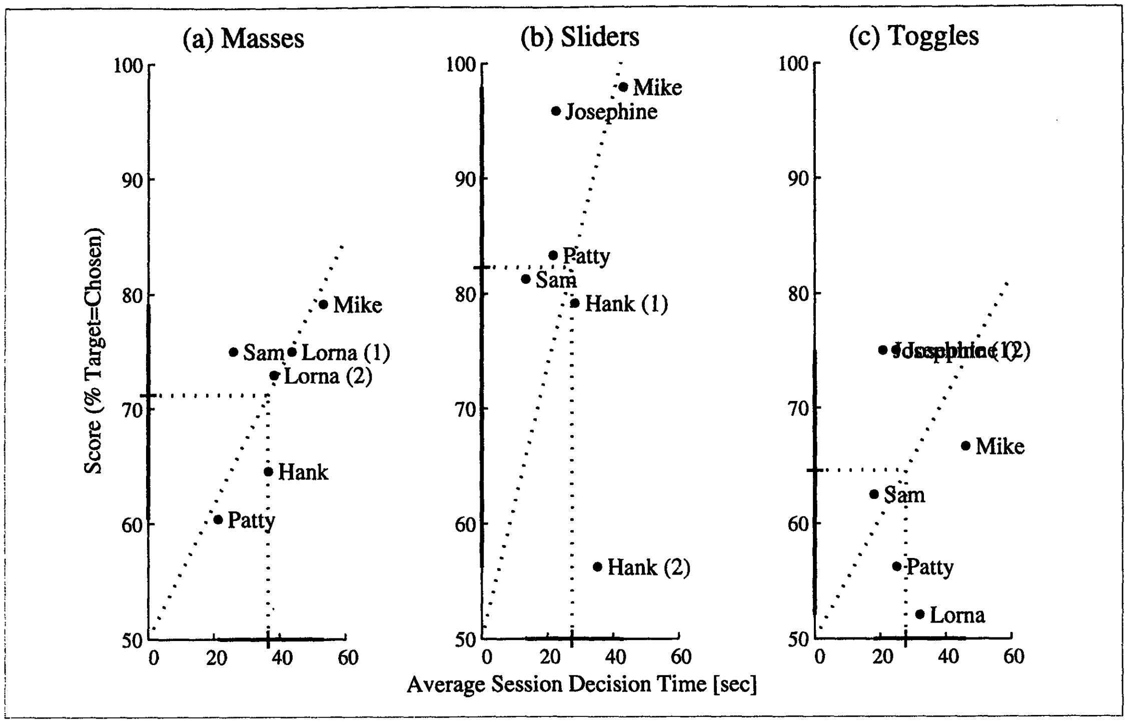


Figure 9-33: Session scores related to $T_{D_{avg}}$, arranged by family.

scores occurring with any regularity below 50% would signify something seriously wrong with the targeting, and render suspect other results. Fortunately, that does not seem to be the case here, and 50% seems a reasonable score baseline to use.

What to Look For in Score=func(Decision Time) Plots: The point in the following plots is to look for any of several different things happening:

- Is $\frac{\text{Score}}{T_{D_{avg}}}$ constant — i.e., does taking longer generally result in a higher score?
- Or conversely, is the (Score× $T_{D_{avg}}$) product constant — i.e., do shorter decision times ("gut reactions") do better?
- Do some subjects characteristically fall in a certain quadrant of the (score $-$ $T_{D_{avg}}$) plane (the plane's centroid being described by the intersection of the two axis means, indicated by the two orthogonal dotted lines on each plot)?
- Do some families show a preponderance of sessions falling in a certain quadrant of the (score $-$ $T_{D_{avg}}$) plane?
- Are repeat subjects (same device family) correlated in their $\frac{\text{Score}}{T_{D_{avg}}}$ ratio?

Influence of Subject on $\frac{\text{Score}}{T_{D_{avg}}}$ Ratio:

Figure 9-33 shows session scores plotted against session $T_{D_{avg}}$ for each device family: thus within a device family, it is easy to see the degree to which subjects make a difference. For some cases the points do roughly follow the diagonal dotted line representing constant $\frac{\text{score}-50\%}{T_{D_{avg}}}$ ratio; the best fit is for Masses, and without Hank's session 2, Sliders would be a very good fit also (the mean score

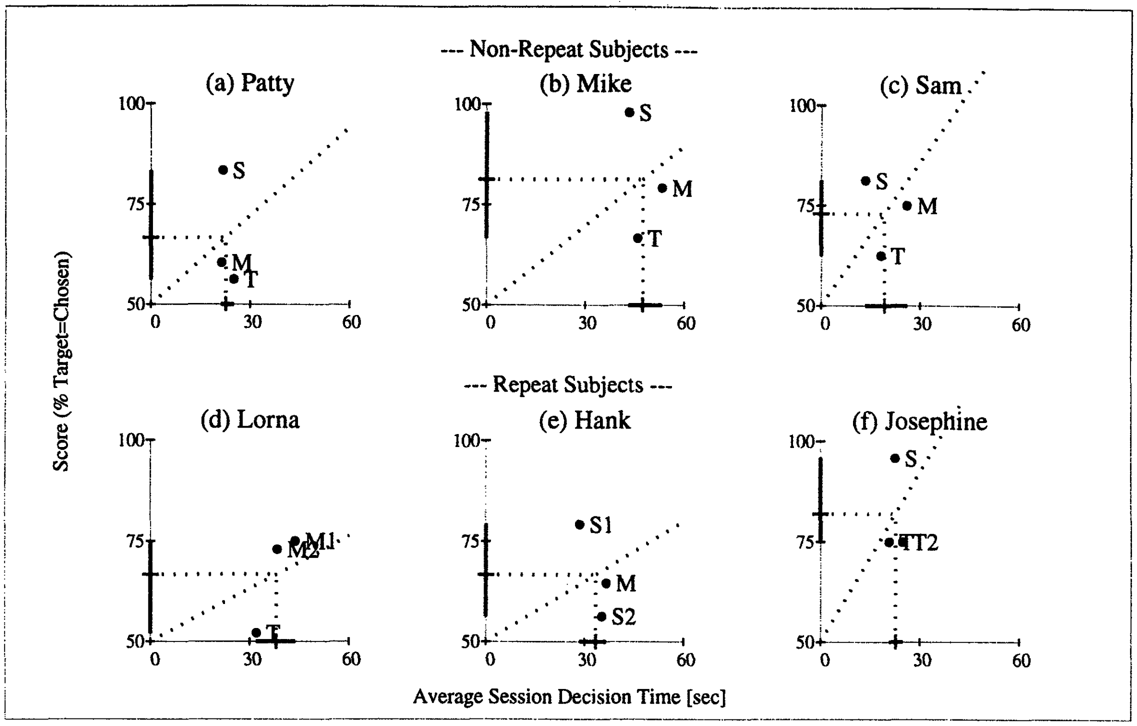


Figure 9-34: Session scores related to $T_{D_{avg}}$, arranged by subject.

would be a little higher without Hank, making the fit for the remaining five sessions even closer than it already is).

There is more spread for Toggles, with the best fit line probably being a vertical one through constant $T_{D_{avg}}$, or even a line with slightly negative slope — implying that there was less variation in score for this family than there was in $T_{D_{avg}}$. This is true for average decision time, not necessarily decision time in general. The raw $T_{Decision}$ plot, Figure 9-32 (Ta–b) shows about as much range in trial-by-trial decision time as the other families, even though the means are a bit lower. This implies that some trials took a long time for every subject, but on average they were quicker for Toggles.

Thus, with the exception of Toggles (the most complex and least accurately emulated family), longer decision times mean higher scores.

Influence of Device Family on $\frac{Score}{T_{D_{avg}}}$ Ratio:

The same data is viewed in a different arrangement in Figure 9-34, where they are sorted by subject and labeled by device family. In this format we can see, from subject to subject, what trends the device family is responsible for.

As we saw before when looking at scores without reference to decision time, Sliders are always highest on the vertical scale and Toggles the lowest for any subject.

New insights available from the introduction of decision time data are the following:

- What horizontal excursion in $T_{D_{avg}}$ for each device family relative to translation in mean $T_{D_{avg}}$ (for all device families) is incurred for each subject?

  The $T_{D_{avg}}$ excursion is indicated by the bold horizontal range-axis, which is seen to vary by a factor of perhaps four from Patty and Josephine (smallest excursions of about 4–5 seconds) to Sam (largest, 15 seconds); mean $T_{Decision}$ for all sessions by that subject is indicated by the location of the bold tick and vertical dotted line. The overall-mean $T_{Decision}$ moves, from one subject to the next, by more than subject excursions vary (19.7 seconds for Sam to 48.8 seconds for Mike, a delta of 29 seconds; these numbers can be seen in a later figure).

Thus individuality seems to be a stronger influence on $T_{D_{avg}}$ than device family.

- Are points for specific device families consistently ordered in the horizontal ($T_{D_{avg}}$) direction?

  No. The respective orders found in subplots (a–f) are:

  M–S–T S–T–M S–T–M T–M–M S–S–M T–S–T

  That is, Patty does does Masses fastest and Toggles slowest — although not by much; Mike and Sam do Sliders fastest and Masses slowest. If one wanted to make a generalization, it would be most valid to do it for the subjects who showed the most variation in $T_{D_{avg}}$ to begin with.

  From that standpoint, Masses seem to take the most time while Sliders often result in the quickest decisions.

- Is there a characteristic quadrant location of data points for each device family, relative to the $\frac{\text{Score}}{T_{D_{avg}}}$ for that subject's sessions?

  Sliders are always above the diagonal average-ratio line (they have to be, since their scores are always highest); that is, they are always in the high-score half of the plane. Their left-right position (fast-slow) is generally barely on the fast side.

  This suggests that Sliders are indeed the easiest of the discriminations — eliciting response which are both fast and accurate.

  Masses and Toggles are always in the low-scoring half-plane and closer to the diagonal dotted line; this is because both their scores and average decision times were generally closer to each other than to Sliders. Regarding fast-slow quadrants, Masses and Toggles are indeterminant as mentioned above; although Masses tend to be slow low-scoring and Toggles faster low-scoring.

- Are repeat session data consistent in quadrant location?

  With the usual "Except for Hank" caveat, yes. Not only are scores repeatable for Masses (Lorna) and Toggles (Josephine), so are the decision times.

Summary of Hypothesis 1 Argument

This analysis supports the claim that over an entire session and probably as a characteristic of an individual, longer decision times mean higher scores for the case of simple environments. Such a claim cannot be made for complex environments.

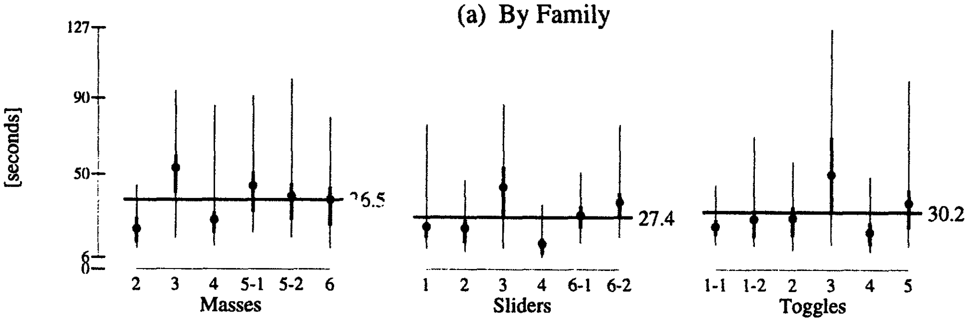
Hypothesis 2: $T_{Decision}$ as Indicator of Decision Difficulty

Under the second hypothesis stated above, we will test the ways in which

1. $T_{Decision}$ (not $T_{D_{avg}}$) might act as an indicator of the general effectiveness — and perhaps "quality" — of the emulation; as implied by evidence that correct decisions take less time on a trial-by-trial basis.

2. $T_{Decision}$, employed as the basis for an analysis response variable in much the same way that $Y_{unweighted}$, Y_{mech}, etc. were used earlier, is associated with different levels of emulation variables (here, this means VE Expression, Damping and Sampling Rate) and by that association indicate that those factor levels result in easier (more obvious and therefore faster) decisions and thus those factor levels are more functional — over and above what the response score indicates.

(a) By Family



(b) By Subject

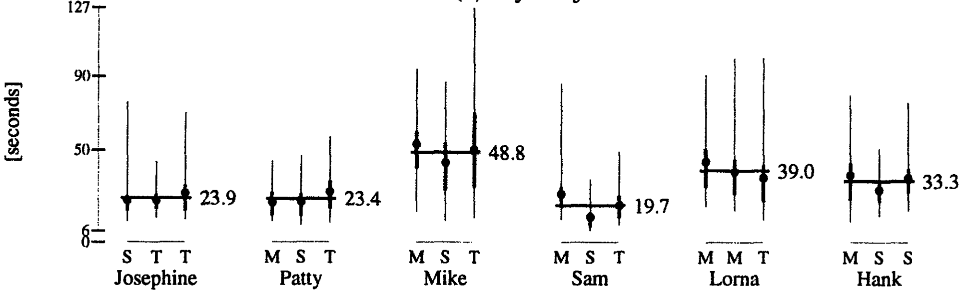


Figure 9-35: Variability in decision time: quartile distributions.

Variability of Decision Time Across Families and Subject

How much $T_{Decision}$ variability is there? This is the starting point — if there's very little variability within a session, then there's no room for $T_{Decision}$ to act as an indicator of anything on a trial-by-trial basis.

Figure 9-35 shows in (a), family-by-family $T_{Decision}$ variation, as well as family $T_{Decision}$ means (all sessions); and in (b) the same thing but by subject, and with subject $T_{Decision}$ means. For all sessions, mean (dot), range (thin line) and 2nd and 3rd quartiles (bold line) are displayed; the horizontal line in each group shows the group mean.

The short answer is that yes, there is a great deal of inner-session variability in decision time: not only the ranges for each session but also the inner quartiles for each session are larger than the differences between group means. The most impressive in this respect is Mike in his Toggles session, with the long-$T_{Decision}$ outlier visible at 127 seconds and a low value of 13 in the same session, for range of 114 seconds or just shy of 2 minutes.

There is an addition a fair amount of movement in subject means (b), with Mike being consistently the slowest and Sam the quickest — a total variation in subject means of about 29 seconds, or 148% of the low value. The family means are more stationary, but still vary by a substantial degree: a difference of 9 seconds, or 33% of the low value.

Once we've established that subjects vary in how long they require to be satisfied with their decisions during a single session, we can go on to look for associations between decision time and positive performance — still on average, but now looking at within-session trends rather than through session-by-session comparisons.

In Figure 9-36, device family is shown by column, subject by row. The horizontal axis for each column is decision time, and two quartiles are shown for each session (1 pair for each family and non-repeat subject, 0 or 2 pair for each family and repeat-subject). The upper quartile in the pair, marked with an '*o*', represents the distribution in $T_{Decision}$ for the trials in which a correct identification was

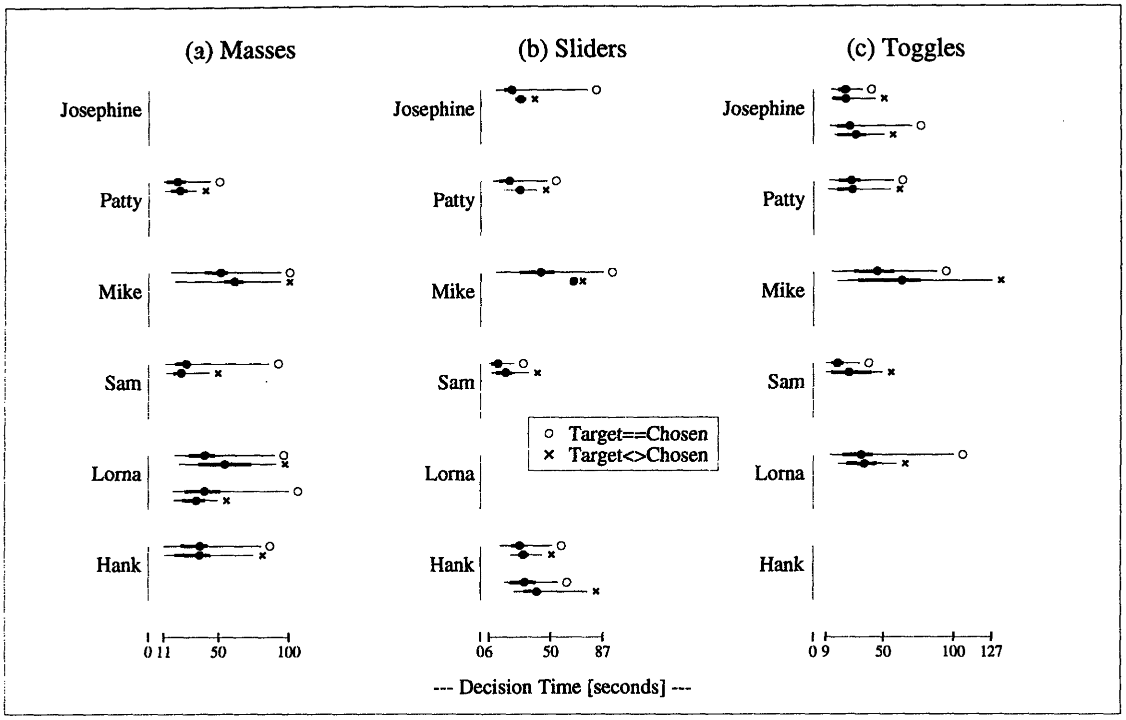


Figure 9-36: Decision time related to correctness of answer within sessions.

made. The lower in the pair, marked with a '×', likewise represents the distribution of $T_{Decision}$ for for the "miss" trials. Data for the unstabilizable Toggle trials was omitted; these would have been counted as "misses", but the time to decision is not known.

The question, then, is whether subjects generally took more time to come up with a "hit" than they took to "miss", or vice versa. This can be seen by the relative horizontal placement of the dense part of the quartile pair members: i.e., given that this is an average assessment, the mean and inner quartile location is more telling than the overall ranges. The differences in inner quartiles is more difficult to characterize than the single-variable mean, but more often than not follows the same pattern.

The graphical presentation is supported by the following simple statistical test. To test the hypothesis that $\bar{T}_{dec_{incorrect}} > \bar{T}_{dec_{correct}}$ (i.e., incorrect decisions take longer), I performed a paired comparisons t-test on the difference between average $T_{Decision}$ for incorrect trials in a session and $T_{Decision}$ for correct trials,

$$\begin{aligned} d_{session\ i} &= \frac{1}{n_{misses}} \sum_{j=1}^{n_{misses}} T_{dec_{misses}} - \frac{1}{n_{hits}} \sum_{j=1}^{n_{hits}} T_{dec_{hits}} \\ n_{trials} &= n_{misses} + n_{hits} = 48 \end{aligned} \tag{9.1}$$

The test statistic was constructed as

$$\begin{aligned} t_o &= \frac{\bar{d}}{S_d/\sqrt{n}} \\ \bar{d} &= \frac{1}{n_{sess}} \sum_{i=1}^{n_{sess}} d_{session\ i} \end{aligned}$$

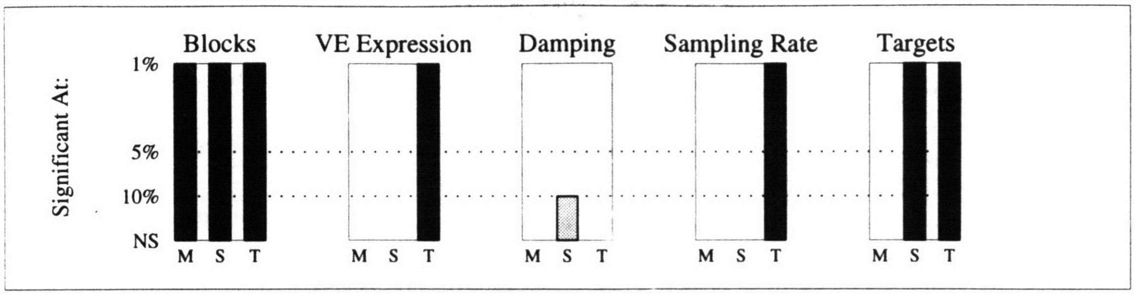


Figure 9-37: Significance of main effects: all subjects averaged, $Y_{T_{Dec}}$.

$$\begin{aligned} S_d &= \left[\frac{\sum_{i=1}^{n_{sess}}(d_i - \bar{d})^2}{n_{sess} - 1}\right]^{1/2} \\ n_{sess} &= n_s ubjects \times n_f amilies = 18 \end{aligned}$$

The values in seconds for $d_{session}$ were as follows:

Session	Masses	Sliders	Toggles
1	1.56	6.22	.569
2	9.99	7.53	4.29
3	-3.75	23.3	.993
4	14.3	5.37	17.3
5	-6.02	2.61	8.47
6	-.078	8.64	2.37
family average	2.66	8.94	5.67

By the above relations,

$$\begin{aligned} \bar{d} &= 5.75 \\ S_d &= 7.30 \\ t_o &= 3.35 \\ \text{significance} &= .0025\% \end{aligned}$$

The hypothesis is therefore supported statistically through a small but consistent trend: *on average, errors take longer.*

The Analysis of Variance Based on Y=$Y_{T_{Dec}}$

We will see now what an analysis based on a response variable computed from $T_{Decision}$ tells us about the relation of $T_{Decision}$ to the levels of the emulation variables. A correlation would mean that changing the emulation variable in some way makes results in subjects taking longer to decide; by extension, we might reason that this implies a more difficult decision. Note that the extension would not be directly supported by the data, but a conjecture based upon it.

The response variable used in the analysis of variance for this case was simply

$$Y_{T_{Dec}} = \frac{T_{Decision}}{\max(T_{Decision})|_{\text{all sessions}}},$$

that is, $T_{Decision}$ for the trial normalized by the maximum $T_{Decision}$ used in any trial in any session in any family — which happens to be 127 seconds for this data. Note that a linear normalization has no effect on the significance test, as long as the same digits of precision are retained; it does, however, help to keep the sums of squares in the ANOVA down to a reasonable size, which makes for nicer tables.

Refer to Appendix Tables G.7 for details of the ANOVA performed using $Y_{T_{Dec}}$; Figure 9-37 displays graphically the results of those significance effects.

The first step is to compare the significance results due to $Y_{T_{Dec}}$ with those of other, choice (hit/miss)-based response variables; see Figure 9-19, which displays the same type of data although arranged in a somewhat different format. There are some differences, but the similarities are stronger:

Blocks: Significant for all families, the same as for all choice-based response variables.

VE Expression: A distinct difference in patterns of significance between $Y_{T_{Dec}}$ and the other response variables. $Y_{T_{Dec}}$ results in only Toggles significant (at 1%); the others find Masses strongly significant, and in some cases Toggles mildly significant.

Damping: Reasonable agreement — general ambiguity for all response variables. Only Sliders come out even mildly significant for $Y_{T_{Dec}}$; Masses and Toggles are usually mildly significant as well for the choice-based response variables.

Sampling Rate: Good agreement — Toggles only are strongly significant.

Targets: Limited agreement — all find Toggles and Sliders strongly significant, but $Y_{T_{Dec}}$ finds Masses non-significant, countering the results from the choice-based response variables.

There is definitely a correlation between $Y_{T_{Dec}}$ and the other response variables, which supports the earlier correlation between $T_{Decision}$ and correctness of answer since large values in other response variables result from erroneous identifications.

Summary of Hypothesis 2 Argument

Based on this dataset, we may claim that errors take longer on average when a subject is allowed unconstrained choices. Proceeding from this observation to using $T_{Decision}$ as a measure of emulation quality is purely conjectural, but the data here indicate that it may be a worthwhile direction to pursue.

9.10 Subject Debriefing

Subjects were asked some questions following the end of each experimental session. This section summarizes and comments on those verbal subject responses. The letters 'NA' in the following tables means 'data not available'; some questions were added to the questionnaire after some sessions had already been run.

Question 1

What percentage of the comparison trials do you think you guessed correctly?

Objective: test hypothesis that the subject's confidence in his/her performance (as indicated by a high anticipated score) is an indicator of emulation quality. This question was asked before the subject had been informed of his/her session score.

Subject	Masses		Sliders		Toggles	
	Predicted	Actual	Predicted	Actual	Predicted	Actual[6]
Josephine	—	—	80	95.8	70,80	83.4, 83.0
Patty	86	60.4	70	83.3	50	64.6
Mike	NA	79.2	95	97.8	70	75.0
Sam	NA	75.0	50	81.3	60	70.8
Lorna	NA, 70	75.0, 72.9	—	—	65	60.4
Hank	NA	64.6	80, 80	79.2, 56.3	—	—

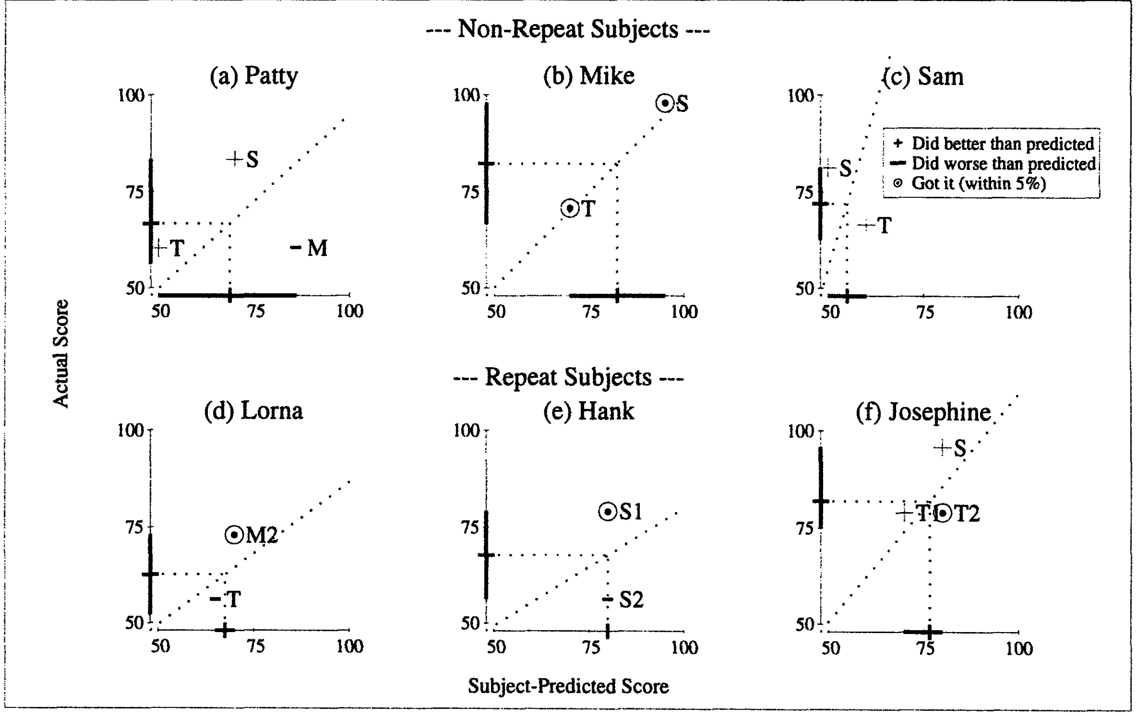


Figure 9-38: Subject score: self-predicted vs. actual. The dotted diagonal line goes through the intersection of mean self-predicted and mean actual scores, and the slope thus indicates the degree of over- or under-confidence of this subject. A bull's eye always lies within 5% of a 45° line; a '+' will always be above it, and a '-' below it. No Masses self-predicted score was available for Mike or Sam.

The tabular data and Figure 9-38 indicate that there is within-subject consistency in degree of over- and under-confidence in performance (compare slopes of dotted lines). Since there are many more "misses" in prediction (+'s or -'s) than "hits" (bulls-eyes), it would seem that self-confidence is not an excellent indicator of performance.

Neither is there a visibly obvious pattern in predicted score of the various device families. Toggles is often leftmost (lowest score expected) but not always.

Question 2

On average, how hard was it to decide (use scale of (1–5) where 1=obvious, 5=random guess) ?

Subject	Masses	Sliders	Toggles
Josephine	—	2	3-3.5, 2
Patty	2	2	3
Mike	NA	1	2
Sam	2-3	4	3
Lorna	2, 2	—	3
Hank	3	3, 3	—
Family Mean	2.3	2.5	2.67

The family average data in the bottom row suggest a small increasing trend in perceived difficulty, with Masses easiest and Toggles harder; this in spite of the fact that most subjects were most experienced by the time they did Toggles and more likely to have a discrimination strategy worked out. The most variability in subject perception of difficulty was for Sliders; one subject found them very hard, others found them easier than the other families.

Whatever the trend, the most important observation is that the difference in perceived difficulty was quite small.

Question 3

Was it generally easy to narrow the choices down to two?

Subject	Masses	Sliders	Toggles
Josephine	—	yes	yes, yes
Patty	yes	yes	yes
Mike	NA	yes	yes
Sam	yes	NO	yes
Lorna	yes, yes	—	NO
Hank	yes	yes, yes	—

Only two subjects found that they frequently could not narrow the choice down easily to two choices. Both of these subjects said the choice was always between three devices or obviously one; in the latter case, the choice was the highest impedance device.

Questions 4–5

Did you notice a variation in the emulation "quality" (the degree to which the emulation felt real) ?

When the quality was not so good, was it harder to make a choice?

Subject	Masses		Sliders		Toggles	
	variation	↑difficult	variation	↑difficult	variation	↑difficult
Josephine	—	—	yes	yes	yes, yes	NO, NO
Patty	yes	yes	yes	yes	yes	NO
Mike	NA	yes	NO	...	yes	yes
Sam	yes	yes	yes	NO	yes	NO
Lorna	yes, yes	yes, yes	—	—	yes	yes
Hank	yes	yes	yes, yes	yes, yes	—	—

Most subjects detected nonuniformity in emulation quality for all device families; the exception was a single subject under Sliders. However, there was more variation in subject reaction to poor quality. Most often in Masses and Sliders, poor quality corresponded to more difficult discriminations in the subject's opinion. But for Toggles, 4/6 subjects felt that degraded quality did not make their decisions more difficult.

Questions 6–7

Was the experiment tiring, either mentally or physically?
Would it have helped to take more of a break?

All subjects said the experiment was not tiring, and it would not have changed their performance to shorten it. However, several subjects commented on feeling bored, particularly in their second and third sessions.

Questions 8–9

At some point, did you find it becoming easier to make choices?
When was this point (approximately)?

Subject	Masses		Sliders		Toggles	
	easier	when?	easier	when?	easier	when?
Josephine	—	—	yes	2/4	no, no	..., ...
Patty	no	...	no	...	no	...
Mike	NA	NA	yes	1/4	no	...
Sam	yes	3/4	no	...	no	...
Lorna	yes, no	2/4, ...	—	—	no	...
Hank	yes	3/4	no, no	..., ...	—	—

For Masses and Sliders, there was substantial variation in subject's ability to and rate of settling into a strategy which permitted them to make decisions more easily. For Toggles, all subjects were consistent in finding the discriminations the same degree of difficulty throughout. We saw from Question 2 responses that this degree was generally higher than for Masses and Sliders.

Question 10–11

At some point, did you find yourself becoming less careful in making your choices (it's okay — this is a long experiment.) ?
When was this point (approximately) ?

Subject	**Masses**		**Sliders**		**Toggles**	
	careless	when?	careless	when?	careless	when?
Josephine	—	—	no	...	no, no	..., ...
Patty	YES	3/4	no	...	no	...
Mike	NA	NA	no	...	no	...
Sam	YES	4/4	no	...	no	...
Lorna	no, no	..., ...	—	—	no	...
Hank	YES	3/4	no, no	..., ...	—	—

3/5 subjects found they slipped in concentration during their first session (Masses). No subjects acknowledged such slip in later sessions. This might be interpreted as improved strategy, a tendency to move towards decisions based on gut reflex (which would require less concentration), or perhaps simply that in later sessions, concentration was at a generally lower level than in the first (novel) session due to boredom.

Question 12

Do you have any comments or suggestions regarding the setup, or how you felt during the experiment?

Masses

Patty:

- Most discriminations were easy; a few were very difficult.
- There was more variation in quality in the lighter Mass emulations.
- The reason it was harder to make a choice in low-quality emulations was a reduced confidence in the choice.
- Zoned out a little bit in the third quarter of the session.
- Keyed on sound of real devices: bearings felt a little different for the different masses.
- The emulator handle seemed a little different than the real devices (perhaps because of the slight roll in the real device handles).
- On low-quality emulations, tended to choose a greater real mass. This is because vibrations feel like additional mass (forces feel slightly higher).

Mike: not available (questionnaire added after this session).

Sam:

- Found decisions very difficult in beginning (rated a 5/5); got very easy towards end (1/5). Was less careful but more confident at end.
- Quality was better for heavy masses; the lighter 'twitchy' masses felt qualitatively different from the real devices.
- Light devices were hard to discriminate, although got better as session progressed. At first, felt distracted by the light emulations having less friction,[7] then learned to "filter out" the friction.

---

[7] I believe the reference to "friction" is observation that there was more friction in the real devices than in emulation.

- Strategy was to toss the device back and forth.[8]

Lorna:

- Requested to offset devices so more hand-rest room for right hand.[9]
- Figured out a strategy by the 2nd quarter.
- Found low quality in emulations distracting.
- In second session, felt less confident than in first.
- In second session, felt "trained" from the first in that she already had a strategy worked out.

Hank:

- Experiment was a little boring towards the end.
- Towards end, focused more on "gut instinct".
- Weights show through the slots if you look closely *(He mentioned this at the beginning of the session, and I requested that he try not to look.)*
- Says he is cueing on the pitch (sound) of the real devices; the bearings sound a little different for different masses.

Sliders

Josephine:

- Worked out a strategy in 2nd session, and became more efficient but not less careful after that.
- Found it was easier to make discrimination when closed eyes.

Patty:

- Higher impedances felt higher quality. It was harder to make choices in low quality cases because the emulations/real devices felt qualitatively different.
- Emulator makes a sound which is distracting.

Mike:

- Felt the emulations were of 'uniform quality', except for the very low-impedance emulation; there was a little stiction in the real device, so the emulation didn't quite correspond.
- Found a strategy after about 3 trials.
- Felt there was generally a better correspondence than with Masses.

Sam:

- Quality was uniform, except for a couple of trials.

---

[8] Note that (a) tossing was also my strategy, but not the method used by most subjects; and (b) Mike scored the highest on Masses. Suggests that due to using the same strategy, he perceived the same thing I did and this showed up in his scores.

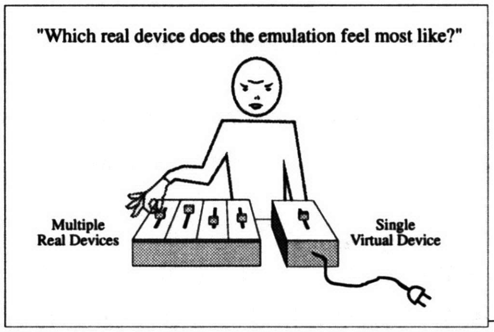
[9] Prior to this session, the second one run, I had arranged the four real devices in the four switchbox slots closest to the emulator. Lorna found this did not give her enough room to rest her right hand in order to manipulate the device in slot 1 (closest to emulator). On her suggestion, I moved all the devices into slots 2–5 and placed a blank plate in slot 1; and used this arrangement for all succeeding sessions.

Lorna:

- – Noise and jitter interfered more with choice in Toggles than in other families, because here are looking at features rather than background.
- – Some emulations felt very much like the real devices.
- – Found was using a strategy of "integrating" many manipulations of both the real and virtual switches in order to become aware of small differences.
- – Could usually eliminate 1 real device, but was then difficult to discriminate between remaining 3.
- – Found it easier to make choices after about halfway, but confidence level remained the same.
- – Judged "quality" by clickiness and force level.

Chapter 10

Discussion



Controlled Variables	Performance Indices
• Real device targeted by emulation • Tested Emulation Variables: 1. Sampling rate (300, 600, 1200 Hz) 2. Physical damping (0.001, 0.07 N-s/mm) 3. Virtual environment expression (impedance, admittance)	• Score (% correct) • Factor significance (ANOVA) • Factor effect (effects analysis) • Decision time

Figure 10-1: Experiments were based on comparisons between a set of real devices and a single emulated device which targeted one of the real ones.

This chapter focuses on three different areas:

10.1 Experiment Results — a data-rooted resolution of the questions which inspired the experiments or arose as the experiments were designed and executed.

1. Experiment effects: sensitivity of subject discriminations to emulation variables and targets
2. The shape of subjects' haptic perceptions: Weber's Law
3. Subjects: idiosyncrasies and trends
4. Decision time as an indicator

10.2 Emulations — issues relating to the design and construction of the emulation system which made the experiments possible, as well as to the process of choosing, creating and targeting the real device set.

1. Adequacy of this emulation system
2. Modeling and emulation of real devices
3. Producing real devices of specified impedance
4. Automatic haptic identification technique

10.3 Critique of Experiment Approach — issues concerning the validity and usefulness of the decisions made in designing, administering and analyzing the experiment itself.

1. Experiment model: was the Fixed Effects model adequate?
2. Quality vs. function: the dichotomy and its rooting in the comparison model
3. Role of calibration trials and duplication
4. Impact of targets: type and distribution
5. Administration: did the protocol accomplish what was intended?
6. Analysis: which aspects yielded the most useful insights?

Review of Basis of Experiment Comparisons

Recall the experiment paradigm (Figure 10-1): in a single trial, the subject compares a single emulated device to a set of real ones. One of the real devices has been targeted by the emulation, and from trial to trial both the target and the emulation variables levels are varied. The response is the difference between the chosen device's index [1:4] and the targeted device's index, also [1:4]. Thus the simplest response variable is

$$Y_{unweighted} = \text{index}(\textit{target device}) - \text{index}(\textit{chosen device}).$$

Other response variables are computed by various weightings of this simple response, described more fully in Methods, Section 8.2.7.

10.1 Experiment Results

10.1.1 ANOVA and Effects Analyses

This section aims to summarize and interpret the actual experiment results. To the extent the data and my own observation of the process permits, I will answer the following questions:

1. Which emulation variables are always important?
2. Which are important sometimes, and under what conditions?
3. Did the different variables act in concert or alone, and can this interaction be represented or modeled?
4. Which results were unexpected; and are these surprises artifacts of hardware and experiment design characteristics, or new information?
5. What are the most *useful* new insights attained as a result of conducting and analyzing these experiments?

Emulation Variables Which Always Mattered

All of the emulation variables were significant at least some of the time; however, only three of the tested factors were always significant at 10% or less. The strongest of the three are Subjects and Targets. These are not "emulation variables" in that they influence the emulation's output "feel", but they are a source of variation in (a) how the output will be perceived and (b) what the output is directed towards — task, if you will. If the goal is to create effective, versatile emulation, these factors are as crucial to understand as are the emulation variables. This design studied them explicitly.

The domination of the significance results by the experimental factors Subjects and Targets over the emulation factors implies that person-to-person variations and difficulty of discrimination are at least as important as any of the emulation variables.

Blocks (Subjects): Significant at 1% for every device family. The explicit study of individuals (summarized and discussed subsequently in Section 10.1.3) showed that in general individuals tended to act in similar, though not identical patterns. The greatest differences were in overall performance, probably due to attention rather than skill or innate ability — this conjecture is based on an observed correlation of high performance to long decision times.

Targets: Significant at 1% for every device family. Calibration trial results showed that not only were the low-impedance devices more closely spaced physically than the high-impedance devices (a deliberate outcome), subjects perceived them as being more closely spaced, contrary to Weber's Law and something of a surprise. Thus these low-impedance discriminations were harder.

In addition, Targets figured heavily in interactions with other effects. In those cases, the influence apparently arose from the variation in impedance level from low to high (different kinds of control problems result) rather than from the nonuniformity in spacing.

Damping: Mildly but consistently significant, 5-10% for every device family. Contrary to expected result, more damping was invariably slightly worse. This mild result actually disguises the whole story, which is that in the emulation creation stage, I found the higher damping levels harder to stabilize and had to adjust gains (at a loss of fidelity) accordingly. In fact, I chose these damping levels as the lowest pair which (a) could be achieved (lower level is minimum possible, no additional damping), and (b) gave a perceptible difference in output. My experience with higher damping levels was simply a continued and heightened degradation in stabilizability as the damping increased.

The reason for this poor performance, which counters theoretical and experimental findings in research by Colgate et al [38], is still uncertain. Although I examined the gross behavior of the damper I built and used here enough to feel it would be acceptable, it was not so straightforward to characterize the subtleties in its behavior and thus this was never done. However, I do know that the damper diverged slightly from the ideal of linear pure viscous damping due to the presence of small amounts of shaft compliance, seal stiction and nonlinear fluid viscosity; and the most likely explanation for the results is that this divergence, which I originally felt was small enough to be acceptable, was in fact enough to do more damage to emulation controllability than it helped. In any case, there is sufficient uncertainty in the matter that this result does not pose a challenge to the strong positive results for damping that others have found.

Emulation Variables Which Sometimes Mattered, and When

We can safely say that all of the variables studied are important, based on ANOVA significance results. Following is a discussion of the distinguishing features of the conditionally-significant factors:

VE Expression: Admittance / Impedance expression of the virtual environment model exerts its influence primarily through interactions with other variables, rather than by itself. As a main effect, it is highly significant (1%) for Masses; otherwise not at all. Further, the results (effects analysis) suggest that this was a case of a targeting bias rather than superior performance of Admittance expression in emulation of inertial environments over Impedance — Y expression generally resulted in positive or zero errors for Masses, Z in negative or zero errors, with about the same *rate* of errors for each controller.

In interactions, VE Expression is highly significant in Damping and Targets for Masses, Sample rate and Targets for Toggles. It even appears in a 1% significant 3-way interaction with Targets and Sample Rate for Toggles.

Despite similarities in end result the Toggle emulations were much harder to stabilize in general, particularly when a high-impedance discontinuous target was combined with lowered sampling rates; this was situation in the unstabilizable Toggle trials. The position-control servo is almost certainly the primary culprit, asking for large jumps in force (the Toggle model was defined in terms of force discontinuities) to be enforced through discontinuous jumps in position, which is not physically realizable with an actuator of finite mass. The dependence on sample rate possibly implicates the admittance integration step as well.

Overall, this Toggles experiential result indicates that perhaps a quite different model for this type of discontinuity would work better in Admittance expression; on the other hand, a better Impedance-expressed model might be a more productive direction to explore.

For all other combinations of variables, different strategies were found for each VE expression which apparently worked (or didn't) about equally well in each case. That is, the emulations based on each expression had a distinct feel to them, but those feels were about equally successful in allowing positive discriminations. The two expressions resulted in emulations which were about equally functional (recognizable), but nothing can be said about their respective qualities.

Sample Rate: Like VE Expression, changing Sample rate did not, with the exception of Toggles, result directly in a reduced functionality even though (a) for this quantitative variable, reduction of sample rate is indisputably a degradation to the extent it has any effect at all (not so with the qualitative variable VE Expression); and (b) reducing sample rate below 500 Hz and sometimes to 600 Hz resulted in emulations which felt different (in a bad way) from higher-rate emulations. That is, they were perceptibly degraded but this did not impact on emulation *functionality*. However, Sample Rate showed up in interactions, with VE Expression and Targets (Toggles) and with Damping (Sliders). The Toggle interactions are the same ones discussed for VE Expression above. The Damping interactions are probably a result of the fact that higher-damping Slider emulations were somewhat harder to stabilize (for all VE expressions), and this tendency caused more problems at low sampling rates.

It is also worthwhile to mention the situations which tend to correlate with variation in the emulation variables:

Importance of Interactions : Factors acting in concert was the norm rather than the exception; most critically, nearly all the emulation variables interacted strongly with Targets. This subject will be explored more fully in Section 10.1.1.

Dependence on Device Family : Different patterns of significance characterize the different device families, implying that there will not be hard and fast rules — the ideal characteristics of the haptic display will depend on the type of environment being emulated.

Some sources of variation among device families are:

Order of environment: compliance-, damping- or inertial-dominated.

The environment order influences which sensors are used (e.g. acceleration for a Z mass, but only position for a Z spring); and bandwidth constraints mean different things at different orders because more integration (a source of phase delay) is required in some cases. The degree to which the virtual environment is similar/dissimilar to the physical characteristics of the emulation system also varies — it is hardest to make the emulator feel like something extremely different than it is naturally. Examples for this are very high stiffness virtual environments, or very low inertia or damping environments.

Complexity of environment being emulated: Number of model parameters required to describe and/or nature of nonlinearity (these characteristics were confounded here because Toggles was the only environment which was both multiparametered and nonlinear)

More complex environments (Toggles, highly nonlinear with several model parameters characterizing each region) were generally more sensitive to more emulation variables than single-parameter, continuous linear systems.

Sample rate is important with nonlinear, discontinuous environments — this is a comforting result to see. Discretization should matter in those circumstances.

Influence of Untested But "Controlled" Emulation Variables : These are the emulation variables which were not in the list tested (comprised of Sampling Rate, VE Expression and Damping), and as such I felt constrained to hold them constant under certain conditions. The problem with this was that strong interactions clearly exist between some untested emulation variables and those tested; and thus the "constant" level used in the experimentation did not permit best possible performance for another emulation variable. This will in general be a serious issue in the study of a complex system such as this, where only a small subset of the many variation sources may be tested at any one time and interactions are ubiquitous and important.

The most glaring example here was the interaction between servo controller gains and Sampling Rate. The choice to hold gains constant for all sampling rates while they were varied in some other cases, as well as the details of this interaction (observed during emulation development as opposed to experiment analysis) were detailed in Section 5.3; and will be further discussed in Section 10.2.2.

Subject : A meticulous subject could often perform significantly better than an apparently careless one (the characterization is based on both my own observation and confirmation with variation in average decision times). This implies that the information content was usually present in the emulation, but in some cases required more effort to extract than others (more on this later, in Section 10.1.3).

Emulation Variable Interactions and Modeling of Effects

Emulation variables acting together to influence subjects' ability to make positive identifications were a dominant feature of the experiment results, to an even greater extent than was expected. The overall impression one takes away from an examination of the data, as well as from time spent working with the emulation system under regularly altered emulation variable conditions, is that changing one variable often causes the whole emulation to shift in some manner which is perhaps explainable after the fact but not always easily predicted.

Some of the interactions which the data proved to be important were the following:

Interaction of emulation variables with Targets: The variables which interacted most strongly with Targets were often most critical in high-fidelity (subtle discrimination) emulations, since the Target effect generally consisted of more errors at low impedance (closely spaced) devices.

This was not always the case — for Toggles, the most difficulties happened at high Target (high impedance), where the Admittance emulations were sometimes unstabilizable for low sample rates and all of the emulations were apparently over-targeted. In this situation, the Target effect was not of discrimination difficulty but of pushing the emulation system's achievable impedance limits.

VE Expression with Sample Rate and Target: Admittance combined with low sampling in high-impedance Toggle emulations was excessively active and in same cases unstabilizable in Toggles.

Surprises and Important New Insights in the Data Analysis

1. The Quality versus Function concept. Emulations which were obviously of degraded quality permitted rates of accurate discrimination similar to those of undegraded emulations; i.e., their functionality was uncompromised.

2. Weakness of Weber Law behavior in perception. Subject perception of real device impedance distributions varied between linear and logarithmic, with a distortion from linear that held an arguable relation to the uniformity and resolution of real device impedance distributions (Section 10.1.2).

3. Sample Rate / Fidelity (gain) interdependence, showing up as the minimal experimental effect of Sample Rate. It was learned through emulation development that high gains (for high fidelity) required high sampling rates; to test low sampling rates while holding gains constant meant using soft gains for high sampling rates. Thus fidelity as a function of sampling rate was uniformly sub-maximum at all sampling rates, due to interaction with an untested emulation variable.

   This result highlights the general ambiguity of interpretation that arises in testing a subset of variables in a complex emulation system which is characterized by strong interactions between many different factors.

4. The limited and indirect nature of VE Expression influence. VE Expression is significant only as a main effect for Masses, a simple environment in which modeling was not at issue, Experimental performance of Admittance and Impedance was virtually indistinguishable where both environments could be stabilized. Qualitative differences might have shown up in more Quality-oriented test but did not impact on discrimination effectiveness.

5. Prominent *interactions* of VE Expression with other factors. The main difference between expressions emerged as degraded admittance expression at low sampling rate and high impedance. It was expected that admittance, relying on a position rather than force servo, would have trouble with force fidelity in high-impedance environments. What was not predicted was the difficulty in simply stabilizing a high-impedance Y environment filled with harsh desired-force discontinuities.

6. The possibility of consistent discrepancies between perception and measurement of real impedances, suggesting there is an unmodeled difference between the two perspectives which may be quantifiable and predictable (Section 10.2.2).

7. Polarity of Damping effect, which counters results from other investigators. It was possibly due to problems with damper, whose true linear viscous damping was corrupted with some compliance and hysteresis due to stick-slip and fluid leakage.

10.1.2 Perceived Impedance Distributions: Weber's Law Upheld?

The correspondence between measured real device impedances and subject's perceived impedance distributions suggests that subjects were perceiving on only a weakly log (Weber) scale; in some cases closer to the (linear) measured scale. This might be argued as a case where Weber's Law perception does not hold. However, such a statement must be qualified by the following observations on the perceived distributions and the method of obtaining them:

- There is a possible distortion of the curves due to the method of obtaining perceived impedance distributions (scale ends pegged to members in the set), which was deliberately oriented towards ascertaining relative differences rather than absolute magnitudes.

- The range of tested real-device impedances might have contributed the shape of perceived distributions; perhaps, like the small-angle theorem, Weber's Law is accurate only as an approximation which does not hold for large (order-of-magnitude) jumps in impedance. In support of this argument is the observation that the least-Weber-like distribution is for the family whose impedance covered the largest range (Sliders, 32:1), the most Weber-like for the family with the smallest impedance range (Masses, 5:1).

- Subjects vary in their Weber Law behavior. The above statements are made with respect to averages, but in fact individual variation exceeds the average difference between families.

- Subjects are distinguished by cross-family consistencies in behavior, confirming that individuality is a strong factor. It also implies that subjects may be keying on different environment features or have slightly different internal scales.

- Duplicates were not always perceived as identical: this is an indication of either inconsistency in subject perception or lack of haptic correspondence between "duplicate" real devices. There was considerable mismatch in two (out of twelve) device duplicates (Section 9-8); however, for the most part the data suggests that subjects could not reliably distinguish the two members of a duplicate pair. Thus it is reasonable to attribute most inconsistency to subject unrepeatability.

10.1.3 Subjects

The starting hypothesis, that

> Variability within subjects is less than variability between subjects,

was supported to some degree by all aspects of the data.

There are also, however, strong consistencies across subjects, indicating that generalizations can be made from studying a representative sample in spite of the presence of strong subject differences.

Idiosyncrasies: The following summarizes the evidence indicating that subjects were different.

- For complex haptic environments, perceived distributions varied enormously. In these cases, subjects could choose what to cue on, and the evidence is that in such a situation they choose different things.
- N=2 (repeat subjects only) score means were often outside, although barely, the first standard deviation of the N=6 scores. This was despite the fact that the 1/3 of the N=6 score consisted of the N=2 data.
- Variability in repeat scores was relatively small. Scores for the repeat subjects were much closer to each other than to those for other subjects (with the exception of Sliders, Hank).
- Scores for the three non-repeat subjects, representing each of the three device families, were characteristic of both family and subject — the same by-family order was repeated within each subject, but at an offset representative of the subject.
- Some (although not all) subjects showed characteristic patterns in their response to different levels of different emulation variables, suggesting that in some cases subjects were using different strategies or noticing different things.

Trends: The following summarizes the evidence indicating that individuals were the same.

- For simple haptic environments, there was good consistency in subjects' perceived impedance distributions for the real device set. This suggests that when people are given little choice in what to cue on and the virtual environment is qualitatively similar to the real one (i.e., few tricks are required to get a semblance of either correspondence or stability), people make the same judgements.

  Taken a step further, this suggests that peoples' perceptual apparatus is telling everyone about the same thing. This is in agreement with the findings of others who have studied perception of simple environments ([182, 190, 188]).
- The Analyses of Variance performed for the N=2 and N=6 populations, respectively, showed similar patterns of significance, the primary differences being for Toggles. Since Toggles were the only haptically complex family tested — meaning in this context that the subject had a choice of what to react to — it is unsurprising but important to confirm that subjects showed the most variation in their reactions to this set of devices.
- Scores for the three non-repeat subjects showed a strong and consistent family dependence. That is, all subjects compared had the same relative difficulty in distinguishing the types of environments tested, finding Toggles the most difficult and Sliders the easiest.

- Inattention was probably an important source of subject variation, in terms of comparisons between effects and scores. This conclusion is made through observing that despite cross-family score consistency of a given subject, the same subject was often (but not always) inconsistent in patterns of effects. It seems that those subjects didn't necessarily have a set strategy or perceptual offset which was more or less effective than another subject, and inattention is the only other likely explanation.

Interpretation

The data suggest that people have similar low-level perceptual abilities and tendencies, but differentiation occurs when there is opportunity for choice in (a) overall attention allocated to this task, as opposed to looking out the window or thinking about lunch; (b) how the task attention budget is allocated to various features in the environment. Thus little attention might be required for simple, single-parametered environments, and that attention can be exclusively allocated to the single feature or characteristic which dominates the environment. For complex environments, more attention is required to begin with; and then the subject gets to choose what to notice. When can we assume that individuals will react in similar ways, and when would such an assumption be unwise?

Conditions under which it is reasonable to assume homogeneity of individual response:

1. Emulation of simple systems, described by a single parameter and emulated with a minimum of non-physically based haptic "tricks".

2. High-quality emulations, possessing a strong qualitative resemblance to a real environment or simply passive, realistic virtual environments when no comparisons need be made, make it more likely that subjects will react in the same way they would react to the corresponding real environment.

   To the extent that reactions are homogeneous for the corresponding real environments, that homogeneity is likely to hold here. However, this statement can only be loosely supported by the data, since quality and complexity were confounded in the emulations tested here: simple systems are easier to emulate at high quality.

Conditions under which we should presume that individuals will react differently until the reverse has been proven:

1. Emulations of complex haptic environments. A rule of thumb: if it's tricky to emulate, it's quite likely that people will interpret it differently.

2. Low-quality emulations, characterized by distracting control artifacts; for example, activeness on contact with virtual discontinuities, jitter in high-gain, low-impedance regions, or slip-through in rapid manipulation through small or subtle features.

10.1.4 Decision Time as an Indicator

Hypothesis 1: Subject $T_{D_{avg}}$ as Indicator of Performance

In the case of simple haptic environments (here, Masses and Sliders) there is a positive correlation between decision time and score: longer decision times resulted in more accurate decisions on a session, rather than trial-by-trial, basis. That is, given a set task (in this case, discriminations for a given device family) with a given average decision time for all sessions, the session with the longest average decision time was also the session with highest score (average performance). This statistic is thus an attribute of the session and as it turns out, of the individual as well:

A slow *session* is a low-error session, suggesting a correspondence between score and attention.

The exception to this observation is Toggles, the most complex haptic environment tested here. While there was a similar degree of variation in both decision times and scores for this family and the other families, there was not a strong correlation; the raw data combined with Target effect shows shortest decision times for A and D, and also the lowest error rates for those devices, and therefore by family (not by subject) shorter decision times were correlated with correct responses. However, this is not a strong result.

The most likely reconciliation of these conflicting observations is that the slower subjects were being more careful, which in turn implies something much larger: that given enough time and attention on the part of the subject, *the correct answer is available.* That is, the information about the real target is present in the emulation, as long as the subject makes the effort to extract it; sometimes it is harder to extract, and there the less conscientious subjects (who are probably more representative of eventual users of this technology) often err. For complex environments, or those in which the targeting is suspect and by extension, the information content not present, taking more time does not improve the result.

This is a very different interpretation than the converse conclusion that under certain conditions, and much of the time, the emulation was simply "wrong" and misleading. While it is evident that this situation was occasionally the case (e.g. for the high-impedance Toggle targets) the data suggest it was not the norm.

Hypothesis 2: $T_{Decision}$ as Indicator of Emulation Quality and Decision Difficulty

The data suggests that on a trial-by-trial basis, errors usually take longer.

That is, given a single session with a given average performance and average $T_{Decision}$, for any given trial a slow decision will more often be associated with an error than with a positive identification. Note that $T_{Decision}$ differs from the average time taken for all decision in the session, which in the previous section we noted tended to be longer for sessions with higher average performances.

A statistic which counters this hypothesis is that there was little difference in average decision time across different families, even though subjects' verbal responses indicated that Toggles generally a posed a more difficult discrimination task than the others. In fact, there was more correlation between subjects and $T_{D_{avg}}$ than there was between device families and $T_{D_{avg}}$.

However, this statistic could be interpreted in different ways. Recall that within each family, there were subtasks of varying difficulty (discriminations between closely spaced devices versus those between widely spaced devices); and there were some of each in each family. Further, $T_{D_{avg}}$'s for Toggles were brought down for every subject by the "obvious" (though very low quality) Device D targets. Even though the other discriminations for Toggles might or might not have taken longer on average, the whole-session averages were not particularly high.

The supporting correlations, which seem stronger than the nonsupporting data, might result from at least two opposite causal relations which are not necessarily mutually exclusive:

1. Mistaken identifications are a consequence of harder discriminations, which naturally take longer; in which case, $T_{Decision}$ might plausibly be an indicator of decision difficulty. It is not necessarily an inverse indicator of emulation quality, since we saw that apparently some of the most difficult discriminations were due to close spacing of real devices rather than poor emulation quality or emulation targeting.

2. The longer you take, the more likely you are to make a mistake — the "agony" factor. If valid, this might mean that it makes sense to limit decision times, on the grounds that quick reactions

are more indicative of a subject's real perception.

It is difficult to formally validate either of these explanations without more data. However, reference to the Hypothesis 1 result — that from a whole-session view, decisions which are slower on average tend to result in higher average performance — suggests that the first explanation is more plausible. Agonizing tends to improve performance, at least for simple environments; thus the correlation noted here is most likely with difficulty of discrimination.

From the experimentalists' point of view, this might still mean that it makes sense to control decision time, but to do this by requiring longer decisions rather than limiting them. However, requiring slow decisions will not necessarily improve attention; in fact, it might have the reverse effect of exacerbating boredom. Even worse, long and painstaking decisions are a poor model for future applications of haptic devices — at least, we hope so. From the standpoint of modeling real-world situations, intuitive snap judgements through short decision times are likely to be the most informative.

Summary

I conjecture that, in the case of simple haptic environment comparisons, the results of the decision time investigation mean that

> Trial by trial, a slow decision means the decision was hard;
>
> on average for an individual, slow average decision time means the subject was being careful and this care showed as improved average performance.

But the only statement I can make for comparisons of haptically complex environments is that some subjects were more careful than others. Their performance was not necessarily improved as a result of the pains they took, because they might not have been using the same criteria for judgement that the emulations were based upon. It is quite likely that this result would change if learning were allowed; i.e., subjects were told when they had made a "correct" identification. In this case, it is possible that the careful subjects would show a higher rate of learning.

10.2 Emulations

In this section, I will discuss issues relating to the design and construction of the emulation system, as well as to the process of choosing, creating and targeting the real device set.

10.2.1 Adequacy of Emulation System

The following is meant as a quick summary; refer to Chapter 4 for details.

System strengths:

- Fine for continuous environments (Masses, Sliders).
- Excellent high-saturation force output.
- Accelerations adequate for many discontinuous environments (e.g. walls).
- Sensor noise levels acceptable.
- Bandwidth for remaining system elements fine (sampling, controller).
- Good range in achieved impedances for all device families.

System limitations:

- Motor inertia limited performance in discontinuous environments.
- A small but perhaps noticeable amount of compliance between the force sensor and actuator.
- Physical damper was non-ideal.
- In terms of testing these emulation variables, a lower minimum level of physical damping would have aided experiment comparisons.

10.2.2 Modeling and Emulation of Real Devices

In this section I comment on the success or failure of the modeling and parameterization techniques used in targeting the set of real devices used in this experiment. Slightly different modeling and parameterization methods were used for each family, although in all cases a great deal of iteration was involved. Thus, each family will be discussed separately below. In each case, I will (a) briefly review the targeting method used for each family (the procedures were covered more extensively in Chapter 6, Real Devices); then (b) discuss how this targeting influenced subject ability to make accurate discriminations.

When considering the targeting process, it is important to once again note the difficulty encountered in both creating real devices of the desired impedance and impedance distribution, and in accurately parameterizing the real devices (or finding real devices which could be easily parameterized). Much of the error in targeting was a direct consequence of inadequacies in the real device production and parameterization step, particularly in the Toggles family.

Characteristics of the target set and indications in the data which are relevant to the following discussion include:

- Impedance range covered by the real device set ($[Z_{\max} - Z_{\min}]$ as well as $\frac{Z_{\max}}{Z_{\min}}$).
- Shape of the impedance distribution, both measured and perceived by subjects; that is, the degree of distortion from a linear distribution, in the direction of logarithmic spacing.

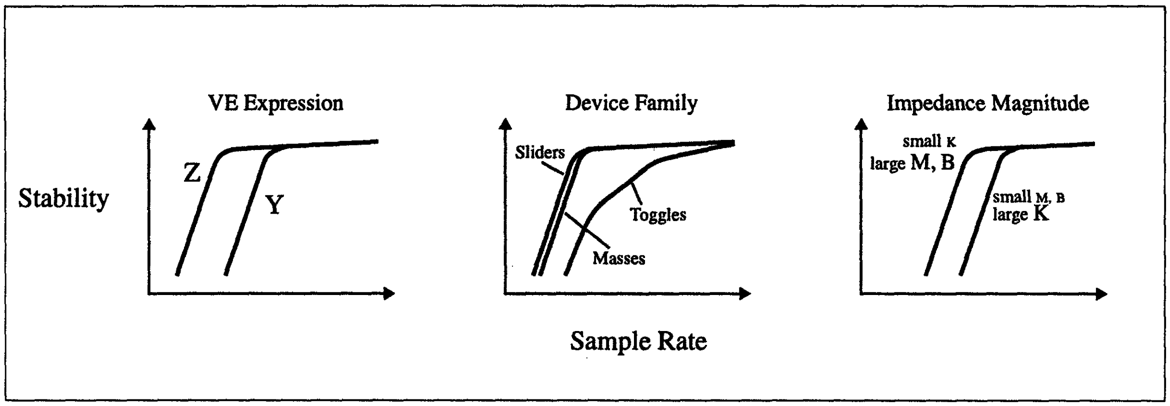


Figure 10-2: Influence of sample rate on emulation stability.

- Method of targeting used.
- Success of targeting, based on

  Scores: subjects' ability to make positive identifications.

  Target Effect: rate of errors as a function of which device is targeted.

  This indicates more specifically than overall score which real devices are problems, although the Target effect is also confounded with the fact that closely-spaced devices (low impedance here) are harder to discriminate regardless of quality of targeting.

  Other observed anomalies in data, such as the very low error rate for Toggles, device D; and the lack of significance of Sample Rate for Toggles, largely because gains had to be reduced, resulting in low-fidelity (i.e., poorly targeted) devices for all sample rates.

Novelty of Effort to Accurately Emulate Real *Environments* : To my knowledge, this is the first time that careful comparisons have been made between people's perceptions of real device impedances and of emulated impedances, when the real device impedances have been well-characterized and the emulated impedances are high fidelity (i.e., based on an accurate model and implemented with a verifiably low-error servo).

It will be interesting to see if observations such as that of the perceived mass offset, described immediately below, is repeatable with other emulation systems and haptic environments, or turns out to be an artifact of the experimental setup used here.

The Story about Sampling Rate that the Experiment Results Didn't Tell : Sample Rate did not figure as an important player in the experiment significance results; however, a great deal was learned about the importance of sampling rate during emulation development. These insights are displayed qualitatively in Figure 10-2. The point to take away from these sketches is not the exact location of a critical sample rate, which is an approximately measured characteristic of this emulation system, but the relation of the curves to one another. This relation is probably a more general result which will hold true in other emulation systems and tasks.

A relation between sample rate and fidelity such as that hypothesized in Figure 10-3 could be argued based on my observations as I created the emulations. Higher sampling permits higher gains and thus higher fidelity; the cost of stability at low sampling rate is sloppy gains and poor discrimination performance.

This was built into some of the emulation variables for certain families; most notably, VE Expression in Toggles. In order to stabilize Y Toggle emulations at higher Target impedances and low

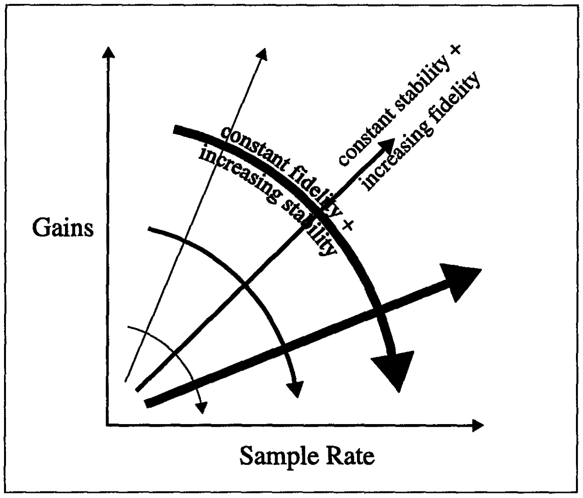


Figure 10-3: Interrelatedness of sampling rate, servo gains and emulation fidelity.

sample rates, I had to reduce gains; this in turn compromised the higher-sampling rate Y emulations of the same (higher-impedance) devices, and may have hurt their performance in the experiments.

Thus the overall factor (sampling rate) didn't show up as significant in the experiment results, but that is because another (untested) factor — servo gains — was held constant at a level that didn't optimize performance at all sampling rates.

Masses

Modeling Effort and Assumptions: With the assumption that the real devices were reasonably well described as a pure mass, the model was trivial. Only a single parameter identification was required, and that parameter was easily measured.

Method of Parameterization: The real device was weighed, and emulator sensor calibration verified.

Comments: This should have been the most accurate method, and aside from the observed mass offset, it did seem to be a reasonably well-modeled device in that subjects more or less agreed on the distributions. If the subjective trial and error parameterization approach used for Sliders had been used here as well, it is possible that agreement between subject perception and "measurement" would have been as good as for Sliders.

Perceived Mass Offset : The single device set which I was able to parameterize accurately felt, in emulation, too heavy by a constant amount at all Z levels considered (details on the offset and possible causes in Section 5.3). For the purpose of obtaining useful data from the experiments I added 70 grams (measured offset magnitude) to all of the Mass emulations.

The experimental results did not confirm the actual degree, or even the existence, of the perceived mass offsets (see Results chapter, Section 9.6). A fairly consistent bias was seen in responses to Y and Z Mass emulations; subjects on average found the Y emulations lighter than the targets, and the Z emulations heavier than the targets. Further, they found the lowest impedance Mass (device A) emulation consistently heavier than the real device A.

Thus, the question remains open. The pilot-study offset finding was persuasive, and the sensors were well calibrated; but the correction did not seem to produce the expected uniformly accurate identification.

Sliders

Modeling Effort and Assumptions: With the assumption that the real devices were well described as pure viscous damping, the model was trivial. This assumption was probably a good one; there was a little friction, but it came across more as a texture than as a source of hysteresis.

Method of Parameterization: A trial and error approach comparing the emulation and the real devices was used, with matching based on my own and pilot subjects' perception.

Comments: The result was excellent. This was the best modeled device; subjects were in good agreement in their impedance distribution perceptions, and positive identifications were highest for all the families (two subjects approached 100% correct).

It was also the most widely spaced device, bottom to top (1:32), so each discrimination may not have been as difficult. However, the distribution was highly dished (logarithmic), meaning that the low-impedance devices were still quite close together and most of the range was taken up by the difference between Devices C and D. Subjects claimed about the same feeling of difficulty in choice as with the Masses, but got it right more often.

Toggles

Modeling Effort and Assumptions: This family was by far the hardest to model. My first attempt was a physically-based approach, but this proved too difficult to stabilize. It is therefore somewhat irrelevant to talk about assumptions, since correspondence to the real device's structure was thrown out at the start.

Method of Parameterization: The end result was largely a collection of emulation stabilization tricks, with a few physically measured parameterizations thrown in (geometric displacements and maximum forces); see Section 6.2 for details.

"Slipthrough" (a phenomena described in Section 5.3, wherein a finite force slew rate meant that sometimes model discontinuities could be passed through before they were felt) made the targeting hit-and-miss. An emulation was correct in impedance magnitude only for a certain speed of manipulation, and it proved impossible to completely enforce that speed during experiments. A further complication was encountered in that the real devices were hard to produce exactly to specification; thus the "target" moved.

Comments: The zero experiment error rate observed for Device D indicated a problem in targeting for that device. The combination of over-targeting (most emulations apparently felt harder than the real devices, at the speeds of manipulation most subjects eventually used) and the unintentionally wide impedance gap between the two highest impedance devices (C and D) made the correct response for some comparisons obvious, even though the emulation was quite low fidelity.

In addition, the high rates of confusion between Devices B and C confirmed the more direct (calibration data) observation that those family members were not accurately duplicated.

10.2.3 Producing Real Devices

Problems in producing and parameterizing the real device sets complicated interpretation of the experiment results, particularly for the Toggles family; and the degree to which satisfactory real device sets were created (Masses and in most cases, Sliders) came at a great deal of effort. The problems could conceivably have been avoided or at least minimized by use of different device sets; for more thoughts on that topic, see Section 10.3.4.

The problems with the real device production process used here are:

1. Testable device families had to be chosen from those which could be modified in an experimentally useful way, over an adequate range and with adequate resolution and accuracy.
2. The range of impedances in a given device family was in some cases limited to the range of modifications possible (e.g. Sliders were a problem because with the first method of modification, described in Section 6.1, I could get a range for to accommodate only three distinct set members. To get the fourth member (D), a different modification method was required).
3. It was difficult to hit a specific desired impedance, and the method of modification was often not highly repeatable.
4. It was also difficult to quantify the result, except perceptually ("Do they feel about the right difference from each other?"). In fact, the method finally used was actually to target the device in question with an emulation, iterate on the emulation until the model parameterization felt as much as possible like the real device (according to both myself and various informal subjects), and then use the emulation model parameterization as the impedance measurement of the real device. i.e., it was quite indirect.
5. It was difficult to achieve accurate duplications; this was particularly a problem with the middle Toggle devices (B and C).

Thoughts for Future Iterations : The complications inherent in real device creation is a general and serious shortcoming of this experiment design.

Some of the complications encountered in this iteration were a result of the particular characteristic which was modified here (impedance magnitude) while attempting to hold all other characteristics constant. Since the modified characteristic was quantitative, the magnitude specification came into play. Further, with four real devices in the set (ideally, even more would have been used) a large range of impedances was required. Again, this complicated matters.

If a qualitative characteristic (such as the use of coulomb friction versus viscous damping in the targeting of a single real slider switch, or the efficacy of various non physically based modeling strategies tried in the targeting of Toggles) were studied instead, there would be one less constraint to satisfy although the analysis would be quite different, and possibly a [Single Real:Multiple Virtual] comparison model would be more appropriate. On the other hand, it is not clear that varying a qualitative characteristic while holding constant impedance magnitude would have been any easier; and studying impedance magnitude was highly informative.

Converting this design to a random effects model where a different device set is used for each trial only makes the situation worse. A much more convenient means of device production will clearly be required. This will have to happen on an ad hoc basis, with a different design for each class of environments to be tested. Designs based on "dialing in" an impedance (or other characteristic) out of a continuous range will be the most successful; for example, a rotary inertia whose magnitude is modified by moving a weight in and out along a radial arm, changing its moment arm.[1] This approach could be further automated by instituting computer control of the real device variation; e.g., using a small motor to move the weight along the arm.

In summary, the situation would probably improve if any of the following experiment and/or real device modifications occur:

---

[1] This design could be made to work for the linear case also, by using a very large arm and making the point of contact the end of the radial arm instead of its axis of rotation. The motion would be approximately linear for small excursions.

- More versatile design of real device (dial-in based).
- Replacement of the [MR:SV] or [SR:MV] comparison model with a "paired comparisons" approach, or [Two Real:Single Virtual] (Section 8.1.4) to get away from the need for a whole set of devices which are properly distributed. For some implementations, only the distance between pairs of real devices would have to be precisely specified. In other cases, the absolute magnitude (location) of the pair might also have to be specified.
- Qualitative environment characteristics studied.

10.2.4 Automated Technique for Identifying Impedances of Real Environments

The following merely summarizes key points of Section 7.4. The "Haptic Camera" identification technique was developed to address the need to objectively and conveniently record and play back the haptics of a real environment. Its algorithm was based on the partitioning a model of the real environment impedance into components (e.g. mass, compliance, damping and nonlinear characteristics) and piecewise-continuous geometrical regions, and then parameterizing successively higher-order model components with a computer-controlled force probe.

10.2.5 Applications

There are numerous potential applications for an automated haptic record/playback utility such as this. Section 7.4 details a few; in summary, they are:

- emulation of real environments;
- improving many types of virtual environments;
- development of mapping between virtual and real environments;
- production quality control;
- psychophysics research.

10.3 Validation of Experiment Approach

The most important question asked in this thesis was not the specifics of how sampling rate, damping and VE expression impact on an emulation's fidelity, but rather

> *What are the components of a high fidelity emulation, and how can they be characterized?*

Likewise, the "results" only begin with the experiment data which was discussed in the Results chapter and earlier in this one. The general question addressed here is:

> *Did this approach for experimentation work or didn't it, and how could it be made to work better?*

The short answer is Yes:

> it worked in that carrying out the experiment and carefully analyzing both the data and the process afforded useful insights about the emulation variables chosen to study here;

and, Yes:

> it also worked in that we have learned many ways in which the characterization process could be improved, and for this knowledge a first iteration was necessary.

In the following, I critique the various experimentation, hardware (both virtual and real devices), emulation and analysis decisions that were made throughout the process of designing and carrying out these experiments. I discuss the major strengths and weaknesses of the approach and consider whether and how the weaknesses are reparable.

10.3.1 Fixed Effects Experiment Model

Was the Fixed Effects[2] experiment model adequate and/or the best possible model to have used here?

Although the [Multiple Real:Single Virtual] (MR:SV) comparison mode worked well with the exception of real device production difficulty, the experiments exposed a need for either a modification of the fixed effects model or replacement altogether with another experimental model such as random effects.

The problem was the small size (four members) and discreteness of the real device set. In any trial, there was a limited number of responses that the subject was likely to make: the target, or one of the near neighbors of the target. This assumes that only next-neighbor errors are made, which for this experiment was always true.

This had two important implications:

1. The response were biased. This was because (a) an end member had only one next neighbor and therefore only two possible responses, whereas a middle member had two neighbors and three possible responses; (b) in half of the comparison trials, an end member was targeted, and in the other half a middle member was targeted.
2. Given that in the fixed effects design the targets are maintained in a fixed relation to one another in terms of relative impedance, the discrete target spacing led to a response which is not very information-dense, although it is unambiguous in interpretation.

---

[2] A "fixed effects" experiment model, discussed in Section 8.1.4 and in any experiment design text, refers to the use of a fixed set of experiment factor levels rather than choosing factor levels at random for each trial (a "random effects" model). Here, the factor which caused problems was Targets, fixed at $N_{targets}$ real devices.

Modification of Fixed Effects Model :

Some of the possibilities for improving the situation without throwing out the [MR:SV] comparison mode would be to:

1. Employ a larger real device set, which would reduce the percentage of biased trials. It would not, however, eliminate them altogether and would incidentally necessitate huge designs if the full-factorial approach was to be maintained
2. Include a "dummy" real device at each end of the scale which is never targeted (but the subject does not know this).
3. Use a richer but more ambiguous response variable, e.g. a magnitude estimation of the "distance" between the emulation and each of the real devices. In many cases (defined by quality of emulations achieved), for Functionality-type trials this would be a comparison of apples and oranges; it would perhaps be more valid in the case of Quality measurements, where the desire is to measure to what degree the apples differ from the oranges. Even in that case, however, there would be uncertainty as to what was being estimated.
4. Use a real device set whose impedances aren't fixed relative to one another — that is, a random effects model. This would mandate much more accurate control over real device impedances, as well as the ability to alter and target them quickly and arbitrarily.

Random Effects Model : As opposed to the mode of comparisons (many real devices compared with a single emulation), the experimental model is concerned with how levels of the experiment factors are chosen. With a fixed effects model, the same levels are used throughout the experiment — for instance, here Sampling Rate was always either 300, 600 or 1200 Hz, and never 423 Hz. With a random effects model, the variable levels used in a given trial are chosen at random from the experimental range; thus the sample rate *ranges* would be set at 300 and 1200 Hz, and any value within that range might appear in a given trial.

Obviously this is a great difference; the whole factorial-based analysis used here would be inapplicable. Mixed models may be used; thus a single factor — Targets — could be varied randomly, while the others were held fixed. However, this is little help because of all the factors, Targets is perhaps the most problematic to vary randomly as the experiment is currently designed, because of the difficulty in creating real devices of specified haptic characteristics.[3] Note that not only would real devices have to appear randomly, the targeting of those real devices would also have to vary randomly.

10.3.2 Quality versus Function

Early in this thesis (Chapter 3 and Section 8.1.3), I laid out arguments for using a [Multiple Real: Single Virtual] comparison approach. My observations of the experiments and their results confirm the judgement that the [MR:SV] approach was both best suited to the challenges of measuring emulation "resolution" and of system physical realizability (single emulation system required).

But as a consequence of performing the experiments and observing the sometimes surprising results, we have a deeper understanding of the illuminating power of these two approaches ([MR:SV] vs [SR:MV]): the [MR:SV] paradigm emphasizes determination of "Ability to Discriminate" (Function, as it is defined here) over "Qualitative Likeness" which includes things like how real it feels, how

---

[3] Were the comparison mode [Single Real:Multiple Virtual], the random effects model for at least the Targets factor would probably be both the most informative choice and little more effort to implement than the fixed effects model; because in this case the emulation would be varied randomly, and the modifications much more simple than the building of an arbitrarily large set of real devices.

Qualitative Attributes:

- Realism
- Passiveness
- Resemblance of features and textures of emulation to an arbitrary real environment
- Resemblance to the real environment that the emulation is intended to represent
- Individual aesthetic preferences
- Culturally conditioned expectations and preferences

Functional/Quantitative Attributes:

- Intuitive feature coding from emulation to real environment (e.g., matching of max force to convey sense of a detent of particular stiffness; or one-dimensional emulation of two- or three- dimensional texture, a la Minsky)
- Range of impedances achievable
- Type of impedances achievable (continuous environments and discrete features)
- Resolution of impedances achievable

Table 10.1: Attributes of a haptic emulation.

closely features resemble one another, etc.. Thus, in choosing this paradigm for unrelated and entirely justifiable (both pre- and post-experiment) reasons, we unknowingly walked into a previously unrecognized but crucially important dichotomy between Quality and Function (Table 10.1).

The tension was surprising in that there was no reason to suppose ahead of time that subjects would be so good at perceptual "coding",[4] at ignoring degradations and feeling features through them; and that by their very robustness in accomplishing the task set before them, they would dull this particular experiment design's power in measuring effect of the degradations on emulation "quality".

On the other hand, the power of the experiment which was simultaneously unveiled was its keenness in measuring functionality in the face of severe degradations. Perhaps the most valuable result of the experiments was learning just how functional a degraded emulation can still be; or conversely, how severely an emulation may be degraded before it becomes dysfunctional.

There are times when one part of the whole haptic-emulation picture, now subdivided into "Quality" and "Function", will be of more value than the other. If no connection to the real world is required — the application is for exploring an entirely virtual environment, for example, or directing a computerized operation — then functionality means that the control should allow the user to do his task well, and feel "nice" as well. However, there is no need for the controls to resemble anything physically based; the fidelity constraint is gone. In other cases, the fidelity requirement may be paramount; for instance, in surgical simulations of delicate operations which are meant to train surgeons for the real thing in a risk-free environment. However, in general it will be desirable to know both.

In these cases it is necessary to augment this methodology with one which explicitly takes note of emulation quality. The experiments described here, based on a [MR:SV] comparison model, assume a single VE model and alter only the model's parameterization; thus different features are

---

[4] From a functional standpoint, emulated and real environment features can correspond without being qualitatively similar. That is, a perceptual *coding* can occur which is relatively natural (not requiring training and repeatable from subject to subject) but which does not pass the haptic Turing Test; subjects consider the two features dissimilar in other respects and would never confuse them.

not compared for intelligibility, but rather a single feature is tested for intelligibility as a function of other emulation variables.

A [Single Real:Multiple Virtual] comparison model will be a good place to begin testing qualitative correspondence of the emulation to the real environment. such experiments could be done tomorrow with the same design and experiment setup. For example, to determine which features in a complex haptic environment contribute most to intelligibility, different virtual environment models could be used as levels of an emulation variable. The Toggles model could have been modified to test the relative importance of maximum actuation force versus switch throw length for importance in intelligibility.

10.3.3 Role of Calibration Trials and Real Device Duplication

A set of calibration trials executed at the beginning of each experiment session provided a dataset revealing how the subjects perceived the impedance distributions of the real device set they were tested on for that session. Explicit conclusions were drawn earlier from this data (Section 10.1.2) regarding subject perceptual distributions.

Duplication of Real Devices

Each device family member was duplicated, such that eight real devices were used in every session (the same eight for each of the six sessions executed for that family), four for each comparison trial. Of these, there were ostensibly four different values each with two identical representatives. The intent was to get an idea of subject consistency by offering them a duplicate as well as four distinct values in each calibration trial without telling them there were duplicates present, and seeing whether they could detect a duplicate or at least assign them very close magnitudes. This would in theory give estimates of (a) each subject's consistency in magnitude estimation; (b) subject ability to resolve small differences in a single trial; and (c) investigator error in duplication, exposed through different subjects consistently placing the duplicates in the same order (if duplicates were identical, they would be ordered randomly) and giving them similarly large magnitude differences. The last was important because the duplicates were also used in the comparison trials, randomly swapped into the set of four used in each trial in order to decrease likelihood of subjects recognizing physical characteristics (e.g. scratches or other flaws) of a small device set.

I found that while in most cases an adequate degree of duplication was achieved, in a few cases duplication was violated (Toggles, Devices C and D), and this in turn compromised some results. However, the fact that I was able to document the violation through the calibration trials was obviously useful.

Use in Response Variables

The use of the calibration data in the response variables was of limited value here because of the low information content in the comparison trial responses themselves (see Section 10.3.6). Thus this function of the calibration data was not as valuable as the others.

However, if the experiment model is altered so that the response variable is more informative, this will change. In particular, if a random effects model or otherwise nonuniform real device spacing is used in the future, the alternative response variables will be most important to the analysis.

10.3.4 Impact of Targets

Choice of Real Device Families

Was this a good collection of real devices to test? Was as much learned through the testing of Toggles, Sliders and Masses as might have been through another collection?

Difficulty of Creating and Targeting Families: The real devices families were sometimes quite difficult to build to specification, to model and to target. Minimizing the degree of difficulty was certainly a concern in the device selection process, yet still these were the most easily modified devices I could locate which also exhibited an interesting range and degree of complexity.

A deliberate tradeoff was made in this case, therefore, between the anticipated robustness of the experiment data and the richness of the environments studied. In two cases (Masses and Sliders), simple, easily-modeled-and-emulated systems were used and the result was solid and reliable data.

In the third case I took a chance and tried something much harder. The result somewhat sacrificed trustworthy quantitative results in the interest of practical insights and glimpses of the challenges we face in eventually accomplishing such difficult emulations successfully.

Confounding of Family Characteristics : One important confounding occurred by making only one device family both "complex" in that it had multiple characterizing parameters and was discontinuous and nonlinear in other ways as well. Thus we don't have explicit information about the effect of system order (inertial, damping or stiffness) on emulation performance, or what sort of performance might result for an emulated environment that was both complex and high order.

In future iterations, it might be useful to compare on one hand uniformly simple devices and on the other, uniformly complex and nonlinear devices. However, for the present purpose the approach I did take was fortuitous in that it probably exposed the most useful issues and questions.

Impedance Distribution of Real Devices

Did the real device impedance spacings (both range and distribution) ultimately prove to be good ones?

Note first that (a) different ranges and distributions were used for each real device family; and (b) the actual ranges and distributions used were partly specified by myself and partly enforced by what was possible to achieve with a finite effort (it was difficult to create real devices of specified impedance). I unintentionally achieved a set of impedance scales of a variety of ranges and distributions, and in hindsight the variation was informative. Even though the variation was confounded with the limited number of real device families, meaning that no strong conclusions may be drawn now, it maximized the hypotheses extractable for future testing.

The real device distribution was specified as logarithmic on the presupposition that subjects would perceive a linear distribution logarithmically; I wished to achieve linear spacing in subject perception. The possible impact of the actual distribution characteristics on the subjects' perceived distributions has already been discussed (Section 10.1.2). It would appear that subjects perceived the real device impedances on scales that diverged most markedly from Weber-like (logarithmic) distributions when the devices were most widely spaced on average, and were most logarithmic when the devices were set close together at nonuniform intervals.

However, even if I had known ahead of time that subjects would be perceiving in a mode often more linear than logarithmic, there would still be arguments for using a non-uniform spacing of devices. Although there is thus a confounding between targets and their distance from a neighbor, it is useful for some of the discriminations to be more difficult than others. In hindsight, mixing up the spacing might have been the best solution — and taking this to its logical extension, a random effects experimental model might be the best of all.

Thus, the logarithmic spacing of real device impedances turned out to be useful in this experiment, but in future experiments either a linear spacing or a random effects model would probably give more informative results.

10.3.5 Experiment Administration

Limits on Decision Time

A more explicit study would be needed to see if decision time limits would make a difference. My sense is that giving subjects freedom to choose decision time resulted in a more interesting and useful measure of the variation in individual perception and motivation in the population. However, limiting decision times and thus enforcing a brisk decision pace would probably have made for more intuitive decisions on the part of all subjects, which may be useful for future iterations.

Double-Blind Protocol

The efforts to organize the protocol such that neither subject nor investigator was aware of emulation variable settings appeared to be adequate. There was no indication either in the data or from observation of the experiments that subjects were aware of a critical subset of the emulation variables (including emulation target) except through the ease of discrimination or characteristics of emulation; I can confirm that although I did, after a few sessions, learn the color coding and thus generally knew which target was where in the box, I (a) did not know which of those real devices was targeted, (b) did not know which emulation variables were in effect except when it was obvious from emulation behavior or subject response, and (c) through part intention and part sheer boredom, did not generally pay attention to any of the above.

Physical damping level was known to me at all times; it was probably obvious to at least some subjects (those most familiar with the aims of my research, although not the details of the experiment protocol) as well. However, only one subject (Josephine) had any understanding of the possible implications of damping level; and not knowing which device was targeted, her performance was unlikely to have been biased.

Randomization

We cannot know from this data whether the restriction on randomization posed by Damping (changed only every 12 trials, rather than altered randomly throughout the session) was a problem; this is a shortcoming of the analysis. Observation of the experiments suggests that it was not a significant source of corruption. The arguments used to justify that damping was not a source of data corruption for the double-blind arrangement above holds here as well; although they are less strong because they do not rule out the possibility of confounding Damping with other, e.g. temporal, factors.

Physical Layout

This consisted of such geometrical factors as switch box, emulator positioning (with impact on subject reach), seating, consistency of hand position and of subject's approach to emulator and real devices. Based on observation and questioning of subjects as to their comfort and ability to reach all points in the workspace, it seems that this was satisfactory.

Subject Boredom and/or Fatigue

Fatigue was not a problem, based on subject responses and observation. Boredom, while it definitely happened, probably did not often corrupt data based on the absence of temporal trends in the data.

It was almost certainly a factor in the case of one session (Sliders, Hank); but from the experimental standpoint, this was actually a good test of inherent population variability.

10.3.6 Analysis

Trial-by-Trial Plots of Pre-Analyzed Responses

These consisted of the time-history (session sequence) and standard order plots of unanalyzed (showing simply the target and chosen devices for each trial) or minimally analyzed ($Y_{unweighted}$) data. They proved useful for locating temporal trends and biases, and locating outliers; and eye balling patterns which later were identified more quantitatively through the ANOVA and Effects analysis. Raw data examinations are an essential part of any analysis; you cannot claim to be familiar with your data without looking at it in this form, before analysis has camouflaged protocol artifacts.

Analysis of Variance

The ANOVA was highly useful in obtaining a quick overview of what matters and what does not. As in any application, it may be misleading to use it in isolation from more painstaking analysis techniques, primarily the Effects study.

Effects

The effect plots sorted subsets of the response data according to the level of a particular factor (e.g. for a single session, dividing the data into the 24 trials in which Damping was low and the 24 trials where Damping was high) and then comparing the average responses for each group. Because of the control it permitted over the level of detail in which the responses could be studied (at subject, sets of subjects, all-device-family or all-data subdivisions), and the intimate low-level grasp of the data such study afforded, this was unquestionably the richest form of analysis used. However, by the very magnitude of information present, it was also useful to have the ANOVA significance results to guide the study, indicating the most fruitful avenues of investigation.

The following rules of thumb were used in combining the ANOVA and Effects analysis, to good reward:

1. Confirm what the ANOVA says is significant;
2. When there's a surprise in the ANOVA significance — e.g., something didn't come up significant which was expected to, based on observation of experiments — make sure it can be explained and believed through the Effects analysis.

Scores

Study of this simplest-possible response variable, the percentage of session comparison trials in which a subject's choice agreed with the emulation target, was simple (the results could be and were obtained within moments of concluding a session, in order to compute the subject's financial compensation) and informative as long as one kept in mind exactly what the score did and did not mean:

- It did not mean that the emulation was always "right", in turn implying that a "miss" meant that the subject had made a mistake.
- It did mean that in a "miss", the emulation had failed through a targeting (or degradation) which made it difficult for the subject to make the correct identification.

- Comparing scores by different subjects for similar experimental situations, however, did indicate that some subjects consistently had a much higher hit rate than others.

  This means that *the information content was present* in the emulations, even when it was well-disguised. The same result could be found in the Effects analysis, but the Scores analysis made it much more graphically clear, and also made it easy to view the data in many ways which would have been difficult with Effects data (e.g., scores by family, by subject, versus decision time, etc.).

Usefulness of Response Variables

The most basic response variable, $Y_{unweighted}$, was obvious given the form of the experiments (given a selection of real devices, subject was asked to choose one; that was the response). Had we used an experiment structure which relied on, for example, magnitude estimation, the response variable would have naturally taken a different (and less discrete) form. The reason for using this experiment structure was extensively discussed in Methods; and the actual experimentation and analysis has not undermined that reasoning — the simplicity and ease of interpretation of the "Which One?" response was useful. However, the experimentation has exposed a need for more finely-grained information, both in terms of the Function analysis (through either a more extensive real device set or additional response information, e.g. a magnitude estimation in addition to the choice), and information of a different type for the Quality analysis which did not happen here.

Predictions of the potential value and limitations of the analysis response variables used here were made in Section 8.2.7.

In short, the amount of additional information that may be gleaned by the various weightings and transformations is limited at best, because it is all in post-processing. The limit is imposed by the fixed-effects nature of the experiment design, and in how that design influences the choice a subject makes in any given trial (because it limits the number and nature of choices offered the subject) Once the choice is made and a pattern of choices instituted, there is only so much a transformation can do with it.

Chapter 11

Conclusions and Recommendations

The purpose of this chapter is to summarize and emphasize the significance of the most important points made in Chapter 10, and based on those points, make recommendations for the continuation of this line of inquiry.

11.1 Inferences Drawn from Emulation and Experiments

Quality versus Function

"Quality" in a virtual environment is not the same thing as "Functionality". This dichotomy, as well as the possible merit of functionality-based performance criteria, has been largely overlooked.

Further, improving Quality does not automatically correspond to improvements in Functionality, and vice versa. Perhaps the most valuable result of these experiments was learning just how functional a degraded emulation can still be; or conversely, how severely an emulation may be degraded before it becomes dysfunctional.

Impact of Emulation Variables

I found that the influence of most emulation variables on performance (subjects' ability to make positive discriminations) varied substantially in the context of environment type (Masses, Sliders and Toggles). This was true of both experiment results and observations in emulation development.

The most important experiment trends showed that:

1. Sampling rate had surprisingly small influence on discrimination ability even in the face of a severely *qualitatively* degraded emulation, indicating that:

    > Functionality is far less sensitive to low sampling rates (300 Hz) than is Quality.

2. Increasing physical damping made emulations more difficult to stabilize, and tended to increase the rate of subject discrimination error. This result counters conclusions from the research of others, and may be an artifact of nonideal damping. However, the potential discounting of this effect due to hardware inadequacy is in itself an important result. It implies that:

    > Regardless of the true nature of perfect viscous damping on emulation performance, perfect viscous damping is an extremely difficult condition to attain or even to measure.

3. Virtual environment expression did not have a strong direct effect on discrimination error. It acted more strongly in interactions:

> The Impedance expression used was more robust (in a stability sense) to degradations in other emulation variables than was Admittance.

The Z and Y expressions also produced emulations which often felt qualitatively different from each other, and Y expressed environments were sometimes harder to stabilize.

Perceptual Scales

Subjects' perception of real device impedance distributions varied between linear and logarithmic, with a distortion from linear that held an relation to the uniformity and resolution of real device impedance distributions.

Individual Differences

The hypothesis that variability in an individual is less than variability between individuals was supported to some degree by all aspects of the data. There were also strong consistencies across individuals in some situations, indicating that generalizations can be made from studying a representative sample in spite of the presence of individual differences.

The most likely interpretation of the experiment data is that while people have similar low-level perceptual "tools", differentiation occurs when there is opportunity for choice in amount of attention devoted to a task and its allocation to individual features in an environment. One expression of this hypothesis is that individuals showed much greater consistency in performance and perception for simple environments (Masses and Sliders) than for complex (Toggles) environments.

Decision Time as Indicator

The results support the interpretation that in the case of simple environments,

1. Trial by trial, a slow decision means the decision was hard.
2. Slow average decision times for an individual mean that the subject was being careful, and this care showed as improved overall performance.

11.2 Evaluation of Experiment and its Philosophy

What were the major successes of the first iteration of this experiment methodology, and what were exposed as potential weaknesses?

Each of the following lists is ranked in order of criticality, from "Most" to "Least".

Nailed

1. Comparing multiple real to single virtual devices: the [MR:SV] comparison model proved valuable. Doing it the other way (Single Real:Multiple Virtual) would not have revealed as much in a first pass, in either philosophical or procedural senses.
2. Calibration trials: having a solid knowledge of subject perception of real devices was invaluable, both as a basis for the response variables and in interpretation of other analyses.
3. The emulations worked: for most members of all types of haptic environments tested, the emulations felt very much like the real devices they were emulated. The measure ("very much") is necessarily vague because we have not yet developed a means of quantifying this.
4. Range of experiment impedances achieved: for both real and virtual devices, the range achieved and tested proved appropriate. The approximately logarithmic spacing between adjacent devices was also informative, but introduced new questions (Section 10.3.3).
5. Experiment protocol: in general the protocol was demonstrated to be free of injurious biases or other contaminating effects (Section 10.3.5).

Weaknesses

1. Finite size of real device set:
   The fact that the real devices in a given set were always fixed in their impedance relation to each other resulted in a lower response information content than might have been available from another experiment model. Likewise, because two devices in the set had only one next neighbor while the others had two neighbors, there was a difference in the type and number of response possible for the end devices versus the middle devices.

   Other models considered have flaws of their own. The problem was discussed more fully in Sections 8.1.4 and 10.3.1.
2. Lack of control over real device impedance:
   It was difficult to produce real devices of a desired impedance (Section 10.2.2 and Chapter 6). In some cases (Masses and Sliders) the emulations felt "better" (more like the ideal phenomena they were intended to target) than did the real devices: it is as difficult to build real examples of idealities as it is to emulate subtle divergences from the ideal.
3. Subjectivity of real device parameterizations:
   It was also difficult to objectively parameterize the impedances of the real devices (Section 10.2.2 and Chapter 7).
4. Complications inherent in "controlling" the experiment:
   Holding constant all factors not explicitly tested can lead to distorted results for those tested, because of interactions between the tested and untested factors (for example, the interaction between servo gains and sampling rate).

5. Need better high-end complex environments:
   The non-degraded emulation should be as high performance as possible; the "best" that the emulation can do should bear a quite high fidelity resemblance to the targeted real device. In the case of complex (multiparametered and/or discontinuous) target haptic environments, this can be a challenge since these types of environments may be very difficult to model, parameterize and stabilize.

11.3 Recommendations for Future Experiments

Directions for future refinement come out of the problem list from the first iteration; meanwhile, we would like to retain the elements of the methodology which proved most valuable.

Experiment Model

The experiment design used here was based on a fixed effects model,[1] which had both positive and negative features; and on a Multiple Real: Single Virtual comparison model. Both might be improved upon, depending on the experiment objective.

Avoid Fixed Effects Model

The Fixed Effects model was chosen because I wished to maximize objectivity of subject response, by posing a "multiple choice" style problem rather than a more subjective magnitude estimation — subjective because it is difficult to know the subjects' perceptual reference at all times, and thus the data's reliability can be suspect when the right precautions are not taken. In this case, it was difficult to see what the right precautions could be. A tradeoff was thus made between richness of information content and data reliability.

While the Fixed Effects Model was informative, I predict that other multiple-choice experiment models in which the "multiple" components (emulations or real devices) are not held in a fixed quantitative relation to one another would yield richer and more reliable results for the same amount of experiment energy and data reliability.

This due to several shortcomings of the Fixed Effects model:

1. Subject choice is influenced by the distance to the nearest neighbor:
   The distribution of the devices is confounded with response, no matter what distribution is chosen or what distribution subjects perceive.

2. Low information content of response:
   A subject has a limited number (2 or 3) of choices in any given trial. Further, in the Fixed Effects model these 2–3 choices result in the same values for response variable in every trial. Even after weighting or transforming by measured or subject perceived impedance (or other haptic characteristic being varied) distributions, there are only $[1+(N_{dev}-1)]$ *different* response values possible. This does not permit highly graded responses. With a random effects model, the number of different response values is a function of the number of trials, rather than the number of real devices.

---

[1] "Fixed effects" is the experiment design field's designation for an experiment model in which a fixed number of factor levels, generally small, are used throughout the experiment — e.g., sampling rate was always one of 300, 600 or 1200 Hz, as opposed to the random effects model wherein at each trial, a different value of sampling rate might be chosen randomly between preset bounds of 300 and 1200 Hz.

3. Vulnerable to learning effects:
   Subjects learn to recognize the different members of a fixed family of environments (here, real environments; but I expect the same to be true of fixed families of emulated environments).

The cost of using other models may be greater difficulty in preparing for experiments, and possible a larger number of trials. The same tradeoff made here might be legitimately be repeated in future iterations, whenever time spent is valued over richness of results; i.e., only a general "lay of the land" view is required, rather than a detailed and high precision mapping.

Note that experiment model (e.g. Fixed Effects versus Random Effects) may be mixed within an experiment, and that Targets is the factor which is most influenced by experiment model. If more convenient it might make sense to use a model where targets are varied randomly while other emulation variables comprise a fixed set.

Other experiment models to consider include:

Random effects: requires construction of a set of real devices whose haptic properties may be easily and quickly changed on the fly.

Paired comparisons: in a given trial, only two real devices needed. From trial to trial, need rich set of real devices, as in Random Effects, or else this model will suffer from most of the same shortcomings as the fixed effects.

If none of these prove convenient, recommendations for improving the fixed effects model (targets) situation were given in Section 10.3.1.

Dependence of Comparison Model on Experiment Goal

For every application, the comparison model — e.g., "Multiple Real compared to Single Virtual" — should be carefully considered. Table 11.1 is a representative list of possible experiment objectives, accompanied by a recommendation for comparison model for that objective:

Augmenting Functionality Test with Quality Test

For some experiment objectives, the experiment design here was limited in that while it was good at elucidating performance of an emulation on a set of Functional indices, it did not do well at measuring the less tangible Quality performance aspect.

Conversely, most previous efforts to systematically evaluate perceptual response to aspects of virtual haptic environments has focused on Quality performance evaluation — "How good does this feel?" rather than "How much like ___ does this feel?", or "How well does ___ let you do ___?" [71, 135, 139, 167].

In most cases, both criteria will be important and both need to be tested. Thus an improvement in the design would be to augment the present Functional measure testing with an element which focuses on Quality. In the design presented here, a likely augmentation might be to request a quantitative evaluation (magnitude estimation) of the qualitative difference between the chosen real device and the emulation in addition to the choice itself, with appropriate reference pinnings.

This augmentation was considered for these experiments and rejected because of the additional experiment session time required to make such a cognitively difficult and consequently time consuming and mentally fatiguing assessment. In order to include the Quality augmentation, at least one and perhaps two emulation variables would have had to be dropped, reducing the richness of the interaction study.

Objective	Comparison	Justification
I. To learn how emulation variables improve emulation performance, in a task-based metric.	MR:SV these experiments	Comparing a single emulation to a set of real devices will reveal the most about the emulation's power in aiding subject discernment at a given level of emulation variables.
II. Both qualitative and task-based characteristics of an emulation are to be tested, but quality is more important.	SR:MV	By comparing emulations directly with each other in reference to a real device, this model is best suited to demonstrating which features in a virtual environment contribute most to intelligibility, as opposed to which emulation variables do so. Disadvantage: requires multiple emulation systems or one emulation system where changing emulation variables is fast (e.g. subject has N buttons to select N different emulations).
III. Measure only qualitative correspondence of emulation to real environment.	SR:SV magnitude estimation	Since only un-like environments are compared (real to virtual, with no component of R-R or V-V as in I. and II.), the nature of the differences quantified by the subject is unclear (what does the estimate refer to?). However, their relative magnitudes are known so a closer qualitative correspondence can be homed in on. Advantage: explicitly does away with magnitude "end effects" problem.
IV. Systematically target an emulation to a specific real environment.	2R:SV or SR:2V paired comparisons	When implemented as a "directed learning" algorithm, where the pair presented in a trial is computed as a result of the response to the previous trial, the best answer is homed in on in the same way that an optician homes in on the right set of glasses by comparing two sets at a time.

Table 11.1: Summary of possible experiment objectives and the comparison models to which they are best suited

Real Device Parameterizations

Whatever the experiment model and comparison mode, any experiment design whose intent is to investigate fidelity of an emulation to a real world analog should in some way rely on comparisons between real and virtual environments, with the virtual targeting the real. Thus, modeling and parameterization of real devices will continue to be an important issue. It will be solvable with one or both of the following approaches:

1. Choose and/or design real device(s) such that they are easily modeled and parameterized. This may constrain the types of haptic environments which may be studied.

2. Develop better means of accurately and precisely modeling and parameterizing real environments (a start on this objective was described here).

Building real devices to specified haptic characteristics will continue to be a challenging and constraining element of the experiment design and preparation; and will have to be considered on an ad hoc basis since every type of environment and characteristic studied within that environment (here, generalized impedance magnitude) will potentially be different.

Controlling of Untested Emulation Variables

During the experiment design phase, pilot experiments should be conducted to make sure dependencies of the emulation variables of primary interest on those which are being held constant are fully understood; and tradeoffs in experiment size versus constancy of non-tested emulation variables examined carefully.

Better High End

It is a bit of a chicken-or-the-egg problem, but in general the best (most informative and reliable) results will be obtained when the high-end (undegraded) emulations are closest to passing the haptic Turing Test, and degradations move the emulations away from this high level of performance.

The second-best situation will be when at least one of the Quality/Functionality pair is achieved at a high level for at least some emulations; in this case, which of Quality/Functionality is best achieved should be the performance characteristic of greatest interest to the investigator.

11.4 Guidelines for Specifications of Haptic Virtual Environment Hardware and Control

Because there is little information in the literature on haptic display hardware requirements, the specifications for this system were arrived at through a combination of successful estimation and trial and error. An important objective of the project was to reveal those emulation parameters in which further investment (at potentially higher system cost) is most likely to offer a payoff of significantly improved emulation performance. That is, the study was to indicate *directions* of improved performance, rather than limiting itself to an analysis of what is possible with this system in its current configuration.

While a definitive statement of emulation variable levels required for a given emulation task is not yet available and will change as technology advances, some insight has been obtained through extensive manipulation of the haptic interface described here. The following comments are offered as guidelines, which this field will continue to supplement in the future:

- Casual observation suggests that sample rates in excess of 500 Hz are necessary for emulations of discontinuous, high bandwidth haptic phenomena (such as walls) to feel at all convincing. 1 kHz is substantially better, while up to 10 kHz, performance gains appear to diminish but are still perceptible to the author in certain demanding emulation situations.[2]

  Thus 1 kHz may be a good minimum target for higher-end emulations that include discontinuous features, beyond which improvement in other emulation variables — namely hardware bandwidth — will have higher return on investment.

  Depending on the model and control strategy used, the presence of a velocity sensor and consequently a smooth displacement derivative may reduce the sample rate requirement; likewise for a low-noise operating environment, permitting the use of unfiltered sensor output.

- Sample rates in excess of 200-300 Hz appear necessary for stable emulation of most continuous physical systems which are to be disturbed at velocities characteristic of human operators (10-20 Hz).

---

[2] This is presumed to be a combination of the effects of a human force perception cutoff of about 1kHz, and of the 100 Hz closed loop force bandwidth of this system.

- Relatively low hardware position-servoing bandwidth (10-15 Hz) is not necessarily an obstacle to convincing emulator performance; but force-servo performance appears to be more critical. Thus while high actuator inertia is a handicap, it is possible to achieve good performance in its presence and may productively be traded off for other desirable actuator characteristics such as maximum output force and force slew rate. However, to emulate high-frequency phenomena at high performance a lower inertia or coupled macro/micro design will be necessary [145].
- A linear motor force of 10-20 N is sufficient to produce the illusion of crisp, audible walls on initial contact. However, higher force levels are necessary to maintain the illusion in response to sustained and penetrating exploration; that is, a surface which is to be leaned against.
- A low-drift accelerometer is required to emulate inertia with the Z expression used here, because double-differentiation of a position signal tends to accentuate sensor noise. It should be noted that other methods of Z control approach the inertia emulation problem differently [66, 82, 127]. Only a force sensor is required to emulate inertia in Y control.

11.5 Contributions

Primary Contributions

1. The Experiment Design Works:

Development and validation of a functional comparisons experiment approach for systematic analysis of virtual environment performance.

The concept of ascertaining emulation fidelity, as opposed to different aspects of virtual environment "quality" — realness, passiveness, etc.— by comparing emulated environments to specific real environments whose haptic variability has been explicitly controlled is, to my knowledge, new.

Likewise, the idea of minimizing subjectivity (e.g. magnitude references) in subject responses by posing the question as a multiple choice —

> "*Which* [real/virtual] device in the set feels most like the single [virtual/real] device?"

rather than a magnitude-oriented question ("How different do they feel?"). I feel that the objective nature of the question increased the reliability of the results, regardless of the direction of the comparisons (i.e. whether the comparison paradigm is [Multiple Real:Single Virtual], or [Single Real:Multiple Virtual]).

2. Learned What's Important:

The most important results to emerge from this first iteration of the new experiment design are threefold:

Recognition and documentation of the important dichotomy between Quality and Function and their respective performance indices.

It appears that researchers, including myself prior to observing the results of these experiments, generally assumed that Quality is equivalent to Functionality; and that improving or optimizing one would automatically improve and/or optimize the other. This work has demonstrated that this is not always the case; and further, it has raised the provocative and commercially important concept that high Quality is not always what's needed, and by extension, the best place to put one's money.

A new research subfield has thus been opened to investigation: *When* is Quality needed or not, and how little can be gotten away with in what situations?

Proposal and preliminary characterization of an "Emulation Variable / Emulation Performance" mapping.

Prior to this work, there has been little attempt to relate the disparate elements of our gradually emerging understanding of how to create high performance haptic virtual environments, and how to relate emulation components to different measures of emulation performance. I have proposed a structure for the systematic organization of this knowledge; a structure which will act not only as a data repository, but which by its organization can act on the data to predict performance based on a virtual environment system's hardware and control specifications, or conversely to specify characteristics for a system based on desired performance indices in specified task domains.

Initiated the process of filling in the Emulation Variable / Performance Mapping, by testing the effect of three important emulation variables and their interactions on emulation effectiveness.

Controller Sampling Rate, Physical Damping and Virtual Environment Expression were tested for their effect on subjects' ability to use an environment emulation to make a positive discrimination (an inherently Functionality-oriented task).

3. Hardware and Control:

Design and development of versatile haptic emulation testbed.

I designed and built an emulation testbed which is high performance in terms of present state of the art, and in a large subset of all possible and often conflicting performance indices; and which is in many ways ideally suited to the systematic exploration of the Emulation Variable Space. Perhaps more importantly, I used this testbed extensively (independently of the usable explicitly devoted to the described experiments) to advance our collective knowledge of how to create high performance emulations.

Secondary Contributions

4. Technique for Measuring and Playing Back "Feel"

I developed and automated a technique for the identification of haptic characteristics of real physical environments. The "Haptic Camera" objectively and conveniently records and plays back the haptics of a real device.

5. Expressions of Impedance / Admittance Control Compared

In the context of haptic fidelity, similar virtual environment models expressed in examples of Force Out (impedance) and Position Out (admittance) structures were tested for effect on functionality of the emulations they produced. I believe that this is the first time environment expression has been objectively tested for this kind of performance index.

I found that while there were minor "functional" differences in that subjects had similar error rates with either expression, the two differed in qualitative results for complex environments, and Admittance was generally more difficult to stabilize in face of degradations of other emulation variables.

6. Library of Haptic Emulation Insights and Strategies

In the course of learning to control the emulation hardware used here, exploring its limits and developing the emulations used in the experiments here, I developed techniques, insights and strategies (often non-physically based) for emulating a wide variety of haptic phenomena.

For example (see Chapter 5 for detail):

- "Slipthrough"
- Interactions of servo gains and environment performance
- Perceived mass offset
- Use of negative springs
- Different role of various virtual-environment characteristics (such as "B", virtual damping) in admittance, impedance expression. In impedance, B feels like damping. In an admittance expressed, 1st-order system, B stabilizes because it makes the denominator bigger.

7. Hardware and Control Specification Criteria

I produced preliminary guidelines for specifying haptic virtual environment system hardware and control characteristics based on desired performance and emulation tasks for any haptic virtual environment application; and laid the cornerstone for the production of many more (see Section 11.4 for more detail).

11.6 Future Work

Virtual Product Prototyping

Stage I

Design of "Quality" Aspect of Experimentation

Study of Y / Z VE Model Expressions
Expanded theoretical / experimental study of impedance, admittance VE control approaches, including other strategies of expression than the ones I have used here.

Multisensory Augmentation
Add sound and vision to the emulation; test them as additional families of emulation variables.

Expansion of Automatic Haptic System ID Technique
Build hardware and develop stimulation and analysis techniques to accommodate a larger range of systems that can be identified. Test the results, compare with other methods of system ID. Consider using stochastic input.

Extension of Emulation Variable Space Characterization
Look at more variables, and look at other real haptic phenomenon. Walls, for instance ...

Stage II

Extension of Emulation Variable/Performance Mapping to Other Emulation Systems
Some of the results here were influenced by characteristics of this emulation system, some factors of high performance and some limiting its performance — e.g. its relatively large mass and large force saturations, single DOF constraint, linear action, flaws in physical damper.

See how results compare with a different system which has a different set of strengths and weaknesses.

Development of Desktop Virtual Product Prototyping Tools

Haptic Emulation Wish List

This is a summary of the prerequisites I perceive for the concept of emulating real haptic environments at a high fidelity, and in some cases haptic emulation in general, to substantially progress. Some items are hardware or control; some are semantics and some are metrics. Some relate to conducting experiments in the style of those performed here; some to any application of haptic emulation. All are important.

- A field-wide agreed-upon definition of and means of measuring Emulation Quality (in the present work, I hope I have made a start at defining and measuring Emulation Fidelity).
- A field-wide agreed-upon definition of and means of measuring Emulation System Performance.

- A way to characterize an arbitrary real device haptically, by means of a simple probe.
- Designs for building an Ideal Dashpot (both linear-acting and rotary versions needed)
- High-bandwidth, high-strength, high-range of motion, and ideally low-cost linear actuators
- Servo amplifiers with high power output that don't generate electronic noise
- Real haptic environments whose haptic characteristics can be arbitrarily and quickly modified.

Part V

Appendices

Appendix A

List of Haptic Human Interfaces

Auto Cockpit

- brake, clutch pedals
- steering wheel
- control stalk
- horn
- door handles (inside and out)
- stereo volume, tuning, balance controls
- clock/calendar programming controls
- heater controls
- radar detector
- theft alarm
- seat adjustment control
- interior rear-view mirror (position, night vision)
- side-mounted rear-view mirror remote adjustment

Home Entertainment Electronics

- audio/video remote controls
- stereo equalizer controls
- handheld stereo/cassette players for joggers
- portable stereos
- tape deck eject
- CD eject
- television controls
- programmable VCR
- Nintendo joystick

Around the House

- shower and sink faucets
- door knobs and handles (indoor and outdoor)
- doorbell
- door lock
- vacuum cleaner
- clock and clock/radio (mechanical and electronic)
- hose spigot
- hose watering gun
- thermostat
- home alarm system
- sewing machine pedal and hand controls
- gas barbecue
- piano (conventional)
- piano (electric)

Kitchen

- pepper grinder
- conventional oven and stove controls (dials)
- electric mixer (handheld and counter-top)
- pop-up toaster
- toaster oven
- refrigerator temperature control
- microwave oven
- food processor
- blender
- flour sifter
- egg timer

Personal

multi-function programmable watches
hair dryer
electric razor
electric toothbrush

Garage and Machine Shop

socket/ratcheting wrenches
handheld drill
circular saw
nail driver
table saw
milling machine
drill press
lathe
belt sander
digital multi-meters
digital calipers
car jack

Sport and Recreational

cycling computer
bicycle gearshift levers
bicycle brake levers
bicycle clip-less pedals
motorcycle hand grips
downhill ski bindings
motor boat controls
mountaineering and climbing tools

Switches

dimmer
push-button
toggle light switch
membrane touch controls
power on/off toggles
kill buttons
generic MIT light switch
lamp switch

Telephones

programmable office type
cordless
wall-mounted home variety
car cellular
telephone cradle
answering machine

Computer and Computer Interface Devices

keyboard
mouse
floppy disk insert and eject
power toggle switch
data glove
space ball
track ball
Phantom

Office

handheld calculator
copying machine
fax machine
stapler
three-hole paper punch
adding machine
cash register
door entry combination-type locks
security card systems
intercom

Complex Interfaces

airplane cockpit (see auto cockpit)
nuclear power plant control room
medical instrumentation
stereolithography machine
NC milling machine / lathe

Appendix B

Hardware Specifications

Information on vendors, specifications, etc. of all important purchased components.

Real Device Components

Specifications for key purchased elements of real devices used in experimentation.

Family	**Vendor**	**Part**	**Description**
Masses	*see Appendix C*		
Sliders	All Electronics	ASP-2K	2k ohm audio taper slide pot, 90 mm long (66 mm travel)
Toggles	Carlingswitch	2FA53-73	SPST On-None-Off
		2GK50-73	DPST On-None-Off
		HK250-73	3PST On-None-Off
		IK251-73	4PST On-None-Off

Purchased Components of Emulation System

Specifications for all key emulation system elements which were not constructed in-house.

Component	Vendor	Part	Specifications	
voice coil motor	Control Data	BJ717	mechanical stroke	90 mm
	Arden Hills, MN	(disk drive)	coil diameter	57 mm
			total moving weight	0.25 kg
			R_{coil}	2.3 Ω
			I_{coil}	0.27 mH
			peak current	8 A
			cont. stall current	3.5 A
			peak force at 8A, 150W	60 N
			continuous stall force	25 N
			mechanical time constant	17.7 msec
			electrical time constant	∼0.1 msec
			force constant (K_t)	6.16 N/A
			back EMF (K_e)	5.5–6.9 V-s/m
PWM	Performance	DCM 1000	switching frequency	25 kHz
servo amplifier	Controls, Inc.		nominal power	480 W
			nominal voltage	24 VDC
	Horsham, PA		maximum voltage	32 VDC
			continuous current	20 A
			peak current	20 A
			load current settle time (for 15A step)	0.2 msec
			current slew (-3 dB)	6 kHz
amplifier	Performance	SPM 500	output voltage	24 V
power supply	Controls, Inc.		continuous current	20 A
			peak current	30 A
linear	Waters Mfg. Inc.	LT134	travel length	100 mm
potentiometer	Wayland, MA		resistance	40 Ω /mm travel
			operation force	30 grams max
force transducer	Revere Transducers	FT170-10	type	strain gauge
	Cerritos, CA		nominal capacity	±45 N
			maximum load	4500 N
			deflection	0.18 mm, FS load
			weight	30 grams
accelerometer	Entran Devices	EGAX-10	type	miniature piezoresistive
	Fairfield, NJ		range	±5 g
			overrange protection	±10,000 g
			weight	630 mg
signal	Analog Devices	IB31AN	(signal amplifiers)	for force,
conditioning	Norwood, MA	AC122	(mounting cards)	acceleration
data acquisition	Analog Devices	RTI-815	ADC channels	8 differential
	Norwood, MA		DAC channels	2
			bidirectional digital I/O	8 bits
			A/D resolution	12 bits
			A/D conversion time	25 μsec
			realtime clock	3μsec–656 sec

Appendix C

Silicone Viscous Fluid Specifications

The following pages are manufacturer information and Material Safety Data Sheets for the Dow Corning silicone fluids used in the physical damper and in the construction of some of the real devices used in the experiments (Sliders and Toggles).

The fluid was purchased through Nye Lubricants, P.O Box 8927, New Bedford, MA 02742, (508) 996-6721.

Information about Silicone Fluids

DOW CORNING

DOW CORNING® 200 Fluid, 50 cSt.
DOW CORNING® 200 Fluid, 100 cSt.
DOW CORNING® 200 Fluid, 200 cSt.
DOW CORNING® 200 Fluid, 350 cSt.
DOW CORNING® 200 Fluid, 500 cSt.
DOW CORNING® 200 Fluid, 1000 cSt.

DESCRIPTION

DOW CORNING® 200 fluids, 50-1000 centistokes (cSt.), are medium viscosity polydimethylsiloxane polymers manufactured to yield essentially linear polymers with average kinematic viscosities ranging from 50 to 1000 cSt.

COMPOSITION

Linear polydimethylsiloxane polymers characteristically have the following typical chemical composition:

$(CH_3)_3SiO[SiO(CH_3)_2]nSi(CH_3)_3$

Commercial bulk-polymerized dimethyl silicone fluids, such as DOW CORNING 200 fluids, 50-1000 cSt., typically contain trace amounts of process impurities.

INCOMING INSPECTION

Dow Corning recommends that incoming inspection tests be performed to confirm product identity and condition on arrival. Suggested tests include viscosity and infrared identification, and any other tests deemed necessary for the application. Such tests may or may not be run routinely by Dow Corning as lot acceptance tests. Obtain the Sales Specifications for lot acceptance tests and test limits conducted on DOW CORNING 200 fluids, 50-1000 cSt.

SAFE HANDLING INFORMATION

A Materials Safety Data Sheet, as required under existing federal regulations, is available upon request from Dow Corning Corporation, Midland, Michigan 48640.

Note: May cause temporary eye discomfort.

SHELF LIFE

Shelf life is the period of time during which a material may be stored under specified conditions in its original unopened container (except for inspection) while retaining the material's sales specifications. Shelf life starts with the date of shipment (unless otherwise specified), and ends on a given date. Continued storage beyond the designated shelf life does not necessarily mean that the material may not be used. However, after the expiration of the designated shelf life, testing of critical properties and redetermination of suitability for contemplated use of the product is imperative.

Dow Corning certifies that DOW CORNING 200 fluids, 50-1000 cSt., will meet sales specification requirements for a period of 12 months from date of shipment.

Storage temperature: ambient.

PACKAGING

DOW CORNING 200 fluids, 50-1000 cSt., are supplied in 40- and 440-lb (18.1- and 199.6-kg) containers, net weight. Smaller containers are available from repackagers.

Caution! Containers will have product residues when emptied. Follow precautions recommended for handling these products when disposing of the container. Containers are not intended for reuse.

SALES SPECIFICATIONS

Sales specifications information, including detailed test methods and analysis procedures used by Dow Corning, is available upon request. Since Dow Corning reserves the right to update sales specifications information without prior notice, users should periodically request this information.

PRODUCT CHARACTERISTICS

DOW CORNING 200 fluids, 50-1000 cSt., have the following product characteristics:

- Clear
- Essentially Nontoxic
- Nonbioaccumulating
- Nonbioactive
- Nongreasy
- Nonocclusive
- Nonrancidifying
- Nonstinging on Skin
- Tasteless

DOW CORNING 200 fluids, 50- cSt., when compared to other materials that may be substitu a given application, may offe more of these comparative characteristics:

- High Compressibility
- High Dampening Actior

Information about Silicone Fluids

DOW CORNING

DOW CORNING® 200 Fluid, 10,000 cSt.
DOW CORNING® 200 Fluid, 12,500 cSt.
DOW CORNING® 200 Fluid, 30,000 cSt.

DESCRIPTION

DOW CORNING® 200 fluids, 10,000-30,000 centistokes (cSt.), are high viscosity polydimethylsiloxane polymers manufactured to yield essentially linear polymers with average kinematic viscosities ranging from 10,000 to 30,000 cSt.

COMPOSITION

Linear polydimethylsiloxane polymers characteristically have the following typical chemical composition:

$(CH_3)_3SiO[SiO(CH_3)_2]nSi(CH_3)_3$

Commercial bulk-polymerized dimethyl silicone fluids, such as DOW CORNING 200 fluids, 10,000-30,000 cSt., typically contain trace amounts of process impurities.

INCOMING INSPECTION

Dow Corning recommends that incoming inspection tests be performed to confirm product identity and condition on arrival. Suggested tests include viscosity and infrared identification, and any other tests deemed necessary for the application. Such tests may or may not be run routinely by Dow Corning as lot acceptance tests. Obtain the Sales Specifications for lot acceptance tests and test limits conducted on DOW CORNING 200 fluids, 10,000-30,000 cSt.

SAFE HANDLING INFORMATION

A Materials Safety Data Sheet, as required under existing federal regulations, is available upon request from Dow Corning Corporation, Midland, Michigan 48640.

Note: May cause temporary eye discomfort.

SHELF LIFE

Shelf life is the period of time during which a material may be stored under specified conditions in its original unopened container (except for inspection) while retaining the material's sales specifications. Shelf life starts with the date of shipment (unless otherwise specified), and ends on a given date. Continued storage beyond the designated shelf life does not necessarily mean that the material may not be used. However, after the expiration of the designated shelf life, testing of critical properties and redetermination of suitability for contemplated use of the product is imperative.

Dow Corning certifies that DOW CORNING 200 fluids, 10,000-30,000 cSt., will meet sales specification requirements for a period of 12 months from date of shipment.

Storage temperature: ambient.

PACKAGING

DOW CORNING 200 fluids, 10,000-30,000 cSt., are supplied in 40- and 440-lb (18.1- and 199.6-kg) containers, net weight. Smaller containers are available from repackagers.

Caution! Containers will have product residues when emptied. Follow precautions recommended for handling these products when disposing of the container. Containers are not intended for reuse.

SALES SPECIFICATIONS

Sales specifications information, including detailed test methods and analysis procedures used by Dow Corning, is available upon request. Since Dow Corning reserves the right to update sales specifications information without prior notice, users should periodically request this information.

PRODUCT CHARACTERISTICS

DOW CORNING 200 fluids, 10,000-30,000 cSt., have the following product characteristics:

- Clear
- Essentially Nontoxic
- Nonbioaccumulating
- Nonbioactive
- Nongreasy
- Nonocclusive
- Nonrancidifying
- Nonstinging on Skin
- Tasteless

DOW CORNING 200 fluids, 10,000-30,000 cSt., when compared to other materials that may be substituted in a given application, may offer one or more of these comparative characteristics:

- High Compressibility
- High Dampening Action
- High Dielectric Strength
- High Oxidation Resistance*
- High Shearability without Breakdown
- High Temperature Serviceability*
- High Water Repellency
- Low Environmental Hazard
- Low Fire Hazard*
- Low Odor
- Low Reactivity*
- Low Surface Energy

---

*See "C

TYPICAL PHYSICAL PROPERTIES

Physical properties vary from lot to lot and should not be used for writing sales specifications.
SPECIFICATION WRITERS! Before writing a specification, obtain Dow Corning's Sales Specification on the product.
Contact "Specification Coordinator, Dow Corning, Box 1767, Midland, MI, 48640."

As Supplied	*DOW CORNING 200 Fluid, 10,000 cSt.*	*DOW CORNING 200 Fluid, 12,500 cSt.*	*DOW CORNING 200 Fluid, 30,000 cSt.*
Appearance	Crystal clear liquid free from suspended matter and sediment.		
Specific Gravity @ 25°C	—	—	0.971
Refractive Index @ 25°C	1.4036	1.4036	1.4037
Color, APHA	5	5	5
Flash Point, open cup, °F	>620	>620	>620
Acid Number, BCP	trace	trace	trace
Pour Point, °C	—	-46	-43
Surface Tension @ 25°C, dynes/cm	—	21.5	21.5
Volatile Content, @ 150°C, percent	0.27	0.23	0.29
Viscosity Stability, @ 25°C, after 16 hr exposure @ 150°C, % change	-0.7	-1.6	-2.0
Viscosity Temperature Coefficient	—	0.61	0.61
Coefficient of Expansion, cc/cc/°C	—	0.00096	0.00096
Thermal Conductivity @ 50°C, gm cal/cm · sec · °C	—	0.00037	—
Specific Heat @ 25°C, cal/gm/cm	—	—	—
Solubility Parameter*	7.4	7.4	7.4
Solubility in Typical Solvents,			
Chlorinated solvents	High	High	High
Aromatic solvents	High	High	High
Aliphatic solvents	High	High	High
Dry alcohols	Poor	Poor	Poor
Water	Poor	Poor	Poor
Fluorinated propellents	Poor	Poor	Poor
Dielectric Strength @ 25°C, volts/mil	—	400	400
Volume Resistivity @ 25°C, ohm-cm	—	1.0x10^{15}	1.0x10^{15}

*Fedors Method: R.F. Fedors, Polymer Engineering And Science, Feb. 1974.

Dow Corning does not routinely test all these physical properties. Users should independently test these properties when they are critical in the application.

- Low Temperature Serviceability
- Low Toxicity
- Low Vapor Pressure
- Good Heat Stability*

APPLICATION INFORMATION

DOW CORNING 200 fluids, 10,000-30,000 cSt., are not intended for food or medical use. They are intended for use by industrial manufacturers. Typical end uses include:

- Coatings Additive
- Dampening Fluid
- Elastomer and Plastics Lubricant
- Electrical Insulating Fluid
- Foam Preventative or Breaker
- Household Products Ingredient
- Mechanical Fluid
- Mold Release Agent
- Petroleum Refinery Defoamer
- Plastics Additive
- Polish Ingredient
- Specialty Chemical Products Ingredient
- Surface Active Agent

---

*See "Contamination and Fire Prevention."

CONTAMINATION AND FIRE PREVENTION

At elevated temperatures, DOW CORNING 200 fluids, 10,000-30,000 cSt., are sensitive to contamination by strong acids, bases, some metallic compounds and oxidizing agents. These contaminants may cause an accelerated rate of volatile by-product formation. Oxidizing agents can also cause an increase in fluid viscosity. When these conditions may exist, it is recommended that the flash point of the fluids be checked periodically to monitor operational safety. Also, ignitable conditions may exist if the fluid is giving off smoke.

Note: For answers to any questions regarding the efficacy, safety, health or environmental aspects of using DOW CORNING 200 fluids, 10,000-30,000 cSt., in any application, contact your nearest Dow Corning sales office or call the Dow Corning Customer Service.
Telephone:
In Michigan (800) 292-2323
Outside Michigan (800) 248-2345

The information and data contained herein are based on information we believe reliable. You should thoroughly test any application, and independently conclude satisfactory performance before commercialization. Suggestions of uses should not be taken as inducements to infringe any particular patent.

DOW CORNING CORPORATION
MIDLAND, MICHIGAN 48640

DOW CORNING

Information about Silicone Fluids

DOW CORNING

DOW CORNING® 200 Fluid, 60,000 cSt.
DOW CORNING® 200 Fluid, 100,000 cSt.

DESCRIPTION

DOW CORNING® 200 fluids, 60,000-100,000 centistokes (cSt.), are high viscosity polydimethylsiloxane polymers manufactured to yield essentially linear polymers with average kinematic viscosities of 60,000 and 100,000 cSt.

COMPOSITION

Linear polydimethylsiloxane polymers characteristically have the following typical chemical composition:

$(CH_3)_3SiO[SiO(CH_3)_2]_nSi(CH_3)_3$

Commercial bulk-polymerized dimethyl silicone fluids, such as DOW CORNING 200 fluids, 60,000-100,000 cSt., typically contain trace amounts of process impurities.

INCOMING INSPECTION

Dow Corning recommends that incoming inspection tests be performed to confirm product identity and condition on arrival. Suggested tests include viscosity and infrared identification, and any other tests deemed necessary for the application. Such tests may or may not be run routinely by Dow Corning as lot acceptance tests. Obtain the Sales Specifications for lot acceptance tests and test limits conducted on DOW CORNING 200 fluids, 60,000-100,000 cSt.

SAFE HANDLING INFORMATION

A Materials Safety Data Sheet, as required under existing federal regulations, is available upon request from Dow Corning Corporation, Midland, Michigan 48640

Note: May cause temporary eye discomfort.

SHELF LIFE

Shelf life is the period of time during which a material may be stored under specified conditions in its original unopened container (except for inspection) while retaining the material's sales specifications. Shelf life starts with the date of shipment (unless otherwise specified), and ends on a given date. Continued storage beyond the designated shelf life does not necessarily mean that the material may not be used. However, after the expiration of the designated shelf life, testing of critical properties and redetermination of suitability for contemplated use of the product is imperative.

Dow Corning certifies that DOW CORNING 200 fluids, 60,000-100,000 cSt., will meet sales specification requirements for a period of 12 months from date of shipment.

Storage temperature: ambient.

PACKAGING

DOW CORNING 200 fluids, 60,000-100,000 cSt., are supplied in 40- and 440-lb (18.1- and 199.6-kg) containers, net weight. Smaller containers are available from repackagers.

Caution! Containers will have product residues when emptied. Follow precautions recommended for handling these products when disposing of the container. Containers are not intended for reuse.

SALES SPECIFICATIONS

Sales specifications information, including detailed test methods and analysis procedures used by Dow Corning, is available upon request. Since Dow Corning reserves the right to update sales specifications information without prior notice, users should periodically request this information.

PRODUCT CHARACTERISTICS

DOW CORNING 200 fluids, 60,000-100,000 cSt., have the following product characteristics:

- Clear
- Essentially Nontoxic
- Nonbioaccumulating
- Nonbioactive
- Nongreasy
- Nonocclusive
- Nonrancidifying
- Tasteless

DOW CORNING 200 fluids, 60,000-100,000 cSt., when compared to other materials that may be substituted in a given application, may offer one or more of these comparative characteristics:

- High Compressibility
- High Dampening Action
- High Dielectric Strength
- High Oxidation Resistance*
- High Shearability without Breakdown
- High Temperature Serviceability*
- High Water Repellency
- Low Environmental Hazard
- Low Fire Hazard*
- Low Odor
- Low Reactivity
- Low Surface Energy
- Low Temperature Serviceability
- Low Toxicity
- Low Vapor Pressure
- Good Heat Stability*

---

*See "Contamination and Fire Prevention."

CAL PHYSICAL PROPERTIES

cal properties vary from lot to lot and should not be used for writing sales specifications.
IFICATION WRITERS! Before writing a specification, obtain Dow Corning's Sales Specification on the product.
act "Specification Coordinator, Dow Corning, Box 1767, Midland, MI, 48640."

pplied	*DOW CORNING 200 Fluid, 60,000 cSt.*	*DOW CORNING 200 Fluid, 100,000 cSt.*
arance	Crystal clear liquid free from suspended matter and sediment.	
ific Gravity @ 25°C	—	—
active Index @ 25°C	1.4036	1.4037
r, APHA	5	5
Point, open cup, °F	>620	>620
Number, BCP	trace	trace
Point, °C	-41	-33
ce Tension @ 25°C, dynes/cm	21.5	—
tile Content, @ 150°C, percent	0.23	0.30
osity Stability, @ 25°C, after 16 hr exposure @ 0°C, % change	-1.6	-2.4
osity Temperature Coefficient	0.61	0.61
fficient of Expansion, cc/cc/°C	0.00096	0.00096
rmal Conductivity @ 50°C, gm cal/cm · sec · °C	—	—
cific Heat @ 25°C, cal/gm/cm	—	—
bility Parameter*	7.4	7.4
bility in Typical Solvents,		
hlorinated solvents	High	High
romatic solvents	High	High
liphatic solvents	High	High
ry alcohols	Poor	Poor
ater	Poor	Poor
uorinated propellents	Poor	Poor
ectric Strength @ 25°C, volts/mil	400	—
me Resistivity @ 25°C, ohm-cm	1.0×10^{15}	—

ors Method: R.F. Fedors, Polymer Engineering And Science, Feb. 1974.
Corning does not routinely test all these physical properties. Users should independently test these properties when they are critical in the application.

LICATION INFORMATION

W CORNING 200 fluids, 60,000-,000 cSt., are not intended for d or medical use. They are nded for use by industrial manuurers. Typical end uses include:

- oatings Additive
- ampening Fluid
- lastomer and Plastics Lubricant
- lectrical Insulating Fluid
- oam Preventative or Breaker
- lousehold Products Ingredient
- ubricant and Grease Additive
- lechanical Fluid
- lold Release Agent
- il Field Chemical Defoamer
- etroleum Refinery Defoamer
- lastics Additive
- pecialty Chemical Products Ingredient
- urface Active Agent

CONTAMINATION AND FIRE PREVENTION

At elevated temperatures, DOW CORNING 200 fluids, 60,000-100,000 cSt., are sensitive to contamination by strong acids, bases, some metallic compounds and oxidizing agents. These contaminants may cause an accelerated rate of volatile by-product formation. Oxidizing agents can also cause an increase in fluid viscosity. When these conditions may exist, it is recommended that the flash point of the fluids be checked periodically to monitor operational safety. Also, ignitable conditions may exist if the fluid is giving off smoke.

Note: For answers to any questions regarding the efficacy, safety, health or environmental aspects of using DOW CORNING 200 fluids, 60,000-100,000 cSt., in any application, contact your nearest Dow Corning sales office or call the Dow Corning Customer Service.

Telephone:

In Michigan	(800) 292-2323
Outside Michigan	(800) 248-2345

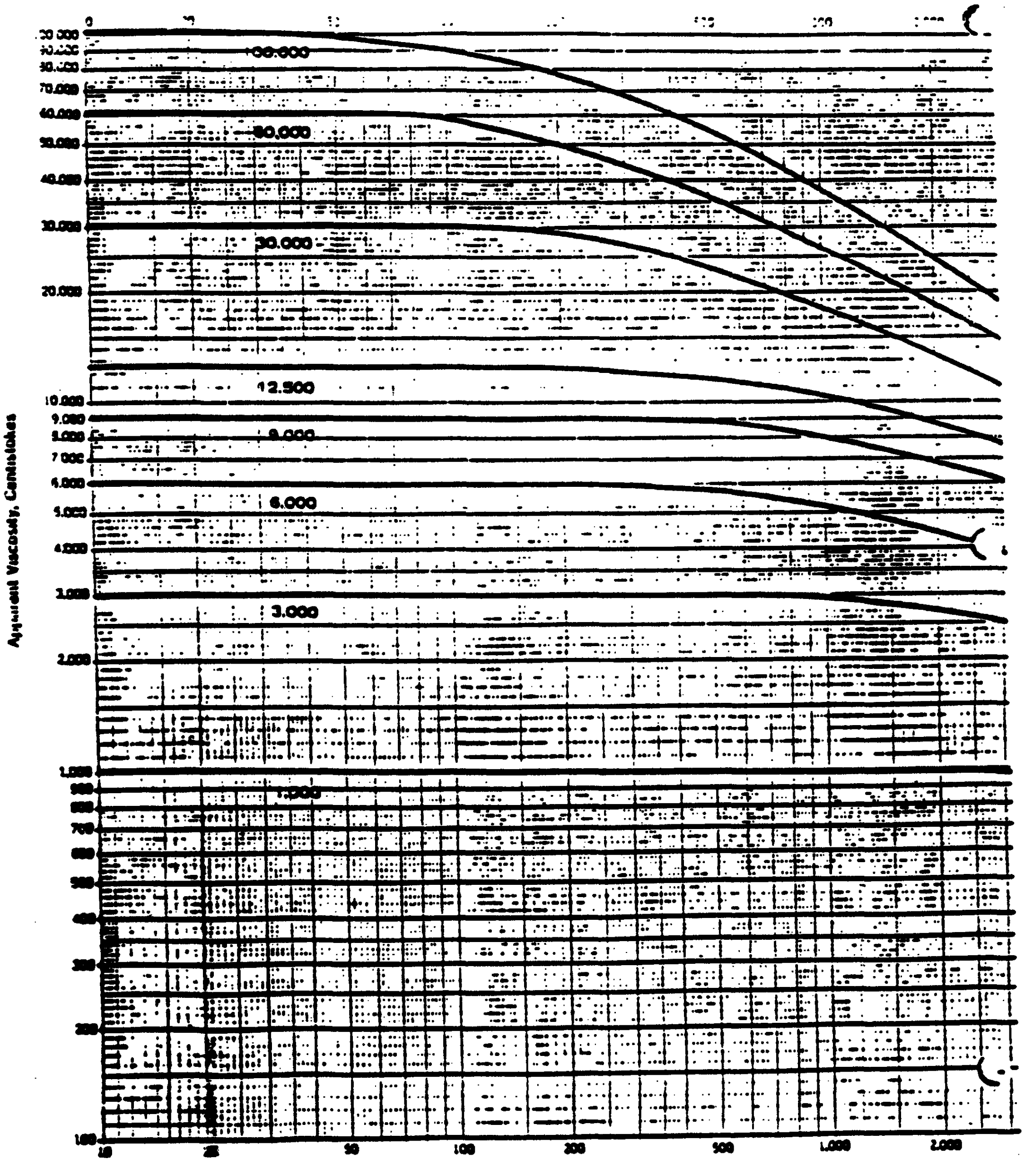
The information and data contained herein are based on information we believe reliable. You should thoroughly test any application, and independently conclude satisfactory performance before commercialization. Suggestions of uses should not be taken as inducements to infringe any particular patent.

OW CORNING CORPORATION
IDLAND, MICHIGAN 48640

ted in U.S.A. Form No. 22-930-82

DOW CORNING

Apparent Viscosity vs. Shear R
(Capillary Vis



DOW CORNING CORPORATION
MATERIAL SAFETY DATA SHEET

OW CORNING 200(R) FLUID, 10,000 CS

SECTION 1. CHEMICAL PRODUCT AND COMPANY IDENTIFICATION

Dow Corning Corporation
South Saginaw Road
Midland, Michigan 48686

24 Hour Emergency Telephone: (517) 496-5900
Product Information: (517) 496-6000
Product Disposal Information: (517) 496-5813
Transportation Information: (517) 496-8577
CHEMTREC: (800) 424-9300

MSDS No: 02113520 **Print Date: 11/12/93** **Last Revised: 09/28/93**

Generic Description: Silicone
Physical Form: Viscous Liquid
Color: Colorless
Odor: Odorless
NFPA Profile: Health NA Flammability 1 Reactivity 0

Note: NFPA = National Fire Protection Association

SECTION 2. HAZARDOUS COMPONENTS

CAS Number	Wt%	Component	Exposure Limits
None Known			

Comments: None present. This is not a hazardous material as defined in the OSHA Hazard Communication Standard.

SECTION 3. EFFECTS OF OVEREXPOSURE

Acute Effects

Eye: Direct eye contact may cause temporary discomfort with mild redness and dryness similar to windburn.

Skin: A single prolonged exposure (24 to 48 hours) causes no known adverse effect.

Inhalation: No irritation to eyes and respiratory passages. No injury is likely from relatively short exposures of less than 8 hours.

Oral: Small amounts transferred to the mouth by fingers during use, etc., should not injure. Swallowing large amounts may cause digestive discomfort.

Repeated Exposure Effects

Skin: None Known.

Inhalation: None Known.

Oral: None Known.

Special Hazards

This material contains the following components with the special hazards listed below.

Carcinogens

None Known

Teratogens

None Known

Mutagens

None Known

None Known

Sensitizers

None Known

Comments: When heated to temperatures above 150 degrees C in the presence of air, product can form formaldehyde vapors. Formaldehyde is a potential cancer hazard; a known skin and respiratory sensitizer; and an irritant to the eyes, nose, throat, skin, and digestive system. Safe handling conditions may be maintained by keeping vapor concentrations within the OSHA Permissible Exposure Limit for formaldehyde. Please read the additional information below.

The above listed potential effects of overexposure are based on actual data, results of studies performed up similar compositions component data and/or expert review of the product.

SECTION 4. FIRST AID MEASURES

Eye: Immediately flush with water.

Skin: No first aid should be needed.

Inhalation: No first aid should be needed.

Oral: No first aid should be needed.

Comments: Treat symptomatically.

SECTION 5. FIRE FIGHTING MEASURES

Flash Point (Method): 212.00 DEGREE F / 100.00 DEGREE C
Autoignition Temperature: Not Determined

Flammability Limits in Air: Not Determined

Extingushing Media: Carbon dioxide (CO2). Water. Water fog (or spray). Dry chemical. Foam.

Unsuitable Extinguishing Media: None

Fire Fighting Procedures: Self-contained breathing apparatus and protective clothing should be worn in fighting fires involving chemicals.

Unusual Fire Hazards: None

Hazardous Decomposition Products: Silicon dioxide. Carbon oxides and traces of incompletely burned carbon compounds. Formaldehyde.

SECTION 6. ACCIDENTAL RELEASE MEASURES

Containment/Clean-up: Disposal of collected product, residues, and cleanup materials may be governmentally regulated. Observe all applicable local, state, and federal waste management regulations. Mop up, or wi up, or soak up with absorbent and contain for salvage or disposal. For large spills, provide diking or other appropriat containment to keep material from spreading. Clean any remaini slippery surfaces by appropriate techniques, such as: several moppings or swabbings with appropriate solvents; washing with mild, caustic detergents or solutions; or high pressure steam f large areas. For nonsilicones, use typical industrial cleaning materials. Observe any safety precautions applicable to the cleaning material being used. Observe all personal protection equipment recommendations described in Sections 5 and 8. Local state, and federal reporting requirements may apply to spills o releases of this material into the environment. See applicable regulatory compliance information in Section 15.

ECTION 7. HANDLING AND STORAGE

andling: No special precautions.

torage: No special precautions. Use reasonable care.

ECTION 8. EXPOSURE CONTROLS/PERSONAL PROTECTION

ngineering Controls

ocal exhaust: None should be needed
eneral Ventilation: Recommended

ersonal Protective Equipment For Routine Handling

yes: Use proper protection - safety glasses as a minimum.

Skin: Washing at mealtime and end of shift is adequate.

Suitable Gloves: No special protection needed.

Inhalation: No respiratory protection should be needed.

Suitable Respirator: None should be needed.

Personal Protective Equipment For Spills

Eye: Use proper protection - safety glasses as a minimum.

Skin: Washing at mealtime and end of shift is adequate.

Inhalation/
Suitable Respirator: No respiratory protection should be needed.

Precautionary Measures: Avoid eye contact. Use reasonable care.

Comments: None

Note: These precautions are for room temperature handling. Use at elevated temperature, or aerosol/spray applications, may require added precautions.

SECTION 9. PHYSICAL AND CHEMICAL PROPERTIES

Physical form:	Viscous Liquid
Color:	Colorless
Odor:	Odorless
Specific Gravity @ 25C:	0.97
Viscosity:	10000.00 CST
Freezing/Melting Point:	Not Applicable.
Boiling Point:	Not Determined.
Vapor Pressure @ 25C:	Not Determined.
Vapor Density:	Not Determined.
Solubility in Water:	None.
pH:	Not Applicable.
Volatile content (Wt%):	Not Determined.

Note: The above information is not intended for use in preparing product specifications. Contact Dow Corning before writing specifications.

Chemical Stability:	Stable.
Hazardous Polymerization:	Hazardous polymerization will not occur.
Conditions to Avoid:	None.
Materials to Avoid:	Oxidizing material can cause a reaction.

Comments: None

SECTION 11. TOXICOLOGICAL INFORMATION

OPTIONAL SECTION - Complete information not yet available.

SECTION 12. ECOLOGICAL INFORMATION

OPTIONAL SECTION - Complete information not yet available.

SECTION 13. DISPOSAL CONSIDERATIONS

OPTIONAL SECTION - Complete information not yet available.

Call Dow Corning Environmental Mgmt. (517)496-6315, if more information is desired.

SECTION 14. TRANSPORT INFORMATION

DOT Information (49CFR 172.101)

Proper Shipping Name:	Not Available
Hazard Technical Name:	Not Available
Hazard Class:	Not Available
UN/NA Number:	Not Available
Packing Group:	Not Available

Call Dow Corning Transportation, (517)496-8577, if additional information is required.

SECTION 15. REGULATORY INFORMATION

Contents of this MSDS comply with the OSHA Hazard Communication Standard 29CFR 1910.120

TSCA Status: All chemical substances found in this product comply with the Toxic Substances Control Act inventory reporting requirements.

EPA SARA Title III Chemical Listings:

Section 302 Extremely Hazardous Substances:
None

Section 304 CERCLA Hazardous Substances:
None

Section 312 Hazard Class:

Acute:	N
Chronic:	N
Fire:	N
Pressure:	N
Reactive:	N

Y = Yes N = No

Section 313 Toxic Chemicals:
None present or none present in regulated quantities.

Supplemental State Compliance Information

CAS Number	Wt%	Component
Massachusetts		
No ingredient regulated by MA Right-to-Know Law present.		
New Jersey		
063148629	100	Polydimethylsiloxane
Pennsylvania		
063148629	100	Polydimethylsiloxane

SECTION 16. OTHER INFORMATION

Prepared by: Dow Corning Corporation

This information is offered in good faith as typical values and not as a product specification. No warranty, expressed or implied, is hereby made. The recommended industrial hygiene and safe handling procedures are believed to be generally applicable. However, each user should review these recommendations in the specific context of the intended use and determine whether they are appropriate.

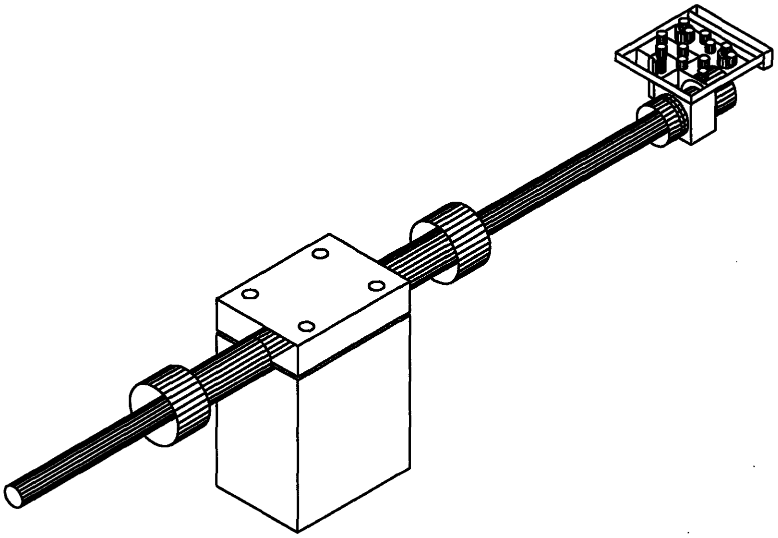
(R) indicates Registered or Trademark of the Dow Corning Corporation.

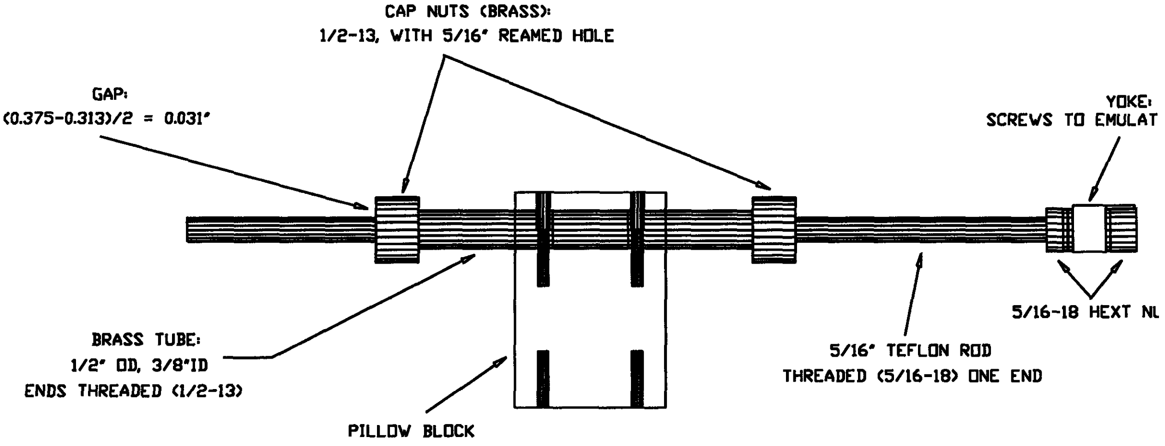
*** * * * *   This is the last page.   * * * * ***

Appendix D

Working Drawings

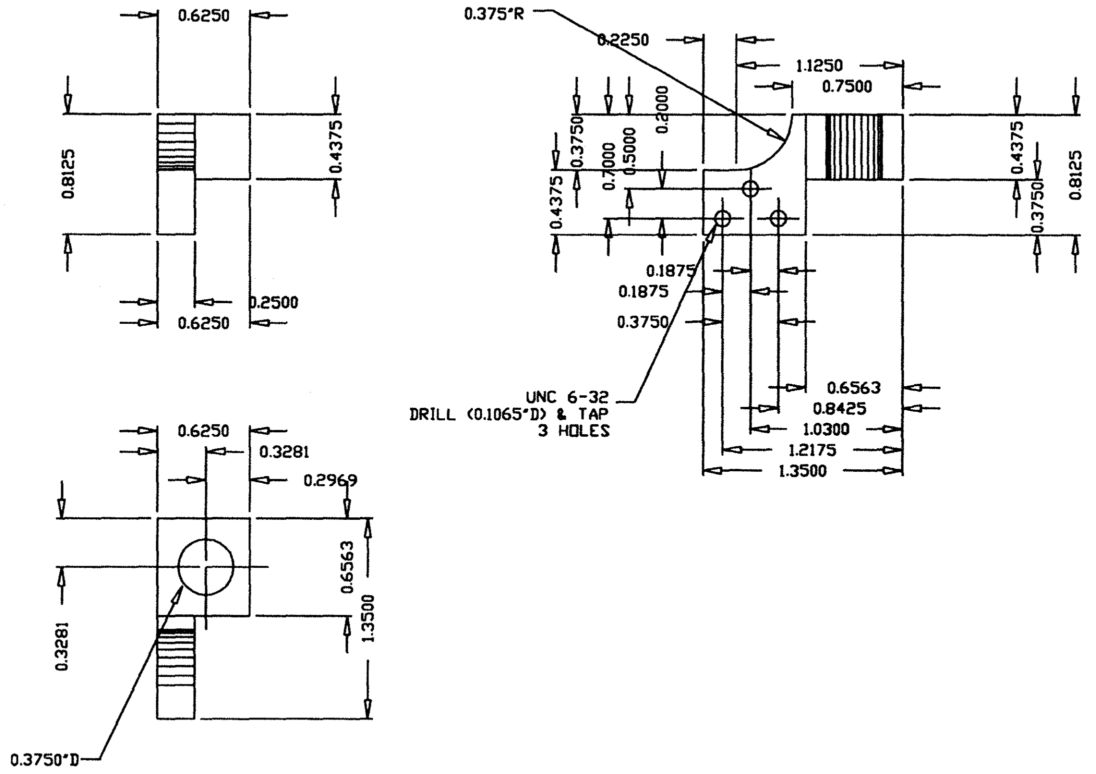
D.1 Physical Damper



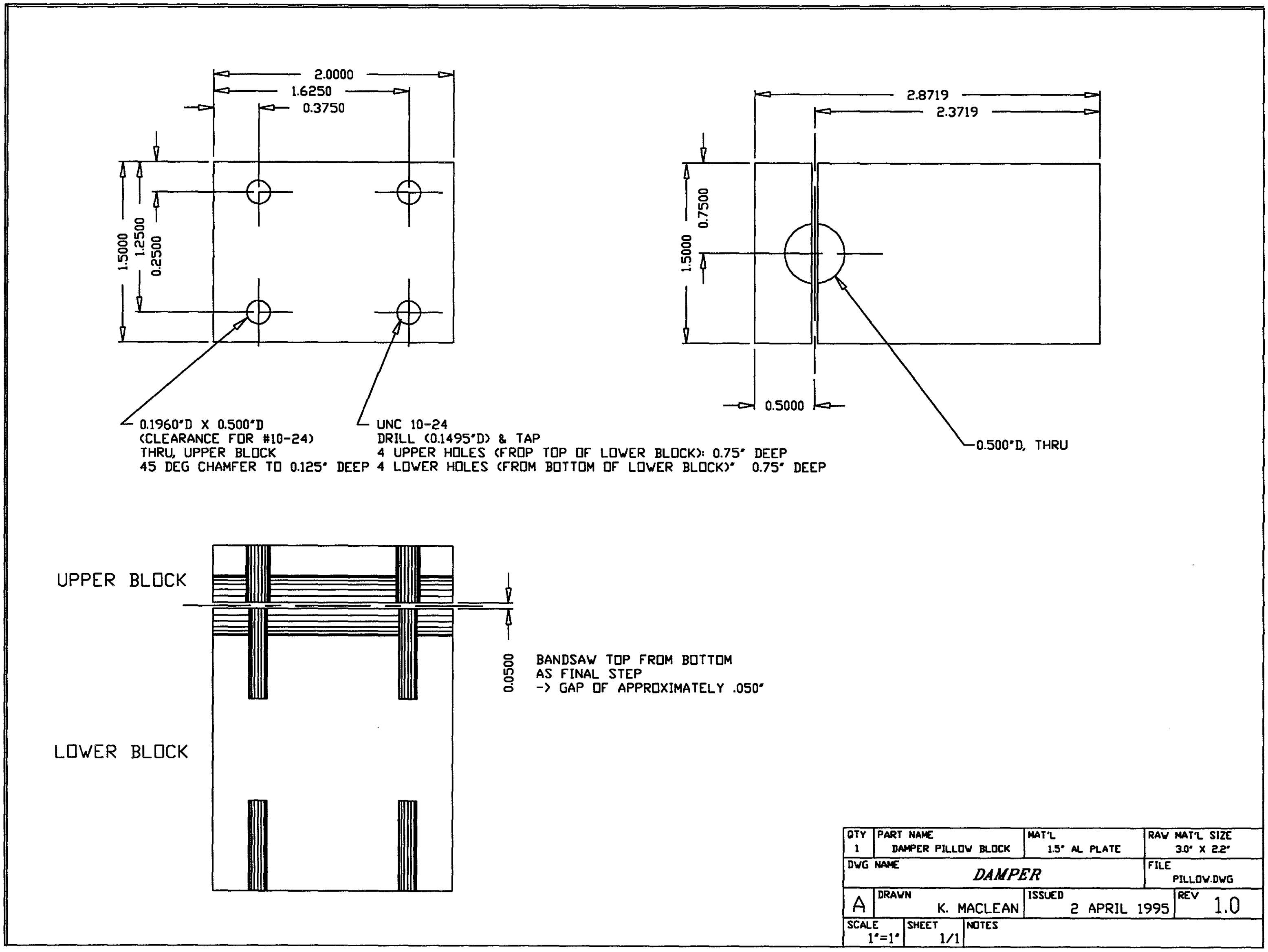


An assembly view of the mechanism designed and built to add physical damping to the emulation system. The damper is based on an inner shaft shearing high-viscosity fluid through a thin gap, maintained between a thin-walled outer stationary shell and a moving inner shaft. The fluid is retained inside the damper by a pair of cap nuts screwed onto either end of the outer shell, with clearance holes reamed to permit passage of the inner shaft. The inner shaft is constructed of teflon to reduce friction.

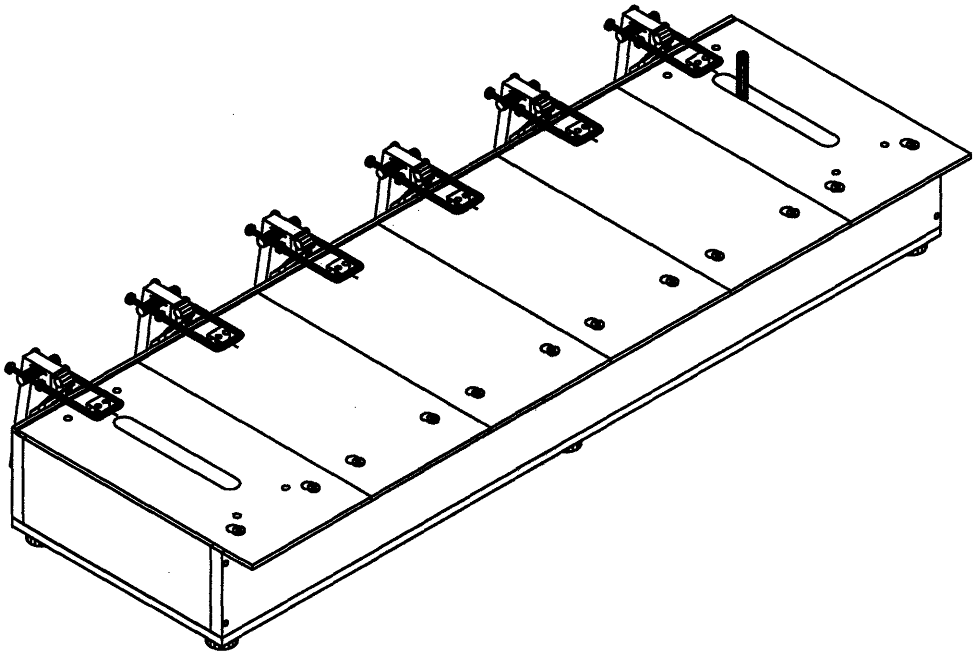
The outer shell is restrained in a pillow block mounted on the emulator base plate. The inner shaft is connected to the moving emulator carriage via a yoke. Working drawings for the yoke and pillow block are shown on the next two pages. All damper raw materials are commonly available, and were purchased through Small Parts, Inc.



QTY 1	PART NAME DAMPER YOKE	MAT'L 5/8" AL PLATE	RAW MAT'L SIZE 1.45" X 0.90"
DWG NAME	*DAMPER*		FILE YOKE.DWG
A	DRAWN K. MACLEAN	ISSUED 31 MARCH 1995	REV 1.0
SCALE 1"=1"	SHEET 1/1	NOTES	

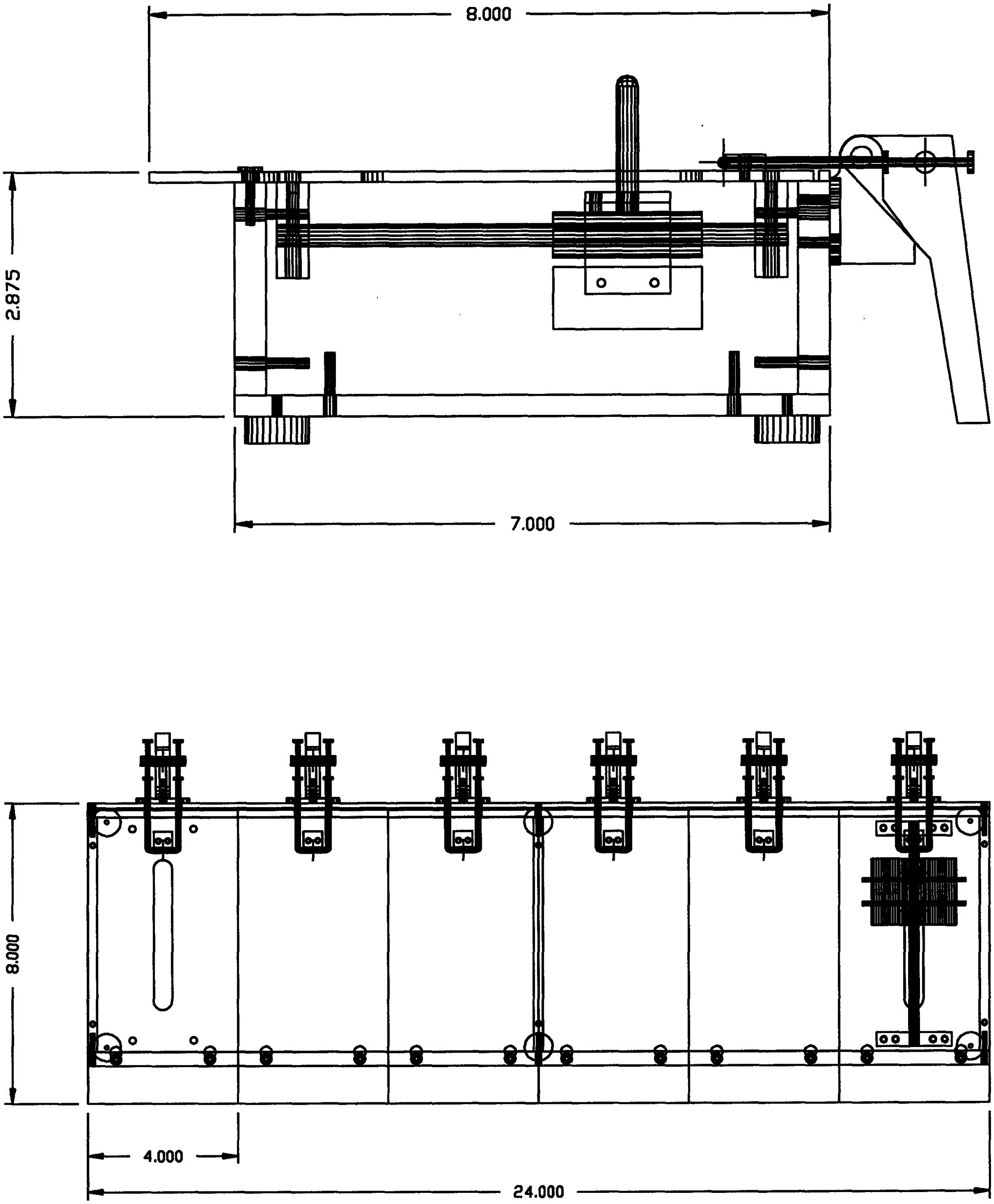

2.0000
1.6250
0.3750
1.5000
1.2500
0.2500
0.1960"D X 0.500"D
(CLEARANCE FOR #10-24)
THRU, UPPER BLOCK
45 DEG CHAMFER TO 0.125" DEEP
UNC 10-24
DRILL (0.1495"D) & TAP
4 UPPER HOLES (FROP TOP OF LOWER BLOCK): 0.75" DEEP
4 LOWER HOLES (FROM BOTTOM OF LOWER BLOCK)" 0.75" DEEP
2.8719
2.3719
1.5000
0.7500
0.5000
0.500"D, THRU
UPPER BLOCK
LOWER BLOCK
0.0500
BANDSAW TOP FROM BOTTOM
AS FINAL STEP
-> GAP OF APPROXIMATELY .050"
QTY 1
PART NAME DAMPER PILLOW BLOCK
MAT'L 1.5" AL PLATE
RAW MAT'L SIZE 3.0" X 2.2"
DWG NAME DAMPER
FILE PILLOW.DWG
A
DRAWN K. MACLEAN
ISSUED 2 APRIL 1995
REV 1.0
SCALE 1"=1"
SHEET 1/1
NOTES

D.2 Switch Box

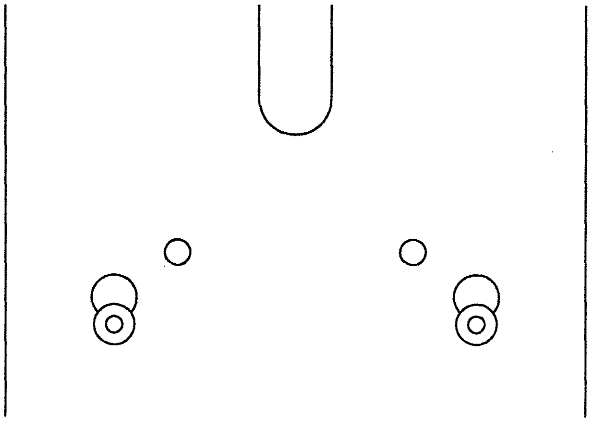


The switch box is shown partially loaded: a Masses module is installed in the far right slot, and an empty slotted switch-plate in the far left slot. The remaining slots have blank plates. A catch (approximately 1/4×1/2×3/4″) is mounted on the latch end of each switch-plate, and each plate hooks onto a pair of pegs mounted on the front wall of the box. It is held in tension by the latch which closes over the catch and pulls on the pegs. For rigidity, the switch box frame is constructed entirely out of aluminum.

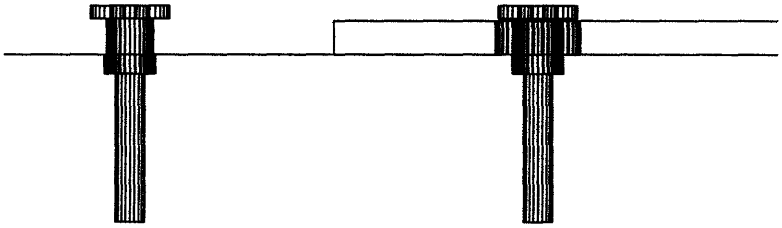
The switch box latches (pull-action toggle, U-hook style 324) were purchased from De-Sta-Co, an MSC vendor, part # 6966717 ($9.90 each) All other parts were generic or constructed in-house.



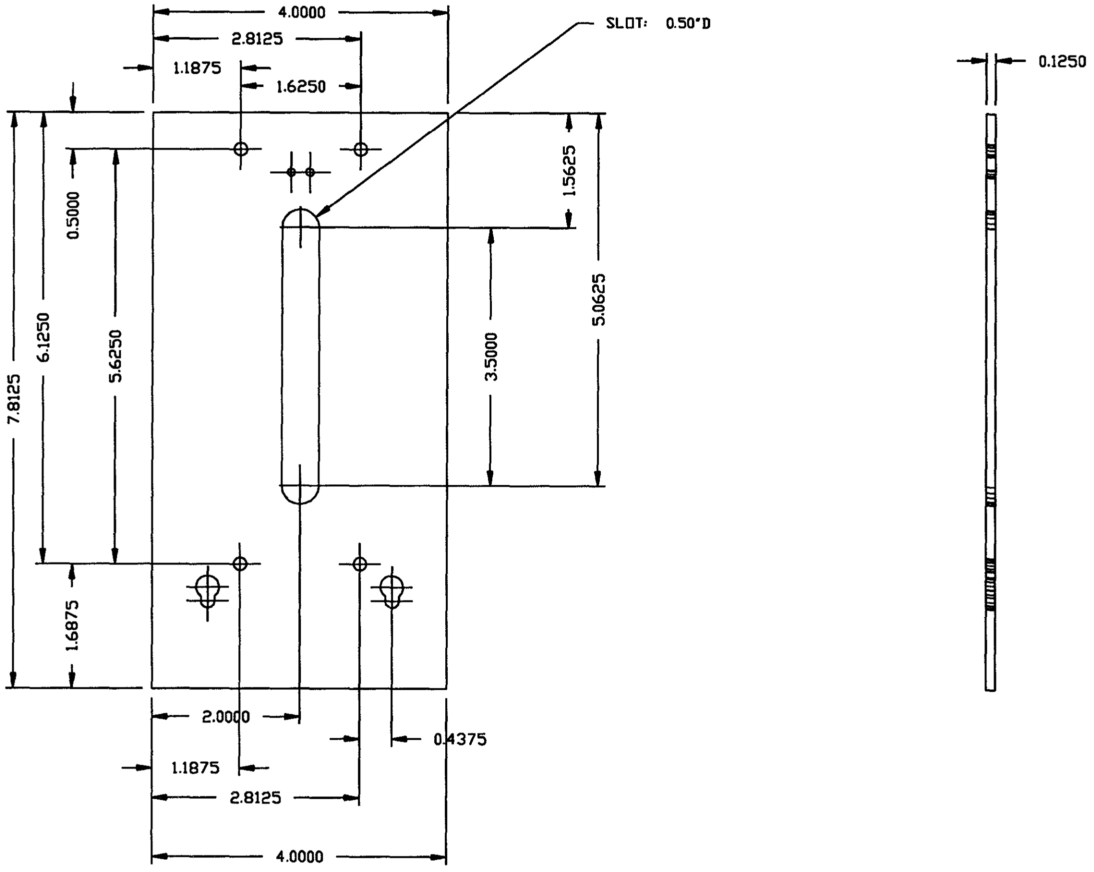
Top and end views of the switch box, illustrating key dimensions and latching mechanism (note that hidden lines are not removed).



The keyhole/post structure is illustrated in close-up in top (above) and front (below) views. A pair of keyhole-shaped holes are machined into each switch plate; these holes fit over a purchased nylon spacer which has a wider top than base. The retaining posts are inset into a recess in the switchbox's front panel and held with screws. The keyhole's larger diameter fits over the post's top, but its smaller diameter only fits around the post's base.

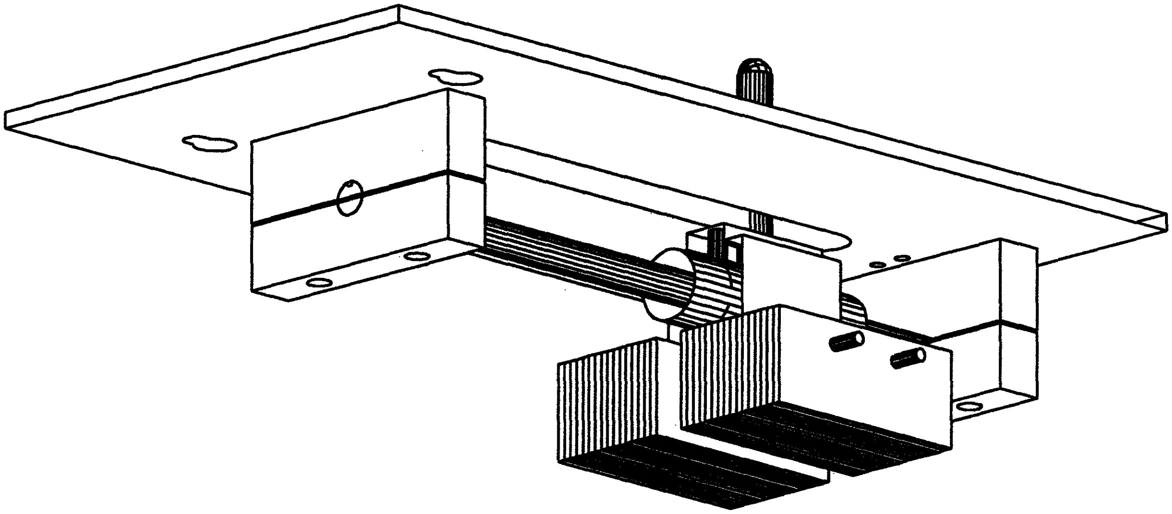


NOTE: SEE ST_SPBL.DWG FOR BLANK PLATE DIMENSIONS
(SHOWN ARE MODIFICATIONS FOR MASS MODULE, ONLY)



QTY 8	PART NAME MASS MODULE PLATE	MAT'L 1/8" AL PLATE	RAW MAT'L SIZE 4.25" X 8.0"
DWG NAME **PURE MASS MODULE**			FILE MS_SWPL.DWG
A	DRAWN K. MACLEAN	ISSUED 12 FEBRUARY 1995	REV 1.0
SCALE 1"=2"	SHEET 1/1	NOTES BEGIN WITH SWITCHPLATE BLANK	

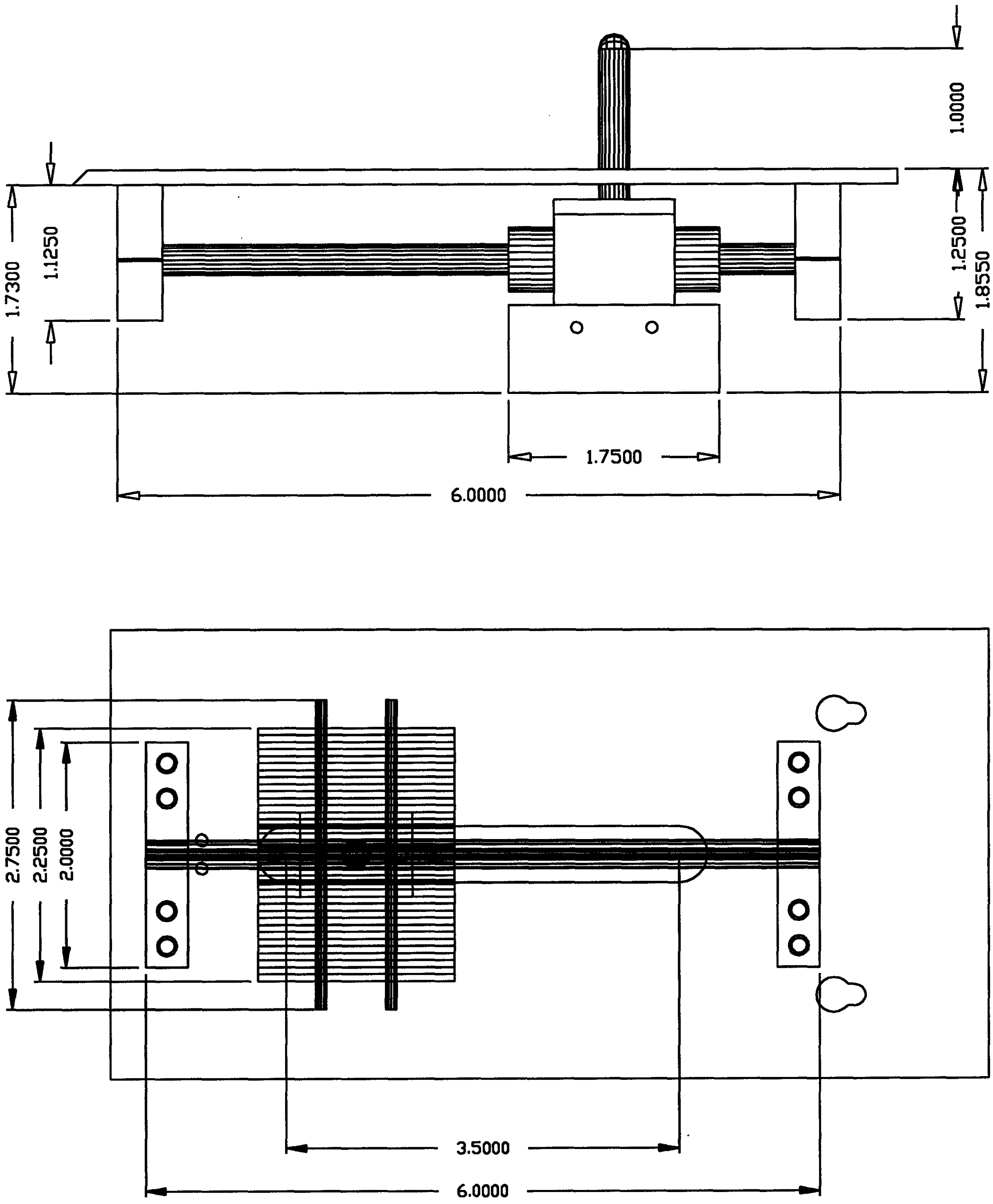
D.3 Real Masses



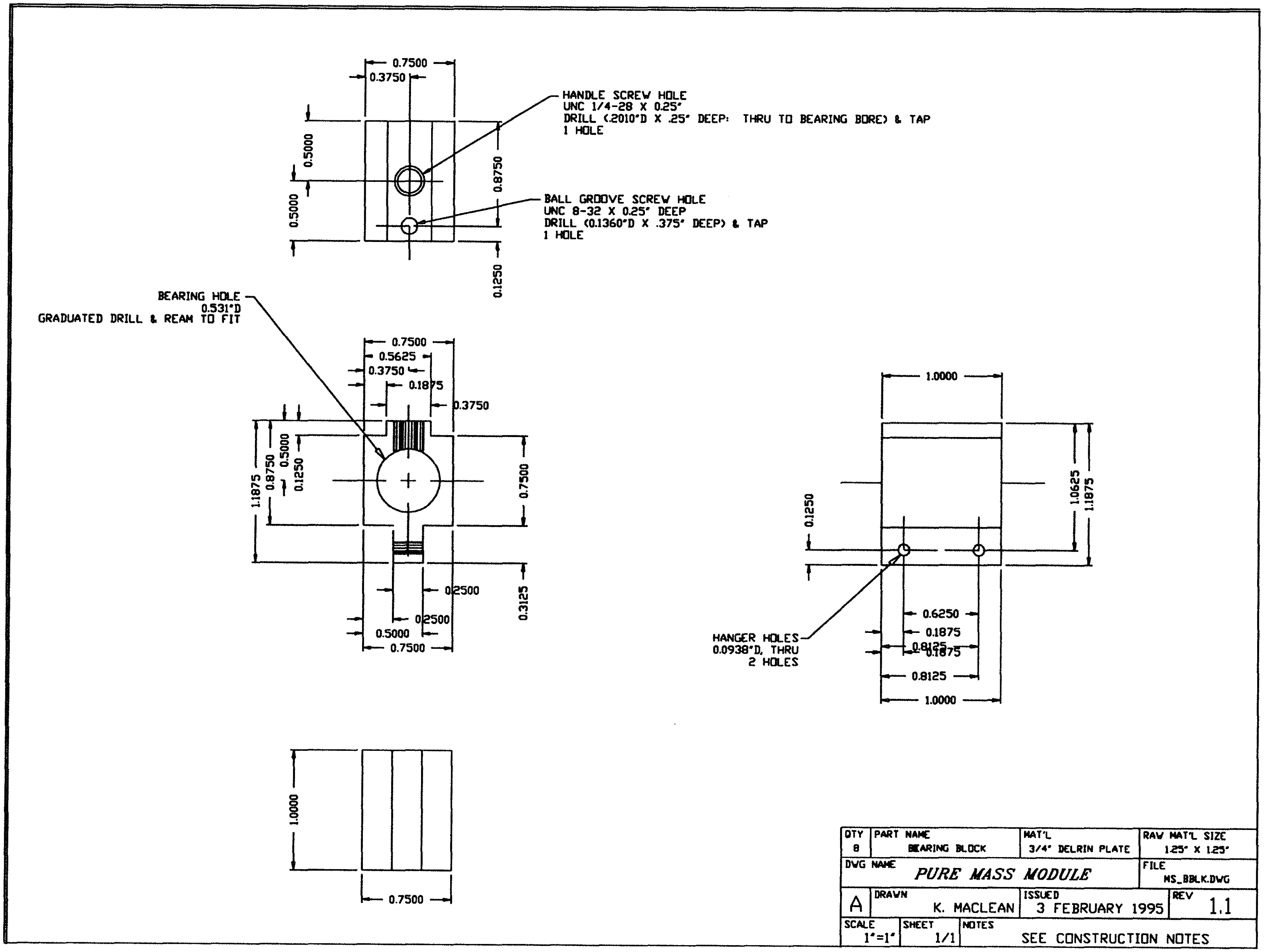
The mass module's operation is based on a cartridge of rolling ball bearings which roll on a single horizontal ball-groove shaft, supporting a variable number of lead slugs which can be adjusted to set the module's total mass. The bearing/shaft assembly is supported by shaft supports screwed onto the bottom of the switch plate, and the ball groove prevents lateral rotation of the bearing assembly on the shaft. The entire moving mass with no lead slugs is about 35 grams. Key assembly dimensions are shown on the next page, and working drawings for several constructed components (the bearing block, shaft supports and lead slugs) follow.

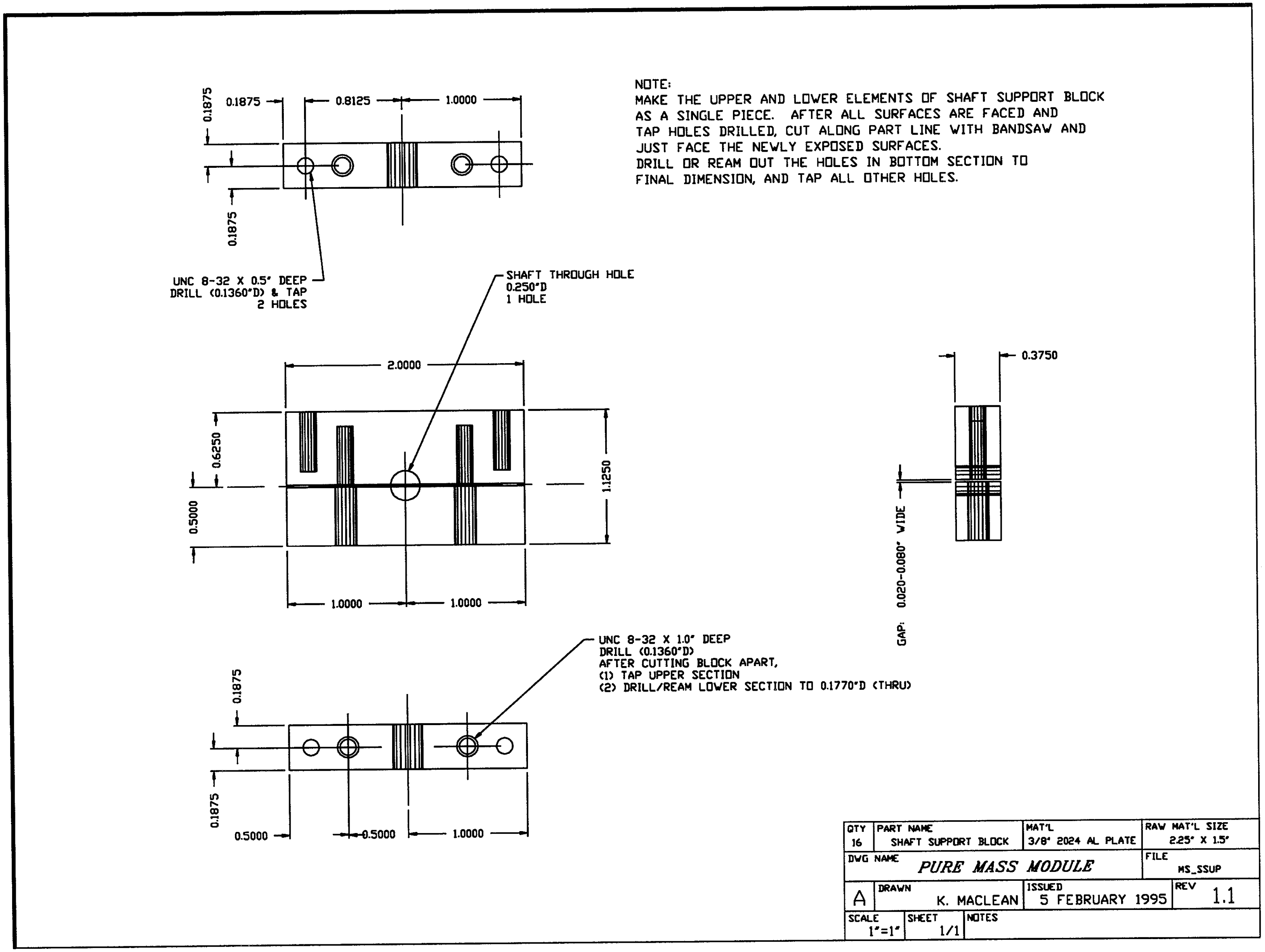
Important purchased elements include:

Item	Description	Vendor	Part No.	Unit Cost
Bearings	Twin 1/4″ linear bearing cartridges	Thomsen	SCB-4-TWN	27.08
Ball groove shaft	1/4″×6″ hardened case steel	Thomsen	—	4.50
Ball groove screws	1/4″ shaft, 8-32	Thomsen	BGS-1	0.38
Lead	0.062″×3″×6″ sheet	Small Parts	Q-XR-62	2.56



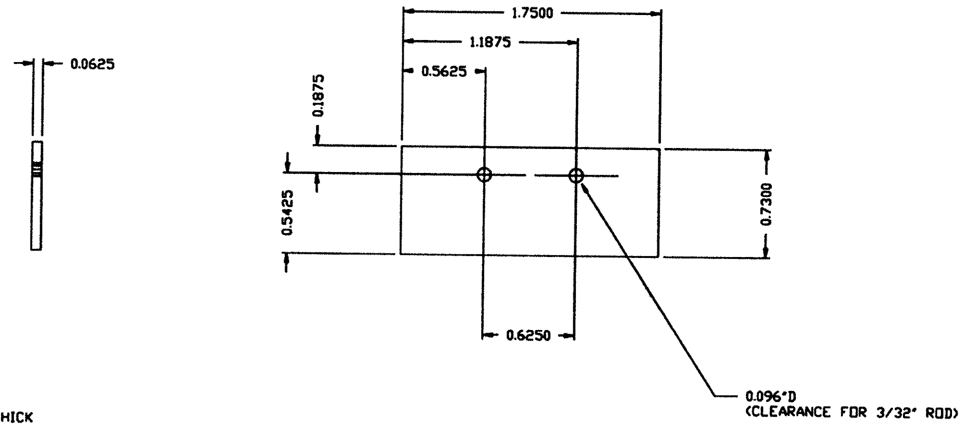
Side and top views of the mass module assembly, showing key mechanisms and dimensions. Note that the two views do not correspond in direction; the latch end is at right for the upper drawing, and at left for the bottom drawing.


HANDLE SCREW HOLE
UNC 1/4-28 X 0.25"
DRILL (.2010"D X .25" DEEP: THRU TO BEARING BORE) & TAP
1 HOLE
BALL GROOVE SCREW HOLE
UNC 8-32 X 0.25" DEEP
DRILL (0.1360"D X .375" DEEP) & TAP
1 HOLE
BEARING HOLE
0.531"D
GRADUATED DRILL & REAM TO FIT
HANGER HOLES
0.0938"D, THRU
2 HOLES
0.7500
0.3750
0.5000
0.5000
0.8750
0.1250
0.7500
0.5625
0.3750
0.1875
0.3750
1.1875
0.8750
0.5000
0.1250
0.7500
0.2500
0.2500
0.3125
0.5000
0.7500
1.0000
0.7500
1.0000
0.1250
1.0625
1.1875
0.6250
0.1875
0.8125
0.1875
0.8125
1.0000
QTY
8
PART NAME
BEARING BLOCK
MAT'L
3/4" DELRIN PLATE
RAW MAT'L SIZE
1.25" X 1.25"
DWG NAME
PURE MASS MODULE
FILE
MS_BBLK.DWG
A
DRAWN
K. MACLEAN
ISSUED
3 FEBRUARY 1995
REV
1.1
SCALE
1"=1"
SHEET
1/1
NOTES
SEE CONSTRUCTION NOTES



NOTE:
MAKE THE UPPER AND LOWER ELEMENTS OF SHAFT SUPPORT BLOCK AS A SINGLE PIECE. AFTER ALL SURFACES ARE FACED AND TAP HOLES DRILLED, CUT ALONG PART LINE WITH BANDSAW AND JUST FACE THE NEWLY EXPOSED SURFACES.
DRILL OR REAM OUT THE HOLES IN BOTTOM SECTION TO FINAL DIMENSION, AND TAP ALL OTHER HOLES.

QTY	PART NAME	MAT'L	RAW MAT'L SIZE
16	SHAFT SUPPORT BLOCK	3/8" 2024 AL PLATE	2.25" X 1.5"
DWG NAME	*PURE MASS MODULE*		FILE MS_SSUP
A	DRAWN K. MACLEAN	ISSUED 5 FEBRUARY 1995	REV 1.1
SCALE 1"=1"	SHEET 1/1	NOTES	



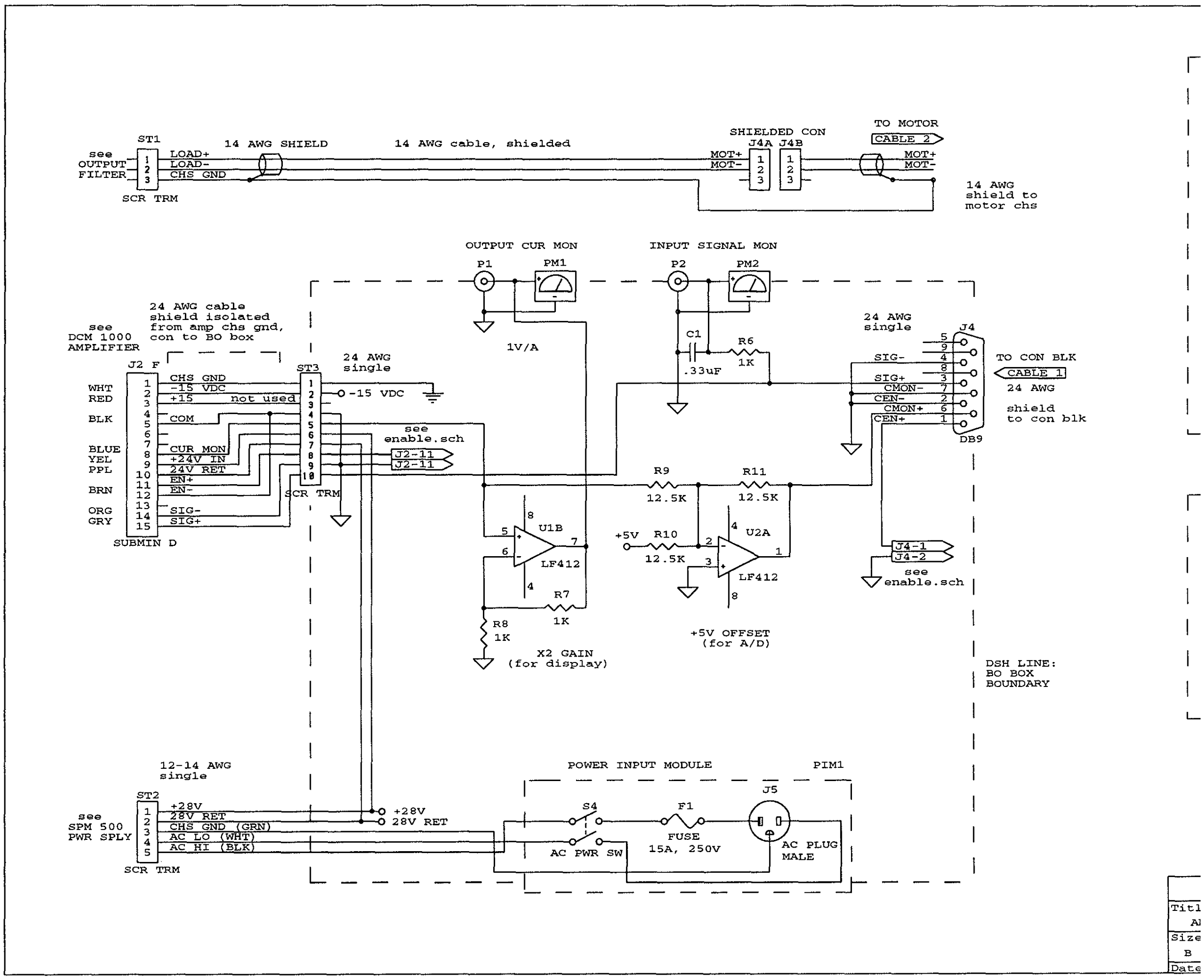
NOTE:
(1) STACK SHEETS UP TO 1" THICK
(2) BANDSAW TO SIZE, USING HIGH SAWSPEED, HIGH FEEDRATE, WOOD UNDERNEATH
(3) DRILL IN MILL WITH HIGH DRILL SPEED, VERY SLOW FEED RATES
(4) HAND FINISH (FILE) AND HAMMER FLAT

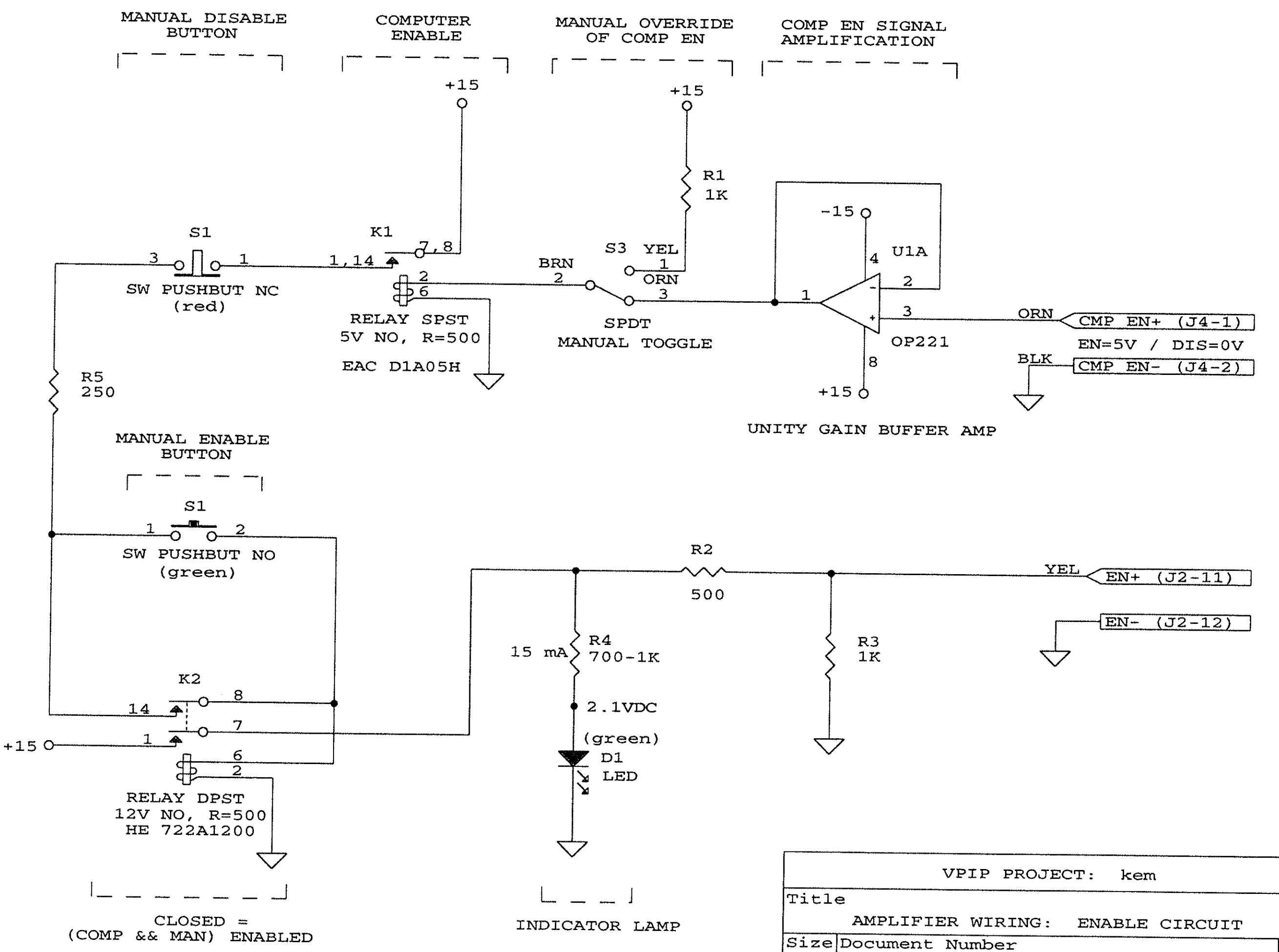
QTY 160	PART NAME 1/2 OZ LEAD SLUG	MAT'L 1/16" LEAD SHEET	RAW MAT'L SIZE 3" X 6"
DWG NAME	*PURE MASS MODULE*		FILE MS_SLUG.DWG
A	DRAWN K. MACLEAN	ISSUED 31 MAY 1995	REV 1.0
SCALE 1"=1"	SHEET 1/1	NOTES 12 SLUGS PER 3X6" SHEET	

Appendix E

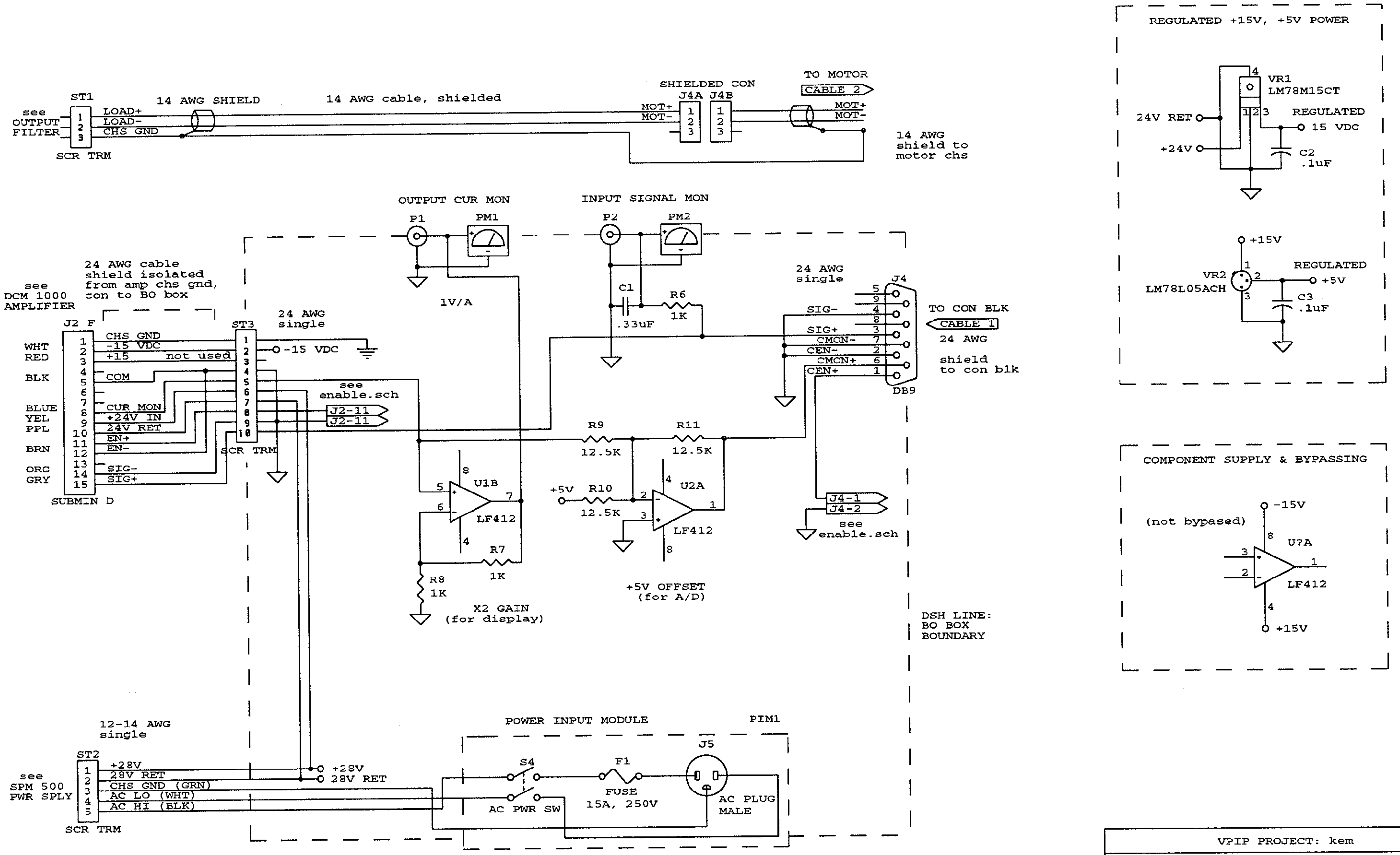
Electronic Schematics

E.1 PWM Servo Monitor and Enabling Circuitry


ST1
14 AWG SHIELD
14 AWG cable, shielded
SHIELDED CON
J4A J4B
TO MOTOR
CABLE 2
see
OUTPUT
FILTER
LOAD+
LOAD-
CHS GND
SCR TRM
MOT+
MOT-
MOT+
MOT-
14 AWG
shield to
motor chs
OUTPUT CUR MON
INPUT SIGNAL MON
P1
PM1
P2
PM2
24 AWG cable
shield isolated
from amp chs gnd,
con to BO box
see
DCM 1000
AMPLIFIER
J2 F
ST3
24 AWG
single
1V/A
C1
.33uF
R6
1K
24 AWG
single
J4
TO CON BLK
CABLE 1
24 AWG
shield
to con blk
DB9
WHT
RED
BLK
BLUE
YEL
PPL
BRN
ORG
GRY
CHS GND
-15 VDC
+15
not used
COM
CUR MON
+24V IN
24V RET
EN+
EN-
SIG-
SIG+
SUBMIN D
-15 VDC
see
enable.sch
J2-11
J2-11
SCR TRM
SIG-
SIG+
CMON-
CEN-
CMON+
CEN+
R9
12.5K
R11
12.5K
U1B
LF412
+5V
R10
12.5K
U2A
LF412
J4-1
J4-2
see
enable.sch
R7
1K
R8
1K
X2 GAIN
(for display)
+5V OFFSET
(for A/D)
DSH LINE:
BO BOX
BOUNDARY
12-14 AWG
single
ST2
see
SPM 500
PWR SPLY
+28V
28V RET
CHS GND (GRN)
AC LO (WHT)
AC HI (BLK)
SCR TRM
+28V
28V RET
POWER INPUT MODULE
PIM1
S4
F1
J5
AC PWR SW
FUSE
15A, 250V
AC PLUG
MALE
Titl
A
Size
B
Date



VPIP PROJECT: kem		
Title AMPLIFIER WIRING: ENABLE CIRCUIT		
Size A	Document Number AMPWIRE\ENABLE.SCH	REV 1.0
Date: October 9, 1994	Sheet 1 of 1	



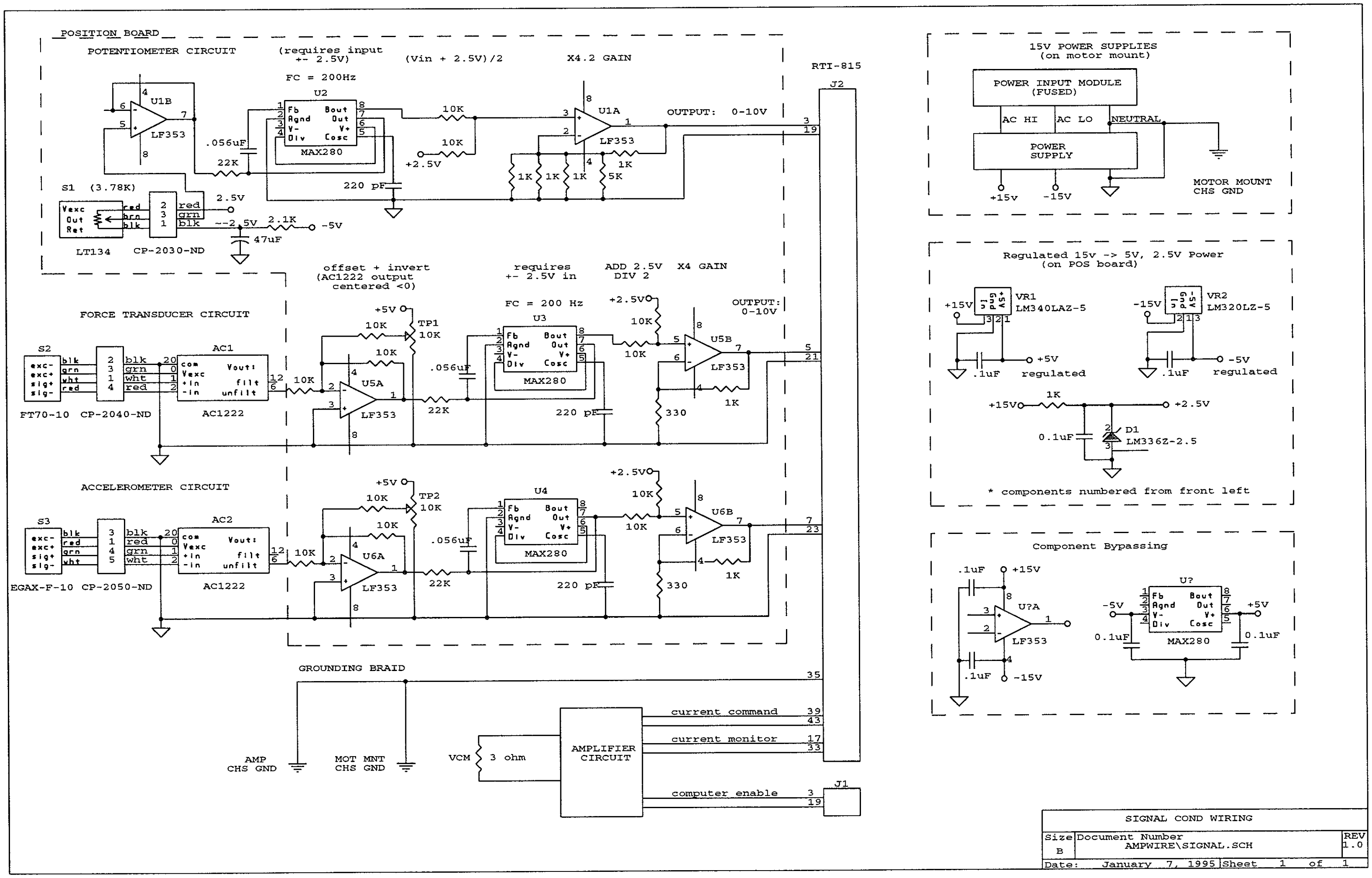
VPIP PROJECT: kem

Title
AMPLIFIER WIRING: BREAKOUT CONNECTIONS

Size	Document Number	REV
B	AMPWIRE\BREAKOUT.SCH	1.0

Date: January 21, 1995 Sheet 1 of 1

E.2 Signal Conditioning


POSITION BOARD
POTENTIOMETER CIRCUIT
(requires input +- 2.5V)
FC = 200Hz
(Vin + 2.5V)/2
X4.2 GAIN
U1B
LF353
U2
Fb Agnd V- Div
Bout Out V+ Cosc
MAX280
.056uF
22K
10K
10K
+2.5V
220 pF
U1A
LF353
OUTPUT: 0-10V
1K
1K
1K
5K
1K
S1 (3.78K)
Vexc Out Ret
red brn blk
red grn blk
2.5V
--2.5V
2.1K
-5V
47uF
LT134
CP-2030-ND
RTI-815
J2
offset + invert (AC1222 output centered <0)
requires +- 2.5V in
ADD 2.5V DIV 2
X4 GAIN
FORCE TRANSDUCER CIRCUIT
S2
exc- exc+ sig+ sig-
blk grn wht red
FT70-10
CP-2040-ND
AC1
com Vexc +in -in
Vout: filt unfilt
AC1222
+5V
TP1
10K
U5A
LF353
FC = 200 Hz
U3
MAX280
U5B
LF353
OUTPUT: 0-10V
330
ACCELEROMETER CIRCUIT
S3
blk red grn wht
EGAX-F-10
CP-2050-ND
AC2
AC1222
TP2
U6A
U4
U6B
GROUNDING BRAID
AMP CHS GND
MOT MNT CHS GND
VCM
3 ohm
AMPLIFIER CIRCUIT
current command
current monitor
computer enable
J1
15V POWER SUPPLIES (on motor mount)
POWER INPUT MODULE (FUSED)
AC HI
AC LO
NEUTRAL
POWER SUPPLY
+15v
-15V
MOTOR MOUNT CHS GND
Regulated 15v -> 5V, 2.5V Power (on POS board)
VR1
LM340LAZ-5
+5V regulated
VR2
LM320LZ-5
-5V regulated
.1uF
+15V
1K
+2.5V
0.1uF
D1
LM336Z-2.5
* components numbered from front left
Component Bypassing
U?A
LF353
U?
0.1uF
SIGNAL COND WIRING
Size B
Document Number AMPWIRE\SIGNAL.SCH
REV 1.0
Date: January 7, 1995
Sheet 1 of 1

Appendix F

Experiment Documents

F.1 COUHES Application, Consent Form and Approval

Following are the application made for these experiments to MIT's Committee on the Use of Humans as Experimental Subjects (COUHES) and COUHES' approval. The informed consent form which all subjects signed is included in the COUHES application. Actual signed forms have been kept on file.

Application No.__________

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Committee on the Use of Humans as Experimental Subjects

Application for Approval to Use Humans as Experimental Subjects

Part I.

DATE: 31 January, 1995

Title of Study:	Emulation of Haptic Feedback in Manual Interfaces
Principal Investigator:	Karon E. MacLean (Doctoral Candidate)
Department:	Mechanical Engineering
Room No.:	3-147
Telephone No.:	253-8114
Associated Investigators:	Prof. William K. Durfee *(formerly of MIT, Mechanical Engineering)* Dept. of Mechanical Engineering University of Minnesota 111 Church St., SE Minneapolis, MN 55455 (612) 625-0099
Collaborating Institutions:	N/A
Financial Support:	Brit and Alex d'Arbeloff Endowment Fund:

Purpose of Study:

Definition: "Haptic" is the perception of touch, body forces and motion.

Background: Haptic aspects of human-machine interfaces comprise an important but often neglected source of sensory feedback. Mechanical feedback to the haptic sense remains more difficult to understand in the course of designing a user interface than other interface aspects. In part this is because the computer tools which have promoted an iterative prototyping design strategy in other respects do not allow hands-on manipulation of the prototype. Such physically interactive prototyping tools will depend upon the new technique of computer-controlled *emulation* of haptic feedback.

Objectives: Understanding haptic feedback will facilitate its systematic design in electromechanical manual interfaces. This study will be used to identify the parameters which influence the *fidelity* of a haptic emulation - that is, the degree to which an emulated haptic phenomena is perceived by human subjects to be similar to the real (non-emulated) environment being imitated - and quantify their interdependence with each other and with human cognition.

Approach: An experimental haptic emulation system has been designed and built for the purpose of this study. This *haptic emulator* consists of a computer-controlled linear motor, with which the human user interacts via a mounted handle. The motor is computer-controlled to emulate specific physical environments (e.g. a linear spring or mass; a toggle or slider switch) which can be "felt" by the user through the handle.

Nature of experiments: The subject is asked, in a systematic manner, to compare the feel of these emulations with that of the real devices (e.g. light switches and miniature power switches) which the emulations are intended to represent. To do this, the subject manipulates the real devices and the handle of the haptic emulator, all of which are mounted on a panel in front of him or her.

Part II.

Please provide an outline of the actual experiments to be performed, including, where applicable, detailed information as to .., experimental devices and procedures, ..., physical or emotional stress, and the appropriate measures you are planning to take.

Experimental Protocol

Description of Experimental Apparatus:

The experimental hardware consists of a linear motor with a handle mounted on it, sensors and signal conditioning, servo amplifier and a computer controller. The computer controller samples the sensors in realtime and outputs a command to the servo amp and actuator, based on an internal model of the physical virtual environment being emulated. A user interacts with the motor via the handle, and through the forces exerted on his/her hand, feels the modeled environment. The total travel of the handle is about 4 inches.

Nature and Aim of Experiments:

The objective of this study is to understand what combinations and levels of *emulation parameters* - e.g. sampling rate, controller structure, model structure, various hardware variables - increase fidelity in the emulation of manual feedback. Thus, the experimental model used is to model and emulate a specific real device, and then ask a subject to compare the emulation with a set of real devices which are more or less similar in manual feedback characteristics to each other and to the emulation.

Execution of Experiments

In successive trials, emulation parameters are randomly varied. The subject's ability to discriminate which of the real devices is being emulated during this variation of emulation parameters is an indication of the emulation's degradation or improvement, and implies the influence of the given emulation parameter on the emulation fidelity.

A given trial will last about 1 minute, and an entire session will last no more than 60 minutes. With the subject's permission, some sessions maybe be tape or video recorded.

Part III.

Please answer each question below, and indicate "NA" where not applicable to your application. Positive answers should be briefly explained, with detailed information included in Part II.

1.	How will subjects be obtained? Number of subjects needed? Age(s) of subjects?	Verbal recruitment 8-15 18-65 yrs.
2.	Will subjects receive any payment or other compensation for participation?	No
3.	Will your subjects be studied outside MIT premises?	No
4.	Will the facilities of the Clinical Research Center be used?	No
5.	Will drugs be used?	No
6.	Will radiation or radioactive materials be employed?	No
7.	Will special diets be used?	No
8.	Will subjects experience physical pain or stress?	No
9.	Will a questionnaire be used?	No
10.	Are personal interviews involved?	No.
11.	Will subjects experience psychological stress?	No
12.	Does this study involve planned deception of subjects?	No
13.	Can information acquired through this investigation adversely affect a subject's relationships with other individuals?	No

14. Please explain how subjects anonymity will be protected, and/or confidentiality of data will be preserved.

The records of the study will be kept private. In published reports and oral presentations, no information will be included which might make it possible to identify a particular subject.

Audio and/or visual taping: Subjects will not be requested to identify themselves by name or circumstance on the tape. The records will be reviewed only by the investigator, and when not in use will be kept locked in the investigator's office. Following the study, the taped records will be destroyed.

Part IV.

A. *Please summarize the risks to the individual subject, and explain how you propose to deal with these risks.*

In this experimental setup, a series of redundant precautions are taken to ensure the subject's absolute safety at all times while he or she manipulates the motor handle. They are detailed as follows:

(a) *Safe usage of hardware:* All emulations are designed, debugged and thoroughly tested prior to use in an experiment, such that any human user will be physically unable to destabilize the controller and/or cause the motor to respond in an unexpected and/or unsafe manner.

(b) *Mechanical design of interaction.* The user will at no point be able to place any part of his/her body in a position where injury (pinching, etc.) might occur, should the motor move in an abrupt and unexpected manner.

(c) *Manual disabling of servo amplifier:* The servo amplifier which supplies electrical power to the actuator can be disabled both manually and through computer control. Instantaneous manual disabling of the servo amplifier is possible at all times, overriding any computer enabling in effect. The manual disable control will be in easy reach of the investigator throughout any interaction of a subject with the emulator.

(d) *Automatic computer disabling of servo amplifier:* Motor current, position, acceleration and applied force are sensed and velocity is derived, and all parameters are processed in realtime. When any of these parameters exceed established safe thresholds, the servo amplifier is immediately (within 0.2 milliseconds) disabled, overriding any manual enabling in effect.

As a result of these precautions, and to the best knowledge and judgment of the investigators, there is virtually no likelihood of harm, either physical or psychological, to subjects either during the course of an experiment or manifested later as a result of participation in the experiment.

B. *Detection and reporting of harmful effects: If applicable, please describe what follow-up efforts will be made to detect any harm to subjects, and how this Committee will be kept informed.*

In the unlikely event of physical harm to a subject (psychological harm is not deemed a concern), the full extent of the harm would be immediately obvious and identifiable. The injury would be treated at M.I.T. Medical and reported to COUHES; and the experimentation modified or discontinued until assurance could be made that such harm would not be repeated.

Part V.

INFORMED CONSENT MECHANISMS: see Appendix A for the proposed Statement of Informed Consent.

Signatures

Signature of Principal Investigator: ____________________ Date: ________

Print Full Name: ____________________

Signature of Department Head : ____________________ Date: ________

Print Full Name: ____________________

Appendix A: Statement of Informed Consent

Study: *Emulation of Haptic Feedback in Manual Interfaces*

Principal Investigator: Karon E. MacLean

Name of Subject:_______________________________ Date(s) of Experiment:_______________________

The experiments that we are performing are concerned with understanding how to make a computer-controlled motor imitate the way manual interfaces (e.g., knobs and switches) feel. The long-term goal of this research is to learn how to design better manual interfaces through advancement of haptic (the sense whereby we perceive body forces and motions) rapid prototyping technology.

Definition: An *emulation* is the computer-controlled motor with a handle mounted on it, being controlled such that when you push on the handle, it feels more or less like some real manual device (such as a switch or slider).

In these experiments, you will be presented with a series of different *emulations*, and asked questions which compare the feel of the emulation with the feel of several real switches or sliders. A given session will last somewhere between a few minutes and an hour.

With your permission, the experimental session may be tape and/or video recorded. The recording will be used and heard solely by the experimenter and P.I., as an aid to interpreting the data taken during the session.

You should not at any point feel any discomfort or fatigue (aside from boredom). If you ever feel discomfort or fatigue, including postural, etc., you are encouraged to report this (as well as any other subjective impression of the experiment) to the experimenter, and the experimental protocol and/or setup will be modified accordingly.

Further information about the experiments will be provided to the subject by the experimenter at any time the subject wishes.

Complete anonymity will be maintained for you when we report the results of the experiments.

Your participation in this research project is voluntary. You are free to withdraw your consent and to discontinue participation in the project at any time without prejudice.

CONSENT:

I have read and understood the above statement and agree to participate as a subject in these experiments.

In the unlikely event of physical injury resulting from participation in this research, I understand that medical treatment will be available from the M.I.T. Medical Department, including first aid emergency treatment and follow-up care as needed, and that my insurance carrier may be billed for the cost of such treatment. However, no compensation can be provided for medical care apart from the foregoing. I further understand that making such medical treatment available; or providing it, does not imply that such injury is the Investigator's fault. I also understand that by my participation in this study I am not waiving any of my legal rights[1].

I understand that I may also contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T. 253-6787, if I feel I have been treated unfairly as a subject.

Date: ______________________________________

Signature of Subject: ______________________________________

Signature of Witness: ______________________________________

---

[1] **Further information may be obtained by calling the Institute's Insurance and Legal Affairs Office at 253-2822.**

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
77 MASSACHUSETTS AVENUE
E23-230 3-6787 **CAMBRIDGE, MASSACHUSETTS 02139**

COMMITTEE ON THE USE OF HUMANS AS EXPERIMENTAL SUBJECTS

COMMITTEE APPROVAL

DATE: 02/16/1995

TO: Karon E. MacLean
3-147

FROM: H. Walter Jones, Jr., M.D. H Walter Jones
Chairman

APPLICATION NO.: 2215

TITLE: Emulation of Haptic Feedback in Manual Interfaces

RENEWAL DATE: 02/16/1995

Your application has been approved by the Committee on the Use of Humans as Experimental Subjects at its meeting on 02/16/95. This approval is valid until one year from the above renewal date, at which time your entire application will be due for annual review.

It is expected that you will promptly notify the Committee if your subjects experience any undesirable effects. Please inform the Committee of any changes and when your project is terminated.

The COUHES number assigned to your project is 2215.

In the future, please note this number on all correspondence referring to this project.

cc: T. Duff, OSP

F.2 Investigator Script and Subject Questionnaire

Following is the "script" which the investigator read from when conducting an experiment. The script was used to ensure that all subjects received identical instructions and no component of the experiment session was varied or omitted. It includes a copy of the debriefing questions asked of the subject at the end of the session.

Introduction of Subject to Experiment

This section necessary only for a first session.

Informed Consent

Give subject an informed consent statement and request subject to read it, ask questions, and sign. Offer to make them a copy if they want one.

Get Subject Information

Ask profile questions.

Get Subject Comfortable

Make sure that the stool fits — small people sometimes want taller stool. Move emulator, switch box left/right if necessary.

Describe Experiment to Subject

Your job, as a subject, is to tell me how good the emulation is. You will do this by interacting with the emulation and comparing it with a set of real devices which feel different from each other; and telling me which real device the emulation feels most like.

One of these emulation-real device comparisons consists of a single comparison trial. The majority of the time in this session will be spent doing a series of comparison trials. In between each trial, I need to change the lineup of real devices; this will take about half a minute.

Before the comparison trials, we need to do two other things.

First, I'll give you instructions on how to interact with the emulation, and we'll have a practice session for that.

Then, I will ask you to do a short series of calibration trials where you interact only with the real devices. In each of these trials, I'll ask you to rank the real devices, and then to assign a magnitude to each of them (I'll tell you more about that later).

The experiment is run double-blind; that means that I, the experimenter, don't know what experimental parameters are set at, or what's being targeted.

The entire session should take between 1.5 and 2 hours.

Emulation-Interaction Instructions

Ground Rules

This section necessary only for a first session.

Run first practice emulation in practice batch file, and demonstrate the following instructions.

In each trial, you will be comparing the emulation and the set of real devices, then answer the question:

"Which real device does the emulation feel most like? (Tell me the slot number).

Making a Choice:

1. You have to choose one.
2. Take the time you feel you really need, but don't dawdle: go with your gut reaction.
3. You may re-visit both the real devices and the emulator to make your choice.
4. Make your choice based on the middle of the emulation, not the feel of the edges. Don't be distracted if the location and feel of emulator and real edges are not the same.

It is crucial that you maintain the same posture and approach the emulation and the real devices in the same way for each trial. Therefore you need to follow these rules:

1. Don't touch the emulator except when I tell you to.
2. Start your comparison after the second beep from the computer. Try to start deciding as soon as possible after that second beep, because I'll be timing your decision time.
3. Use your right hand to handle BOTH the emulation and the real devices. Thus, you can only touch one device (real or emulation) at a time.
4. Always rest the heel of your hand on the table and manipulate the handle with your thumb and forefinger.
5. Hold the handle lightly. In some cases, squeezing the handle tightly will make it vibrate. If this happens, just relax your grip.
6. Hold the handle near the tip, not at the base.
7. Always sit on the stool while performing a trial. Keep the stool in the same place.
8. You are free to get up and move around between trials, as long as you stay away from the bench where I'm working. Request a break if you need one at any time.
9. *For toggles only:* When manipulating the toggle switches, always grasp the exposed metal tip of the handle gently between thumb and forefinger, and snap it gently back and forth *(demonstrate)*. Don't flick it with one finger, and don't switch it extremely quickly.

Practice Session with Emulation

This section necessary for all sessions; use a different practice batch file for each device family.

Run a batch file with 3-4 sample devices, which includes good performance, and examples of all the severely degraded performance they will encounter in the session.

Calibration Trials

Have first calibration trial set up already

Ranking

Here are the devices for the first calibration trial. We're going to do four of these.

In the first part, I ask you to rank the devices in order of how "hard" they are.

By "hardness", I mean the physical force you must use to make the device move. The hardest device will be the one which is stiffest, or heaviest, or otherwise presents the greatest resistance to motion.

If two devices feel the same after a reasonable attempt to distinguish them, pick their relative order arbitrarily.

Take the time you feel you really need, but don't dawdle: go with your gut reaction.

You may re-visit the devices to make your choice.

You can make notes on a piece of paper if that will help you organize your thoughts.

Give subject time to interact with the devices and rank them; record responses.

Magnitude Estimation

Now, tell me how hard you think each of those devices is, by assigning it a number between 1 and 10. '1' means it's the easiest in the set; '10' means it's the hardest in the set.

Thus for each trial, you're always going to give me a '1' and a '10' ranking, and the others should be distributed depending on how hard you think they feel.

If you can't tell the difference between two or more devices, give them the same number.

emphasize This goes for the '1' and '10' levels; there can be more than one of them as well.

Be aware that you are not necessarily getting the same set of devices in each trial.

Request magnitude for device that was ranked 1, 2, etc.

Shuffle devices and repeat calibration trials

Comparison Trials

Brief review of emulator instructions

Set up first experiment trial.

We are ready to do our first experiment trial.

Again, I'll give you as much time as you feel you really need, but try to make the decision quickly. I want your gut reaction; don't over-analyze the options.

Periodically I will have to change something in the hardware. At these times, I will ask you to leave the room for a minute, so I can make the change without you watching.

INVESTIGATOR SHOULD NOT TALK DURING COMPARISONS

Incentive Description

At the end of the session, I'm going to pay you in cash based on how well you have performed; this is in addition to the Toscinini's certificate that I already told you about.

Each selection that you make correctly is worth 25 cents. There are 48 trials, so you will get $12 if you are 100 percent correct. However, I'm not going to tell you if you're choosing correctly during the experiment (in fact, I won't know the results until the end of the experiment, either).

Trials

Perform calibration trials (4 sets of 12 with 4-minute break between each set)

Subject Questionnaire

1. What percentage of the comparison trials do you think you guessed correctly?
2. On average, how hard was it to make your decisions (on a scale of 1-5, 1 = obvious, 5 = impossible, had to guess at random)?
3. Was it generally easy to narrow the choices down to two?
4. Did you notice a variation in the 'quality' (the degree to which the emulation felt like real)?
5. *(if yes to previous question)* When the quality was not so good, was it harder to make a choice?
6. Was the experiment tiring, either mentally or physically? Would it have helped to take more of a break?
7. At some point, did you find it becoming easier to make choices? Where was this point (approximately)?
8. At some point, did you find yourself becoming less careful in making your choices? Where was this point (approximately)? (Don't be uncomfortable saying yes — this is a long and demanding experiment).
9. Do you have any comments or suggestions regarding the setup, or how you felt during the experiment?

Subject Payment

Compute subject's performance (percentage "correct") and pay subject after have asked debriefing questions.

Notes on script

Points to deliberately leave out:

- Don't caution subjects to concentrate on "feel" instead of "sound". That would encourage them to dissect the emulation rather than give a gut reaction.
- Don't say anything about safety ("I have my hand near the kill switch at all times, so you don't need to worry about anything") because it will cause concern when none is needed. The emulations seem to be very difficult, if not impossible, to destabilize; and if they are, I do have my hand near the kill switch, and the person can't get pinched easily.

Appendix G

Tabular Presentation of Analyzed Data

This appendix contains the numerical results of the 36 analyses of variance conducted for these experiments; the results were discussed and presented graphically in Section 9.5.

These ANOVA tables include:

G.1 A full-subject (N=6) analysis of each response variable for each real device family (15 analyses in all).

G.2 A repeat-subject (N=2) analysis of each response variable for each real device family (15 analyses).

G.3 N=6 and N=2 analyses for the response variable based on comparison trial decision times, $Y_{T_{Dec}}$, for each real device family (6 analyses).

G.1 All Performance-Based Response Variables: N=6

Y=index(*target device*) – index(*chosen device*)

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	5.642	5	1.128	5.786	3.125	1%
C	7.031	1	7.031	36.051	6.783	1%
D	1.253	1	1.253	6.427	3.896	5%
S	0.028	2	0.014	0.071	2.335	
T	16.177	3	5.392	27.648	3.899	1%
DC	3.781	1	3.781	19.388	6.783	1%
SC	0.083	2	0.042	0.214	2.335	
SD	0.028	2	0.014	0.071	2.335	
TC	2.594	3	0.865	4.433	3.899	1%
TD	1.872	3	0.624	3.199	2.656	5%
TS	0.583	6	0.097	0.498	1.805	
SDC	0.583	2	0.292	1.495	2.335	
TDC	0.788	3	0.263	1.347	2.115	
TSC	0.583	6	0.097	0.498	1.805	
TSD	0.972	6	0.162	0.831	1.805	
TSDC	0.806	6	0.134	0.688	1.805	
Err	45.833	235	0.195			
Tot	82.997	287				

$\mathbf{Y}=Z_{mech_T} - Z_{mech_C}$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	2.218	5	0.444	5.039	3.125	1%
C	0.779	1	0.779	8.855	6.783	1%
D	0.126	1	0.126	1.431	2.738	
S	0.086	2	0.043	0.491	2.335	
T	3.695	3	1.232	13.991	3.899	1%
DC	0.771	1	0.771	8.753	6.783	1%
SC	0.079	2	0.039	0.446	2.335	
SD	0.075	2	0.038	0.426	2.335	
TC	0.402	3	0.134	1.523	2.115	
TD	0.620	3	0.207	2.347	2.115	10%
TS	0.345	6	0.058	0.654	1.805	
SDC	0.352	2	0.176	2.001	2.335	
TDC	0.054	3	0.018	0.203	2.115	
TSC	0.279	6	0.047	0.529	1.805	
TSD	0.438	6	0.073	0.829	1.805	
TSDC	0.730	6	0.122	1.382	1.805	
Err	20.687	235	0.088			
Tot	29.518	287				

$\mathbf{Y}=Z_{perc_T} - Z_{perc_C}$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	5.357	5	1.071	4.696	3.125	1%
C	7.144	1	7.144	31.316	6.783	1%
D	0.840	1	0.840	3.682	2.738	10%
S	0.080	2	0.040	0.176	2.335	
T	8.566	3	2.855	12.516	3.899	1%
DC	3.440	1	3.440	15.081	6.783	1%
SC	0.005	2	0.003	0.011	2.335	
SD	0.077	2	0.039	0.169	2.335	
TC	2.485	3	0.828	3.631	2.656	5%
TD	1.665	3	0.555	2.433	2.115	10%
TS	0.319	6	0.053	0.233	1.805	
SDC	1.005	2	0.503	2.203	2.335	
TDC	0.327	3	0.109	0.478	2.115	
TSC	0.695	6	0.116	0.508	1.805	
TSD	0.963	6	0.161	0.704	1.805	
TSDC	1.174	6	0.196	0.857	1.805	
Err	53.611	235	0.228			
Tot	82.398	287				

$\mathbf{Y}=\ln(Z_{mech_T}) - \ln(Z_{mech_C})$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	7.784	5	1.557	5.667	3.125	1%
C	9.429	1	9.429	34.320	6.783	1%
D	1.486	1	1.486	5.409	3.896	5%
S	0.058	2	0.029	0.106	2.335	
T	17.082	3	5.694	20.725	3.899	1%
DC	4.880	1	4.880	17.762	6.783	1%
SC	0.085	2	0.043	0.155	2.335	
SD	0.032	2	0.016	0.058	2.335	
TC	3.152	3	1.051	3.825	2.656	5%
TD	2.239	3	0.746	2.717	2.656	5%
TS	0.694	6	0.116	0.421	1.805	
SDC	0.707	2	0.354	1.287	2.335	
TDC	0.700	3	0.233	0.850	2.115	
TSC	0.746	6	0.124	0.453	1.805	
TSD	1.180	6	0.197	0.716	1.805	
TSDC	1.120	6	0.187	0.679	1.805	
Err	64.562	235	0.275			
Tot	108.153	287				

$\mathbf{Y}=\ln(Z_{perc_T}) - \ln(Z_{perc_C})$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	3.246	5	0.649	4.941	3.125	1%
C	5.029	1	5.029	38.275	6.783	1%
D	1.259	1	1.259	9.579	6.783	1%
S	0.009	2	0.005	0.035	2.335	
T	13.591	3	4.530	34.477	3.899	1%
DC	2.485	1	2.485	18.912	6.783	1%
SC	0.060	2	0.030	0.229	2.335	
SD	0.105	2	0.053	0.401	2.335	
TC	3.363	3	1.121	8.532	3.899	1%
TD	1.456	3	0.485	3.693	2.656	5%
TS	0.295	6	0.049	0.374	1.805	
SDC	0.423	2	0.211	1.609	2.335	
TDC	1.407	3	0.469	3.569	2.656	5%
TSC	0.338	6	0.056	0.429	1.805	
TSD	0.548	6	0.091	0.695	1.805	
TSDC	0.275	6	0.046	0.348	1.805	
Err	30.879	235	0.131			
Tot	61.522	287				

Table G.1: Analyses of Variance for Masses: all subjects included.

$\mathbf{Y=index}(target\ device) - \mathbf{index}(chosen\ device)$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	6.142	5	1.228	8.449	3.125	1%
C	0.031	1	0.031	0.215	2.738	
D	0.781	1	0.781	5.373	3.896	5%
S	0.028	2	0.014	0.096	2.335	
T	4.455	3	1.485	10.214	3.899	1%
DC	0.087	1	0.087	0.597	2.738	
SC	0.083	2	0.042	0.287	2.335	
SD	1.333	2	0.667	4.585	3.049	5%
TC	0.122	3	0.041	0.279	2.115	
TD	0.538	3	0.179	1.234	2.115	
TS	0.972	6	0.162	1.114	1.805	
SDC	0.194	2	0.097	0.669	2.335	
TDC	0.622	3	0.207	1.425	2.115	
TSC	0.306	6	0.051	0.350	1.805	
TSD	1.722	6	0.287	1.974	1.805	10%
TSDC	0.806	6	0.134	0.923	1.805	
Err	34.167	235	0.145			
Tot	46.247	287				

$\mathbf{Y}=Z_{mech_T} - Z_{mech_C}$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	1.165	5	0.233	6.671	3.125	1%
C	0.018	1	0.018	0.506	2.738	
D	0.226	1	0.226	6.478	3.896	5%
S	0.016	2	0.008	0.231	2.335	
T	0.686	3	0.229	6.545	3.899	1%
DC	0.013	1	0.013	0.362	2.738	
SC	0.068	2	0.034	0.969	2.335	
SD	0.312	2	0.156	4.471	3.049	5%
TC	0.067	3	0.022	0.635	2.115	
TD	0.072	3	0.024	0.687	2.115	
TS	0.224	6	0.037	1.072	1.805	
SDC	0.044	2	0.022	0.625	2.335	
TDC	0.083	3	0.028	0.791	2.115	
TSC	0.177	6	0.029	0.844	1.805	
TSD	0.244	6	0.041	1.164	1.805	
TSDC	0.109	6	0.018	0.522	1.805	
Err	8.206	235	0.035			
Tot	10.564	287				

$\mathbf{Y}=Z_{perc_T} - Z_{perc_C}$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	1.558	5	0.312	8.613	3.125	1%
C	0.018	1	0.018	0.501	2.738	
D	0.173	1	0.173	4.772	3.896	5%
S	0.020	2	0.010	0.275	2.335	
T	0.700	3	0.233	6.447	3.899	1%
DC	0.007	1	0.007	0.205	2.738	
SC	0.070	2	0.035	0.964	2.335	
SD	0.336	2	0.168	4.637	3.049	5%
TC	0.053	3	0.018	0.493	2.115	
TD	0.054	3	0.018	0.500	2.115	
TS	0.211	6	0.035	0.974	1.805	
SDC	0.049	2	0.024	0.674	2.335	
TDC	0.062	3	0.021	0.568	2.115	
TSC	0.162	6	0.027	0.747	1.805	
TSD	0.265	6	0.044	1.218	1.805	
TSDC	0.120	6	0.020	0.550	1.805	
Err	8.504	235	0.036			
Tot	10.804	287				

$\mathbf{Y}=\ln(Z_{mech_T}) - \ln(Z_{mech_C})$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	5.163	5	1.033	8.673	3.125	1%
C	0.027	1	0.027	0.227	2.738	
D	0.561	1	0.561	4.710	3.896	5%
S	0.033	2	0.017	0.141	2.335	
T	3.616	3	1.205	10.122	3.899	1%
DC	0.057	1	0.057	0.481	2.738	
SC	0.036	2	0.018	0.151	2.335	
SD	0.964	2	0.482	4.050	3.049	5%
TC	0.086	3	0.029	0.240	2.115	
TD	0.341	3	0.114	0.953	2.115	
TS	0.776	6	0.129	1.087	1.805	
SDC	0.148	2	0.074	0.621	2.335	
TDC	0.436	3	0.145	1.222	2.115	
TSC	0.223	6	0.037	0.312	1.805	
TSD	1.169	6	0.195	1.636	1.805	
TSDC	0.679	6	0.113	0.950	1.805	
Err	27.980	235	0.119			
Tot	37.132	287				

$\mathbf{Y}=\ln(Z_{perc_T}) - \ln(Z_{perc_C})$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	5.137	5	1.027	7.872	3.125	1%
C	0.046	1	0.046	0.356	2.738	
D	0.752	1	0.752	5.760	3.896	5%
S	0.012	2	0.006	0.046	2.335	
T	3.981	3	1.327	10.166	3.899	1%
DC	0.104	1	0.104	0.794	2.738	
SC	0.074	2	0.037	0.283	2.335	
SD	1.107	2	0.553	4.239	3.049	5%
TC	0.079	3	0.026	0.202	2.115	
TD	0.531	3	0.177	1.357	2.115	
TS	0.837	6	0.139	1.068	1.805	
SDC	0.174	2	0.087	0.666	2.335	
TDC	0.524	3	0.175	1.339	2.115	
TSC	0.237	6	0.039	0.302	1.805	
TSD	1.419	6	0.236	1.812	1.805	10%
TSDC	0.867	6	0.145	1.107	1.805	
Err	30.672	235	0.131			
Tot	41.415	287				

Table G.2: Analyses of Variance for Sliders: all subjects included.

$$Y=\mathbf{index}(\mathit{target\ device}) - \mathbf{index}(\mathit{chosen\ device})$$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	4.319	5	0.864	4.796	3.125	1%
C	0.056	1	0.056	0.308	2.738	
D	0.500	1	0.500	2.776	2.738	10%
S	7.840	2	3.920	21.761	4.736	1%
T	10.417	3	3.472	19.275	3.899	1%
DC	0.222	1	0.222	1.234	2.738	
SC	7.382	2	3.691	20.489	4.736	1%
SD	0.146	2	0.073	0.405	2.335	
TC	2.694	3	0.898	4.986	3.899	1%
TD	0.194	3	0.065	0.360	2.115	
TS	9.104	6	1.517	8.423	2.912	1%
SDC	0.049	2	0.024	0.135	2.335	
TDC	0.917	3	0.306	1.696	2.115	
TSC	13.618	6	2.270	12.599	2.912	1%
TSD	0.243	6	0.041	0.225	1.805	
TSDC	0.729	6	0.122	0.675	1.805	
Err	42.333	235	0.180			
Tot	96.444	287				

$$Y=Z_{mech_T} - Z_{mech_C}$$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	1.780	5	0.356	4.807	3.125	1%
C	0.305	1	0.305	4.117	3.896	5%
D	0.204	1	0.204	2.752	2.738	10%
S	7.623	2	3.811	51.474	4.736	1%
T	8.976	3	2.992	40.407	3.899	1%
DC	0.043	1	0.043	0.577	2.738	
SC	8.646	2	4.323	58.385	4.736	1%
SD	0.106	2	0.053	0.714	2.335	
TC	1.967	3	0.656	8.856	3.899	1%
TD	0.148	3	0.049	0.667	2.115	
TS	8.766	6	1.461	19.730	2.912	1%
SDC	0.006	2	0.003	0.039	2.335	
TDC	0.222	3	0.074	0.999	2.115	
TSC	12.718	6	2.120	28.626	2.912	1%
TSD	0.191	6	0.032	0.429	1.805	
TSDC	0.185	6	0.031	0.416	1.805	
Err	17.401	235	0.074			
Tot	67.504	287				

$$Y=Z_{perc_T} - Z_{perc_C}$$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	4.043	5	0.809	4.600	3.125	1%
C	0.098	1	0.098	0.559	2.738	
D	0.920	1	0.920	5.233	3.896	5%
S	7.346	2	3.673	20.898	4.736	1%
T	8.915	3	2.972	16.907	3.899	1%
DC	0.548	1	0.548	3.119	2.738	10%
SC	6.636	2	3.318	18.877	4.736	1%
SD	0.158	2	0.079	0.450	2.335	
TC	2.481	3	0.827	4.704	3.899	1%
TD	0.495	3	0.165	0.938	2.115	
TS	9.607	6	1.601	9.110	2.912	1%
SDC	0.094	2	0.047	0.269	2.335	
TDC	1.960	3	0.653	3.716	2.656	5%
TSC	14.772	6	2.462	14.007	2.912	1%
TSD	0.286	6	0.048	0.272	1.805	
TSDC	1.028	6	0.171	0.975	1.805	
Err	41.305	235	0.176			
Tot	96.649	287				

$$Y=\ln(Z_{mech_T}) - \ln(Z_{mech_C})$$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	2.338	5	0.468	4.337	3.125	1%
C	0.365	1	0.365	3.387	2.738	10%
D	0.352	1	0.352	3.263	2.738	10%
S	6.202	2	3.101	28.757	4.736	1%
T	5.027	3	1.676	15.540	3.899	1%
DC	0.138	1	0.138	1.281	2.738	
SC	5.452	2	2.726	25.282	4.736	1%
SD	0.045	2	0.023	0.209	2.335	
TC	2.546	3	0.849	7.870	3.899	1%
TD	0.155	3	0.052	0.478	2.115	
TS	6.679	6	1.113	10.324	2.912	1%
SDC	0.038	2	0.019	0.178	2.335	
TDC	0.609	3	0.203	1.883	2.115	
TSC	9.475	6	1.579	14.646	2.912	1%
TSD	0.098	6	0.016	0.152	1.805	
TSDC	0.529	6	0.088	0.818	1.805	
Err	25.340	235	0.108			
Tot	63.051	287				

$$Y=\ln(Z_{perc_T}) - \ln(Z_{perc_C})$$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	8.391	5	1.678	7.054	3.125	1%
C	0.836	1	0.836	3.514	2.738	10%
D	0.404	1	0.404	1.699	2.738	
S	5.360	2	2.680	11.265	4.736	1%
T	7.569	3	2.523	10.605	3.899	1%
DC	0.471	1	0.471	1.978	2.738	
SC	4.807	2	2.403	10.102	4.736	1%
SD	0.025	2	0.013	0.053	2.335	
TC	2.814	3	0.938	3.943	3.899	1%
TD	0.310	3	0.103	0.434	2.115	
TS	7.424	6	1.237	5.201	2.912	1%
SDC	0.209	2	0.104	0.439	2.335	
TDC	1.959	3	0.653	2.745	2.656	5%
TSC	10.393	6	1.732	7.281	2.912	1%
TSD	0.063	6	0.010	0.044	1.805	
TSDC	0.462	6	0.077	0.323	1.805	
Err	55.909	235	0.238			
Tot	99.014	287				

Table G.3: Analyses of Variance for Toggles: all subjects included.

G.2 All Performance-Based Response Variables: N=2

Y=index(*target device*) − **index**(*chosen device*)

Var	SS	df	MS	F_o	$Test$	Sig
Blks	0.010	1	0.010	0.140	2.822	
C	4.594	1	4.594	61.688	7.229	1%
D	0.010	1	0.010	0.140	2.822	
S	0.021	2	0.010	0.140	2.422	
T	6.198	3	2.066	27.743	4.247	1%
DC	3.010	1	3.010	40.426	7.229	1%
SC	0.062	2	0.031	0.420	2.422	
SD	0.271	2	0.135	1.818	2.422	
TC	0.365	3	0.122	1.632	2.212	
TD	2.281	3	0.760	10.211	4.247	1%
TS	0.146	6	0.024	0.326	1.909	
SDC	0.646	2	0.323	4.336	3.202	5%
TDC	0.281	3	0.094	1.259	2.212	
TSC	0.604	6	0.101	1.352	1.909	
TSD	0.562	6	0.094	1.259	1.909	
TSDC	0.687	6	0.115	1.539	1.909	
Err	3.500	47	0.074			
Tot	23.240	95				

$Y = Z_{mech_T} - Z_{mech_C}$

Var	SS	df	MS	F_o	$Test$	Sig
Blks	0.011	1	0.011	0.312	2.822	
C	0.658	1	0.658	18.272	7.229	1%
D	0.150	1	0.150	4.165	4.052	5%
S	0.005	2	0.002	0.068	2.422	
T	0.425	3	0.142	3.927	2.812	5%
DC	0.442	1	0.442	12.266	7.229	1%
SC	0.001	2	0.000	0.011	2.422	
SD	0.063	2	0.031	0.872	2.422	
TC	0.132	3	0.044	1.221	2.212	
TD	0.382	3	0.127	3.535	2.812	5%
TS	0.023	6	0.004	0.104	1.909	
SDC	0.090	2	0.045	1.245	2.422	
TDC	0.119	3	0.040	1.102	2.212	
TSC	0.041	6	0.007	0.190	1.909	
TSD	0.165	6	0.028	0.764	1.909	
TSDC	0.153	6	0.025	0.706	1.909	
Err	1.694	47	0.036			
Tot	4.542	95				

$Y = Z_{perc_T} - Z_{perc_C}$

Var	SS	df	MS	F_o	$Test$	Sig
Blks	0.006	1	0.006	0.054	2.822	
C	7.096	1	7.096	68.498	7.229	1%
D	0.130	1	0.130	1.255	2.822	
S	0.034	2	0.017	0.166	2.422	
T	7.790	3	2.597	25.066	4.247	1%
DC	4.125	1	4.125	39.820	7.229	1%
SC	0.027	2	0.014	0.132	2.422	
SD	0.838	2	0.419	4.044	3.202	5%
TC	0.861	3	0.287	2.770	2.212	10%
TD	3.232	3	1.077	10.399	4.247	1%
TS	0.319	6	0.053	0.514	1.909	
SDC	1.388	2	0.694	6.701	5.110	1%
TDC	0.250	3	0.083	0.805	2.212	
TSC	0.833	6	0.139	1.340	1.909	
TSD	1.827	6	0.305	2.940	2.308	5%
TSDC	1.783	6	0.297	2.869	2.308	5%
Err	4.869	47	0.104			
Tot	35.404	95				

$Y = \ln(Z_{mech_T}) - \ln(Z_{mech_C})$

Var	SS	df	MS	F_o	$Test$	Sig
Blks	0.002	1	0.002	0.016	2.822	
C	5.879	1	5.879	60.174	7.229	1%
D	0.085	1	0.085	0.868	2.822	
S	0.035	2	0.017	0.177	2.422	
T	6.661	3	2.220	22.727	4.247	1%
DC	3.702	1	3.702	37.895	7.229	1%
SC	0.039	2	0.019	0.197	2.422	
SD	0.507	2	0.253	2.594	2.422	10%
TC	0.368	3	0.123	1.257	2.212	
TD	2.757	3	0.919	9.406	4.247	1%
TS	0.225	6	0.037	0.383	1.909	
SDC	0.995	2	0.497	5.092	3.202	5%
TDC	0.139	3	0.046	0.476	2.212	
TSC	0.721	6	0.120	1.229	1.909	
TSD	1.158	6	0.193	1.975	1.909	10%
TSDC	1.170	6	0.195	1.995	1.909	10%
Err	4.592	47	0.098			
Tot	29.030	95				

$Y = \ln(Z_{perc_T}) - \ln(Z_{perc_C})$

Var	SS	df	MS	F_o	$Test$	Sig
Blks	0.054	1	0.054	1.417	2.822	
C	3.041	1	3.041	79.586	7.229	1%
D	0.052	1	0.052	1.350	2.822	
S	0.016	2	0.008	0.215	2.422	
T	6.185	3	2.062	53.957	4.247	1%
DC	1.976	1	1.976	51.724	7.229	1%
SC	0.083	2	0.042	1.091	2.422	
SD	0.147	2	0.073	1.920	2.422	
TC	1.004	3	0.335	8.759	4.247	1%
TD	1.692	3	0.564	14.764	4.247	1%
TS	0.110	6	0.018	0.481	1.909	
SDC	0.439	2	0.220	5.746	5.110	1%
TDC	0.768	3	0.256	6.697	4.247	1%
TSC	0.543	6	0.091	2.370	2.308	5%
TSD	0.281	6	0.047	1.225	1.909	
TSDC	0.488	6	0.081	2.131	1.909	10%
Err	1.796	47	0.038			
Tot	18.623	95				

Table G.4: Analyses of Variance for Masses: repeat subjects only.

Y=index(*target device*) − **index**(*chosen device*)

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	3.010	1	3.010	14.894	7.229	1%
C	0.010	1	0.010	0.052	2.822	
D	0.094	1	0.094	0.464	2.822	
S	0.396	2	0.198	0.979	2.422	
T	7.365	3	2.455	12.145	4.247	1%
DC	0.260	1	0.260	1.288	2.822	
SC	0.021	2	0.010	0.052	2.422	
SD	2 312	2	1.156	5.720	5.110	1%
TC	0.031	3	0.010	0.052	2.212	
TD	0.115	3	0.038	0.189	2.212	
TS	1.104	6	0.184	0.910	1.909	
SDC	0.021	2	0.010	0.052	2.422	
TDC	0.448	3	0.149	0.739	2.212	
TSC	0.812	6	0.135	0.670	1.909	
TSD	1.354	6	0.226	1.117	1.909	
TSDC	0.646	6	0.108	0.533	1.909	
Err	9.500	47	0.202			
Tot	24.490	95				

Y=$Z_{mech_T} - Z_{mech_C}$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	0.816	1	0.816	11.534	7.229	1%
C	0.007	1	0.007	0.104	2.822	
D	0.084	1	0.084	1.189	2.822	
S	0.212	2	0.106	1.496	2.422	
T	0.975	3	0.325	4.592	4.247	1%
DC	0.036	1	0.036	0.510	2.822	
SC	0.089	2	0.044	0.626	2.422	
SD	0.631	2	0.315	4.458	3.202	5%
TC	0.035	3	0.012	0.166	2.212	
TD	0.125	3	0.042	0.590	2.212	
TS	0.468	6	0.078	1.102	1.909	
SDC	0.022	2	0.011	0.152	2.422	
TDC	0.050	3	0.017	0.235	2.212	
TSC	0.523	6	0.087	1.232	1.909	
TSD	0.188	6	0.031	0.443	1.909	
TSDC	0.112	6	0.019	0.264	1.909	
Err	3.325	47	0.071			
Tot	6.882	95				

Y=$Z_{perc_T} - Z_{perc_C}$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	1.088	1	1.088	13.550	7.229	1%
C	0.004	1	0.004	0.048	2.822	
D	0.071	1	0.071	0 884	2.822	
S	0.210	2	0.105	1.310	2.422	
T	1.151	3	0.384	4.778	4.247	1%
DC	0.035	1	0 035	0.436	2.822	
SC	0.091	2	0.046	0.570	2.422	
SD	0.719	2	0.359	4.473	3.202	5%
TC	0.042	3	0.014	0.175	2.212	
TD	0.142	3	0.047	0.588	2.212	
TS	0.486	6	0.081	1.009	1.909	
SDC	0.023	2	0.012	0.145	2.422	
TDC	0.046	3	0.015	0.189	2.212	
TSC	0.520	6	0.087	1.079	1.909	
TSD	0.257	6	0.043	0.534	1.909	
TSDC	0.111	6	0.018	0.230	1.909	
Err	3.776	47	0.080			
Tot	7 685	95				

Y=$\ln(Z_{mech_T}) - \ln(Z_{mech_C})$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	2.441	1	2.441	15.179	7.229	1%
C	0.013	1	0.013	0.083	2.822	
D	0.089	1	0.089	0.554	2.822	
S	0.323	2	0.162	1.005	2.422	
T	7.139	3	2.380	14.795	4.247	1%
DC	0.274	1	0.274	1.705	2.822	
SC	0.020	2	0.010	0.062	2.422	
SD	1.830	2	0.915	5.688	5.110	1%
TC	0.022	3	0.007	0.047	2.212	
TD	0.081	3	0.027	0.167	2.212	
TS	0.900	6	0.150	0.932	1.909	
SDC	0.011	2	0.005	0.034	2.422	
TDC	0.428	3	0.143	0.887	2.212	
TSC	0.645	6	0.107	0.668	1.909	
TSD	0.962	6	0.160	0.997	1.909	
TSDC	0.585	6	0.098	0.607	1.909	
Err	7.559	47	0.161			
Tot	20.882	95				

Y=$\ln(Z_{perc_T}) - \ln(Z_{perc_C})$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	2.430	1	2.430	14.444	7.229	1%
C	0.009	1	0.009	0.054	2.822	
D	0.088	1	0.088	0.521	2.822	
S	0.344	2	0.172	1.023	2.422	
T	7.396	3	2.465	14.653	4.247	1%
DC	0.234	1	0.234	1.389	2.822	
SC	0.004	2	0.002	0.013	2.422	
SD	1.957	2	0.979	5.816	5.110	1%
TC	0.017	3	0.006	0.033	2.212	
TD	0.100	3	0.033	0.198	2.212	
TS	0.966	6	0.161	0.957	1.909	
SDC	0.013	2	0.006	0.038	2.422	
TDC	0.458	3	0.153	0.907	2.212	
TSC	0.575	6	0.096	0.570	1.909	
TSD	1.164	6	0.194	1.153	1.909	
TSDC	0.605	6	0.101	0.600	1.909	
Err	7.907	47	0.168			
Tot	21.836	95				

Table G.5: Analyses of Variance for Sliders: repeat subjects only.

```
ERROR: typecheck
OFFENDING COMMAND: imagemask

STACK:
```

$\mathbf{Y}=\mathbf{index}(\textit{target device}) - \mathbf{index}(\textit{chosen device})$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	0.375	1	0.375	3.525	2.822	10%
C	0.167	1	0.167	1.567	2.822	
D	0.667	1	0.667	6.267	4.052	5%
S	2.583	2	1.292	12.142	5.110	1%
T	1.125	3	0.375	3.525	2.812	5%
DC	1.042	1	1.042	9.792	7.229	1%
SC	2.083	2	1.042	9.792	5.110	1%
SD	0.333	2	0.167	1.567	2.422	
TC	0.833	3	0.278	2.611	2.212	10%
TD	1.000	3	0.333	3.133	2.812	5%
TS	3.000	6	0.500	4.700	3.231	1%
SDC	0.083	2	0.042	0.392	2.422	
TDC	1.125	3	0.375	3.525	2.812	5%
TSC	3.667	6	0.611	5.744	3.231	1%
TSD	0.750	6	0.125	1.175	1.909	
TSDC	0.500	6	0.083	0.783	1.909	
Err	5.000	47	0.106			
Tot	23.958	95				

$\mathbf{Y}=Z_{mech_T} - Z_{mech_C}$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	0.051	1	0.051	1.052	2.822	
C	0.084	1	0.084	1.742	2.822	
D	0.154	1	0.154	3.195	2.822	10%
S	2.413	2	1.207	25.078	5.110	1%
T	0.930	3	0.310	6.444	4.247	1%
DC	0.401	1	0.401	8.333	7.229	1%
SC	2.632	2	1.316	27.352	5.110	1%
SD	0.203	2	0.102	2.111	2.422	
TC	0.691	3	0.230	4.787	4.247	1%
TD	0.169	3	0.056	1.173	2.212	
TS	2.540	6	0.423	8.799	3.231	1%
SDC	0.012	2	0.006	0.125	2.422	
TDC	0.418	3	0.139	2.899	2.812	5%
TSC	3.369	6	0.562	11.671	3.231	1%
TSD	0.412	6	0.069	1.426	1.909	
TSDC	0.106	6	0.018	0.368	1.909	
Err	2.261	47	0.048			
Tot	16.797	95				

$\mathbf{Y}=Z_{perc_T} - Z_{perc_C}$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	0.574	1	0.574	3.866	2.822	10%
C	0.306	1	0.306	2.063	2.822	
D	1.132	1	1.132	7.622	7.229	1%
S	2.438	2	1.219	8.208	5.110	1%
T	1.635	3	0.545	3.671	2.812	5%
DC	1.359	1	1.359	9.154	7.229	1%
SC	1.687	2	0.844	5.683	5.110	1%
SD	0.333	2	0.167	1.122	2.422	
TC	1.015	3	0.338	2.279	2.212	10%
TD	1.539	3	0.513	3.456	2.812	5%
TS	3.312	6	0.552	3.718	3.231	1%
SDC	0.188	2	0.094	0.631	2.422	
TDC	1.603	3	0.534	3.599	2.812	5%
TSC	4.438	6	0.740	4.981	3.231	1%
TSD	1.000	6	0.167	1.122	1.909	
TSDC	0.854	6	0.142	0.959	1.909	
Err	6.979	47	0.148			
Tot	29.819	95				

$\mathbf{Y}=\ln(Z_{mech_T}) - \ln(Z_{mech_C})$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	0.298	1	0.298	4.368	4.052	5%
C	0.507	1	0.507	7.433	7.229	1%
D	0.550	1	0.550	8.068	7.229	1%
S	1.956	2	0.978	14.344	5.110	1%
T	0.875	3	0.292	4.278	4.247	1%
DC	0.634	1	0.634	9.295	7.229	1%
SC	1.488	2	0.744	10.917	5.110	1%
SD	0.134	2	0.067	0.980	2.422	
TC	0.466	3	0.155	2.276	2.212	10%
TD	0.790	3	0.263	3.863	2.812	5%
TS	2.406	6	0.401	5.883	3.231	1%
SDC	0.083	2	0.042	0.611	2.422	
TDC	0.799	3	0.266	3.905	2.812	5%
TSC	2.962	6	0.494	7.241	3.231	1%
TSD	0.417	6	0.070	1.020	1.909	
TSDC	0.376	6	0.063	0.918	1.909	
Err	3.204	47	0.068			
Tot	17.646	95				

$\mathbf{Y}=\ln(Z_{perc_T}) - \ln(Z_{perc_C})$

Var	*SS*	*df*	*MS*	F_o	*Test*	*Sig*
Blks	0.229	1	0.229	4.842	4.052	5%
C	1.027	1	1.027	21.671	7.229	1%
D	0.441	1	0.441	9.295	7.229	1%
S	1.424	2	0.712	15.028	5.110	1%
T	0.937	3	0.312	6.588	4.247	1%
DC	0.330	1	0.330	6.964	4.052	5%
SC	0.986	2	0.493	10.401	5.110	1%

Bibliography

[1] B.D. Adelstein and M.J. Rosen. Design and implementation of a force reflecting manipulandum for manual control research. In H Kazerooni, editor, *Advances in Robotics*, pages 1–12. ASME, 1992.

[2] C. Alphonse, editor. *Ethnic Variables in Human Factors Engineering.* The Johns Hopkins Press, Baltimore, MD, 1975.

[3] R. J. Anderson and M. W. Spong. Hybrid impedance control of robotic manipulators. *IEEE Journal of Robotics and Automation*, 4(5):549–556, 1988.

[4] S.J. Andriole. Methodologies for computer-assisted decision-making: a matrix-based requirements/methods matching strategy. In W.B. Rouse, editor, *Advances in Man-Machine Systems Research*, volume 5, pages 1–46. JAI Press Inc., Greenwich, CONN, 1989.

[5] S. Ashley. The battle to build better products. *Mechanical Engineering*, 113:34–38, November 1990.

[6] S. Ashley. Rapid prototyping systems. *Mechanical Engineering*, 113(4):34–43, April 1991.

[7] R. Bailey. *Human Performance Engineering: Guidelines for Designers.* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1982.

[8] W. Banks and M. Boone. A method for quantifying control accessibility. *Human Factors*, 23:299–303, 1981.

[9] Barfield and Furness. *Virtual Environments and Advanced Interface Design.* Oxford University Press, 1995.

[10] S. Baron and W. Levison. Display analysis with the optimal control model of the human operator. *Human Factors*, 19(5):437–457, 1977.

[11] D. Beaty. *The Human Factor in Aircraft Accidents.* Becker and Warburg, London, 1969.

[12] G.L. Beauregard, M.A. Srinivasan, and N.I. Durlach. The manual resolution of viscosity and mass. In *Dynamic Systems and Control Division*, volume 57-2, pages 657–662. Proceedings of the ASME International Mechanical Engineering Congress and Exposition, San Francisco, CA, 1995.

[13] A. K. Bejczy. Sensors, controls and man-machine interfaces for advanced teleoperation. *Science*, 208:1327–1335, 1980.

[14] F.E. Binet, R.T. Leslie, S. Weiner, and R.L. Anderson. Analysis of confounded factorial experiments in single replications. Technical Report 113, North Carolina Agricultural Experiment Station, October 1955.

[15] K.R. Boff, L. Kaufman, and J.P. Thomas, editors. *Handbook of Perception and Human Performance.* John Wiley & Sons, Inc., New York, NY, 1986.

[16] G.E.P. Box, W.G. Hunter, and J.S. Hunter. *Statistics for Experimenters.* John Wiley & Sons, New York, NY, 1978.

[17] J.V. Bradley. Optimum knob crowding. *Human Factors*, 11(3):227–238, 1969.

[18] G.E. Briggs, P.M. Fitts, and H.P. Barrick. Effects of force and amplitude cues on learning and performance in a complex tracking task. *Journal of Experimental Psychology*, 54:262–268, 1957.

[19] F.P. Brooks, Jr., M. Ouh-Young, J.J. Batter, and P.J. Kilpatrick. Project GROPE — haptic displays for scientific visualization. *Computer Graphics*, 24(4):177–185, August 1990.

[20] P.B. Brown. Ground simulator studies of the effects of valve friction, stick friction, flexibility and backlash on power control system quality. Technical Note 3998, NASA, April 1957.

[21] N.J. Buchaca. Models and mockups as design aids. Technical Document 266, Navy Ocean Systems Center, San Diego, CA, 1979.

[22] Burdea and Coiffet. *Virtual Reality Technology.* John Wiley & Sons Ltd, 1994.

[23] G. Burdea and J. Zhuang. Dextrous telerobotics with force feedback — an overview: Part 1: human factors. *Robotica*, 9:171–178, 1991.

[24] G. Burdea and J. Zhuang. Dextrous telerobotics with force feedback — an overview: Part 2: control and implementation. *Robotica*, 9:291–298, 1991.

[25] G. Burdea, J. Zhuang, E. Roskos, D. Silver, and N. Langrana. A portable dextrous master with force feedback. *Presence*, 1(1):18–28, 1992.

[26] G.C. Burdea and N.A. Langrana. Virtual force feedback: lessons, challenges, future applications. In H Kazerooni, editor, *Advances in Robotics*, pages 41–47. ASME, 1992.

[27] J.H. Burgess. *Designing for Humans: The Human Factor in Engineering.* Petrocelli Books, Princeton, NJ, 1986.

[28] D. Burke and C.B. Gibbs. A comparison of free-moving and pressure levers in a positional control system. *Ergonomics*, 8:23–29, 1965.

[29] A. Burrows. Control feel and the dependent variable. *Human Factors*, 7:413–422, 1965.

[30] P. Buttolo and B. Hannaford. Advantages of actuation redundancy for the design of haptic displays. In *Dynamic Systems and Control Division*, volume 57-2, pages 623–630. Proceedings of the ASME International Mechanical Engineering Congress and Exposition, San Francisco, CA, 1995.

[31] A. Chapanis. *Research Techniques in Human Engineering.* The John Hopkins University Press, Baltimore, MD, 1959.

[32] A. Chapanis and R.G. Kinkade. Design of controls. In H.P. Van Cott and R.G. Kinkade, editors, *Human Engineering Guide to Equipment Design*, chapter 8, pages 345–379. McGraw-Hill Book Co., New York, NY, 1972.

[33] J. Childs. Time and error measures of human performance. *Human Factors*, 22:113–117, 1980.

[34] J. M. Christensen and R. G. Mills. What does the operator do in complex systems. *Human Factors*, 9(4):329–340, 1967.

[35] D. Clausing. *Total Quality Development.* ASME Press, 1994.

[36] J. E. Colgate and J. M. Brown. Factors affecting the Z-range of a haptic display. In *IEEE International Conference on Robotics and Automation*, volume 4, pages 3205–3210, 1994.

[37] J. E. Colgate and N. Hogan. An analysis of contact instability in terms of passive physical equivalents. In *Proceedings of the 1989 IEEE international conference on robotics and automation*, volume 1, pages 404–409, Scottsdale, AZ, 14–19 May 1989. IEEE Comput. Soc. Press, (Cat. No. 89CH2750-8).

[38] J. E. Colgate and G. G. Schenkel. Passivity of a class of sampled-data systems: application to haptic interfaces. In *American Control Conference*, Baltimore, 1994.

[39] J.E. Colgate. Strictly positive real admittances for coupled stability. *Journal of the Franklin Institute*, 329(3):429–444, 1992.

[40] J.E. Colgate, P.E. Grafing, M.C. Stanley, and G. Schenkel. Implementation of stiff virtual walls in force-reflecting interfaces. In *IEEE Virtual Reality Annual International Symposium*, 1993.

[41] R. Curry et al. A design procedure for control/display systems. *Human Factors*, 19:421–436, 1977.

[42] K. Dandekar. *Role of Mechanics in Tactile Sensing of Shape.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, June 1995.

[43] C.R. Dohrmann, H.R. Busby, and D.M. Trujillo. Smoothing noisy data using dynamic programming and generalized cross-validation. *Journal of Biomechanical Engineering*, 110:37–41, 1988.

[44] H. Dreyfuss. *Designing for People.* Paragraphic Books, New York, NY, 1967.

[45] S. Dubowsky, W. Durfee, A. Kuklinski, U. Muller, I. Paul, and J. Pennington. The design and implementation of a laboratory test bed for space robotics: The VES Mod II. In *Proceedings of the ASME Design Technical Conferences*, 1994.

[46] W.K. Durfee, S. Dubowsky, and H. Idris. Real-time control of the MIT vehicle emulation system. In *Proceedings of the American Control Conference*, 1991.

[47] N.I. Durlach, T.B. Sheridan, and S.R. Ellis, editors. *Human Machine Interfaces for Teleoperators and Virtual Environments*, Santa Barbara, CA, March 4–9 1990. NASA Ames Research Center, NASA Conference Publication No. 10071.

[48] J.I. Elkind, S.K. Card, J. Hochberg, and B.M. Huey, editors. *Human performance models for computer-aided engineering.* National Academy Press, Washington, D.C., 1989.

[49] S.R. Ellis. Nature and origins of virtual environments: A bibliographical essay. To appear in *Computer Systems in Engineering* (1991).

[50] EXOS, Inc. Development of a force feedback anthropomorphic teleoperation input device for control of robot hands. Technical Report NASA Phase I SBIR NAS8-38910, EXOS Inc., 1991.

[51] S. EXOS Inc.: Chang, H. Tan, B. Eberman, and B. Marcus. Sensing, perception and feedback for VR. In *Virtual Reality Systems Fall '93 Conference*, New York, NY, October 1993.

[52] E. EXOS Inc.: Chen and B. Marcus. EXOS slip display research and development. In *Dynamic Systems and Control Division*, volume 55(1), pages 265–270. Proceedings of the ASME International Mechanical Engineering Congress and Exposition, San Francisco, CA, 1994.

[53] E.D. Fasse. *On the use and representation of sensory information of the arm by robots and humans.* PhD thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, 1992.

[54] E.D. Fasse and N. Hogan. Quantitative assessment of human perception in virtual objects. In H. Kazerooni, J.E. Colgate, and B.D. Adelstein, editors, *ASME Advances in Robotics, Mechatronics and Haptic Interfaces*, Chicago, IL, 1993.

[55] P. Fisher, R. Daniel, and K.V. Siva. Specification of input devices for teleoperation. In *Proceedings of the 1990 IEEE International Conference on Robotics and Automation*, Cincinnatti, OH, 1990.

[56] S. S. Fisher, M. McGreevy, J. Humphries, and W. Robinett. Virtual workstations: a multimodal, stereoscopic display environment. In D.P. Casasent, editor, *Proceedings of SPIE — The International Society for Optical Engineering: Intelligent Robots and Computer Vision*, Cambridge, MA, 28–31 October 1986. SPIE — The International Society for Optical Engineering.

[57] P.M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 46:381–391, 1954.

[58] P.M. Fitts. Functions of man in complex systems. *Aerospace Engineering*, 21:34–39, 1962.

[59] G. Frost. Man-machine dynamics. In H.P. Van Cott and R.G. Kinkade, editors, *Human Engineering Guide to Equipment Design*, chapter 6, pages 22–309. McGraw-Hill Company, New York, NY, 1972.

[60] T.A. Furness. The supercockpit and its human factors challenges. In *Proceedings of the 30th Annual Meeting of the Human Factors Society*, pages 48–52, Dayton, OH, 1986.

[61] T.A. Furness. Designing in virtual space. In W.B. Rouse and K.R. Boff, editors, *System Design: Behavioral perspectives on designers, tools and organizations.* North-Holland, Amsterdam-New York, 1987.

[62] G.A. Gescheider. *Psychophysics: Method, Theory and Application.* Lawrence Erlbaum Associates, Hillsdale, New Jersey, second edition, 1985.

[63] C.B. Gibbs. Controller design: interaction of controlling limbs, time lags and gains in positional and velocity systems. *Ergonomics*, 5:383–402, 1962.

[64] B. Gillespie. Dynamical modeling of the grand piano action. In *International Computer Music Conference Proceedings*, San Jose, CA, 1992.

[65] B. Gillespie. The touchback keyboard. In *International Computer Music Conference Proceedings*, pages 77–80, San Jose, CA, 1992.

[66] B. Gillespie, 1995. personal communication.

[67] B. Gillespie and Mark Cutkosky. Interactive dynamics with haptic display. In *ASME Advances in Robotics, Mechatronics, and Haptic Interfaces*, pages 65–72, 1993.

[68] G.D. Glosser and W.S. Newman. The implementation of a natural admittance controller on an industrial manipulator. In *Proceedings of the 1994 IEEE International Conference on Robotics and Automation*, volume 2, pages 1209–1215, San Diego, CA, May 1994.

[69] A. A. Goldenberg. Implementation of force and impedance control in robot manipulators. In *Proceedings of the 1988 IEEE International Conference on Robotics and Automation*, pages 1626–1632, Philadelphia, PA, April 1988. IEEE.

[70] A.A. Goldenberg. Analysis of force control based on linear models. In *Proceedings of the 1992 IEEE International Conference on Robotics and Automation*, Nice, France, May 1992.

[71] J.K. Gotow, M.B. Friedman, and M.L. Nagurka. Controlled impedance test apparatus for studying human interpretation of kinesthetic feedback. In *Proceedings of the 1989 American Control Conference*, pages 332–337, Pittsburgh, PA, 1989.

[72] R. Gupta. *Prototyping and design for assembly analysis using multimodal virtual environments.* PhD thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, 1995.

[73] R. Haines. A layout designer's projection recticle. *Human Factors*, 19:567–569, 1977.

[74] A.Z. Hajian and R.D. Howe. Identification of the mechanical impedance of human fingers. In *Dynamic Systems and Control Division*, volume 55(1), pages 319–327. Proceedings of the ASME International Mechanical Engineering Congress and Exposition, San Francisco, CA, 1994.

[75] B. Hannaford. A design framework for teleoperators with kinesthetic feedback. *IEEE Trans. Robot. Autom.*, 5(4):426–434, 1989.

[76] B. Hannaford, L. Wood, D.A. McAffee, and H. Zak. Performance evaluation of a six-axis generalized force-reflecting manipulator. *IEEE Trans. Sys. Man Cyber.*, 21(3):620–633, 1991.

[77] S. Harris et al. A system for the assessment of human performance in concurrent verbal and manual control tasks. *Behavior Research Methods and Instrumentation*, 10:329–333, 1978.

[78] P.D. Hauck and D.L. Taylor. Deriving physical design constraints from functional topologies. *Research in Engineering Design*, 2:81–92, 1991.

[79] V. Hayward and O.R. Astley. Performance measures for haptic interfaces. In *Preprints of the 7th International Symposium on Robotics Research*, Herrshing, Germany, October 1995.

[80] H. Helson. Design of equipment and optimal human operation. *American Journal of Psychology*, 62:473–497, 1949.

[81] R.L. Henneman. Human problem solving in dynamic environments: Understanding and supporting operators in large-scale, complex systems. In W.B. Rouse, editor, *Advances in Man-Machine Systems Research*, volume 4, pages 121–164. JAI Press Inc., Greenwich, Conn., 1988.

[82] N. Hogan. Impedance control: An approach to manipulation: Part I—Theory, Part II—Implementation, Part III—Applications. *ASME Journal of Dynamic Systems, Measurement and Control*, 107:1–24, 1985.

[83] N. Hogan. On the stability of manipulators performing contact tasks. *IEEE Journal of Robotics and Automation*, 4(6):677–686, December 1988.

[84] D. Howland and M. Noble. The effect of physical constraints on control on tracking performance. *Journal of Experimental Psychology*, 46:353–360, 1955.

[85] H. Iwata. Artificial reality with force-feedback: Development of desktop virtual space with compact master manipulator. *Computer Graphics*, 24(4):165–170, August 1990.

[86] S.C. Jacobsen, E.K. Iverson, C.C. Davis, D.M. Potter, and T.W. McLain. Deign of a multiple degree of freedom, force reflective hand master/slave with a high mobility wrist. In *Third Topical Meeting on Robotics and Remote Systems*, Charleston, SC, March 13–16 1989. Sponsored by ANS, IEEE and SME.

[87] L. Jandura and M.A. Srinivasan. Experiments on human performance in torque discrimination and control. In *Dynamic Systems and Control*, volume 55-1, pages 369–375. Proceedings of the ASME Winter Annual Meeting, Chicago, IL, 1994.

[88] W. Jenkins. The discrimination and reproduction of motor adjustments with various types of aircraft controls. *Human Factors*, 60:397–406, 1947.

[89] A. Jennings and W. Chiles. An investigation of time sharing ability as a factor in complex performance. *Human Factors*, 19:535–547, 1977.

[90] H.R. Jex. Four critical tests for control-feel simulators. In *Proceedings of the 23rd Annual Conference on Manual Control*, Cambridge, MA, 1988.

[91] H.R. Jex and W.F. Clement. On defining and measuring perceptual motor workload in manual control tasks. In N. Moray, editor, *Mental Workload: Its Theory and Measurement*, volume 8 of *NATO Conference Series III: Human Factors*, pages 125–178. Plenum Press, New York, NY, 1977.

[92] L.A. Jones. Matching forces: constant errors and differential thresholds. *Perception*, 18(5):681–687, 1989.

[93] L.A. Jones and I.W. Hunter. Analysis of the human operator subsystems. In N.I. Durlach, T.B. Sheridan, and S.R. Ellis, editors, *Human Machine Interfaces for Teleoperators and Virtual Environments*, Santa Barbara, CA, March 4–9 1990. NASA CP91035, NASA Ames Research Center.

[94] L.A. Jones and I.W. Hunter. Influence of the mechanical properties of a manipulandum on human operator dynamics. I. elastic stiffness. *Biological Cybernetics*, 62:299–307, 1990.

[95] L.A. Jones and I.W. Hunter. A perceptual analysis of stiffness. *Experimental Brain Research*, 79:150–156, 1990.

[96] L.A. Jones and I.W. Hunter. Human operator perception of mechanical variables and their effects on tracking performance. *Advances in Robotics*, 42:49–53, 1992.

[97] L.A. Jones and I.W. Hunter. Influence of the mechanical properties of a manipulandum on human operator dynamics. II. viscosity. *Biological Cybernetics*, 69:295–303, 1993.

[98] L.A. Jones and I.W. Hunter. A perceptual analysis of viscosity. *Experimental Brain Research*, 94:343–351, 1993.

[99] S. Karason and M.A. Srinivasan. Passive human grasp control of an active instrumented object. In *Dynamic Systems and Control Division*, volume 57-2, pages 641–647. Proceedings of the ASME International Mechanical Engineering Congress and Exposition, San Francisco, CA, 1995.

[100] H. Kazerooni. Human machine interaction via the transfer of power and information signals. *IEEE Trans. Sys. Man Cyber.*, 20(2):450–463, 1990.

[101] H. Kazerooni. Human induced instability in haptic interfaces. In H. Kazerooni, J.E. Colgate, and B.D. Adelstein, editors, *Advances in Robotics, Mechatronics and Haptic Interfaces*, pages 15–27. ASME, 1993.

[102] H. Kazerooni and M.G. Her. The dynamics and control of a haptic interface device. *IEEE Transactions on Robotics and Automation*, 10(4):453–464, 1994.

[103] H. Kazerooni, P.K. Houpt, and T.B. Sheridan. Robust compliant motion for manipulators. *IEEE Journal of Robotics and Automation*, RA-2(2):93–105, 1988.

[104] S.W. Keele. Motor control. In K.R. Boff, L. Kaufman, and J.P. Thomas, editors, *Handbook of Perception and Human Performance*, volume II: Cognitive Processes and Performance, chapter 30. John Wiley & Sons, New York, NY, 1986.

[105] C.R. Kelley. *Manual and Automatic Control: A Theory of Manual Control and Its Applications to Manual and to Automatic Systems.* John Wiley & Sons, Inc., New York, NY, 1968.

[106] J. Kemeny. The accident at Three-Mile Island. Technical report, Report of the President's Commission, October 1979.

[107] J. S. Kinney and S.M. Luria. Conflicting visual and tactual-kinesthetic stimulation. *Perception and Psychophysics*, 8:189–192, 1970.

[108] D.L. Kleinman, S. Baron, and W.H. Levison. An optimal control model of human response, Part I: Theory and validation. Part II: Prediction of human performance in a complex task. *Automatica*, 6:357–369, 1970.

[109] B.W. Knutson, P.E. Gruenbaum, W.A. McNeely, et al. Implementation of a robotic graphics for a virtual control panel. Technical report, Boeing Defense & Space Group and Boeing Computer Services Research & Technology, Seattle, WA, 1994.

[110] D.A. Kontarinis and R.D. Howe. Tactile display of vibratory information in teleoperation and virtual environments. *Presence*, 1995.

[111] S. Konz and R. Day. Design of controls using force as a criterion. *Human Factors*, 8:121–128, 1966.

[112] K.F. Laurin-Kovitz. Design of components for programmable passive impedance. In *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*, volume 1, pages 1476–1480, Sacramento, CA, April 1991.

[113] D.A. Lawrence. Actuator limitations on achievable manipulator impedance. In *Proceedings of the 1989 IEEE International Conference on Robotics and Automation*, volume 1, pages 560–565, 1989.

[114] D.A. Lawrence and J.D. Chapel. Performance trade-offs for hand controller design. In *IEEE International Conference on Robotics and Automation*, 1994.

[115] D.A. Lawrence and J.D. Chapel. Quantitative control of manipulator/task interactions. *IEEE Control Systems*, 4:14–25, 1994.

[116] S.J. Lederman and R.L. Klatzky. Hand movements: a window into haptic object recognition. *Cognitive Psychology*, 19(3):342–368, 1987.

[117] S.J. Lederman and R.L. Klatzky. Processing haptic features from an initial brief contact. In *Dynamic Systems and Control Division*, volume 57-2, pages 675–679. Proceedings of the ASME International Mechanical Engineering Congress and Exposition, San Francisco, CA, 1995.

[118] S.J. Lederman, R.L. Klatzky, R. Chataway, and C. Summers. Visual mediation and the haptic recognition of two-dimensional pictures of common objects. *Perception and Psychophysics*, 47(1):54–64, 1990.

[119] P.H. Lindsay and D.A. Norman. *Human Information Processing: An Introduction to Psychology.* Academic Press, 1977.

[120] J.M. Loomis and S.J. Lederman. Tactual perception. In K.R. Boff, L. Kaufman, and J.P. Thomas, editors, *Handbook of Perception and Human Performance*, volume II: Cognitive Processes and Performance, chapter 31. John Wiley & Sons, New York, NY, 1986.

[121] L. Love and W. Book. Contact stability analysis of virtual walls. In *Dynamic Systems and Control Division*, volume 57-2, pages 689–694. Proceedings of the ASME International Mechanical Engineering Congress and Exposition, San Francisco, CA, 1995.

[122] H.R. Luxenberg and R.L. Kuehn, editors. *Display Systems Engineering.* McGraw-Hill Book Company, New York, NY, 1968.

[123] K.E. MacLean. Rapid prototyping of physical human interfaces: Preliminary thoughts. Internal report.

[124] K.E. MacLean and W.K. Durfee. An apparatus to study the emulation of haptic feedback. In *Dynamic Systems and Control Division*, volume 57-2, pages 615–621. Proceedings of the ASME International Mechanical Engineering Congress and Exposition, San Francisco, CA, 1995.

[125] S. Mallamad et al. Identifying ability requirements by decision-flow diagrams. *Human Factors*, 22:57–68, 1980.

[126] L.E. Marks. Multimodal perception. *Handbook of Perception*, 8 ch 9, 1978.

[127] T. H. Massie and J.K. Salisbury. The PHANToM haptic interface: A device for probing virtual objects. In *ASME International Mechanical Engineering Exposition and Congress*, Chicago, 1994.

[128] M. Massimino. *Sensory substitution for force feedgack in space teleoperation.* PhD thesis, Department of Mechanical Engineering, Massachusetts Institute of Technology, 1992.

[129] R. Maurant et al. Direct looks and control locations in automobiles. *Human Factors*, 22:417–426, 1980.

[130] E.J. McCormick and M.S. Sanders. *Human Factors in Engineering and Design.* McGraw-Hill Book Co., New York, NY, fifth edition, 1982.

[131] M.W. McGreevy. Virtual reality and planetary exploration. In *29th AAS Goddard Memorial Symposium*, Washington D. C., March 1991.

[132] D. Meister. *Human Factors: Theory and Practice.* John Wiley & Sons, Inc., New York, NY, 1971.

[133] D. Meister. *Human Factors Testing and Evaluation.* Elsevier Science Publishing Company Inc., New York, NY, 1986.

[134] D. Meister and G. Rabideau. *Human Factors Evaluation in System Development.* John Wiley & Sons, Inc., New York, NY, 1965.

[135] P.A. Millman. *Haptic Perception of Localized Features.* PhD thesis, Northwestern University, Evanston, IL, June 1995.

[136] P.A. Millman and J.E. Colgate. Effects of non-uniform environment damping on haptic perception and performance of aimed movements. In *Dynamic Systems and Control Division*, volume 57-2, pages 703–711. Proceedings of the ASME International Mechanical Engineering Congress and Exposition, San Francisco, CA, 1995.

[137] P.A. Millman, M. Stanley, and J.E. Colgate. Design of a high performance haptic inteface to virtual environments. In *IEEE Virtual Reality Annual International Symposium*, pages 216–222, Seattle, WA, 1993.

[138] T. Milner and D.W. Franklin. Two-dimensional endpoint stiffness of human fingers for flexor and extensor loads. In *Dynamic Systems and Control Division*, volume 57-2, pages 649–656. Proceedings of the ASME International Mechanical Engineering Congress and Exposition, San Francisco, CA, 1995.

[139] M. Minsky. *Computational Haptics: The Sandpaper System for Synthesizing Texture for a Force-Feedback Display.* PhD thesis, Massachusetts Institute of Technology, 1995.

[140] M. Minsky et al. Feeling and seeing: Issues in force display. *Computer Graphics*, 24(2):235–243, 1990.

[141] D.C. Montgomery. *Design and Analysis of Experiments.* John Wiley & Sons, New York, NY, third edition, 1991.

[142] T. Moore, M. Broekhoven, S.J. Lederman, and S. Ulug. Q'Hand: A fully automated apparatus for stuying haptic processing of spatially distributed inputs. *Behaviour Research Methods, Instruments and Computers*, 23(1):27–35, 1991.

[143] N. Moray, editor. *Mental Workload: Its Theory and Measurement*, volume 8 of *NATO Conference Series III: Human Factors.* Plenum Press, New York, NY, 1977.

[144] C. T. Morgan, J. S. Cook, and M. W. Lund. *Human Engineering Guide to Equipment Design.* McGraw-Hill, New York, NY, 1963.

[145] J. Morrell and J.K. Salisbury. In pursuit of dynamic range: using parallel coupled actuators to overcome hardware limitations. In *Preprints of the Fourth International Symposium on Experimental Robotics, ISER'95*, Stanford, CA, June 1995.

[146] J.B. Morrell and J.K. Salisbury. Parallel coupled actuators for high performance force control: a micro-macro concept. Artifical Intelligence Laboratory, Massachusetts Institute of Technology: unpublished, 1995.

[147] D. Navon and D. Gopher. On economy of the human processing system. *Psychologcal Review*, 86:217–255, 1979.

[148] W.S. Newman and Y. Zhang. Stable interaction control and coulomb friction compensation using natural admittance control. *Journal of Robotic Systems*, 11(1):3–11, 1994.

[149] D.A. Norman. *The Design of Everyday Things.* Doubleday/Currency, 1988.

[150] D. Orne and J. Mandke. The influence of musculature on the mechanical impedance of the human ulna, an *in vivo* simulated study. *Journal of Biomechanics*, 8:143–149, 1975.

[151] M. Ouh-Young, D. B. Beard, and F.P. Brooks, Jr. Force display performs better than visual display in a simple 6-D docking task. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 1462–1466, Arizona, May 1989.

[152] R. Over. An experimentally induced conflict between vision and proprioception. *British Journal of Psychology*, 57:335–341, 1966.

[153] X.D. Pang, H.Z. Tan, and N.I. Durlach. Manual discrimination of force using active finger motion. *Perception and Psychophysics*, 49(6):531–540, 1991.

[154] S. Parsons et al. Human factors design practices for nuclear power plant control rooms. *Proceedings of the Human Factors Society*, 22nd Meeting, 1978.

[155] H.L. Pick, D.H. Warren, and Hay J.C. Sensory conflict in judgements of spatial direction. *Perception and Psychophysics*, 6:203–205, 1969.

[156] R.M. Pirsig. *Zen and the Art of Motorcycle Maintenance.* Bantam Books, New York, NY, 1974.

[157] G.A. Pratt and M.M. Williamson. Series elastic actuators. Artifical Intelligence Laboratory, Massachusetts Institute of Technology: unpublished, 1995.

[158] W.H. Press. *Numerical Recipes.* Press Syndicate of the University of Cambridge, New York, NY, second edition, 1992.

[159] H. Price. The allocation of functions. *Human Factors*, 27:33–45, 1985.

[160] M.H. Raibert and J.J Craig. Hybrid position/force control of manipulators. *ASME Journal of Dynamic Systems, Measurement and Control*, pages 126–133, June 1981.

[161] J.M. Reising and T.J. Emerson. Research in cockpit controls and displays: Past, present and future. In W.B. Rouse, editor, *Advances in Man-Machine Systems Research*, volume 4, pages 69–119. JAI PRESS INC., Greenwich, Conn., 1988.

[162] P. Rivett. *Model Building for Decision Analysis.* John Wiley & Sons, Inc., New York, NY, 1980.

[163] I. Rock and C.S. Harris. Vision and touch. *Scientific American*, 216:96–107, 1967.

[164] I. Rock and J. Victor. Vision and touch: An experimentally created conflict between the senses. *Science*, 143:594–596, 1963.

[165] L.B. Rosenberg. Perceptual design of virtual haptic sensations. In *VR Systems '93, SIG-Advanced Applications*, New York, NY, 1993.

[166] L.B. Rosenberg. How to assess the quality of force-feedback systems. Immersion Corporation, Santa Clara, CA: unpublished, 1995.

[167] L.B. Rosenberg and B.D. Adelstein. Perceptual decomposition of virtual haptic surfaces. In *Proceedings, IEEE Symposium on Research Frontiers in Virtual Reality*, pages 46–53, San Jose, CA, 1993.

[168] T.A. Ryan. Interrelations of the sensory systems in perception. *Psychological Bulletin*, 37:659–698, 1940.

[169] S.E. Salcudean, N.M. Wong, and R.L. Hollis. A force-reflecting teleoperation system with magnetically levitated master and wrist. In *Proceedings of the 1992 IEEE International Conference on Robotics and Automation*, Nice, France, May 1992.

[170] S.E. Salcudean, J. Yan, Z. Hu, and P.D. Loewen. Performance tradeoffs in optimzation-based teleoperation controller design with applications to microsurgery experiments. In *Dynamic Systems and Control Division*, volume 57-2, pages 631–640. Proceedings of the ASME International Mechanical Engineering Congress and Exposition, San Francisco, CA, 1995.

[171] J.K. Salisbury. Active stiffness control of a manipulator in cartesian coordinates. In *Proceedings of the 19th IEEE Conference on Decision and Control*, Albuquerque, NM, December 1980.

[172] J.W. Senders. The human operator as a monitor and controller of multi-degree of freedom systems. *Transactions of the Professional Group on Human Factors in Electronics, IEEE*, 1:HFE–5, September 1964.

[173] T. Sheridan. *Telerobotics and Human Supervisory Control.* The MIT Press, Cambridge, MA, 1991.

[174] T. Sheridan and W. Ferrell. *Man-Machine Systems: Information Control and Decision Models of Human Performance.* The MIT Press, Cambridge, MA, 1974.

[175] T. B. Sheridan and H. G. Stassen. Definitions, models and measures of human workload. In N. Moray, editor, *Mental Workload: Its theory and measurement*, volume 8 of *NATO Conference Series III: Human Factors*, pages 219–233. Plenum Press, New York, NY, 1977.

[176] T.B. Sheridan. *Telerobotics, Automation and Human Supervisory Control.* MIT Press, 1992.

[177] A. Siegel and J. Wolf. *Man-Machine Simulation Models.* John Wiley & Sons, Inc., New York, NY, 1969.

[178] W.T. Singleton, R.S. Easterby, and D. Whitfield, editors. *The Human Operator in Complex Systems.* Taylor & Francis Ltd., 1967.

[179] Jeff Smith and Jim Leftwich. Applying the interaction design approach to medical devices. *Medical Design and Material*, pages 50–55, April 1991.

[180] M. A. Srinivasan. Tactile discrimination and representations of texture, shape and softness. In N.I. Durlach, T.B. Sheridan, and S.R. Ellis, editors, *Human Machine Interfaces for Teleoperators and Virtual Environments*, Santa Barbara, CA, March 4–9 1990. NASA CP91035, NASA Ames Research Center.

[181] M.A. Srinivasan. Haptic interfaces. In N.I. Durlach, Report on the Committee on Virtual Reality Research A.S. Mavor, and Development, editors, *Virtual Reality: Scientific and Technical Challenges.* National Research Council, National Academy Press, Washington, D.C., 1994.

[182] M.A. Srinivasan and J. Chen. Human performance in controlling normal forces of contact with rigid objects. In *ASME Advances in Robotics, Mechatronics and Haptic Interfaces*, volume 49, pages 119–125, 1993.

[183] S.A. Stansfield. Haptic perception with an articulated, sensate robot hand. *Robotica*, 10, May/June 1992.

[184] S.A. Stansfield. A distributed virtual reality simulation system for situational training. *Presence: Teleoperators and Virtual Environments*, April 1994.

[185] T.L. Stedman. *Stedman's Medical Dictionary.* Williams & Wilkins, Baltimore, MD, 24th edition, 1982.

[186] R.T. Stevens. *Operational Test and Evaluation: A systems engineering approach.* John Wiley & Sons, 1979.

[187] N.P. Suh. *The Principles of Design.* Oxford University Press, New York, NY, 1990.

[188] H.Z. Tan, N.I. Durlach, G.L. Beauregard, and M A Srinivasan. Manual discrimination of compliance using active pinch grasp: the role of force and work cues. *Perception and Psychophysics*, 1995. Accepted for publication.

[189] H.Z. Tan, X.D. Pang, and N.I. Durlach. Manual resolution of length, force, and compliance. *Advances in Robotics*, 42:13–18, 1992.

[190] H.Z. Tan, M.A. Srinivasan, B. Eberman, and B. Cheng. Human factors for the design of force-reflecting haptic interfaces. In *Dynamic Systems and Control*, volume 55-1, pages 353–359. Proceedings of the ASME Winter Annual Meeting, Chicago, IL, 1994.

[191] J.B. Teeple. System design and man-computer function allocation. Presented at ORSA-TMS Meeting, April 19–21 1961.

[192] R.M. Thomson. Design of multi-man–machine work areas. In H.P. Van Cott and R.G. Kinkade, editors, *Human Engineering Guide to Equipment Design*, pages 419–466. McGraw-Hill Co., New York, NY, 1972.

[193] E.R. Tichauer. *The Biomechanical Basis of Ergonomics: Anatomy Applied to the Design of Work Stations.* John Wiley & Sons, Inc., New York, NY, 1978.

[194] W.R. Tobler. The geometry of mental maps. In R. G. Golledge and G. Rushton, editors, *Spatial Choice and Spatial Behaviour.* The Ohio State University Press, Columbus, OH, 1976.

[195] K.Y. Toumi and D.A. Gutz. Impact and force control. In *Proceedings of the IEEE Conference on Robotics and Automation*, pages 410–416, 1989.

[196] A. Treisman. Search similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 17(3):652–676, 1991.

[197] A. Treisman. The perception of features and objects. In A. Baddeley and L. Weiskrantz, editors, *Attention: Selection, Awareness and Control. A tribute to Donald Broadbent.* Clarendon Press, Oxford, 1993.

[198] J.C. Tsai and J.E. Colgate. Stability of discrete time systems with unilateral nonlinearities. In *Dynamic Systems and Control Division*, volume 57-2, pages 695–702. Proceedings of the ASME International Mechanical Engineering Congress and Exposition, San Francisco, CA, 1995.

[199] D. Tyler et al. Monitoring performance across sense modes: An individual differences approach. *Human Factors*, 14:539–547, 1972.

[200] K.T. Ulrich and S.D. Eppinger. *Methodologies for Product Design and Development.* McGraw-Hill, 1994.

[201] H.P. Van Cott and R.G. Kinkade, editors. *Human Engineering Guide to Equipment Design.* McGraw–Hill Book Company, New York, NY, 1972.

[202] Y. Vassiliou. *Human Factors and Interactive Computer Systems.* Ablex Publishing Corp., Norwood, NJ, 1984.

[203] J.E.J. Veeger, F.C.T. Van Der Helm, L.H.G. Van Der Woude, G.M. Pronk, and R.H. Rozendal. Inertia and muscle contraction parameters for musculoskeletal modelling of the shoulder mechanism. *Journal of Biomechanics*, 24:615–629, 1991.

[204] W. Veldhuyzen and H.G. Stassen. The internal model concept: an application to modeling human control of large ships. *Human Factors*, 19:367–380, 1977.

[205] W.L. Verplank. *Is There an Optimal Work-Load in Manual Control?* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, August 1977.

[206] W.L. Verplank. Designing graphical user interfaces. Tutorial 1, Xerox Information Systems Division, Palo Alto, CA, April 1985.

[207] R. Volpe and P. Khosla. An experimental evaluation and comparison of explicit force control strategies for robotic manipulators. In *Proceedings of the 1992 IEEE Conference on Robotics and Automation*, Nice, France, May 1992.

[208] R. Volpe and P. Khosla. A theoretical and experimental investigation of impact control for manipulators. *International Journal of Robotics Research*, 12(4):361–365, 1993.

[209] M.B. Wall, K.T. Ulrich, and W.C. Flowers. Evaluating prototyping technologies for product design. *Research in Engineering Design*, 3:163–177, 1992.

[210] D.R. Wallace and M.J. Jakiela. Computer-automated design of aesthetic injection molded products. Submitted to the 1991 ASME Design Theory and Methodology Conference, Sept 22–25 1991.

[211] D.H. Warren and W.T. Cleaves. Visual-proprioceptive interaction under large amounts of conflict. *Perception and Psychophysics*, 90:206–214, 1971.

[212] A.B. Watson. Towards a model-based human factors. In W.B. Rouse, editor, *Advances in Man-Machine Systems Research*, volume 5, pages 229–276. JAI PRESS INC., Greenwich, Conn, 1989.

[213] R.B. Welch and D.H. Warren. Intersensory interactions. In K.R. Boff, L. Kaufman, and J.P. Thomas, editors, *Handbook of Perception and Human Performance: Sensory processes and Perception*, volume I. Wiley & Sons, New York, 1986.

[214] R.B. Welch, M.H. Widawski, J. Harrington, and D.H. Warren. An examination of the relationship between visual capture and prism adaptation. *Perception and Psychophysics*, 25:126–132, 1979.

[215] A.T. Welford. *Skilled Performance: Perceptual and Motor Skills.* Scott, Foresman and Company, Glenview, IL, 1976.

[216] P. Wellman and R.D. Howe. Towards realistic vibrotactile display in virtual environments. In *Dynamic Systems and Control Division*, volume 57-2, pages 713–718. Proceedings of the ASME International Mechanical Engineering Congress and Exposition, San Francisco, CA, 1995.

[217] D.E. Whitney. Force feedback control of manipulator fine motions. *ASME Journal of Dynamic Systems, Measurement and Control*, pages 91–97, June 1977.

[218] D.E. Whitney. Designing the design process. *Research in Engineering Design*, 2:3–13, 1990.

[219] C.D. Wickens. The effects of control dynamics on performance. In K.R. Boff, L. Kaufman, and J.P. Thomas, editors, *Handbook of Perception and Human Performance*, volume II: Cognitive Processes and Performance, chapter 39. John Wiley & Sons, Inc., New York, NY, 1986.

[220] C.D. Wickens et al. Multiple resource, task hemisphere integrity, and individual differences in time sharing. *Human Factors*, 23:211–229, 1981.

[221] W. Wierwille and S. Connor. Evaluation of 20 workload measures using a psychomotor task in a moving base simulator. *Human Factors*, 25:1–16, 1983.

[222] J.F. Wilcott. Perceptual judgments with discrepant information from audition and proprioception. *Perception and Psychophysics*, 14:577–580, 1973.

[223] A. Witkin, K. Fleisher, and A. Barr. Energy constraints on parameterized models. *Computer Graphics*, 21(4):225–232, 1987.

[224] A. Witkin, M. Gleicher, and W. Welch. Interactive dynamics. *Computer Graphics*, 24(2):11–22, 1990.

[225] J.J. Wlassich. Nonlinear force feedback impedance control. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, February 1986.

[226] W.E. Woodson. *Human Factors Design Handbook.* McGraw-Hill, Inc., 1981.

[227] W.E. Woodson and D.W. Conover. *Human Engineering Guide for Equipment Designers.* University of California Press, Berkeley, CA, 2nd edition, 1964.

[228] W.E. Woodson, M.P. Ranc, Jr., and D.W. Conover. Design of individual workplaces. In H.P. Van Cott and R.G. Kinkade, editors, *Human Engineering Guide to Equipment Design*, chapter 9, pages 381–418. McGraw-Hill Co., New York, NY, 1972.