Lehrstuhl für Steuerungs- und Regelungstechnik
Technische Universität München
Univ.-Prof. Dr.-Ing./Univ. Tokio Martin Buss

Visual-Haptic Presence Systems: Utility Optimization, Compliance Perception, and Data Compression

Martin Kuschel

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr.-Ing. E. Steinbach
Prüfer der Dissertation:	
1.	Univ.-Prof. Dr.-Ing./Univ. Tokio M. Buss
2.	Univ.-Prof. Dr. phil. rer. soc. habil. B. Färber Universität der Bundeswehr München

Die Dissertation wurde am 27.06.2008 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 27.01.2009 angenommen.

The study of the art of motorcycle maintenance
is really a study of the art of rationality itself.
Working on a motorcycle, working well, caring,
is to become part of a process,
to achieve an inner peace of mind.
The motorcycle is primarily a mental phenomenon.

Robert M. Pirsig [1]

Foreword

This text presents a large part of my research on human perception and mechatronics as a PhD-student at the Institute of Automatic Control Engineering, Technische Universität München.
During the last four years human perception and haptic presence systems dominated my life. It ruled not only my workday but also my thoughts before and afterward, commonly called 'private life'. It affected the way I interpreted talks to my family and to my friends as well as the thoughts that aroused when reading books and newspapers. It was a demanding and also wonderful time and I am proud of the outcome! But I am also looking forward to reducing focus and to broadening my scope again.
The work presented has been facilitated by many persons. I would like to thank Prof. Martin Buss, for providing an environment for open-minded, interdisciplinary robotics research and for providing contact to international renowned researchers. I would like to thank Prof. Roberta Klatzky from the Carnegie Mellon University in Pittsburgh for the close and successful collaboration, for the numerous inspiring discussions, and for her cordial support. I would like to thank Marc Ernst, PhD and Massimiliano Di Luca, PhD from the Max-Planck-Institute in Tübingen for introducing me to the fascinating field of human perception, for guiding me through psychophysics, and for the successful collaboration as well. I am grateful to Philipp Kremer who sharpened my body of understanding for passivity, for the inspiring discussions, and for the successful collaboration on haptic data reduction. Thanks to Prof. Olaf Stursberg for the advice in optimization theory and to Dr. Dirk Wollherr for making things work at the institute. Thanks to Dr. Franziska Freyberger for the efficient collaboration in the SFB 453.
Experiments-based research that covers a large part of this thesis is not possible without a team of passionate people. I want to thank all my students, especially Maria-Magdalena, Antonia Glaser, Isabell Borutta, Pit Hoffmann, Andreas Schmid, Norbert Gattinger, Uwe Friedrich, Jan Renner, Liang Gong, Jakob Maier, Amir Solhjoo, Sebastian Deusser, as well as Franziska Becker, Matthias Rungger, Klaas Klasing, Thomas Schauss, Jens Hölldampf, and Michael Scheint for the realization of many experiments. I want to thank my colleagues who provided support and advice as well as perfect targets for my, sometimes exhaustive, humor. Finally, I want to thank Maike for reminding me that there is a life besides research just when research started to seize everything.
This all would be impossible without the support of my parents, Marlis and Wolfgang Kuschel, who I have never had to thank for anything, so many times.

Martin Kuschel, Pittsburgh, 24. June 2008

Scope

Research on presence systems, i.e. telepresence and VR-systems, is two-sided. On the one side, there are the engineering technologies, in particular mechanical engineering, software engineering, and control systems. They deal with technical problems, such as developing hardware and software. On the other side, there are the human sciences, in particular sociology, psychology, and philosophy. They deal with the influence of the presence-mediating technology on the human being and use presence systems for studies on human's perceptual and cognitive capabilities. This dichotomy frequently causes problems. Both sides, engineering and human sciences, need the knowledge of the opposed side to generate high-quality results. Presence systems combine human and technology in a hardly separable manner. This is why this thesis used a consequently interdisciplinary approach. Engineering methods, particularly optimization theory and control theory, were combined with psychological methods, particularly signal detection theory and psychophysics. The combined methodology was applied to different problems of visual-haptic presence systems. As an outcome innovative research results are presented dealing with problems such as presence measurement, visual-haptic perception of mechanical environments, and haptic data compression.

Zum Inhalt

Forschung an Präsenzsystemen, das sind Telepräsenzsysteme und Systeme zum Erleben virtueller Realitäten, ist zweigeteilt. Auf der einen Seite stehen Ingenieurwissenschaftler, insbesondere Regelungstechniker und Maschinenbauer. Sie beschäftigen sich mit den technischen Problemen und entwickeln Hardware und Software. Auf der anderen Seite stehen Humanwissenschaftler, insbesondere Soziologen, Psychologen und Philosophen. Sie beschäftigen sich mit der Rezeption von Präsenzsystemen durch den Menschen und nutzen diese, um sich mit dem Menschen als Forschungsobjekt zu befassen. Diese Zweiteilung ist oft problematisch. Beide Seiten benötigen die Kompetenzen der anderen Seite, um die Probleme ihre jeweiligen Forschungsgebiete mit hinreichender Qualität lösen zu können. Denn Präsenzsysteme vereinen schwer teilbar Mensch und Technik. In dieser Arbeit wurde deshalb ein streng interdisziplinärer Ansatz gewählt. Methoden der Ingenieurwissenschaften, insbesondere Optimierungstheorie und Regelungstechnik, wurden mit Methoden der Psychologie, insbesondere Signalentdeckungstheorie und Psychophysik kombiniert und auf verschiedene Bereiche visuell-haptischer Telepräsenzsysteme angewendet. Als Resultat werden innovative Forschungsergebnisse aus den Bereichen Präsenzmessung, visuell-haptische Wahrnehmung mechanischer Umgebungen und haptische Datenreduktion präsentiert.

Contents

Notations

General

Sets are denoted in upper case letters boldface type, e.g. $\mathbb{X}$. Scalars are denoted in either lower case or upper case letters in italic type, e.g. $f(\cdot), F(\cdot)$, or in several mainly upper case letters in roman (normal) type, eDS, JND$(\cdot)$. Vectors are denoted by lower case letters in boldface type. A vector $\mathbf{x}$ is composed of elements x_i. Partial derivations are denoted, besides the standard notation, by indexed letters, e.g. x_y, where the index defines the variable subject to the derivation. Subscripts and superscripts are declared locally.

$\mathbb{X}$	Set
x, MRS	Scalar
$f(\cdot), F(\cdot)$, JND$(\cdot)$	Scalar function
$\mathbf{x}$	Vector
x_y	Partial derivative
μ	Mean value
σ	Standard deviation
s	Basic stimulus
S	Combined stimulus (e.g. compliance)

Control

C	Controller
f	Force signal
g	Incident wave (g_l wave before entering COM, g_r after COM)
h	Reflected wave (h_r wave before entering COM, h_l after COM)
P	Physical power
S	Impedance stimulus (compliance)
T	Communication delay
v	Velocity signal
Z	Impedance

Presence

COM	Communication channel
HSI	Human system interface
TO	Teleoperator
VR	Virtual reality

Perception

CE	Constant error
DT	Detection threshold
JND	Just noticeable difference
PSE	Point of subjective equality

d'	Perceptual performance
F_Ψ	Gaussian psychometric function
F	Gaussian cumulative distribution
f	Gaussian probability distribution
k_i	Perceptual weightings, index denotes modality
p	Perceptual variable
Ψ	Gaussian sigmoid
S_{ref}	Reference or standard stimulus (compliance)
Z	Z-score transformation
z	Z-score

Multiobjective Optimization

EoS	Elasticity of Substitution
MO	Multiobjective optimization
MOA	Multiobjective analysis
MODA	Multiobjective decision aid
MOP	Multiobjective optimization problem
MRS	Marginal rate of substitution
rUMP	Reduced UMP
UMP	Utility maximization problem

a, b	Distribution parameter
ϵ	EoS
L	Lagrangian vector
$\mathbb{P}, f_p$	Pareto frontier
ϱ	Substitution parameter
u	Utility function
ξ	MRS
$\mathbb{Y}$	Feasible set
$\mathbf{y}$	Alternative vector
$\mathbf{y}^*$	Optimal alternative

Data Compression

CR	Compression ratio
DRS	Data rate savings
eDS	Passive extrapolative downsampling
iDS	Passive interpolative downsampling

List of Figures

1 Introduction

1.1 Presence Systems: Development and Challenges

Presence systems allow humans to operate in two kinds of target environments: *Virtual reality systems* allow humans to immerse in an artificially generated environment. *Telepresence systems* allow humans to immerse in a real, but inaccessible environment. The inaccessibility can be due to distance, scaling or living conditions. A presence system consists of a *human operator* who commands an *avatar/teleoperator* (TO) in the *virtual/remote environment*. A multimodal *human system interface* (HSI) is used for the operator to command the TO and, concurrently, to display the target environment. Signals are exchanged over a *communication channel* (COM).

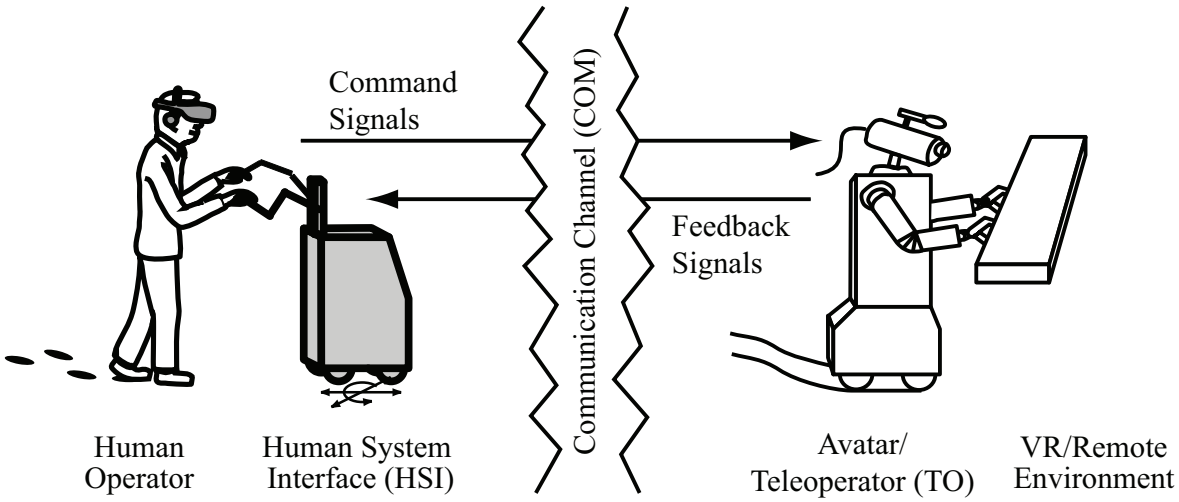


Figure 1.1: Multimodal presence system: A virtual or remote environment is mediated via technological equipment.

Presence systems have stirred up a lot of enthusiasm in the scientific community since their first deliberative beginnings in the 1970s. A decent amount of research on telepresence and VR is available. Engineers have built presence systems in arbitrary configurations. Systems range from small, single degree-of-freedom devices to powerful multi-robot robots systems with redundant kinematics and a huge workspace. Algorithms have been developed to render different target environments acoustically, visually, and haptically. In the course of these developments, prices have decreased significantly and presence systems for scientific use are available today even for low budgets. This has facilitated human sciences to use presence systems for the analysis of human's behavior in any kind of situation and for arbitrary applications. For example, social scientists have analyzed human-robot interactions as a possibility to simplify the daily life of elderly people. Medical scientists have

pushed forward the construction of telepresence systems for minimal-invasive surgery or telemedicine. Psychologists have started to use presence systems to analyze human perception and to run their psychophysical experiments. Also philosophers have found presence systems to be fertile playground. The definition of presence based on the thoughts of famous philosophers like René Descartes or Martin Heidegger is constantly being developed further. Analyzes of how the human mind performs the change of body representation to an artificial agent have been subject to many studies.
However, as it seems, presence systems have not yet performed the leap to leave science and go for commercial applications in industry, economics, or simply, human's daily life. This could be explained by shortcomings in all involved scientific disciplines. For example, presence systems are still lacking an effective, ergonomic structure. Especially, the hardware of HSI and TO is often too cumbersome to enable effective, dexterous exploration and manipulation tasks. Social science and psychology still lack implementable methods to analyze human behavior and perception in fast changing multimodal environments. Furthermore, philosophers still not provide clear guidance for what are the possibilities and the limits of presence systems in society and what are the constituting factors of presence itself. In this sense, the developments of presence systems is stuck.
On the other hand, one can use a more pragmatic view on presence systems. Then, it becomes visible that the principles developed for presence systems can be found in nearly all technical applications that involve the interaction of humans with machines. One of the most prominent examples is the ordinary car. A car perfectly matches the patterns of a presence system, although the standard subsystems are not as obvious as known from theory. A car consists of one or more human operators (driver and passengers), who interact with the HSI (instruments, gas pedal, seats, steering wheel, etc.) to command a TO (chassis, compartment, motor, suspension, steering, etc.). HSI and TO exchange command and feedback information over the the COM (compartment, steering, transmission, wires, etc.). The purpose of a car is to enable the driver and the passengers to be present on the street and to interact with other driver's in the traffic (target environment). The degree to which this objective is realized on a certain car model, distinguishes one car manufacturer from another. In a sportive car the experience of being on the street is very different from a comfortable car, which only provides absolutely necessary information from the roadbed and the traffic situation. Huge effort and a lot of money is spent just to implement a certain driving experience and a certain comfort level.
As automobiles match the structure of presence systems, most of the research conducted for presence systems can be used for automobiles. For example, control algorithms have already found their way into motor, suspensions, or other parts of the car. Ergonomic considerations that apply for HSIs also apply for the drivers cabin. Drive-by-wire methods have nearly found their way into mass production. Last but not least, psychophysical methods for human perception can be deployed to analyze the experience of being-on-the-street for a certain car model. And what holds for cars holds for other vehicles as well, e.g. for aircrafts, boats, motorcycles, trains, wheelchairs, etc.
As a consequence, the hypotheses that development of presence systems is stuck because of its constrained usability for real world purposes cannot be held. This view only arises from a constrained, technocratic attitude. Presence systems need not to consist of detached subsystems each represented by a certain machine or robot. They can also be represented by a highly integrated system, in which the standard subsystems (HSI, COM, TO) cannot be clearly distinguished from each other or might be scattered throughout the overall system.

In general, presence systems provide an efficient abstraction of any kind of human-machine interaction. They provide the possibility to model highly integrated systems in a detached structure of standardized subsystems. And they provide the possibility to identify problems of these subsystems and to treat them separately.
However, the separation has its limits. As structural separation is an intrinsic part of presence systems, methodological separation, which is common in research on presence systems, can lead to wrongly guided research and waste of resources. The technical part of a presence system can seldom be separated from the way it influences the perception of the human operator. This circumstance forces researchers to use interdisciplinary knowledge. Technical implementations, like control structures or changes on the design, always have to be evaluated for their perception by the human operator. For the evaluation, standardized methods from psychophysics and psychology have to be deployed that normally result in extensive experiments. Even more, psychophysical models of the human operator should be developed and used to parametrize algorithms and design changes. Therefore, engineers and human scientists are forced to closely collaborate and to understand each others contributions. This is the main challenge of research on presence systems.
A field in which the interdisciplinary character of presence systems has already emerged to be irreplaceable is research on human perception. The human's perceptual system is a very complex structure that interweaves sensory, sensorimotor, and cognitive processes and is only observable to a small extent. Research on human perception has a short and slowly developing history. Psychophysical models and methods (Weber fraction, method of constant stimuli, etc.) initially indicated the start of decisive research on perception, only 150 years ago. And they are still present to a huge extent in contemporary studies. Methods are based on experiments that are characterized by many repetitions and many participants. Furthermore, the stimuli have to be under full control of the experimenter, since the results emerge from the input/output relations. To provide these stimuli presence systems have become a great means. Providing artificial stimuli is superior to providing real stimuli for mainly three reasons. Firstly, presence systems allow to render nearly arbitrary stimuli with highest accuracy. Secondly, by presence systems stimuli can be presented veridically and non-veridically. Thirdly, they can be presented according to any scheduling mechanism and subject's reports can be recorded conveniently.
An example the analysis of visual-haptic perception of compliance with a 2AFC-method. Without the use of presence systems the experimenter would have to rely on mechanical specimens that consist of a flexible material, e.g. rubber. Therewith, it is difficult to exactly control the compliance of the object. Hence, the experimenter is already uncertain about the stimulus he presents. Furthermore, the visual stimulus cannot be decoupled from the haptic stimulus. Hence, the experimenter is constrained in the bimodal stimulus combinations and can test his hypotheses only under normal, veridical but not under irregular, non-veridical perceptual conditions. Finally, the 2AFC-method demands a complex stimulus scheduling such that each stimulus is influenced by the subject's reported percept. Hence, the experimenter has to define and protocol all stimuli by himself, which can lead to mistakes in the stimulus presentation.
All three problems are solved elegantly by presence systems. The haptic control system is able to render the compliance of a virtual object at very high accuracy, e.g. by an admittance control scheme. The visual modality can be decoupled from the haptic modality and conflicts between visual and haptic information can be deliberately induced, also at very high accuracy. Finally, any stimulus scheduling procedure can be implemented

conveniently, whether it depends on the subjects answer or not, since answers can be electronically feed into the scheduling algorithm by an input device attached to the presence system. For the reasons of accuracy, multimodality, and scheduling convenience, presence systems are the means to propel research on human perception with the goal to develop better models of the perceptual system. However, for the successful use of presence systems in perceptual research an interdisciplinary proceeding is mandatory.

1.2 Outline and Contributions

The work presented deals with visual-haptic presence systems and uses a consequent interdisciplinary approach. The contributions aim at improving or describing the human operator's experience of a target environment. Methods and knowledge developed were evaluated in terms of psychophysical parameters. Furthermore, contributions are based on control and optimization theory. In Chapter 2 detailed introductions to all methods used throughout this theses are provided as well as a philosophical introduction to presence systems.

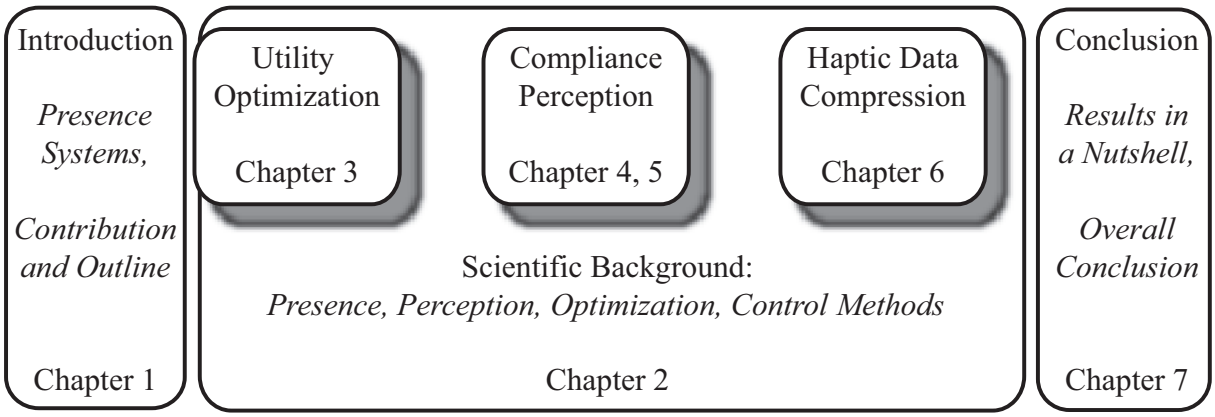


Figure 1.2: Contributions and outline: Based on an interdisciplinary background three different scientific contributions are presented in four chapters. Each chapter can be read independently. State-of-the-art information is presented at the beginning of each innovation chapter.

All contributions involve extensive experiments with state-of-the art psychophysical methods implemented on high fidelity visual-haptic presence systems. In the course of this thesis seven experimental studies are presented. Thereby, 189 participants were tested who completed 38 different conditions during 298 h experimental time.
The scientific contributions can be divided into three groups: Utility optimization is presented in Chapter 3, compliance perception is presented in Chapters 4 and 5, and haptic data compression is presented in Chapter 6. The innovation chapters are arranged according to the purpose of the psychophysical analysis. At first, psychophysics are integral to the method presented, then psychophysics are used to analyze human perception, eventually, psychophysics are used to parametrize the compression algorithms. The structure is illustrated in Figure 1.2.

The contributions can be read independently. However, I recommend to start from the beginning to assure a broad understanding of the main topic. State-of-the-art information is presented at the beginning of each innovation chapter. Further references are given where needed. In each chapter a discussion of the scientific benefits and the open problems is presented at the end. Since the experiments were conducted with only two presence systems, a detailed description is sourced out to the Appendix and only the different configurations are explained locally. A summary of the results and a conclusion of the complete work is given in Chapter 7. The abstract of each contribution is presented below.

Perception-Based Utility Optimization

In Chapter 3 a new method is proposed to identify the optimal trade-off between performance and efforts/resource savings of a presence system. In a first step, performance is measured in terms of immersive experience, i.e. perceived quality of presence. The measurement is conducted using a psychological method (post-test rating) and psychophysical method (magnitude estimation). The identified relation between performance and efforts/resource savings represents the Pareto-optimal combinations for a certain presence system. In a second step, the operator's preferences for the two objectives (performance, efforts/resource savings) are quantified in terms of utility functions. Different utility functions are proposed that represent different, general preference structures that arise in presence systems applications. In a third step, utility optimization is performed to maximize the operator's preferences. The utility function is maximized with respect to the Pareto-optimal combinations (equality constraint) of performance and efforts/resource savings.
The method is generic to arbitrary presence measures and to arbitrary performance parameters. It can be applied to all kinds of presence systems, e.g. aircrafts, cars, wheelchairs, or any kind of virtual reality.

Compliance Perception

In Chapter 4 three explorative studies are presented analyzing visual-haptic compliance perception. All studies assessed estimation performance in terms of the difference threshold. However, results were complementary rather than redundant as different kinds of difference thresholds were targeted and different methods were used. In the first study, concurrent vs. sequential estimation performance was measured. The focus was on the crossmodal difference threshold, i.e. one modality was held as reference and the other modality had to be matched by the participant as close as possible. An adaptive staircase method was implemented to adjust the target modality toward the reference modality. The second study focused on sequential crossmodal matching. In contrast to the first study the method of adjustment was used and the participants could actively adjust the reference modality. Unimodal matching as well as bimodal matching was analyzed and passive vs. active exploration. The third study concentrated on the fusion of visual and haptic information when perceiving object compliance. Psychometric functions were recorded using 2AFC-tasks. Congruent and incongruent stimuli were displayed to analyze whether and how participants integrated bimodal information to a coherent percept.
In Chapter 5 a theoretic process model of visual-haptic compliance perception is developed

that conceptualizes the results of the explorative studies of Chapter 4. The model gives clear advice how combination of non-redundant information and fusion of redundant information are related in visual-haptic compliance perception. Two experimental studies were conducted. The method of constant stimuli was used to assess the observer's performance. By the first study the general arrangement of combination and fusion processes was identified. By the second study the single contributions of visual and haptic compliance information to the final percept were analyzed.
The results can be used to parametrize the control in visual-haptic or haptic-only displays. Furthermore, they can be used to facilitate the evaluation of presence systems.

Haptic Data Compression

In Chapter 6 lossy data compression methods are proposed to reduce velocity and force data in haptic presence systems. Based on a classification, energy-consumption criteria for interpolative, extrapolative, and direct compression strategies are developed based on passivity theory. The proposed criteria allow arbitrary implementations of compression algorithms such that they will not affect the stability of the overall system. Two compression algorithms are implemented using the interpolative and the extrapolative compression strategy. The interpolative compression algorithm is called passive interpolative downsampling, the extrapolative strategy is called passive extrapolative downsampling. Simulations were conducted to illustrate the different operation modi. Eventually, experiments were performed and an extensive psychophysical study was conducted, recording psychophysical functions by 2AFC-tasks. Based on the experimental results the algorithms are parametrized perceptually transparent. Furthermore, parametrization advice is provided, when perceptual transparency is not the only objective.
The results are not restricted to haptic data compression. The passivity criteria can be applied for arbitrary signal shaping in closed control loops such that stability is not impaired. Especially the field of networked control systems offers a huge application field, since the COM might be already passivated by the Scattering transformation.

2 Scientific Background

Is this the real life?
Is this just fantasy?
Caught in a landslide
No escape from reality...

Queen, *Bohemian Rapshody* (*A Night at the Opera*, 1975)

This chapter sets the stage for the innovation chapters 3-6. The research presented in these chapters is located in the brackish water of engineering and psychology. My intention is to provide a scientific foundation and to give semi-experienced readers efficient access to the diverging underlying theories. Epistemological, perceptual, optimization, and control concepts used in this thesis are described. However, detailed state-of-the-art information is not given as this is done at the introduction of each innovation chapter.
Section 2.1 introduces and structures the experience of technology-mediated presence. Based on a philosophical analysis main paradigms about experiencing technology-mediated presence are elaborated. Presence measures are introduced, too. Section 2.2 deals with human perception. Probabilistic models are described and psychophysical methods used in the following chapters are introduced. Section 2.3 explains the mathematical method of multicriteria optimization to calculate the optimal utility of a presence system. Finally, Section 2.4 describes control measures to render and stabilize haptic environments. Basic references are given at the beginning of each subsection.

2.1 Technology-Mediated Presence

Presence systems allow humans to operate in two kinds of target environments: *Virtual reality systems* allow humans to immerse in an artificially generated environment. *Telepresence systems* allow humans to immerse in a somehow impenetrable, but real environment. The inaccessibility can be due to distance, scaling or hazardous living conditions. A presence system consists of a *human operator* who commands an *avatar/teleoperator* (TO) in the *virtual/remote environment*. A multimodal *human system interface* (HSI) is used for the operator to command the TO and, on the same time, to display the target environment. Signals are exchanged over a *communication channel* (COM). See Figure 1.1 for an illustration. Applications are virtual training, entertainment, aerospace teleoperation, micro assembly or maintenance in hazardous environments.
The goal is to generate a high degree of presence to realistically enable explorative and manipulative activities in a target environment. A large amount of short definitions about presence systems exist, which are all true to a certain extent and, therefore, partly contradictory or at least confusing. However, a clear philosophical structure is necessary for successful research. A suitable philosophical branch to deal with presence in general is

epistemology dealing with the origins nature and extent of human knowledge. For presence systems, epistemological views about presence as a daily-life experience have to be extended to presence mediated by technology.
In Subsection 2.1.1 philosophical insights are provided on the experience of presence in general. In Subsection 2.1.2 the transition to technology-mediated presence is performed and two principles are elaborated. In Subsection 2.1.3 presence measurement is explained and methods are introduced.

2.1.1 Presence - The Experience of Being in the World

Philosophy of presence is, above all, a philosophy of existence (*ontology*). However, *ontological arguments* are too general for research dealing with presence systems. The crucial question is, how humans experience presence while living in an environment. This is strongly connected to the philosophy of the extent of human knowledge, called *epistemology*. Two major epistemological views have been elaborated that bolster the research about technology-mediated presence. The *rationalistic view* was founded by Rene Descartes. He argues that an understanding of presence is based on an understanding of the connection between the psychological and the physical domain. The second view, the *ecologic view*, is based on the epistemology of Martin Heidegger and James J. Gibson. They argue that presence is tied to our normal, physical interaction with the environment. An approach to unify both philosophies was attempted by Sheridan (1999) using dynamical systems and observer theory.
In the following both philosophical stances are explained and the unifying approach is presented in third part. Further readings should contain [2; 4–12].

The Rationalistic View

The rationalistic view is a theory of human knowledge acquisition. Most densely expressed by Descartes' saying "Cogito ergo sum.". In contrast to the empiricists' stance (knowledge is obtained by experience) the rationalists argue that knowledge is obtained by reasoning. The characteristics of the rationalists's view are:

1. Humans live in an environment with objects having properties that are unique, i.e. *objective*.
2. Perception is the process to register these properties and save them in thoughts and feelings that are *subjective*.
3. Decisions whether an action is performed or not are based on the subjective environment model by *reasoning*.

The kernel of the Rationalist's view is the transduction of the *objective* physical realm to the *mental* or *subjective* realm of the human observer. Or, as Descartes said, the distinction between *res extensa* and *res cogitans*.
Since subjective models can never represent the objective environment in full, important questions arise: To what extend are they valid? What are the causal relationships between the physical and the subjective realms?
This leads to a fundamental challenge of the rationalistic approach. Since the perceiver's

subjective world is a representation of the physical world, the nature of this model has to be analyzed to judge its validity. By now, the state of the art denies that it is possible to analyze the very nature of the mental models. Only general inferences can be made since measurements are spoiled by the probabilistic nature of the psychophysical and neuroscientifical results.

The Ecologic View

The philosophy of Heidegger can be seen as the counterpart to the philosophy of Descartes. While Descartes' philosophy is based on the dualism between subjective and objective, Heidegger's work provides a description of how humans exist in relationship to the environment. He argues that detached steps in the process of being (*Dasein*) do not exist. At least not if the human observer is engaged in *concernful action*. Heidegger's epistemology is expressed by the *hermeneutic circle*. It refers to the idea that one's understanding of a situation as a whole is established by reference to the individual parts and one's understanding of each individual part by reference to the whole. Neither the whole situation nor any individual part can be understood without mutual reference.
The characteristics of the Heideggerian view are:

1. The human must continuously act in the environment in which he is *thrown in*.
2. The environment is continuously changing, i.e. time-variant and prediction is impossible.
3. No *detached analysis* of the environment is possible. The environment is too complex to develop subjective models from its objects. Only goal-directed interpretation is possible.
4. The elements of the environment (things) are perceived according to their usefulness. When engaged in *concernful action* things can become transparent (Tools become *ready-to-hand*.).
5. The human-environment interaction can become instable (*break-down*). Then transparency ceases (Tools become *present-at-hand*.)

The kernel of Heidegger's view is the integral, reciprocal conception of the human existence: Dasein means to continuously act, perceive, and interpret the environment. The basic challenge is its usefulness for technical applications. Heidegger refuses to define the relationship between the subsystems within the human-environment closed loop in detail, insisting that no detachable subsystems exist.
The perceptual theory of Gibson has a lot in common with the Heideggerian view. Gibson states that perception is a process of gathering information from an environment. Thereby, information must not be replaced by the sensed stimuli but rather as a property of the environment in the context of the current action. Gibson termed this action-supportive information *affordance*. Gibson's epistemology is congruent with the view of Heidegger in terms that a percept stems from the reciprocity between perceiver and environment. While according to the rationalist's view subjective percepts are veridical if they match the objective environment, Gibson argues that percepts are veridical if they support successful action in the environment. Mental representations to infer new information are superfluous, since the environment provides all information directly.

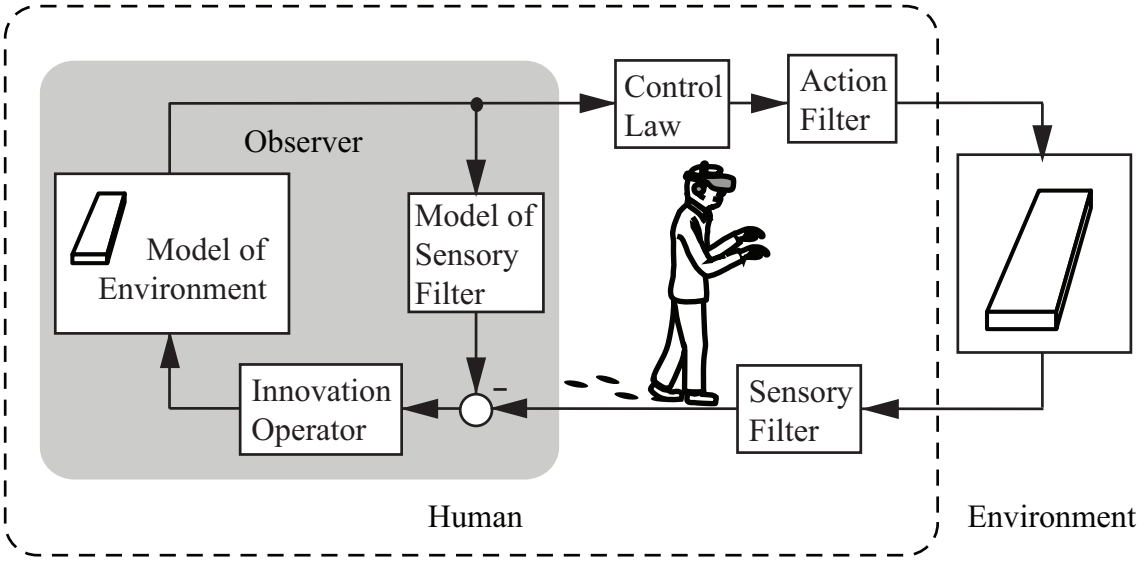


Figure 2.1: Epistemology according to Sheridan (adapted from [2]): An observer structure is the underlying concept of successful perception and action.

Since both epistemologies emphasize the interactions among the human and their environment and have many parallels, they are jointly referred to as the *ecologic view*.

The Unifying Approach by Sheridan

Sheridan developed a unifying approach of presence that comprises the rationalistic and the ecologic view. It is particularly suited for the community of technology-mediated presence and based on observer theory that mediates between imperfect sensors and imperfect internal models. Characteristics of the view are:

1. An environment can never be known, only estimated.
2. Probabilistic estimates are gained by action/perception.
3. Action and sensing are conditioned by filters (humans sensorimotor capabilities).
4. Probabilistic internal models include action and sensory states.
5. Action and sensing changes the environment

The concept is sketched in Figure 2.1.
The unifying consequences of the view are:

1. In the estimation paradigm, objective reality is what one's subjective model converges to in a sufficient, stable environment.
2. Reciprocity between human and environment (throwness) is given by the closed loop characteristic of the overall system.
3. Subjective models/interpretations are more accurate in the state of detached analysis than in concernful action because of extended convergence time.
4. Instability and break-down of the concernful action occurs, when the overall system has dynamics that exceeds human capabilities.

The approach faced critics in that it is only a technically sophisticated form of knowing and evaluating the environment in which the human is immersed in and, therefore, that it is actually a rationalistic approach. On the other hand, it is the only philosophical view that indisputably comprises the kernels of both philosophical stances and that is formulated in a cybernetical commensurable way to be exploited for the topic of interest.
The scientific disciplines of psychology and engineering provide various means to realize this unifying approach. For example, interactive closed loops can be modeled as differential equations with arbitrary complexity; model predictive control algorithms affect the dynamics of the overall closed loop system according to internal representations; the Kalman filter deals with probabilistic information in an optimal manner; Gibbsonian affordances can be represented mechanically by immittances that relate effort and flow variables to each other; etc.

2.1.2 Presence in Technology-Mediated Environments

A sound philosophical foundation enables the derivation of distinctive principles for technology-mediated presence. These principles will provide a helpful guidance for psychological and technical resarch in this field. A general model of technology-mediated presence must be based on the philosophies of presence. In a second step the technological means as the salient criterion for technology-mediated presence have to be taken into account. In a third step the philosophy has to be broken down into different principles that provide a supplemental but distinctive description. In this elaboration, the general model is broken down into two supplemental principles that neatly structure technology-mediated presence. The first principle, *distal attribution*, describes the extension of the human body schema to the target environment. The second principle, *proximal stimulation*, focuses on the display of the target environment on the operator side.
In the following, the general model of technology-mediated presence is presented. Thereafter, the principles of distal attribution and proximal stimulation are derived from this paradigm. Further readings should contain [2; 12–17].

A General Model

A general model of technology-mediated presence is obtained by extending the presence model of Sheridan by inserting technological means between human and environment. Hence, a human observer interactively experiences a target environment through technological means. The sketch depicted in Figure 1.1 is therefore an extension of Figure 2.1. The extension consists of the technological means classified as HSI, COM, and TO. Based on this general model, technology-mediated presence can be defined as:

- Technology-mediated presence is the feeling of being present in a target environment (rationalistical definition).
- Technology-mediated presence is tantamount to successfully supported action in a target environment (ecological definition).

These definitions are well known in the community and derivations of these descriptions are found frequently. However, the proposed definitions are too general to serve as

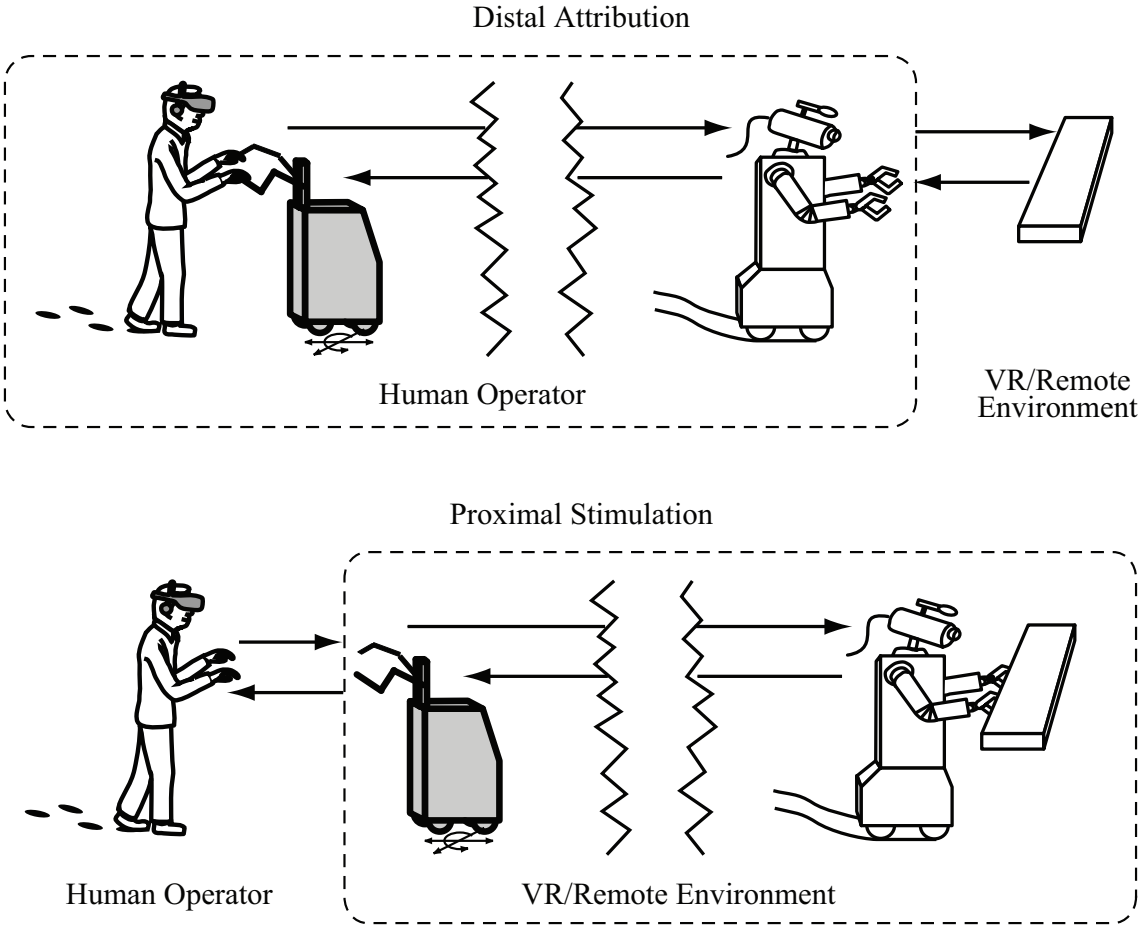


Figure 2.2: Technology-mediated presence from different viewpoints: From the viewpoint of distal attribution (upper figure) the human operator's body schema is extended to the technological means (HSI, COM, TO). From the viewpoint of proximal stimulation the target environment is relayed to the human operator and displayed by the human system interface. The operator's body schema is not changed.

sources for structural research on technology-mediated presence. For example they refuse to separate technological-mediated presence from the kind of presence experienced by reading books or gained by pure imagination. Hence, the definitions need substantial adaptation toward the technological character of the topic under consideration.
Based on the general model two different principles of technology-mediated presence are introduced in the following, what mutually distinguish at the very nature of technological-mediated presence: The way the technology is interpreted within Sheridan's model. The first principle is termed *distal attribution* describing the extension of the human operators body schema to incorporate the technological means. The second model is called *proximal stimulation*, it describes the reproduction of the target environment at the operator side.

Principle 1 - Distal Attribution

Distal attribution describes the effect that the border between human body and environment is not strict. Generally, most of our perceptual experience is referred to external space beyond the limits of the sensory organs where the stimulation occurs. Clear evidence is given by the phantom limb phenomenon, where human's physical body is not congruent with his internal body schema. Hence, the distinction between the human's self and his non-self is obviously not fix. Science states it can vary tremendously. Applied to technological-mediated presence this means, that humans are able to attribute technological devices to their body schema. HSI, COM, and TO become part of human's action and sensory filters (see Figure 2.1). In the ideal case, the technology would be incorporated in full instantly leading to perfect presence. The model of distal attribution is depicted in the upper part of Figure 2.2. A definition for technology-mediated presence based on the principle of distal attribution is:

- Technology-mediated presence is the extension of a human's perceptual and motor capabilities to a target environment.

The definition bolsters the development of measures and strategies for improving the immersion of an operator into the target environment. Developments will be based on the human operator's awareness of the mediating-technology (HSI, COM, TO). Most of the cognitive and psychophysical measures are derived on this definition as well as training of human operators to get immersed into a mediated environment.

Principle 2 - Proximal Stimulation

Proximal stimulation is the counterpart of distal attribution meaning that the target environment is reproduced at the operator side stimulating the human operator locally. The reproduction has to match the original environment as close as possible. The principle is sketched in the lower part of Figure 2.2. The definition derived from this model is straightforward:

- Technology-mediated presence is the sound reproduction of a target environment at the operator side.

The human operator is excluded from the mediating technology. The focus is on the technical congruence between target environment and displayed environment. Technical measures are based on this understanding.

Both principles formally describe a coordination mechanism according to which structural research on presence systems is possible. It is a widespread mistake that it is particularly meaningful trying to obtain a high degree of presence by an operator with a flexible body scheme and a perfect reproduced target environment. This leads to unplanned acting, since no references (either the human operator or the target environment) are given and misfits could be blamed on either both of them.

2.1.3 Evaluation Methods

As described above determinants of presence are the characteristics of the human operator as well as the extent and fidelity of sensorimotor information provided by the technological device. Due to the complex and young research field a single accepted paradigm to evaluate technology-mediated presence still does not exist. The degree of presence obtained by a presence system can be evaluated using e.g. *subjective*, *objective*, *psychophysical*, and *technical methods*. This categorization is not mutual exclusive but partly overlapping. All evaluation methods can be categorized by the help of the elaborated principles (distal attribution, proximal stimulation).
In the following, selected evaluation methods for multimodal presence systems are introduced and briefly described. Further readings should contain [18–24].

Subjective Methods

Subjective information about an operator's experience of presence is given by *post-test ratings* according to Principle 1 and Principle 2. Thereby, different questions concerning immersion and quality of the reproduction are posed to the human operator after he has conducted a task by a presence system. Hence, the questionnaire measures presence according to Principle 1 and Principle 2. Since presence is a highly subjective, situation dependent or interpretative experience, post-test ratings are said to be the basic measurement procedure for evaluating presence systems. This fact and also the straightforward implementation and application of this test make it the predominant evaluation procedure in presence research. Two different post-test questionnaires are described in [21; 23].

Objective Methods

Objective measures give physically revisable, i.e. objective, evidence that a human operator is present in some target environment according to Principle 1 and Principle 2. The underlying concept is that a human operator's performance in a mediated target environment converges toward his performance in a non-mediated environment the more the human operator is immersed in the target environment(Principle 1) and the higher the quality of reproduction by the technological device is (Principle 2). Objective performance measures can be based on reflexive or socially conditioned responses of the operator. Furthermore, task completion time and task completion rate are objective evaluation methods.

Psychophysical Methods

Psychophysical methods strive to objectify the subjective experience by the human operator, hence they are based on Principle 1. The methods can be employed in different experimental paradigms to detect absolute or discrimination thresholds. The basic experimental paradigm for presence system is the paired comparison between target and real environment or the comparison between different presence systems. The experience consists of a stimulus or an environmental property that has to be evaluated by the human

operator through the mediating technology. According to the applied method evaluation occurs in different manners. Using the method of *magnitude estimation* the human operator has to judge the stimuli directly by freely assigning numbers that represent the intensity of the stimulus or the magnitude of a conflict. Using the *method of adjustment* or the *method of constant stimuli*, the operator's detection or discrimination performance is recored indirectly by pairwise comparisons between standard stimulus and deviating comparison stimuli. The following psychophysical methods are applied in the experiments described in this thesis

- Magnitude estimation
- Staircase method
- Method of adjustment
- Method of constant stimuli
- Two-alternative-force-choice task (2AFC-task)

Detailed explanation of the methods is given in Subsection 2.2.

Technical Methods

Technical measures are based on Principle 2 and are termed *transparency* measures. Transparency means, the operator can 'look through' the mediating technology directly perceiving the target environment. According to Yokokohji (1994) transparency requires the equality of position and force measured at the operator and the TO:

$$x_o = x_t \qquad\qquad f_o^e = f_e^t. \tag{2.1}$$

See also Appendix A.2 for a detailed descriptions of the interaction within a presence system. Performance measures can be developed based on the position error and the force error in time or frequency domain. They allow statements of the quality of the technological device in relation to other devices. A similar method to evaluate the technology used in a haptic presence system is the comparison of the dynamics at operator and teleoperator side. According to Lawrence (1993) transparency requires the equality of the mechanical impedance encountered by the TO and the mechanical impedance displayed to the human operator:

$$Z_d = Z_e. \tag{2.2}$$

See also Subsection 2.4 for a detailed description of the control structure of a presence system. Measures derived from this method are impedance error norms mostly applied in the frequency domain. An intuitive interpretation is not possible anymore, but it is straightforward that errors in the lower part of the spectrum are more important than high-frequency errors since these may not be perceivable by the human operator.
A third method, *fidelity*, was introduced by Cavusoglu (2001). According to this method transparency is achieved if the change in the impedance encountered by the TO equals the change in the impedance displayed to the human operator:

$$\mathrm{d}Z_d = \mathrm{d}Z_e. \tag{2.3}$$

Performance measures derived from this method are impedance change error norms.
All measures (2.1), (2.2), and (2.3) describe the impact of the complete technology HSI, COM, HSI. As an advantage these measures can also be applied to evaluate the impact of a certain part of the technology, respectively HSI or COM or TO, just by evaluating the signals produced by the subsystem under consideration.

Other measures based on physiology, posture, etc. exist but play a minor role in current research and are not used throughout this thesis.

2.2 Human Perception

When humans interact with an environment, information about this environment has to be measured and prepared for cognitive and autonomous operations. This process is called *perception*. Humans have multiple sensors to perceive their environment. These sensors are called *modalities* (e.g. *visual, acoustic, haptic, etc.* modalities). The haptic modality can further be divided into *kinesthetic* and *tactile* submodalities[1]. All information perceived from the environment is subject to uncertainty. Hence, a great deal in perceptual research concentrates on how uncertain perception can be modeled and which strategies humans deploy to take uncertainty into account. Another focus is on fusion and combination of multimodal information. The goal is a mathematical model that has a defined scope and that describes perceptual performance in a probabilistic way.
In Subsection 2.2.1 probabilistic models based on signal detection and estimation theory are introduced to describe human perception. The presentation is customized to the research conducted in this thesis. Special emphasis is laid on the theoretical foundations of two-interval perception, since this is the predominant concept of the experimental methods used throughout this thesis. In Subsection 2.2.2 the experimental methods used to record the performance of a human observer[2] are explained in detail.

2.2.1 Probabilistic Nature of Sensory Estimation

Human modalities do not deliver unique information about an environment. The measurement is uncertain because of sensor-inherent noise. Consequently, the observer has to decide if his percept stems from a certain signal or not. This decision can be biased. A mathematical theory that accounts for these aspects is *signal detection theory*. Thereby, human perception is modeled as ambiguous information afflicted with *Gaussian noise*. Operations and parameter definitions are performed in z-score to be comparable. *Bias* toward a certain decision and *sensitivity* (d') are the main statistics provided by signal detection theory. Furthermore, the theory was extended to perception of two-interval stimuli as this is important for many experimental methods. Models of multimodal perception that describe the *fusion of redundant information* or the *combination of complementary information* have also been developed based upon signal detection theory.

[1] This thesis only deals with kinesthetic haptic information, which is henceforth denoted as *haptic information*.
[2] In research branches dealing with perception and cognition the human operator is termed *observer*.

Stimulus	Response (Perceived s_1? - Yes/No.)		
	Yes	No	Total
s_1	x_h (Hits, F_h)	x_m (Misses, F_m)	$x_h + x_m$ ($F_h + F_m = 1$)
s_2	x_{fa} (False Alarms, F_{fa})	x_{cr} (Correct Rejections, F_{cr})	$x_{fa} + cr$ ($F_{fa} + F_{cr} = 1$)

Table 2.1: The 2×2 contingency table: The performance of an observer can be characterized by correct answers (hits, correct rejections) and wrong answers (misses, false alarms).

In the following, signal detection theory and human perception basics are introduced. Bias and sensitivity are explained as well as perception in two-intervals. Eventually, the focus is on multimodal perception. Further readings should contain [3; 25–27].

Signal Detection Theory and the Gaussian Assumption

Signal detection theory (SDT) is a mathematical approach to explain the performance of a human observer. It provides models and experimental paradigms to describe and measure human perception. The underlying assumptions are, firstly, that perception and decision making are inseparable connected to each other and, secondly, that both processes are subject to uncertainty.
Given a one-interval psychophysical experiment, in which subjects are told to respond whether they perceived a stimulus s_1 from a set of stimuli $[s_1, s_2]$, then the outcome can be denoted in the 2×2 *contingency table* depicted in Table 2.1. It shows an observer's performance during the $(x_h + x_m)$ presentations of stimulus s_1 and the $(x_{fa} + x_{cr})$ presentations of stimulus s_2. Thereby, x_h denotes the number of correct detected stimuli s_1 (*hits*), x_m is the number the observer missed to perceive the stimulus s_1 (*misses*), x_{fa} denotes the trials the observer perceived a stimulus s_1 while there was actually a stimulus s_2 (*false alarms*) and x_{cr} is the number the observer correctly identified stimulus s_2 (*correct rejections*). High performance in discriminating both stimuli from each other would be indicated by a concentration of responses along the negative diagonal (many hits and many correct rejections).
Relating the number of the different response types to the total number of responses yields the conditional, cumulative probabilities for a hit $F_h = F(\text{'Yes'}|s_1)$ (*hit rate*), a miss $F_m = F(\text{'No'}|s_1)$ (*miss rate*), a false alarm $F_{fa} = F(\text{'Yes'}|s_2)$ (*false alarm rate*), and a correct rejection $F_{cr} = F(\text{'No'}|s_2)$ (*correct rejection rate*). The sum of hit and miss rate is unity as well as the sum of false alarm and correct rejection rate. The sum of hit and correct rejection rate is called *proportion correct* F_{corr}.
Experimental results give rise to model the underlying probability densities of the stimuli s_1, s_2 as *normal distributions*. Although this assumption is not justified for all situations, it mostly holds. It is supported by the *central limit theorem* which states that a sum of independent and equally distributed random variables converges against the normal distribution. Perceptual processes may be biologically implemented by a chain of neuronal processes providing a normally distributed outcome each. Hence, the density distribution

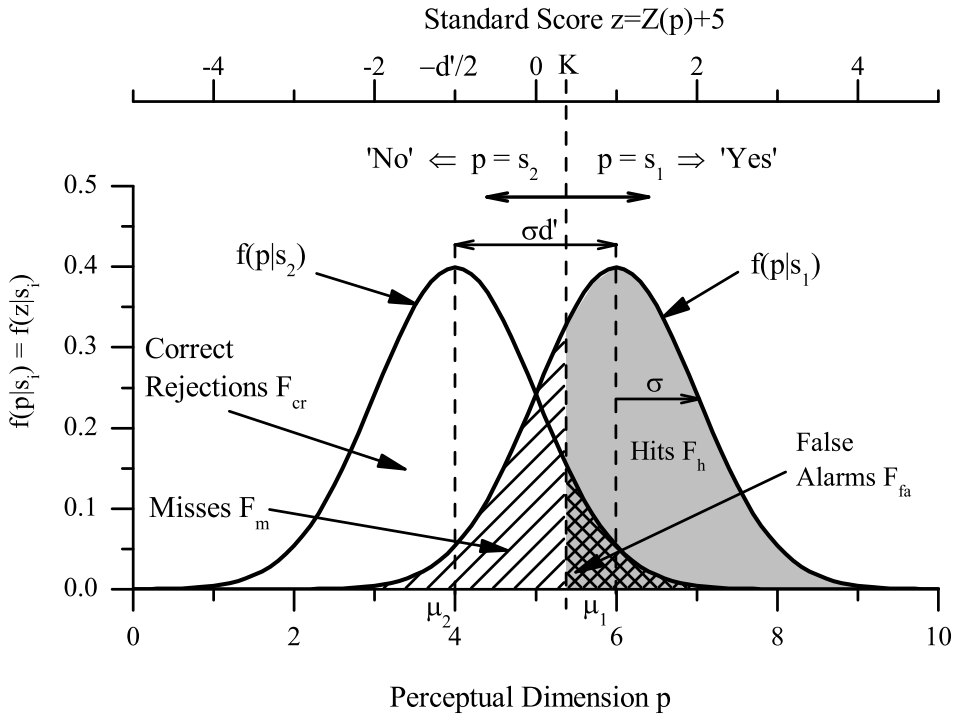


Figure 2.3: Internal problem of an observer in a Yes/No-task: In an ambiguous situation (overlapping distributions) the observer decides by the help of the criterion K and, according to the value of K, is eventually biased. The observer's sensitivity is expresses by the statistics d' which is the distance between the peaks of the density functions related to their common standard deviation, i.e. in z-scores (Since $\sigma = 1$ in this particular example, the perceptual dimension p and the z-scores are directly related by a constant value.).

of a percept f_p can be described by its mean μ and its standard deviation σ,

$$f_p(p|s) = N(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(p-\mu)^2}{2\sigma^2}\}. \tag{2.4}$$

While the mean value may change depending on the stimulus the variance remains the same for a certain sensor. The reliability of a sensor is defined as the inverse of the variance

$$r = \frac{1}{\sigma^2}. \tag{2.5}$$

The larger the variance of the sensor the less its reliability. Reliability is the equivalent to the signal/noise ratio in signal processing theory. However, since in psychophysics noise usually occurs in a different quantity as the stimulus and can have many sources, its influence on the stimulus is better evaluated indirectly using the statistical measure..
Now, the contingency table of a one-interval experiment, as described above, can be represented by two overlapping normal distributions each describing the perception of a sole stimulus (illustrated in Figure 2.3). This is the internal problem an observer faces when asked to report his percept. The observer can assess the perceptual value p but cannot distinguish which distribution led to the value.
The optimal rule to decide for one of the two stimuli is to establish a criterion K that divides the perceptual dimension into two parts. Above the criterion the observer responds

"Yes" and below the criterion the observer responds "No". If a value above the criterion arises from the stimulus s_1 the observer scored a hit otherwise the observer raised a false alarm. If a value below or equal to the criterion arose from the stimulus s_2 the subject performed a correct rejection and missed otherwise. The hit rate is the area under the s_1-curve (gray) constraint by the criterion on the lower side. The false-alarm rate is the area under the s_2-curve, respectively.
Response problems with different normal distributions can be transformed to the *standard normal distribution* f_s ($\sigma = 1$) to be comparable. This is done by the *z-score transformation*. The *z-score* is a dimensionless quantity derived by subtracting the mean from an individual normally distributed score and then dividing the difference by the standard deviation. The z-score transformation of random variable X is

$$z = Z(x) = \frac{x - \mu}{\sigma}. \tag{2.6}$$

The z-score indicates how many standard deviations an observation is above or below the mean. Hence, the standard normal distribution depends on the z-score $f_s(z)$. In Figure 2.3 the axis depicting the z-score is drawn above the diagram. Therein, μ is chosen to be the intersection of the two density functions. The z-score describes the internal problem in unities of its standard deviation σ. The z-score of a cumulative probability F is given by

$$z(F) = \sqrt{2}\text{erfinv}(2F - 1). \tag{2.7}$$

Bias and Sensitivity

Although the separation of the perceptual dimension p by the criterion K yields optimal decisions, the observer can still adopt different decision strategies. The observer's willingness to say "Yes" or "No" is called *response bias*. It can also be expressed as the observer's tendency to favor one response over the other. In general the response bias can be calculated based on hit rate and false alarm rate

$$K = -0.5(z(F_h) + z(F_{fa})). \tag{2.8}$$

When the false alarm rates and miss rates are equal ($z(F_h) = z(1 - F_m) = -z(F_m)$), i.e. on the intersection of the two density functions, the statistics is zero. Negative values arise when the false alarm rate exceeds the miss rate and positive values arise when the miss rate is lower. Three different classes of observer's can be identified.

1. Observers who perform unbiased decisions, $K = 0$, are called *neutral*.
2. Observers who perform negatively biased decisions, $K < 0$, tolerate many false alarms. They are called *liberal*.
3. Observers who perform positively biased decisions, $K > 0$, tolerate many misses. They are called *conservative*.

Varying the criterion $-\infty < K < \infty$ for a certain pair of stimuli yields the characteristic of the observer under all possible decision behaviors. This summary is called *receiver operating characteristic* (ROC).
The sensitivity of the human observer in discriminating the two stimuli from each other can also be inferred from the hit rate F_h and the false alarm rate F_{fa}. A perfectly sensitive

observer has a hit rate of unity and a false alarm rate of zero. A completely insensitive observer is unable to distinguish the two stimuli at all. Normally, sensitivity falls between the two poles. The most important sensitivity measure of SDT is d'[3]

$$d' := z(F_h) - z(F_{fa}). \tag{2.9}$$

When observers cannot discriminate at all, $F_h = F_{fa}$ and $d' = 0$. When the ability to discriminate is nearly perfect, e.g. $F_h = 0.99$ and $F_{fa} = 0.01$ then $d' = 4.65$. Moderate performance implies that d is near unity. Correct performance on 75% corresponds to $d' = 1.35$ and in case of 69% correct $d' = 1$.
Stimuli that are easy to discriminate can be thought of being perceptually far apart, therefore d' measures a perceptual distance. Moreover, d' has all mathematical properties of a distance measure: The distance between the object and itself is zero; the distance between object s_1 and s_2 is same as the distance between s_2 and s_1; the measure fulfills the triangle equality, i.e distances between objects can be inferred from related distances; finally, d' is unbounded. More illustratively the statistics d' is the distance between the s_1 and s_2 means in units of their common standard deviation

$$d' = \frac{|\mu_1 - \mu_2|}{\sigma}. \tag{2.10}$$

The sum of different distances d'_i is called *cumulative* d'. Equations (2.9) and (2.10) only hold for equal distributions for stimulus s_1 and s_2.
Proportion correct F_{corr} is a simple performance measure for unbiased observers. It is defined as the average of hit and correct rejection rate or in case of equal number of trials for s_1 and s_2 the difference between hit and false alarm rate

$$F_{corr} = F_h - F_{fa}. \tag{2.11}$$

If the observer is unbiased, $K = 0$, the equality of hit and correct rejection rate is sufficiently compelling that investigators measure proportion correct, $F_{corr} = F_h = 1 - F_{fa}$, only. Substituting into equation (2.9) and using the relation $z(1 - F) = -z(F)$ yields

$$d' = 2z(F_{corr}). \tag{2.12}$$

Latter equation only serves as a valid performance measure in case of an unbiased observer.

Perception in Two-Interval Experiments

Experiments with two intervals differ from one-interval experiments in having more than one stimulus presented in either spatially or temporally subsequent steps. Experimental paradigms based on two intervals are the *two-alternative forced-choice task* (2AFC-task)and the *same-different task*, from which the *reminder task* is derived.
The decision problem can be depicted in a two dimensional space as illustrated in Figure 2.4 for a reminder experiment. The first stimulus serves as the reminder or *standard* and the observer has to decide if the second stimulus, the *comparison*, is different from the first. Hence, the observer looks for two different stimulus combinations $< s_1, s_1 >$, $< s_1, s_2 >$.

[3] pronounced 'dee-prime'

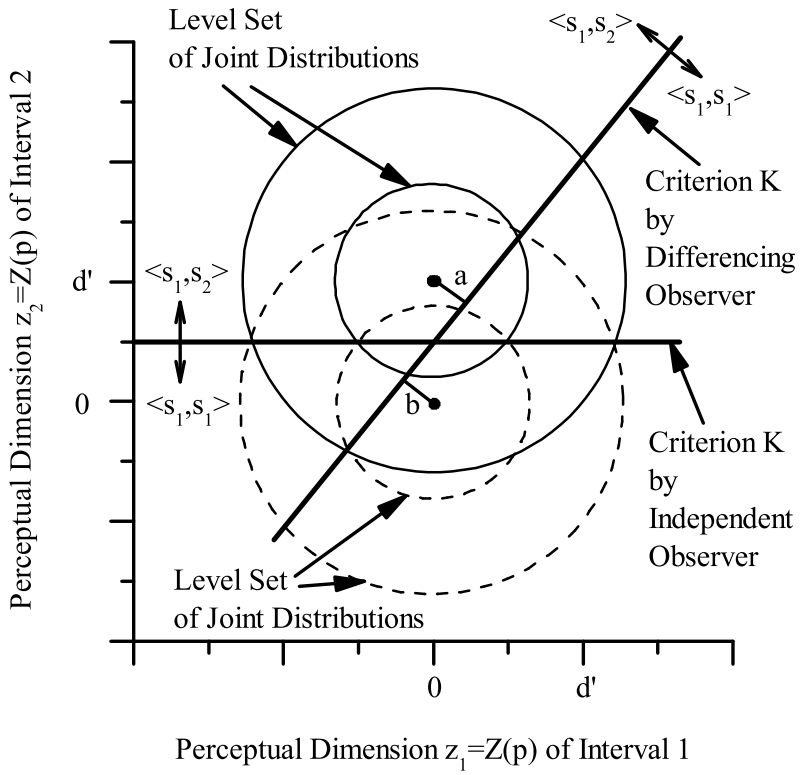


Figure 2.4: Graphical representation of a two-interval discrimination problem (adapted from [3]): The independent observer does only rely on the stimulus presented in the second interval. The differencing observer relies on both stimuli using a criterion that is perpendicular to the line $z_1[p] + z_2[p] = \text{const.}$. Thereby he reduces the distance from the peak values to the criterion line by a factor $1/\sqrt{2}$ in comparison to the independent observer.

For an analysis each interval is treated as a separate dimension that represents the observer's perceptual variable for the stimulus presented in this interval. A single dimension resembles a common Yes/No-task with stimuli peaks having a distance of d' in z-score. An observation resulting from a single trial is a point on the two-dimensional space. The set of possible observations of one stimulus combination results in a two-dimensional normal distribution. Their level sets form concentric circles around a peak value since perception of both stimuli is statistically independent.
Discrimination in two-interval tasks can be ruled by two different strategies. The first strategy, called *independent observer strategy*, has already been analyzed for the one-interval experiment. The observer places his criterion K independently of the stimulus in the first interval implying that the observer ignores the reminder stimulus s_1. The distance between the distribution peaks in unities of standard deviation is d'. The second strategy is called *differencing observer strategy*. Thereby, the observer places his criterion according to the difference between the perceptual variable of interval 1 and interval 2. In case of two equally distributed random variables for both stimuli, the criterion is inclined by 45°. Using this criterion the distance, $a+b$, between the peaks of the two distributions is $d'/\sqrt{2}$ since the standard deviation of the joint probability is $\sqrt{2}$ larger than in case of a single

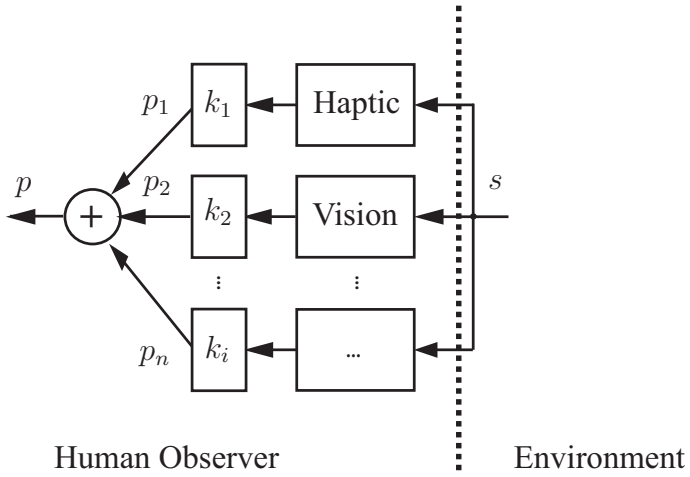


Figure 2.5: Fusion of redundant multimodal information: Weightings that summarize to unity modulate the linear combination of the different sensory estimates according to their reliability.

stimulus set. A further treatment of the reminder task, which is frequently used in this thesis, is conducted in Subsection 2.2.2.

Perception of Redundant and Complementary Multimodal Information

Information from different sensors/modalities about the same environmental property is called *redundant*. For example, the length of an object can be estimated by the visual and the haptic modality. The different redundant sensory estimates p_i may exhibit different means μ_i and different reliabilities (i.e. different standard deviations σ_i). In many cases, especially if the bimodal conflicts ($c := |\mu_i - \mu_{i+1}|$) are not too large, the different sensory estimates are fused to a final percept p by a weighted, linear combination

$$p = \sum_{i=1}^{m} k_i(\sigma_1, \sigma_2, .., \sigma_m)p_i, \tag{2.13}$$

where k_i are the weightings, which sum up to unity

$$\sum k_i = 1. \tag{2.14}$$

It has been shown that for certain cues multimodal fusion by the perceptual system is near the theoretical optimum that can be reached when fusing independent redundant information for maximal reliability. This problem is called *maximum likelihood optimization*. Thereby, the weightings are calculated to minimize the expectation of the quadratic estimation error $e^2 = (p - s)^2$

$$\min\ E[(e)^2], \tag{2.15}$$
$$\text{s.t. } (2.13), (2.14). \tag{2.16}$$

The results are optimal weightings k_i that lead to a final percept with minimal variance, i.e. with highest reliability. The optimal linear weighting represents a fusion mechanism that equals the static *Kalman filter*.
The solution of the bimodal fusion is sketched in the following: Two sensory estimates $p_1 \sim N(\mu_1, \sigma_1)$ and $p_2 \sim N(\mu_2, \sigma_2)$ are combined according to $p = k_1 p_1 + k_2 p_2$ and $k_2 = 1 - k_1$. Hence, the fused estimate can be denoted as

$$p = k_1(p_1 - p_2) + p_2, \tag{2.17}$$

which exhibits the common *predictor/corrector structure* of a *Luenberger observer* and k_1 represents the *Kalman gain*. The estimation error is given by

$$e = p - s = \underbrace{s - p_2}_{e2} + \underbrace{k_1 s - k_1 p_1}_{k_1 e_1} \underbrace{-k_1 s + k_1 p_2}_{-k_1 e_2} \tag{2.18}$$

and finally by

$$e = k_1(e_1 - e_2) + e_2. \tag{2.19}$$

The optimal value of the Kalman gain is calculated by differentiating the the expectation of the quadratic estimation error e^2 with respect to k_1

$$\frac{dE[(e)^2]}{dk_1} = \frac{dE[(k_1(e_1 - e_2) + e_2)^2]}{dk}. \tag{2.20}$$

The well known result is

$$k_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \tag{2.21}$$

Because of equation (2.14) the remaining weighting is

$$k_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}. \tag{2.22}$$

The optimal weightings fuse the different estimates according to their reliability (equation (2.5)): Information with low reliability and, therefore, high standard deviation is weighted less than information with high reliability and small standard deviation. The final percept

$$p_{12} \sim N(\mu_{12}^*, \sigma_{12}^*) \tag{2.23}$$

can be described by its mean

$$\mu_{12}^* = k_1 \mu_1 + k_2 \mu_1 \tag{2.24}$$

and its standard deviation

$$\sigma_{12}^* = \sqrt{\frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}}. \tag{2.25}$$

Latter equation shows that the fused percept has the reliability of both sensory estimates

$$r_{12} = r_1 + r_2. \tag{2.26}$$

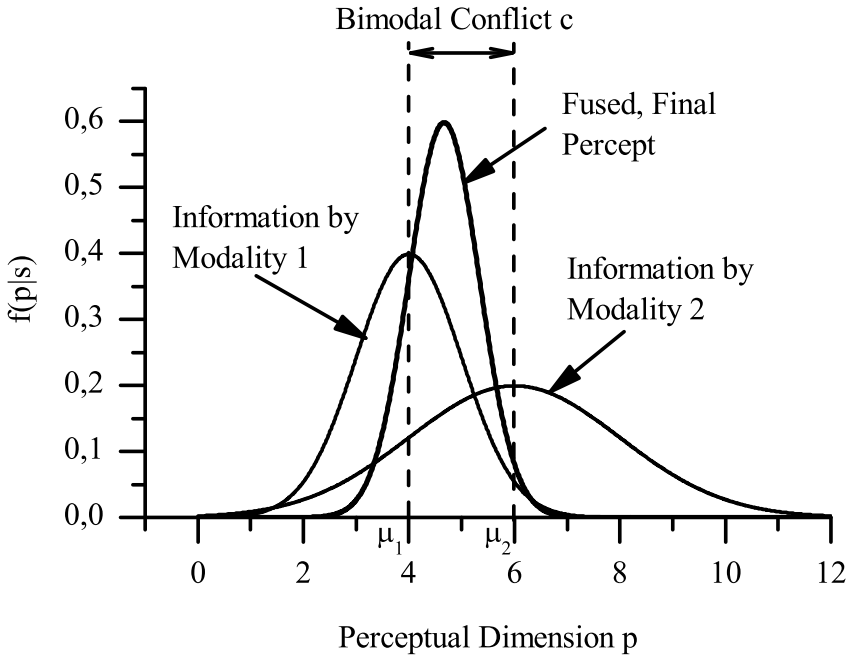


Figure 2.6: Optimal fusion of bimodal redundant information: The final percept has an optimal variance by incorporating the reliabilities of both sensory estimates.

Non-redundant or *complementary* information is usually perceived differently than redundant information. As combinations are manifold in reality the perceptual system might also use manifold techniques combining non-redundant information to infer new information. A model to *combine* incommensurable information and to infer new information is provided by *Bayes' theorem*

$$F(p_1|p_2) = \frac{F(p_2|p_1)F(p_1)}{F(p_2)}. \tag{2.27}$$

$F(p_1)$ is the prior joint probability of the complementary information p_1. It is *prior* in the sense that it does not take into account any information about p_2. $F(p_1|p_2)$ is the conditional probability of p_1, given p_2. It is also called the posterior probability because it depends upon the specified value of p_2. $F(p_2|p_1)$ is the conditional probability of p_2 given p_1. $F(p_2)$ is the prior or marginal probability of p_2, and acts as a normalizing constant. This model describes how complementary information, which is not mutually depending, can be combined to infer new knowledge about the environment. When p_1 contains the complementary information and $F(p_2|p_1)$ encodes how this complementary information gives rise to the new information p_2 then equation (2.27) states how likely the measured, complementary information p_1 is due to the inferred information p_2. For an example, if p_1 ='green and round' and $F(p_2|p_1)$ encodes how likely 'green' and 'round' at the same time is to be an apple, then $F(p_1|p_2)$ provides the possibility that 'green and round' measurements are caused by an apple. The combination of the complementary information p_1 can be performed over their occurence probabilities very simple. In the example given, if both events, 'green' and 'round', are statistically independent, e.g. because the properties

are measured by different modalities, than the joint probability is calculated as

$$F(p_1) = F(p_g)F(p_r), \tag{2.28}$$

where $F(p_g)$ is the probability that 'green' is measured and $F(p_r)$ is the probability of 'round'.
However, this model does not describe the perception of environmental properties that can only be inferred by combining complementary information, such as mechanical impedances. A mechanical impedances can either be a mass, a damper or a spring or any combination of these elements. To perceive mechanical impedances humans have to combine position-based information (position, velocity, acceleration, etc.) with force information. According to the impedance to be perceived a certain mathematical process like division or convolution is necessary to infer the underlying structure. The process also involves fusion of redundant information since position-based information e.g. can be estimated by the visual and the haptic modality. Which impedances humans are able to perceive and how their perceptual system organizes the combination and fusion processes is subject to current research, and also tackled in this thesis.

2.2.2 Experimental Assessment

The set of experimental methods to record an observer's performance is termed *psychophysics*. Psychophysical experiments explore *detection* or *discrimination* performance. Methods are further classified according the number of intervals (one-, two, n-interval). Especially the *reminder procedure,* a two interval Yes/No-experiment is predominantly used for the methods in this thesis. Whether *constant* or *adaptive* stimuli are presented during these intervals provide a third classification criterion. An underlying concept of nearly all experimental methods is the *psychometric function*, that relates an observer's performance to different stimuli. The main performance properties provided are the *empirical threshold* and the *perceptual variability*. Experimental methods used in this thesis are *magnitude estimation*, *method of adjustmen*t, *staircase method*, and *method of constant stimuli*.
In the following the reminder experiment is introduced and analyzed. Afterward, the psychometric function is described and their relation to the perceptual performance (sensitivity) is explained. Lastly, the experimental methods are described and the calculation of threshold and variability is given. Further readings should contain [3; 28; 29].

Reminder Tasks

In common Yes/No-tasks it is sometimes difficult to remember how the stimulus looked or felt like. Placing a reminder or *standard* stimulus before each observation interval may improve performance. This two-interval procedure consists of the standard stimulus followed by the *comparison* stimulus $< s_{ref}, s_c >$ and the observer has to decide whether the comparison stimulus was different from the standard stimulus. Hence, the observer looks for the stimulus combinations $< s_{ref}, s_{ref} >$ or $< s_{ref}, s_c >$. The standard stimulus is not essential to the judgment process since observers can also memorize this part of the comparison. The comparison stimulus is either adjusted by the observer actively (method of matching) or presented to the observer out of a set of different comparison stimuli

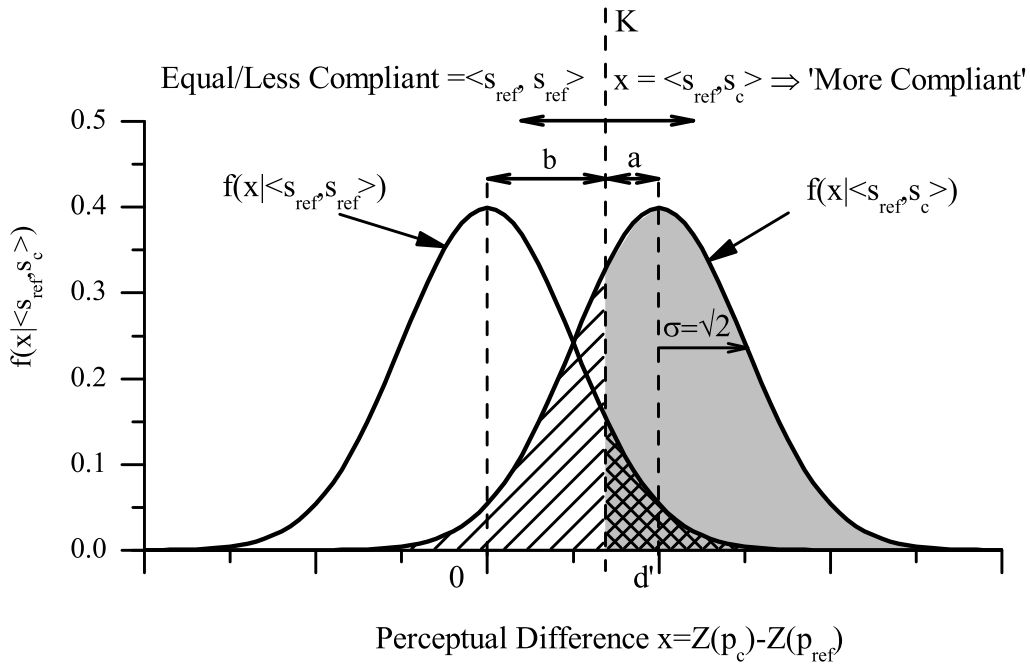


Figure 2.7: Discrimination in a Yes/No-reminder task by using the differencing observer strategy: The decision axis is the perceptual difference between the observations in the two intervals (as introduced in Figure 2.4). Since the observer's result is based on two estimations the standard deviation is $\sqrt{2}$ higher than the standard deviation of a single estimate (The abscissa is not in z-score!). This leads to poorer discrimination performance by the same factor.

(method of constant stimuli). A two dimensional interpretation of the discrimination process is already depicted in Figure 2.4 showing the two possible discrimination strategies as discussed in Subsection 2.2.1.
Empirical evidence bolsters the assumption that observers in reminder experiments apply the differencing observer strategy rather than the independent observer strategy. Using the independent observer strategy implies perfect memorization of the standard stimulus from the training trials. This is not reasonable. Observers are assumed to use the differencing criterion in the experiments described in this thesis. A one dimensional interpretation of the differencing discrimination problem is illustrated in Figure 2.7. Since the reminder stimulus adds variance to the discrimination performance of a differencing observer the difference between hit and false alarm rate is smaller than in a one-interval Yes/No-task and, hence, performance is poorer. According to equations (2.9) and (2.10) the performance in a Yes/No-reminder task is

$$d'_R := z(F_h) - z(F_{fa}) = \frac{d'}{\sqrt{2}}. \tag{2.29}$$

This equals the distance $a + b$ in Figures 2.4, 2.7 as the decision axis is perpendicular to the criterion line of the differencing observer in the first plot. For a comparison, the performance of an observer in a 2AFC-task is

$$d'_{2AFC} := \sqrt{2}d'. \tag{2.30}$$

That means, the Yes/No-reminder task exhibits poorest performance when compared to the Yes/No- and the 2AFC-task. That, means the 2AFC-task is the easiest task. Therefore, 2AFC-tasks permit measurement of sensitivity to smaller stimulus differences than would be possible with Yes/No- or reminder tasks. However, in practice this theoretical result is affected by the imperfect memory of humans dispelling the Yes/No-task without reminder and by the difficult nature of 2AFC-tasks that will be affected by confused subjects. Furthermore, the 2AFC-task is more difficult to implement and the implementation is more difficult to verify.

The Psychometric Function

The standard problem in signal detection theory, as explained in Subsection 2.2.1, is the assessment of the dependent variable d' quantifying an observer's performance of perceiving a certain stimulus difference. The inverse problem is to define a stimulus difference that yields a certain performance, i.e. a certain value of d'. Such a stimulus difference is called *empirical threshold*. Empirical thresholds are not related to the thresholds defined by the threshold theories that preceded signal detection theory and outdatedly used F_{corr} as performance measure.
Bringing to light empirical thresholds requires the display of a series of stimuli that envelop the expected difference level and their impact on the observer's performance. The relation between different stimuli (independent variable) and performance as proportion of detected stimuli (dependent variable) is called *psychometric function* (PF), $F_\Psi(s)$. Basically, two statistics can be inferred from a PF:

1. The empirical threshold
2. The performance variability

The empirical threshold is the inverse of $F_\Psi(s)$ at a particular performance level. In detection tasks it is called the *empirical detection threshold* (DT) and in discrimination tasks the *point of subjective equality* (PSE). Furthermore, in discrimination tasks the difference between the PSE and the center of the displayed stimulus set is called *constant error* (CE). The CE states if subjects are biased in their response behavior. The stimulus difference between PSE and a certain performance level is called the *just noticeable difference* (JND). The JND is be chosen to yield certain sensitivity d' and is then the solution of the inverse problem. Throughout this thesis the JND is chosen to yield

$$\text{JND} = d' := 1. \tag{2.31}$$

Since the value of the JND depends on the slope of the PF it is the measure of performance variability. In detection tasks this performance variability is not specially termed. A Psychometric function of a Yes/No-reminder task is depicted in the upper right diagramm of Figure 2.8.
The shape of the psychometric function does not follow from detection theory. While the distributions that depict an observer's decision problem are plotted against his internal perceptual dimension p (Figure 2.3), the PF is plotted against a physical stimulus s. The shape of the PF can only be inferred from the shape of the internal distributions when based on a model of stimulus transduction joint to signal detection theory. Many theories about stimulus transduction have been proposed such as the ideal observer model by Green/Swets

(1966) or the quantum detection by Cornsweet (1970). Philosophical relations have been pointed out in Section 2.1.1. In its simplest case, the transduction process T linearly maps the stimuli $s \in S$ to the perceptual dimension $p \in P$, $T : \mathbb{S} \rightarrow \mathbb{P}$. That means, the transduction satisfies the following conditions

$$p_1 + p_2 = T(s_1) + T(s_2) \tag{2.32}$$

and

$$ap = T(as). \tag{2.33}$$

According to the Gaussian assumption laid down in Subsection 2.2.1 the shape of the psychometric function is a cumulative normal distribution. Now, performance variability (JND) or slope of the PF can be mathematically defined by the variability parameter of the distribution, i.e. the standard deviation σ_Ψ. Furthermore, the standard deviation of an observer's internal distributions can be inferred by standard deviation of the PF σ_Ψ. In case of a unity mapping $a = 1$ both standard deviations are equal. Hence, in a Yes/No-task the standard deviation of the internal distributions, which is the sensor variability, is equal to the standard deviation of the PF. In a Yes/No-reminder task or a 2AFC-task, however, the sensor variability is $\sqrt{2}$ smaller than the variability of the PF because each decision is based on two sensor distributions and the joint probability has double variance

$$\sigma_\Psi^2 = \sigma^2 + \sigma^2 \tag{2.34}$$

and therefore

$$\sigma = \frac{\sigma_\Psi}{\sqrt{2}}. \tag{2.35}$$

Generally, the PF is denoted as follows:

$$F_\Psi(s) = \gamma + (1 - \lambda - \gamma)\Psi(s), \tag{2.36}$$

where γ is the lower asymptote corresponding to the base rate of performance in absence of a stimulus and $1-\lambda$ is the upper asymptote remarking the observer's performance to an arbitrary large stimulus. The two latter parameters can be due to stimulus independent errors, so called *lapses*. The function $\Psi(s)$ describes the underlying psychological mechanism and is defined to be the normal distribution according to the Gaussian assumption. The stimulus value s typically goes from 0 to a large stimulus for detection and from large negative to large positive values for discrimination. A PF and its relation to signal detection theory is depicted in Figure 2.8.
Three distinctions need to be made for the experimental design set up to record a PF:

1. Experimental Goal: Detection vs. discrimination
2. Procedure: Yes/No - vs. forced-choice methods
3. Test levels: Adaptive methods vs. constant methods

The detection/discrimination distinction is based on the reference stimulus. If it has a natural zero point, such as the intensity of a signal then the investigator strives to record an empirical detection threshold. On the other hand, if the reference pattern is bipolar and has high contrast, such as the largest negative and positive difference to a standard stimulus, then the task is discrimination. The choice of procedure is mainly influenced by

the investigator's experience and preference. Normally, forced-choice tasks are preferred because they promote unbiased decisions and F_{corr} can be taken as performance measure. However, bias can also be eliminated in Yes/No-tasks by mathematically applying a *correction for guessing*. Furthermore, they are said to be more easily to understand for the subject. Finally, the distinction between constant vs. adaptive methods is based on the presentation of stimuli and influences the outcome. While constant methods have preassigned, i.e. constant, test levels provided randomly, adaptive methods provide customized test levels according to the subject's last response. The output of a constant method is a full PF. Adaptive methods provide only its parameters (DT, PSE, JND).
Estimation of the PF is a complex endeavor. Empirical threshold and performance variability are sensitive to different values for γ and λ (lapses). An widely used estimation method that takes the difficulties into account is the maximum likelihood estimation combined with Monte-Carlo simulations to cover a huge range of possible parameters. The method is implemented in the freely available *psignifit toolbox* for Matlab, which was used to calculate the PFs in this thesis.

Method of Magnitude Estimation

The method of *magnitude estimation* proposed by Stevens (1956) is a very simple method to quantify an observer's percept. The experimenter tells a subject that a standard stimulus is s units and asks him to to quantify a comparison stimulus in the same scale, however the scale is up to the subject. By suitable averaging and normalizing these estimates of magnitudes a relation between the physical and the perceptual scale can be obtained easily over a wide variety of stimuli. Subjects concentrate on ordering the comparison stimuli according to their magnitude. Accurate judgments aiming at unveiling the perceptual distance (d') of different stimuli and order them in a cardinal scale are not possible by this method.

Method of Adjustment

The *method of adjustment* is a comparison task proposed by Fechner (1860). A subject has to continuously adjust a comparison stimulus s_c that matches a standard stimulus s_{ref}. if the matching is within a certain modality, i.e. on the same perceptual dimension, the stimuli are presented sequentially, i.e. the standard stimulus is a reminder. If the matching is crossmodal the stimuli are presented concurrently. Matching is more accurate than fixed pairwise comparisons. Within-modal matching is more accurate than between-modal (crossmodal) since additional variance arising from the translation between the perceptual dimensions disturbs the decision process. The PSE is calculated as the mean of the n matched comparisons s_c to a certain standard stimulus s_{ref}

$$\text{PSE} := \frac{1}{n}\sum_{1}^{n} s_{c,n}. \tag{2.37}$$

The JND is computed as the 'between' variable, i.e. it is calculated for each subject independently as the standard deviation of his i matched comparison stimuli s_c to a certain

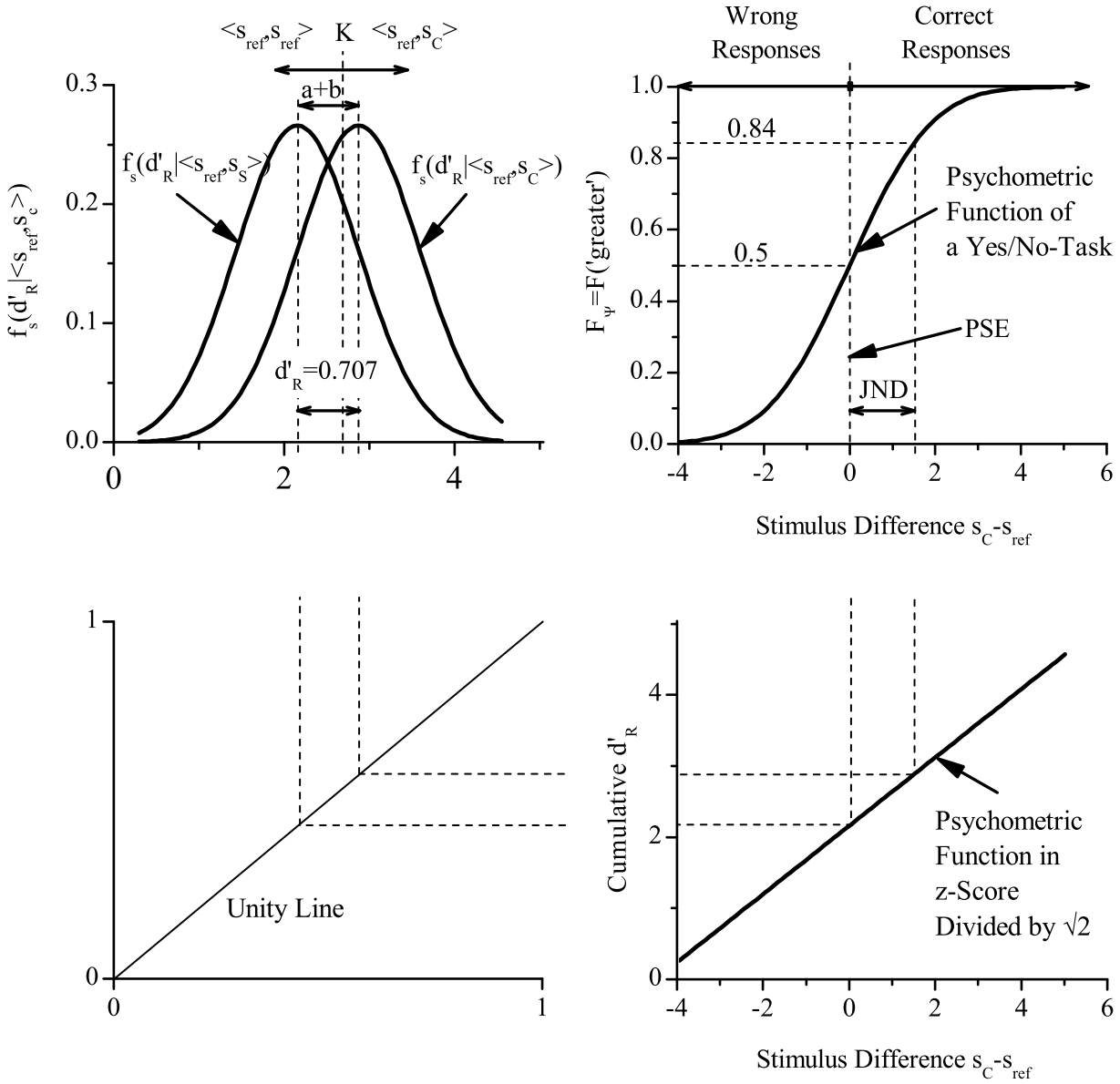


Figure 2.8: Empirical threshold detection by the PF in a Yes/No-reminder task: The PF is traced out from 0 to 100% if no lapses occur. The PSE is defined as the stimulus value that can be correctly identified by 50% of the subjects. The JND is defined to be the difference between the PSE and the 84%-level yielding a performance of $d'_R = 0.707$. The transformation is conducted by calculating the PF in z-score (d') and dividing the result by $\sqrt{2}$ to account for the two statistically independent stimuli the decision is based on.

standard stimulus s_{ref}

$$\mathrm{JND}_j := \sqrt{\frac{1}{i}\sum_{1}^{i}(s_{c,i} - \mathrm{PSE}_j)^2}, \tag{2.38}$$

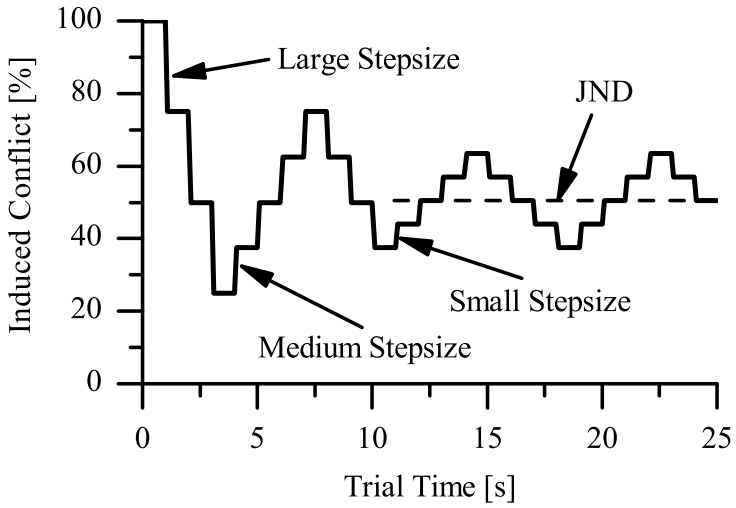


Figure 2.9: Adaptive staircase method for difference thresholds: The JND is the mean of the limit cycle of a subject's answers swaying between 'detected' and 'not detected'. The stimulus is adjusted upwards when the stimulus difference (conflict) is not detected and is adjusted downwards if detected. To save trial time adjustment steps are large at the beginning and are decreased after each lower turning point.

where j is the number subjects and the PSE_j is the point of subjective equality for the subject under consideration. The overall JND is given by the mean value of all individual JNDs

$$\mathrm{JND} := \frac{1}{j} \sum_{1}^{j} \mathrm{JND}_j. \tag{2.39}$$

The Method is also termed *method of average error*.

Adaptive Staircase Method

The *adaptive staircase method* is a comparison task proposed by Cornsweet (1962). For detection thresholds, a variable stimulus is presented repeatedly and is adjusted upwards whenever it is not perceived and downwards whenever it is perceived. The DT is the mean value of the limit cycle the subject settles for.
For difference thresholds a variable stimulus is adjusted to increase its absolute difference from a standard stimulus whenever the difference is not discriminated or is adjusted to decrease its absolute difference from the standard stimulus whenever the difference is discriminated. The PSE is calculated as the DT in the detection task. The JND is calculated according to the equations (2.38) and (2.39). An illustration is given in Figure 2.9.

Method of Constant Stimuli

The *method of constant stimuli* is a comparison task and more accurately a Yes/No-reminder task proposed by Fechner (1860). A standard stimulus (the reminder) precedes

a comparison stimulus $< s_{ref}, s_c >$ and afterward the subject has to indicate if the comparison stimulus had a higher or a lower magnitude. For detection the comparison stimuli cover a range from 0 to a stimulus that can be detected nearly 100% of the subjects. For discimination comparisons are chosen to cover a range that the subject can correctly identify the smallest comparison and the largest comparison stimulus from the standard being in the middle of this range. The DT or the PSE is calculated as

$$\text{DT} = \text{PSE} := F_\Psi(0.5). \tag{2.40}$$

The CE is given as

$$\text{CE} = |\text{PSE} - s_{ref}|\,. \tag{2.41}$$

The detection variability is given by

$$\text{VARAB} := \sigma_\Psi. \tag{2.42}$$

The discrimination variability or JND is given as

$$\text{JND} := \frac{\sqrt{2}\sigma_\Psi}{s_{ref}}, \tag{2.43}$$

which leads a perceptual performance of $d' = 1$ and is normally denoted in [%].

Two-Alternative Forced-Choice Task

An experimental procedure designed to eliminate bias resulting from the observer's criterion is the *two-alternative forced-choice task* (2AFC-task). The task is similar to the method of constant stimuli, but with the position of the standard stimulus s_{ref} randomly differing. The participant has to estimate whether the stimulus in one particular interval was larger or louder or higher or more compliant, etc. The theory assumes that the same level of sensory activation occurs in each observation interval. This is in contrast to the method of constant stimuli in which the standard stimulus might be memorized because of its fixed position. Discrimination performance is therefore higher as with the preceding method. The DT or the PSE is calculated as

$$\text{DT} = \text{PSE} := F_\Psi(0.75). \tag{2.44}$$

The CE is given as

$$\text{CE} = |\text{PSE} - s_{ref}|\,. \tag{2.45}$$

The detection variability is given by

$$\text{VARAB} := \sigma_\Psi. \tag{2.46}$$

The discrimination variability or JND is given as

$$\text{JND} := \frac{\sigma_\Psi}{\sqrt{2}s_{ref}}, \tag{2.47}$$

which leads a perceptual performance of $d' = 1$ and is normally denoted in [%]. The 2AFC-task is easier than the reminder task (method of constant stimuli) and, therefore, permits measurement of sensitivity to smaller stimulus differences. However, it is more difficult to understand and it is more difficult to implement, too.

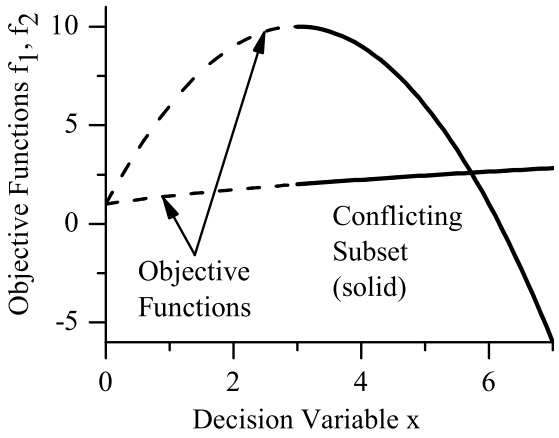
2.3 Multiobjective Optimization

Multiobjective optimization (MO) is a quantitative approach to evaluate decision problems that involve multiple variables (objectives). The first goal is to identify the optimal alternatives. The second goal is to choose a *good* or *best* alternative among the optimal alternatives.
The mathematical foundations of considering multiple objectives in an optimization problem have their origin in economics. Hermann H. Gossen and Vilfredo Pareto accepted the effect of personal utility as a common experience and started to mathematically model decision making processes of individuals. They assumed a number of individuals trading with each other, each with the objective of maximizing its personal ophelimity (economic satisfaction). What distribution of goods would they settle for? Pareto assumes that each individual strives to "*climb the hill of pleasure, to increase his ophelimity*" and Gossen added the concept of diminishing marginal ophelimity (*Gossen's First Law*). Furthermore, Pareto assumed that maximum ophelimity must be reached in a certain position when it is impossible to find a way of moving from that position very slightly in such a manner that the ophelimity enjoyed by each of the individuals of that collectivity increases or decreases." (cited according to [30]). This is the main idea of what is called Pareto-optimium or an efficient solution. Gossen solved the problem to find a final solution amongst the technical efficient alternatives by the help of a utility function. This solution is commonly referred to as *Gossen's Second Law*. By now MO is a matured mathematical method which still has its main application in microeconomic theory [31; 32] but has also moved to operations research and technology-dominated fields.
Subsection 2.3.1 deals with *multiobjective analysis*. It is described how to identify a set of alternatives not dominated by other alternatives, the Pareto frontier. In Subsection 2.3.2 the focus is on *multiobjective decision aid*[4] to make a final choice between the Pareto-optimal alternatives.

2.3.1 Multiobjective Analysis

Multiobjective Analysis (MOA) structures optimization problems with multiple objectives. These problems can be treated in the *decision space* explicitly, i.e. the objectives are functions depending on the alternatives called the *feasible set*. A more illustrative treatment, however, is possible when transforming the problem to the *objective space*, where each dimension represents one objective and the feasible set is given implicitly. Different notions of optimality can be defined and illustrated in objective space such as *Pareto optimality* as the set of alternatives in which no objective can be improved without deteriorate the remaining objective.
In the following decision and objective space are introduced. Pareto optimality is explained in detail. Further readings should contain [33–35].

[4] Since terminology is not unique in literature, the reader might encounter these two theories under deviating titles.



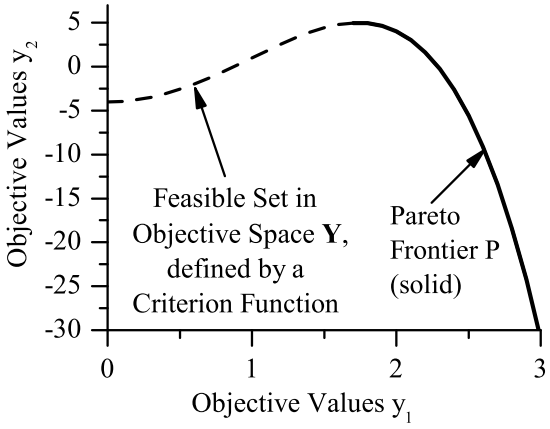


Figure 2.10: (Left) Multiobjective optimization problem in decision space: Objective functions $f_1(x), f_2(x)$ should be minimized over the feasible set of alternatives x. They are at least partly conflicting in the non-trivial case.

Figure 2.11: (Right) Multiobjective optimization problem in objective space: The feasible set $\mathbb{Y}$ is represented by the criterion function. Pareto frontier $\mathbb{P}$ and lexicographic optimum are two possible notions of optimality for decision problems with conflicting objectives.

Decision Space and Objective Space

Consider a mathematical problem with two criteria and a continuous set of alternatives. The two-dimensional[5] *multiobjective optimization problem* (MOP) is defined by

$$\begin{aligned} &\max\ (f_1(x), f_2(x)), \\ &\text{s.t. } x \in \mathbb{X}, \end{aligned} \tag{2.48}$$

where $f_i : \mathbb{X} \to \mathbb{Y}$ are called criteria or *objective functions*. The *decision variable* x belongs to the non-empty *feasible set* $\mathbb{X}$, which contains all possible alternatives and which is a subset of the *decision space* $\mathbb{R}$. The goal is to maximize the objective functions $f_1(x), f_2(x)$ which we want to maximize simultaneously over the feasible set. An example ($f_1 = \sqrt{x+1}$, $f_1 = -x^2 + 4x + 1$) is depicted in Figure 2.10 in decision space. It shows the objective functions f_1, f_2 depending on a non-negative decision variable x. The conflicting subset starts where the slope of the functions becomes opposite.
The image of $\mathbb{X}$ under $f = (f_1, f_2)$ is denoted by $\mathbb{Y} = f(\mathbb{X}) := \{y \in \mathbb{R}^2 : y = f(x)\}$ for $x \in \mathbb{X}$ and called *feasible (objective) set*. The non-negative elements y_1, y_2 of the *alternative* $\mathbf{y}$ are called *objective (function) values*. The set from which the objective values are taken is called *objective space* $\mathbb{R}^2$. The optimization problem in objective space is depicted in Figure 2.11 for the example introduced above.
To obtain the image of the feasible set in criterion space, $\mathbb{Y}$, we substitute y_1 for f_1 and y_2 for f_2 and eliminate the decision variable x. Hence, we obtain the a function $y_1 = g(y_2)$ called *criterion function*. The graph of the criterion function is shown in Figure 2.11. For the example given the feasible objective set is $y_2(y_1) = -y_1^4 + y_1^2 - 4$.

[5]This introduction is restricted to two objectives only.

The feasible objective set $\mathbb{Y}$ can also be described by its constraints if a functional representation is not possible:

$$\mathbb{Y} = \{\mathbf{y} | -\mathbf{h}(\mathbf{y}) \geq 0\}, \tag{2.49}$$

where the vector $\mathbf{h}$ contains is the functions that mark the limits of the feasible set. It can be of arbitrary order. In the two-objectives case the feasible set has three constraints. Since the lowest possible objective function values are zero, the feasible set is constraint by $y_1 \geq 0$ and by $y_2 \geq 0$. The third constraint is the set of the non-dominated alternatives, which is considered in Subsection 2.3.1.
An important characteristic of $\mathbb{Y}$ is its topological structure. For optimization upon this set it is desirable that it is convex:

Definition 2.3.1 *(**Convexity of a set**) A set $\mathbb{Y}$ is convex if whenever alternatives $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ are in $\mathbb{Y}$ so too is the combination $t\mathbf{y}^{(1)} + (1-t)\mathbf{y}^{(2)} \in \mathbb{Y}$ with $0 < t < 1$.*

This means that any straight line between $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(2)}$ is also in $\mathbb{Y}$.

Special attention has to be given to the optimization concept of a MOP. The word 'maximize' generally means that we want to maximize all objective functions simultaneously. If there is no conflict between the objectives than a solution can be found where all objective functions attain their minimum. This is the trivial case. However, we assume that there is no single solution that is optimal with respect to every objective function. That means the objective functions are at least partly conflicting. They may also be incommensurable (i.e. in different units). Furthermore, the word 'maximize' defines how the different objective functions $f_1(x), f_2(x)$ have to be compared for different alternatives $x \in \mathbb{X}$. Different optimization concepts exist such as Pareto optimality, which is introduced in the next subsection.
If an objective function is not to be maximized but to be minimized, it is equivalent to minimize the functions $-f_i$.

Pareto Optimum and Efficiency

In single objective optimization the focus is on the decision space. In MO the criterion space is often more illustrative. Because of possible incommensurability and contradiction of the objective functions, it is not possible to find a single solution that is optimal for all objectives simultaneously. Furthermore, there is no natural ordering in the objective space because it is only partially ordered (i.e. (3,3) is less than (9,9) but the order of (3,9) and (9,3) is not defined). However, there are certain objective vectors that show a distinctive characteristic: None of their objective values can be increased without decreasing the remaining objective value, which means without leaving the feasible objective set $\mathbb{Y}$. This feature is called *Pareto optimality* [36].
To define Pareto optimality, the notion of dominance is introduced:

Definition 2.3.2 *(**Dominance**) An alternative $\mathbf{y}^* \in \mathbb{Y}$ dominates another alternative $\mathbf{y} \in \mathbb{Y}$ if $y_i^* \geq y_i \ \ \forall \ \ i$ and $y_i^* > y_i$ for some i.*

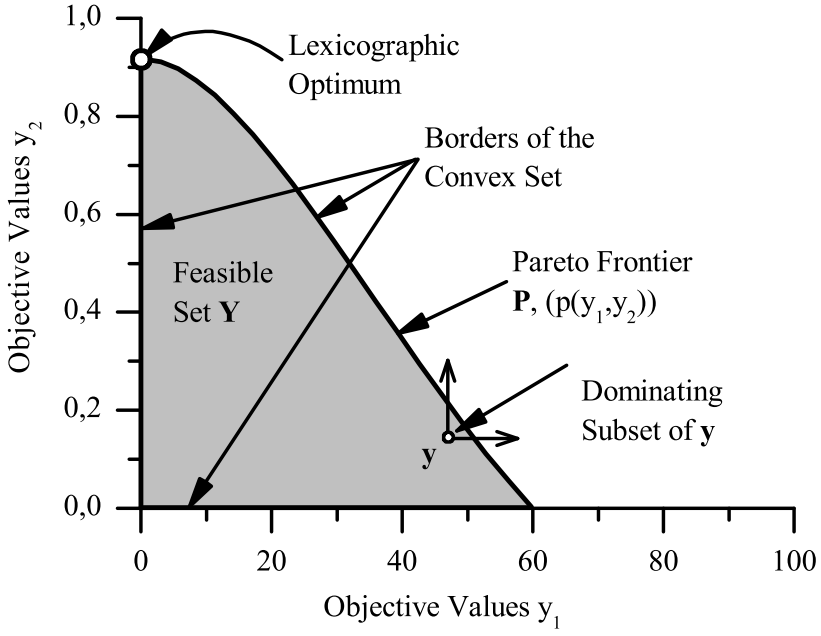


Figure 2.12: Multiobjective optimization problem in objective space: The feasible set $\mathbb{Y}$ is defined by inequalities that mark the borders. In this case the Pareto frontier $\mathbb{P}$ defines the right border of the feasible set. The rectangular cone helps finding dominant alternatives.

Dominance can be illustrated an rectangular triangle as shown in Figure 2.12: If no other alternative lies within the rectangular cone, the alternative in the origin of the cone is non-dominated and therefore Pareto-optimal.

Definition 2.3.3 *(**Pareto Optimality**) An alternative $y^* \in \mathbb{Y}$ is called Pareto optimal, if it is not dominated by another alternative.*

The set of non-dominated alternatives is called Pareto optimal set or *Pareto frontier* and denoted by $\mathbb{P}$. It is a subset of the feasible objective set: $\mathbb{P} \in \mathbb{Y}$. The Pareto frontier of the example problem is depicted in Figure 2.11. It can be non-convex and non-connected. Pareto-optimal alternatives are also called *efficient*[6].
The general analytical formulation of the two-criteria Pareto frontier is given by

$$\mathbb{P} = \{\mathbf{y} | p(\mathbf{y}) = 0; \mathbf{y} \geq \mathbf{0}\}, \tag{2.50}$$

where $p(\cdot)$ is a function $p : \mathbb{R} \rightarrow \mathbb{R}$.

[6]Especially in economics a deviating definition of Pareto-optimality exists. It assumes that an individual's utility of an alternative has to be non-dominated. Non-dominated alternatives are not efficient but only *technical efficient*. However, the general case is used throughout this text since the utility function could also be an identity mapping. Hence, technical efficiency is the same as efficiency throughout this text.

2.3.2 Multiobjective Decision Making

Multiobjective Decision Aid (MCDA) systematically supports a decision maker in finding the optimal solution amongst the 'equally optimal' alternatives on the Pareto frontier. Therefore, the decision maker has to define his *preferences* in a mathematical form called *utility function*. The utility function is then maximized with respect to the Parteo frontier, which is a single objective optimization problem (*utility maximization problem*). Two characteristics of the utility function, the *marginal rate of substitution* and the *elasticity of substitution* describe the decision maker's weighting of the multiple objectives.
In the following, the concept of preferences is explained and the derivation of a utility function. The utility maximization problem is introduced an its solution is described. Further readings should include [34; 37; 38]

Preference

The concept of *preferences* assumes a choice (real or imagined) between alternatives and the possibility of ranking these alternatives, based on happiness, satisfaction, gratification, enjoyment, ophelimity. More generally, it can be seen as a source of motivation or utility. The concept of preferences builds the formal basis for a mathematical modeling of utility. If a decision maker has a preference ranking of alternatives, and if his choices are made in accordance with these preferences, than the preferences can be said to rationalize his choices. The choices are rational with respect to their utility (happiness, satisfaction, etc.) he is driven by. Hence, the notion of rationality is the underlying concept of a formalization of preferences.
In modern theory of consumer choice it is assumed that a rational decision maker performs choices amongst the alternatives in the feasible objective set $\mathbb{Y}$ and that he chooses the alternatives that are optimum with respect to his preference. Preferences are binary relations on the feasible objective set $\mathbb{Y}$ in a way that the decision maker can compare an alternative $\mathbf{y}^{(1)}$ to an alternative $\mathbf{y}^{(2)}$ so that one and only one of the following holds:

$$\begin{array}{lcll} \mathbf{y}^{(1)} & \text{is indifferent to} & \mathbf{y}^{(2)} & (\text{written } \mathbf{y}^{(1)} \sim \mathbf{y}^{(2)}), \\ \mathbf{y}^{(1)} & \text{is preferred to} & \mathbf{y}^{(2)} & (\text{written } \mathbf{y}^{(1)} \succ \mathbf{y}^{(2)}), \\ \mathbf{y}^{(1)} & \text{is less preferred to} & \mathbf{y}^{(2)} & (\text{written } \mathbf{y}^{(1)} \prec \mathbf{y}^{(2)}). \\ \mathbf{y}^{(1)} & \text{is at least as good as} & \mathbf{y}^{(2)} & (\text{written } \mathbf{y}^{(1)} \succeq \mathbf{y}^{(2)}). \end{array}$$

A preference relation must fulfill the following properties:

P1. *Completeness*
This means, the decision maker is able to compare any two alternatives $\mathbf{y}^{(1)}$, $\mathbf{y}^{(2)}$ within the feasible set.

P2. *Transitivity*
A preference relation $\succeq$ is *transitive* if $\mathbf{y}^{(1)} \succeq \mathbf{y}^{(2)}$ and $\mathbf{y}^{(2)} \succeq \mathbf{y}^{(3)}$, then $\mathbf{y}^{(1)} \succeq \mathbf{y}^{(3)}$.
This means, if alternative $\mathbf{y}^{(1)}$ is ranked higher than alternative $\mathbf{y}^{(2)}$ and this is ranked higher than alternative $\mathbf{y}^{(3)}$ then alternative $\mathbf{y}^{(1)}$ must also be ranked higher than alternative $\mathbf{y}^{(3)}$. Hence, the order of alternatives must be consistent.

A preference relation with the properties P1-P2 is called weak order or *total preorder* and also called *rational*. It is denoted by $\succeq$.
Further properties could be:

P3. *Reflexivity*
A preference relation $\succeq$ is *reflexive* if $\mathbf{y}^{(1)} = \mathbf{y}^{(1)}$ then $\mathbf{y}^{(1)} \sim \mathbf{y}^{(2)}$.
This means, the decision maker must be indifferent between two identical alternatives.

P4. *Monoticity*
A preference relation $\succeq$ is *monotone* if $\mathbf{y}^{(1)}, \mathbf{y}^{(2)} \in \mathbb{Y}$ and $\mathbf{y}^{(1)} >> \mathbf{y}^{(2)}$ implies $\mathbf{y}^{(1)} \succ \mathbf{y}^{(2)}$. It is *strictly monotone* if $\mathbf{y}^{(1)} \geq \mathbf{y}^{(2)}$ and $\mathbf{y}^{(1)} \neq \mathbf{y}^{(2)}$implies $\mathbf{y}^{(1)} \succ \mathbf{y}^{(2)}$.

P5. *Convexity*
A preference relation $\succeq$ is *convex* if for any alternatives $\mathbf{y}^{(1)} \succeq \mathbf{y}$ and $\mathbf{y}^{(2)} \succeq \mathbf{y}$ (with $\mathbf{y}^{(1)} \neq \mathbf{y}^{(2)}$) it is $t\mathbf{y}^{(1)} + (1-t)\mathbf{y}^{(2)} \succeq \mathbf{y}$ with $0 < t < 1$.
It is *strongly convex* if for any alternatives $\mathbf{y}^{(1)} \succeq \mathbf{y}$ and $\mathbf{y}^{(2)} \succeq \mathbf{y}$ (with $\mathbf{y}^{(1)} \neq \mathbf{y}^{(2)}$) it is $t\mathbf{y}^{(1)} + (1-t)\mathbf{y}^{(2)} \succ \mathbf{y}$ with $0 < t < 1$.
This means, the decision maker prefers alternatives that are a combination of two indifferent alternatives, or, in other words the decision maker prefers averages over extremes. In two-objective optimization with strict monotonic preferences strict convexity means that any alternative on a straight line between two indifferent alternatives $\mathbf{y}^{(1)} \neq \mathbf{y}^{(2)}$ has a higher preference ranking.

P6. *Continuity*
A preference relation $\succeq$ is *continuous* if, given $\mathbf{y}^{(1)} \succ \mathbf{y}^{(2)}$ and $\mathbf{y}^{(2)} \succ \mathbf{y}^{(3)}$, there is an alternative $\mathbf{y}^{(4)} \sim \mathbf{y}^{(2)}$ on the connection line between $\mathbf{y}^{(1)}$ and $\mathbf{y}^{(3)}$.
This means, there are no discontinuous but only continuous changes in preferences from one alternative to another.

P7. *Non-Saturation*
This means, the decision maker prefers an alternative $\mathbf{y}^{(1)}$ in relation to another alternative $\mathbf{y}^{(2)}$ if $\mathbf{y}^{(1)}$ at least one objective value is greater but no objective value is smaller than in alternative $\mathbf{y}^{(2)}$ (More is better.).

Utility Function

To solve the decision maker's problem his preferences can be mathematically described by means of a utility function. A utility function associates a real number $u(\mathbf{y})$ representing the decision maker's preferences structure to each alternative $\mathbf{y}$.

Definition 2.3.4 *(**Utility function**[7]) A function $u : \mathbb{Y} \to \mathbb{R}$ representing the decision makers's preferences among the objectives is called utility function if, for all $\mathbf{y} \in \mathbb{Y}$,*

$$\mathbf{y}^{(1)} \succeq \mathbf{y}^{(2)} \Leftrightarrow U(\mathbf{y}^{(1)}) \geq U(\mathbf{y}^{(2)}).$$

[7]Sometimes the term *value function* is used and the term utility function is reserved for stochastic problems. However, in this text the term utility function is used for deterministic problems also.

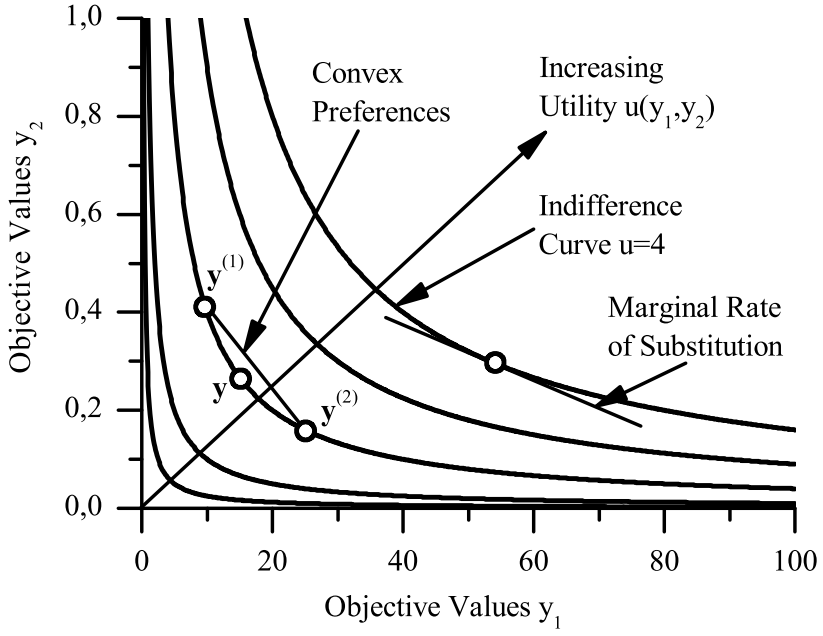


Figure 2.13: Indifference curves of a utility function projected in objective space: The curves show places of constant utility. If they are bent to the origin the utility function is quasiconcave and the represented preference relation $\succeq$ is convex. Generally, utility increases strict monotonically. The marginal rate of substitution of a certain alternative is the slope of the indifference curve at this alternative.

A utility function that represents a preference relation $\succeq$ is not unique. For any strictly increasing function $f : \mathbb{R} \to \mathbb{R}, v(x) = f(u(x))$ is a new utility function representing the same preferences as $u(\cdot)$. It is only the ranking of the alternatives that matters. Properties of utility functions that are invariant for any strictly increasing transformation are called *ordinal*. Those properties that are not invariant to a strictly increasing transformation are called *cardinal*. Consequently, the preference relation associated with an utility function is an ordinal property.
A utility function is totally a decisions maker's concept. Different decision makers might have different utility functions. For a decision maker it might be difficult to mathematically define his utility function. Generally, a utility function is assumed to be strictly monotonically increasing. This means, that the preference of the decision maker increases if the value of an objective function increases while all the other objective values remain unchanged. Many properties of a preference structure and a representing utility function can be illustrated by *indifference curves*. Indifference curves are the projection of the utility function into the objective space representing a constant utility each. An example, $u = \sqrt{y_1 y_2}$, showing the indifference curves of $u = (0.5, 1, 2, 3, 4)$ is depicted in Figure 2.13.
A preference relation can be represented by a utility function only if it is rational, that means the preference relation $\succeq$ must be complete and transitive. Not any rational preference relation $\succeq$ can be represented by a utility function. A preference relation $\succeq$ that should be represented by a utility function must be *continuous*. Continuity is a sufficient condition for the existence of a utility function. In fact, it guarantees the existence of a continuous utility function. Furthermore, characteristics on preferences translate into

characteristics of the representing utility function. A strictly monotone increasing preference relation $\succeq$ implies that the utility function is strictly monotonically increasing too. In Figure 2.13 this can be seen at the fact that an indifference set does never cover an area, i.e. for a given $\mathbf{y}^{(1)}$ there would be more than one $\mathbf{y}^{(2)}$ to combine at the same utility and vice versa. A non-saturated preference relation $\succeq$ is represented by a non-saturated utility function. This means all indifference curves to the right always represent a greater utility than the indifference curves to the left. Convex preferences imply that $u(\cdot)$ is *quasiconcave*. Strict convexity implies *strict quasiconcavity*, respectively. Increasingness, saturation and quasiconcavity are ordinal properties of a utility function. The indifference curves of a convex preference relation $\succeq$ are always bent to the origin. For analytical purposes it is also convenient if $u(\cdot)$ can be assumed to be twice continuously differentiable. It is possible, however, for continuous preferences that $u(\cdot)$ is not differentiable.
Different forms and classes of utility functions exist. An extensive overview is given in [39]. A main property of utility function is the *marginally diminishing utility*. That means, the gain of utility which is created by an additional value of an objective is the smaller the higher the objective value already is. This property is also named "Gossen's first law" according to its inventor. It is denoted by:

$$\frac{\partial^2 u}{\partial y_1^2} < 0 \quad \text{and} \quad \frac{\partial^2 u}{\partial y_2^2} < 0. \tag{2.51}$$

Another main property of multiobjective utility functions is their ability to weight the different objectives. If the value of one objective y_1 is increased by a certain amount, by which amount the value of the second objective y_2 has to be changed in order for the decision maker to remain indifferent? This property of every multiobjective utility function is called *marginal rate of substitution*. Assume the utility function has a smooth second derivative, then the marginal rate of substitution is defined as:

Definition 2.3.5 *(**Marginal Rate of Substitution**) The marginal rate of substitution (MRS) is the slope of the indifference curve or the negative, reciprocal relation of the marginal utilities:*

$$\xi = \frac{dy_2}{dy_1} = -\frac{u_{y_1}}{u_{y_2}}, \tag{2.52}$$

where U_y denotes the partial derivative (marginal utility).

Two objectives are *complements* when they satisfy the same utility, and the marginal utility of one of the will increase when the objective value of the other increases. Complements complement utility such as letter and stamp. Two objectives are *substitutes* when the marginal utility of one of them decreases as the objective value of the other, which satisfies the same utility, increases. Substitutes have the same utility such as butter and magarine. The MRS denotes at which rate objective y_1 can substitute for objective y_2 without affecting utility. The MRS is constant for perfect substitutes, since their indifference curves are linear. The MRS is infinite or zero for perfect complements since their indifference curves are rectangular. In the case of a strict convex preference relation $\succeq$ the MRS is negative $\text{MRS} < 0$.

Definition 2.3.6 ***(Elasticity of Substitution)*** *The elasticity of substitution (EoS) is defined to be the quotient of percentaged difference of y_2/y_1 and dy_2/dy_1:*

$$\epsilon(y_2, y_1) = \frac{d(y_2/y_1) : (y_2/y_1)}{d(dy_2/dy_1) : (dy_2/dy_1)} \tag{2.53}$$

The elasticity of substitution shows to what degree two objectives can be substitutes for one another. In the case of perfect objective mobility the elasticity of substitution is $\epsilon(y_2, y_1) = \infty$ since $d(dy_2, dy_1) = 0$. According to the statements above, the more L-shaped the indifference curves are, the less complementary the objectives and the higher the elasticity of substitution.

Utility Maximization

The decision maker's choice problem can now be determined as the maximization of his utility under consideration of all feasible alternatives. Therefore, it is assumed that the decision maker's rational, continuous, strictly monotonically increasing, and strictly convex preference is represented by a twice differentiable utility function. The decision maker's problem of choosing the most preferred alternative can then be stated as the following *utility maximization problem* (UMP).:

$$\max u(\mathbf{y}), \tag{2.54}$$
$$\text{s.t. } p(\mathbf{y}) \leq 0, \tag{2.55}$$
$$-\mathbf{y} \leq 0, \tag{2.56}$$

where $\mathbf{y}$ denotes the alternatives and the *inequality constraints* define the feasible set in objective space $\mathbb{Y}$. This is a single-objective optimization problem with inequality constraints. A general solution is obtained using the *Karush-Kuhn-Tucker Necessary Conditions*. See [40; 41] for a comprehensive treatment of optimization problems with inequality constraints. The solution is sketched in Figure 2.14
A more straightforward, however, not general approach that relies on the theory of equality constraints is introduced here. It applies for problems where the optimal alternative $\mathbf{y}^*$ is on the boundary of the feasible set. The inequality which defines this boundary is called the *active* inequality constraint. Assumed that at any feasible alternative $\mathbf{y}$ the active inequality is $p(\mathbf{y})$ and the inequalities $-\mathbf{y} \leq 0$ are inactive. If $\mathbf{y}^*$ is a local maximum of the UMP then it is also a local maximum of the UMP without the inactive inequalities. Hence, the inactive inequalities can be ignored in the statement of the optimality conditions. Furthermore, at a local maximum the active inequality constraints can be treated to a large extent as equality constraints [41]. Hence, if $\mathbf{y}^*$ is a local maximum of the UMP then $\mathbf{y}^*$ is also a local maximum for the *reduced utility maximization problem* (rUMP):

$$\max u(\mathbf{y}), \tag{2.57}$$
$$\text{s.t. } p(\mathbf{y}) = 0, \tag{2.58}$$

The necessary condition for an optimum of the rUMP can be derived by the Lagrangian equation

$$L(\mathbf{y}, \mu) = u(\mathbf{y}) + \mu p(\mathbf{y}). \tag{2.59}$$

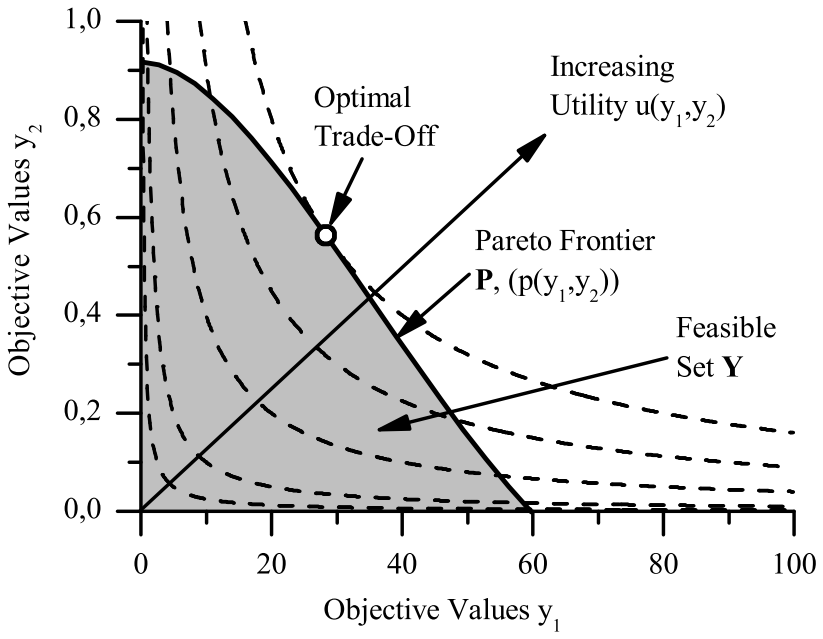


Figure 2.14: Solution of the UMP in objective space: The grey area denotes the feasible set $\mathbb{Y}$. The Pareto frontier $\mathbb{P}$ is defined by the right edge of the feasible set. The optimal alternative $\mathbf{y}^*$ is defined by the intersection point of the Pareto frontier and the indifference curve with the highest utility level.

Then, if $\mathbf{y}^*$ is a local maximum the necessary conditions of first order hold:

$$L_{y_1}(\mathbf{y}^*, \mu^*) = 0, \tag{2.60}$$
$$L_{y_2}(\mathbf{y}^*, \mu^*) = 0, \tag{2.61}$$
$$L_{\mu}(\mathbf{y}^*, \mu^*) = 0, \tag{2.62}$$

where L_{y_1} is the partial derivation of the Lagrangian equation with respect to y_1, L_{y_2} the partial derivative with respect to y_1, and L_{y_μ} the partial derivative with respect to μ. The solution is only valid if the elements represent a feasible alternative, hence, are non-negative.
Since it is obvious by the problem statement that the optimum is a maximum and that this maximum is constraint, sufficient conditions are not necessary. See [40; 41] for a comprehensive treatment of optimization problems with equality constraints.
The existence of a global solution of the rUMP depends on the topological characteristics of the feasible set $\mathbb{Y}$ and the utility function. Especially the form of the indifference curves is decisive for the existence of a global maximum. If the indifference curves are bent to the origin the curves are strictly convex. If, additionally, $\mathbb{Y}$ is a convex set, then a maximum will be unique and global as illustrated in Figure 2.14. If the indifference curves are only convex than a maximum will be global and the set of maxima will be convex, i.e. on a straight line. Utility functions with (strictly) convex indifference curves are (strictly) quasiconcave (Not only all concave functions but also some non-concave functions have convex indifference curves, e.g. $u = y_1 y_2$. The class of functions comprising both groups

is called quasiconcave) [42]. Hence, quasiconcavity of the utility function u and convexity of the feasible set $\mathbb{Y}$ are important features of the rUMP.

2.4 Control Methods

Control methods are an essential part of the subsystems of a presence system. They assure stable realization of the commanded and reflected information at high accuracy. To render virtual, haptic environments, impedance and admittance control schemes are mostly applied. To prevent that COM-delays lead to instability of the overall presence system the passivity paradigm was used. The Scattering transformation transforms a COM afflicted by constant delays into a passive transmission line.
In Subsection 2.4.1 impedance and admittance control are introduced. In Subsection 2.4.2 passivity and the Scattering transformation are explained.

2.4.1 Impedance/Admittance Control

In electronics the relation between voltage and current is termed *immittance*. The quotient between voltage and current, in frequency domain, is called impedance and the inverse is called admittance. Analog to electronics the terms can also be deployed to mechanics where force and velocity take the same place as voltage and current, force being the effort variable (as voltage) and velocity being the flow variable (as current). Hence, in frequency domain, the quotient between force and velocity,

$$Z = \frac{f}{v}. \tag{2.63}$$

is called *impedance* and denoted by Z. The inverse, Z^{-1}, is called *admittance*. In time domain immittances are expressed by linear differential equations. Based on these relations different control strategies have been developed. These strategies are especially suited to render virtual environments to be displayed to an operator by a haptic HSI.
In the following the two basic concepts *impedance control* and *admittance control* (used in the following chapters) are introduced. Further readings should contain [43–45]

Impedance Control

In an impedance-controlled VR the operator commands position based information (velocity) and the device reacts with a force according to impedance Z that has to be rendered. The force is calculated as result of the movement within the desired impedance. This structure is commonly called *open-loop force control*. Additionally, an explicit force control can be implemented by measuring the exerted force on the HSI. Assuming a one dimensional, linear system, the displayed impedance becomes

$$Z_d = \frac{f_o^e}{v_o} = Z + \frac{Z_h}{1 + C_f}, \tag{2.64}$$

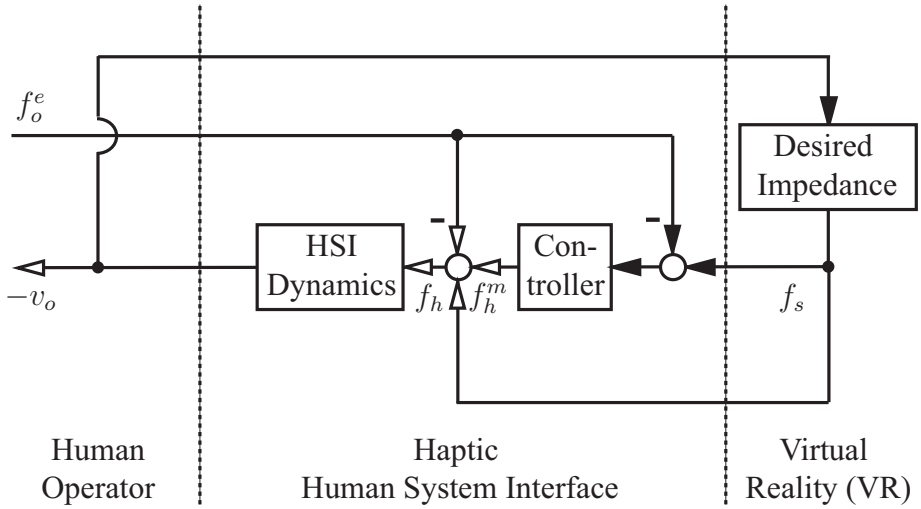


Figure 2.15: Impedance control: Velocity applied to the HSI by the human operator results in a force reaction of the interface. Force feedback can be used to suppress the disturbing influence of the HSI dynamics (Hollow arrows represent physical interaction, filled arrows represent signal processing.)

where f_o^e denotes the exerted force of the operator, v_o denotes the velocity of the operator that is imprinted by the HSI, Z denotes the desired impedance of the VR, Z_h denotes the impedance of the HSI, and C_f denotes the force controller. Without force feedback ($C_f = 0$) the closed loop impedance is the desired impedance Z disturbed by the addition of the HSI-impedance Z_h. Hence, open loop force control should only be used if the device has negligible momentum. Deploying force feedback reduces the influence of the device dynamics proportional to the factor $(1 - C_f)$. Consequently, the gain of the controller C_f should be increased as much as possible but is bounded by stability. Although the interaction with the human operator creates another feedback loop this loop will not affect the stability since the dynamics of the operator is assumed to be adaptive never causing instable behavior.
The advantages of impedance control are a large bandwidth and, in case of open-loop force control, no force measurement. The disadvantages are the constraint performance especially in case of non-compensated, large non-linearities of the HSI. Furthermore, the performance of an impedance controlled robot is heavily depending on the kinematical design of the device. The impedance control architecture is depicted in Figure 2.15.

Admittance Control

In admittance-controlled VRs the operator commands a force and the device reacts with a position change. The position change is calculated as result of the force exerted to the desired admittance Z^{-1} and is imprinted by position control on the HSI. In case of low admittances that only take low forces, the natural force feed-through can be compensated. The displayed impedance becomes

$$Z_d = \frac{f_o^e}{v_o} = \frac{C_x}{Z_h + C_x} Z, \tag{2.65}$$

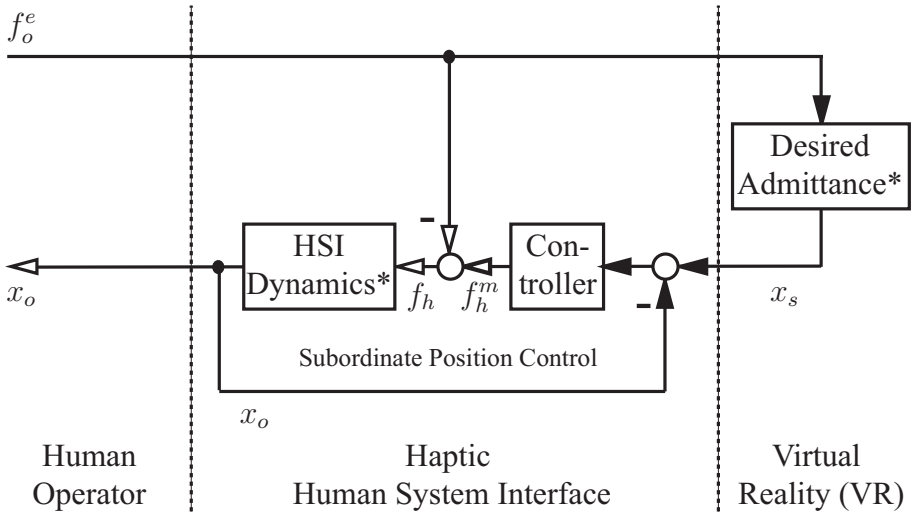


Figure 2.16: Admittance control: Force exerted by the human operator and measured by the HSI results in a position change according to the admittance desired. For low forces the natural force feed-through can be compensated leading to a highly accurate result. (Hollow arrows represent physical interaction, filled arrows represent signal processing. *Desired admittance** and *HSI-dynamics** represent general admittances as the integral operation $1/s$ is attached providing position instead of velocity output.)

where C_x denoted the position controller, Z_h the linear dynamics of the HSI, and Z the desired impedance. High gain control ($x_s \approx x_h$) will not only suppress the linear dynamics (impedance) of the HSI, but also non-linear dynamics, like friction or dynamical properties due to a rotational-kinematic design. However, the position control loop constraints the bandwidth of the VR, rigid environments will appear softer. Force and position measurement induce additional dynamics that can lead to instabilities. Although the interaction with the human operator creates another feedback loop this loop will not affect the stability since the dynamics of the operator is assumed to be adaptive and to never generate instable behavior.
The advantages of admittance control are the high accuracy in case of HSI non-linearities and large HSI dynamics compared to the emulated VR. Disadvantages are the constraint bandwidth due to the additional poles added by the position control loop and the investment for force measurement. The admittance control architecture is depicted in Figure 2.16.

2.4.2 Passivity

The passivity formalism is an energy-based approach to synthesize high performance, large scale dynamical systems. A passive system dissipates energy and is therefore incapable of power gain at any moment. Passive systems are BIBO-stable and Lyapunov-stable. The method is applicable to complex systems since internal models can be avoided and an analysis can be reduced to input and output signals. Constant delay of energy-exchanging signals, as occurring in haptic telepresence systems, can be passivated using the Scattering transformation.

In the following the passivity formalism is introduced, its application to haptic presence systems is described, and the Scattering transformation is explained. Further readings should contain [18; 46–49].

Passivity and Stability

The power of a mechanical system is defined by the scalar product of force and velocity

$$P_{in}(t) = \dot{\mathbf{x}}^T(t)\mathbf{f}(t), \tag{2.66}$$

where $\dot{\mathbf{x}}$ denotes the velocity vector and $\mathbf{f}$ denotes the force vector. Velocity and force are called *power variables*. The time integral of power variables yields the energy consumption of a system

$$E_{in} = \int_0^t P_{in}(\tau)\, d\tau. \tag{2.67}$$

Definition 2.4.1 *(**Passivity**) A system with as many inputs as outputs is passive, if a positive definite storage function $V(\mathbf{x})$ exists such that the 'integral passivity inequality'*

$$\int_0^t P_{in}\, d\tau + V(\mathbf{x}(0)) \geq V(\mathbf{x}(t)), \qquad \forall t \geq 0 \tag{2.68}$$

holds or such that the 'differential passivity inequality'

$$P_{in} \geq \dot{V}(\mathbf{x}(t)), \qquad \forall t \geq 0 \tag{2.69}$$

holds for all initial values $\mathbf{x}(0)$ and all input values[8]*.*

For a passive system there is always a non-negative P_{diss} that denotes the power dissipated into states not considered in the passive system, e.g. friction losses or gear backdrivability, in mechanical problems. A system is *strictly passive*, if the dissipated power is larger than zero, i.e. the equal sign in the equations (2.68), (2.69) does not hold. It is termed *lossless* if the dissipated power is exactly zero, i.e. exclusively the equal sign holds.
Two passive systems retain passive if they are connected in parallel or feedback connection. Hence, complex passive systems must be reducible to a network of passive subsystems in parallel and feedback connections.
Passive systems are *BIBIO-stable*. A serial connection of passive systems is also BIBO-stable since the poles of the comprised system remain in the open right-half plane. According to the inequality (2.69) the time derivative of the storage function is negative semidefinite $\dot{V}(\mathbf{x}) \leq 0$ if the input is zero or constant, hence, passive systems are *Lyapunov-stable*. Passivity of a system can be hard to proof since the characteristics of the storage function $V(\mathbf{x})$ has to be analyzed to evaluate the passivity criteria (2.68), (2.69). A sufficient condition for a system to be passive is *reachability* (*controllability* in linear case, i.e. it is possible to steer the states from any initial value to any final value within certain time) together with *positive realness*. Since positive realness is an input/output property a common way to proof passivity is to verify positive realness for a reachable system.

[8] The input value can either be velocity (in case the system is an impedance) or force (in case the system is an admittance).

Definition 2.4.2 ***(Positive Realness)*** *A system with as many inputs as outputs is positive real if*

$$\int_0^t P_{in}\, d\tau \geq 0, \qquad \forall t \geq 0 \tag{2.70}$$

holds for all initial states $\boldsymbol{x}(0) = 0$.

Positive realness is a necessary condition for passivity. In contrast to passivity, applicable rules to identify positive real systems exist. For example, a linear time-invariant system $G(s)$ is positive real if

- $G(s)$ has no poles in the open complex right-half plane.
- the phase shift is not larger than 90°: $\arg[G(s)] \leq 90°$.
- $G(s)$ has a relative degree of 0 oder 1.

Besides from the initially stored energy, positive realness is only defined by the input power, i.e. exclusively by input/output properties of the system.

Passive Haptic Telepresence

The passivity concept has been widely used to stabilize haptic presence systems especially presence systems in telepresence architecture. Especially the destabilizing effect of communication delay can effectively be processed by passivity measures. To apply the passivity criterion the presence system is divided into different subsystems. As illustrated in Figure 2.17 a presence system can be interpreted as a connection of two-ports (HSI, COM, TO) terminated by one-ports (operator, target environment) at each end. For a passive presence system the subsystems have to be passive and interconnected one-ports in feedback and parallel connections. If serial connections of one ports occur the system is no longer passive but remains stable. Normally, the COM is not connected by parallel or feedback connection to the HSI and the TO. A common way is therefore, to serially connect the COM to HSI and TO. If the operator passivates the HSI and the subsystem comprised of TO and environment is passive then the system shows no instable behavior (although a stability proof is complex).
For the sequel of this thesis it is assumed that the subsystem of human operator handling the HSI is passivated by the operator. The subsystem consisting of the TO working in the remote environment is controlled to be passive as well. This can be done by appropriate control measures. The overall system is connected for stable and high performance.

Passivation of Constant Delay by the Scattering Transformation

The COM of a bilateral haptic presence system (velocity-force architecture) can become instable if afflicted with communication delay. Delay has an active dynamics shifting the phase more than 90°. A passivation of the COM afflicted with *constant* delay T is conducted using the *Scattering transformation*. It maps power variables like velocity and force $(\dot{x}, f)$ to *wave variables* (g, h) that can be transmitted in presence of communication

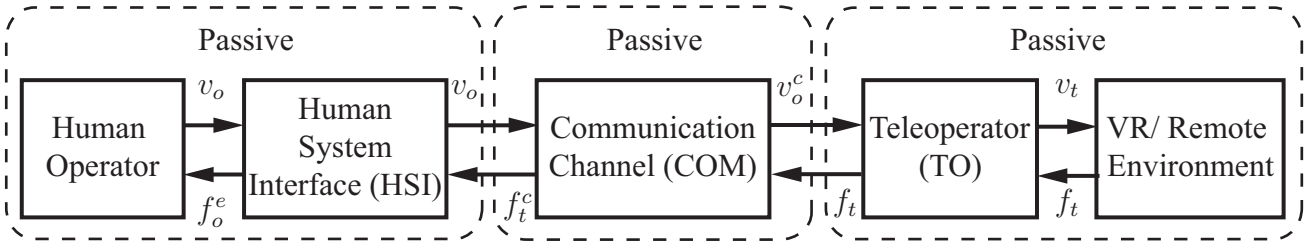


Figure 2.17: Presence system in two-port architecture: If the subsystems (operator-HSI), COM, and (TO- environment) are passive and connected in parallel or feedback connection, the overall system is passive, i.e. stable, as well. If serial connections occur the system is not passive but remains stable. Sensor dynamics in HSI and TO are omited.

delay without affecting stability.
Generally, the Scattering transformation is defined by

$$g = \frac{b\dot{x} + f}{\sqrt{2b}}, \qquad h = \frac{b\dot{x} - f}{\sqrt{2b}}, \tag{2.71}$$

where g is the incident wave and h is the reflected wave (also called *wave reflections*). The parameter b is a positive constant that can be chosen arbitrarily. The transformation is *bijective*, i.e. unique and invertible. Hence, no information is lost or gained by encoding power variables into wave variables or wave variables into power variables. The back-transformation to power variables is given by

$$b\dot{x} = \sqrt{\frac{b}{2}}(g + h), \qquad f = \sqrt{\frac{b}{2}}(g - h). \tag{2.72}$$

The COM of a bilateral haptic presence system is depicted in Figure 2.18[9] The power input (2.66) is defined as

$$P_{in} = \dot{x}_o f_t^c - \dot{x}_o^c f_t, \tag{2.73}$$

where $\dot{x}_o$ is the velocity commanded by the operator and $\dot{x}_o^c$ is the commanded velocity on teleoperator side and f_t is the force reflected by the teleoperator and f_t^c is the reflected force on operator side. The power entering the system is counted positive and the power leaving the system is counted negative. Applying the Scattering transformation (2.71) to equation (2.73) yields

$$P_{in} = \frac{1}{2}(g_l^2 - h_l^2 - g_r^2 + h_r^2), \tag{2.74}$$

where the index indicates the wave variables on the right and on the left hand side of the delay element.
For the passivity proof of the delayed COM in Scattering domain consider the equations governing the delayed transmission

$$g_r(t) = g_l(t - T) \ , \tag{2.75}$$
$$h_l(t) = h_r(t - T) \ . \tag{2.76}$$

[9] v denoting the velocity $\dot{x}$ in frequency domain to prevent for mixed time-frequency notation, all other variables in the diagram are also in frequency domain but remained unchanged.

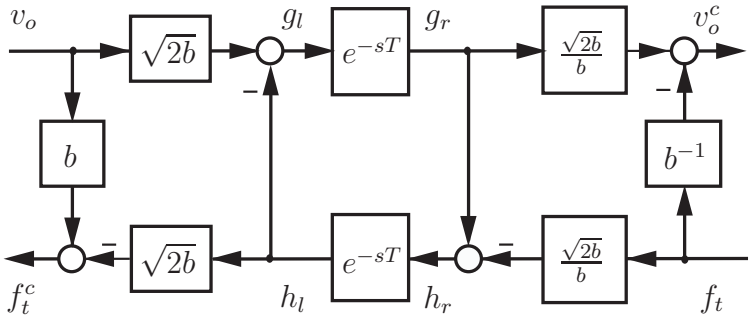


Figure 2.18: Passivated COM by Scattering transformation: Velocity information by the operator v_o and force information by the teleoperator f_t is transformed to wave variables, transmitted over the delayed transmission line, and eventually retransformed to velocity information v_o^c and force information f_t^c.

Substitution of these equations into equation (2.74) yields the input power of the delayed COM in Scattering domain

$$P_{in} = \frac{1}{2}(g_l^2(t) - h_r^2(t-T) - g_l^2(t-T) + h_r^2(t)). \tag{2.77}$$

Integrating latter equation yields

$$\begin{aligned} \int_0^t P_{in} &= \frac{1}{2}\int_0^t (g_l^2(\tau) + h_r^2(\tau))d\tau - \frac{1}{2}\int_0^{t-T} (g_l^2(\tau) + h_r^2(\tau))d\tau & (2.78)\\ &= \frac{1}{2}\int_{t-T}^t (g_l^2(\tau) + h_r^2(\tau))d\tau \geq 0. & (2.79) \end{aligned}$$

The last inequality always holds, for a square operation on real numbers always yields positive results. Since P_{in} is the only source of the stored energy, it is

$$P_{in} = \dot{V}(\mathbf{x}). \tag{2.80}$$

Hence, the delayed COM in Scattering domain with zero initially stored energy is lossless and therefore passive. The Scattering transformation transforms the COM to a lossless transmission line. It originally stems from the solution of the *wave equation*, which is a partial differential equation used e.g. in electromagnetics for modeling the propagation of waves. The parameter b is also called *wave impedance*, respectively. The wave reflections h contain the information about the remote environment but additionally disturbances. The the signal-to-noise-ratio increases with the length of the communication delay T and depends of the dynamics of the target environment. If the remote environment e.g. represents free space the presence system renders a mass proportional to the delay T and if the remote environment e.g. represents a rigid wall the presence system renders a compliance (inverse stiffness) proportional to the delay T. Therefore, in presence systems with HSI and TO having a small momentum communication delay of $T = 1$s can already deteriorate transparency in full.
According to equation (2.79) the energy consumption can be independently considered for the forward and the backward path. In the forward path the energy consumption is defined by the incident wave

$$P_{in,g} = \frac{1}{2}\int_0^t g_l^2(\tau)d\tau \geq 0. \tag{2.81}$$

And in the backward path the energy consumption is defined by the reflected wave

$$P_{in,h} = \frac{1}{2} \int_0^t h_r^2(\tau) d\tau \geq 0. \tag{2.82}$$

This important property illustrates that wave variables carry their own energy also indicated by the unit of measurement $[\sqrt{W}]$. It allows for independent passivity measures, when additionally processing the wave signal.

3 Perception-Based Utility Optimization

"There must be some way out of here," said the joker to the thief,
"There's too much confusion, I can't get no relief.
Businessmen, they drink my wine, plowmen dig my earth,
None of them along the line know what any of it is worth."

Bob Dylan, *All Along the Watchtower* (*John Wesley Harding*, 1967)

Telepresence and *Virtual Reality* (VR) are technology-mediated experiences that enable a human being to be present in a different environment. A *presence system* consists of a *human operator* who commands a multimodal *avatar/teleoperator* (TO) in the *virtual/remote environment*. A multimodal *human system interface* (HSI) is used for the operator to command the TO and, on the same time, to display the target environment. Signals are exchanged over a *communication channel* (COM). See Figure 1.1 for an illustration and Section 2.1 for a precise description. Applications are virtual training, edutainment, aerospace teleoperation, micro assembly, surgery or maintenance in hazardous HSI, COM, and TO enable an operator to immerse into a target environment. However, the same technology spoils the experience adding system-specific disturbances such as kinematical, dynamical and ergonomical restrictions (HSI, TO) as well as communication delay, noise, or bandwidth constraints (COM). The performance can be increased by increasing the quality of these systems, however, efforts would become infinite if the disturbances should be eliminated in full. Due to this restriction, presence systems always represent a trade-off between performance and effort or resource savings. The question must be raised according to the technical efficient trade-offs and according to the most useful trade-offs and how they can be calculated for various applications.

Minsky is usually acknowledged as the beginner of conceptual research on technology-mediated presence stating that the operator must be able "to perform normal human functions" and on the same time "receives sufficient quantity and quality of sensory feedback to provide a feeling of actual presence at the work side" (cited in [50]). Subsequently, different concepts of presence based on different philosophies were elaborated. However, no unified theory of presence could be established by now. In the following, main concepts, determinants of presence, evaluation methods, and effort/resource parameters are introduced and state-of-the-art references are given.

First refinements of Minsky's concept were conducted by dividing presence into two different forms: Subjective presence can be mentally experienced by the human operator individually and objective presence states the physical effectiveness of an operator in a target environment [15; 20; 51; 52] (See Subsection 2.1.1 *The Rationalistic View* for background information.). Criticizing this dualistic view several authors developed concepts of presence based on the epistemology of Heidegger and the perceptual theory of Gibson [2; 8–10; 12; 53]. Both linked perception closely to everyday interaction within an environment. Zahoric and Jenison defined presence as "tantamount to successfully supported action in the environment" [8] (See Subsection 2.1.1 *The Ecologic View* for background

information.). Mantovani and Riva extended this ecological definition by a sociocultural dimension emphasizing that "presence is always mediated by both physical and conceptual tools that belong to a given culture" [8]. Sheridan tried a unification of the opposing philosophical stances using dynamical systems theory and observer theory [2]. Also this view was criticized to actually be a rationalistic view it remains the only concept to effectively guide engineering research (See Subsection 2.1.1 *The Unifying Approach* for more information.). Recent conceptual publications focus on the functional task of presence within the cognitive system. Slater conceived presence as "the selection mechanism that organizes the stream of sensory data into an environmental gestalt or perceptual hypothesis about the current environment" [54]. Additionally, Lee remarked that humans are willingly to accept incoming stimuli in order to react effectively rather than accurately [55]. Floridi calls this concept "Successful Observation" and developed an abstract concept of remote presence to tackle the problems of VR and telepresence to the point avoiding confusion with presence emerging from watching television or reading books [17].
Based upon the different views a general model for presence systems can be based upon Sheridan's unifying approach by inserting the technological means (HSI, COM, TO) between observer and environment (See Subsection 2.1.2 *A general model* for background information.). This model fosters two supplemental principles of technology-mediated presence. The first principle, *distal attribution*, describes the extension of the human body schema to a target environment. It was introduced by [14] and adopted by Biocca who identified presence as a subset of the mind-body problem [12]. To solve the questions 'What is body?" and "What is environment?" matches a key problem of presence research, especially since it is known that the borders between body and technology can vary in human's consciousness [16]. According to this principle presence can be evaluated by the degree to which the presence system becomes part of the human body (See Subsection 2.1.2 *Distal Attribution* for background information.) The second principle, *proximal stimulation* comprises the mostly engineering-driven efforts to reproduce the target environment at the operator side. Seminal contributions that follow this principle were made by Lawrence in [18] and Yokokohji in [19] focussing on haptic presence systems. They state that a presence system is successful if the dynamical properties of the remote environment can be displayed to the human operator by the HSI. According to this principle presence can be evaluated by the degree to which the HSI can reproduce the target environment as encountered by the TO (See Subsection 2.1.2 *Proximal Stimulation* for background information.).
Presence or the performance of presence systems can be observed through different factors ranging from physiological (e.g. heart rate), performance (e.g. task completion time), ratings (e.g. presence questionnaires) to psychophysical measures and measures based on physical quantities: Lawrence proposed to compare the dynamical characteristics of the displayed environment to the target environment to analyse the quality of presence systems for mechanical environments [18]. Similarly, Yokokohji defined the equality of position and force on both sides as 'ideal response' to quantify haptic telepresence [19]. Sheridan proposed three different measures for multimodal presence. Reflexive responses to external stimuli, rating experiments dealing with several factors, and discrimination between real vs. artificial should quantify the realism of the environmental representation [15]. Presence questionnaires to assess the degree of immersion and involvement in the target environment are proposed by different authors [21; 52]. Measures based on task performance are proposed by Schloerb in [20]. Mantovani and Riva refused to measure presence by comparison between real and artificial environment but proposed to measure the realism of the user

interaction with the human system interface instead [10]. IJsselsteijn et al. introduced different objective measures: measures based on postural and physiological responses of the human operator and dual task measures based on a distractive cognitive load [22]. Physiological measures are also proposed in [56]. Experimental studies that quantify the perceived presence predominantly use presence questionnaires and are constraint to one modality. Only few studies analyzed the perception of multimodal feedback with respect to the presence generated. Visual presence is analyzed e.g. in [23; 52; 56–58]. Visual-haptic presence is analyzed e.g. in [12].

Efforts to improve a presence system can be classified according to its subsystems HSI, COM, and TO. A further classification can be performed according to the modalities the presence system provides. Efforts are mostly abstract parameters like development time, performance expenditures or the adaptation of the system to certain individualities of the human operator. *Resources* are more tangible such as resolution of a display in vision, acoustic or haptic parameters, energy consumption, stability margin or bandwidth. The haptic (kinesthetic) modality is by now the most sensitive modality since forces and velocities exchanged between operator and target environment close a global feedback loop. The controller has to stabilize the overall system and on the same time provide the envisaged performance. Different control architectures are proposed in [18]. Stability under communication delay was also analyzed. In [59] the presented architecture was extended by local feedback loops (actually introduced in [60]) to compensate for the dynamics of the HSI and the TO. Furthermore, a force control at operator side relates the reflected force to the force felt by the human operator. An analysis of how the dynamics induced by the passivated COM influence the dynamics of a telepresence system was conducted in [49; 61]. A more detailed analysis was presented in [62; 63]. Haptic data communication was introduced in [64], further elaborated in [65] and moreover in Chapter 6. A disadvantage of most of these engineering-driven studies is that they do not provide an experimental evaluation with accepted psychological methods, hence their impact on the performance of a presence system is still not clear.

In this chapter a new method is proposed to identify the optimal trade-off between performance and efforts/resource savings of a presence system. Utility optimization, based on the paradigm of multi objective decision aid, is used to quantify the operator's preferences (See Subsection 2.3.2 for background information.). The utility function is maximized with respect to an equality constraint that represents the Pareto-optimal combinations of performance and efforts/resource savings (See Subsection 2.3.1 for background information.). The equality constraint is derived experimentally by psychophysical methods mapping efforts/resources to a perceptual conflict. Afterward, the performance depending on the chosen conflict is measured. This is done using the psychophysical methods of post-test rating and magnitude estimation (See Subsection 2.2.2 for background information.). The contributions had been partly elaborated in [66; 67] and published to a large extent in [63; 68]. The method is generic to arbitrary presence measures and to arbitrary performance parameters.

The remainder is organized as follows: In Section 3.1 the derivation of the Pareto frontier is described and an example is presented. In Section 3.2 the utilty function is described and their parameters are explained. Section 3.3 introduces the utility optimization problem and discusses its solution. Finally, a conclusion is presented in Section 3.4.

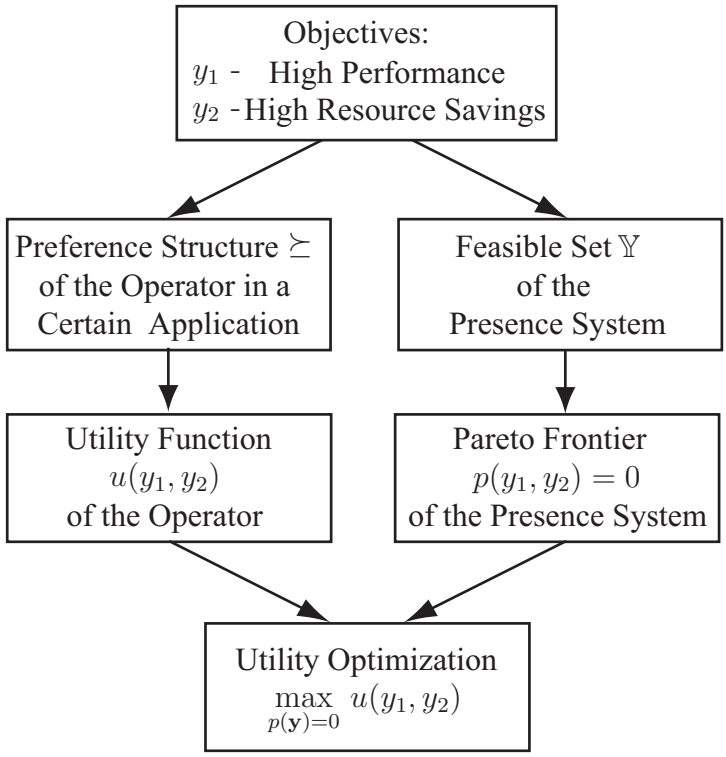


Figure 3.1: Outline of the utility optimization method: A utility function is defined to evaluate the objectives 'high presence' and 'high resource savings' represented by the values y_1 and y_2. The optimization is performed amongst the alternatives on the Pareto frontier, which is a subset of the feasible set.

3.1 Derivation of the Pareto-Frontier

A presence system can be differently tuned by the parameters in HSI, COM, and TO. Each variation claims more or less resources in e.g. energy, bandwidth, or stability margin affecting the performance of the presence system. The operator's naive goal is to have highest possible effort or resource savings y_1 and highest possible performance y_2 at the same time. However, both *objectives* are conflicting: Savings in efforts/resources will normally decrease performance. The main question is which combinations of both objectives, i.e. which *alternatives* $\mathbf{y} = (y_1, y_2)$, are optimal, i.e. not dominated by other feasible combinations. A solution of the two-objective optimization problem can be found using experimental presence evaluation methods. If performance can be modulated by the resource parameter then the impact of a certain value of this parameter on the performance of the presence system can be measured experimentally. Arbitrary presence measures can be used. In this elaboration a new measure based on bimodal conflicts is introduced. An experimental derivation of the Pareto frontier $\mathbb{P}$ is conducted using the new measure and an existing measure, namely a presence questionnaire.
In Subsection 3.1.1 deeper insights are provided about multiobjective optimization and the new measure is formulated. In Subsection 3.1.2 an experiment is described to determine the Pareto-frontier. The following elaboration was published to a large extent in [68].

3.1.1 Theory

Optimization problems with two-objectives can be treated by multiobjective optimization theory. The goal is to identify the Pareto-frontier $\mathbb{P}$ that contains all non-dominated alternatives $\mathbf{y}$. In the case of a presence system the solution can be achieved experimentally if both objectives are dependent and either one objective can be observed by a presence measure. The outcome represents the Pareto-frontier.
In the following multiobjective analysis is explained, the new presence measure is defined, and implications for telepresence and VR systems are pointed out.

Multiobjective Analysis

Optimization problems with multiple objectives are commonly formulated in decision space according to equations (2.48). The goal is to choose 'good' or 'best' alternatives among a feasible set of alternatives. See Subsection 2.3.1 for a detailed introduction. In the following the optimization problem to find the non-dominated alternatives is denoted in the objective space $\mathbb{R}^2$ (See Figures 2.10 and 2.11 for detailed information about decision and objective space). The subspace $\mathbb{Y} \in \mathbb{R}^2$, which contains the possible alternatives is called *feasible set*. The elements y_1, y_2 of the *alternative* vector $\mathbf{y}$ are called *objective values*. Non-dominated alternatives are introduced in Definition (2.3.2). Their elements cannot be increased without decreasing the remaining element to stay in the feasible set. According to Definition (2.3.3) these alternatives are also called *efficient* or *Pareto-optimal*. The set of these alternatives is called *Pareto-frontier* $\mathbb{P}$. The Pareto-frontier is a subset of the feasible set $\mathbb{P} \in \mathbb{Y}$. An illustration of a two-dimensional, feasible set in objective space is given in Figure 2.12. The depicted set is *convex* according to Definition 2.3.1. In this Section the Pareto-frontier of a presence system is identified using psychological and psychophysical methods.

Perceptual Relation of the Two Objectives

The two objectives 'high effort/resource savings' y_1 and 'high performance' y_2 (further referred to as *resource savings* and *performance*) built the basis of the optimization problem to identify the Pareto-frontier. The technology of a presence system (HSI, COM, TO) can be assumed to be tuned to the best possible performance for all parameters except for the parameter under consideration y_1. This parameter represents a system resource such as e.g. stability margin, data reduction rate, energy savings, or workspace savings. General experience unveils that an increase of this resource parameter has a negative effect on performance, i.e. on the capability of the presence system to generate a realistic experience. Hence the relation $y_2(y_1)$ is a decreasing function. This relation can be exploited to solve the optimization problem by presence evaluation methods. Assumed the performance of the presence system is measured depending on the value of the resource parameter ranging from 0 to its highest possible value, while all remaining parameter are tuned for highest possible performance. Then the experimentally derived alternatives $\mathbf{y} = (y_1, y_2)$, related by $y_2(y_2)$, represent the efficient alternatives of the presence system, i.e. the Pareto-frontier.
The presence definition used throughout this article is taken from [12]. It defines pres-

ence in a generally accepted way and can be interpreted according to distal attribution and proximal stimulation. Individual presence means "the phenomenal state by which an individual feels located and active in an environment, and, especially in the case of telepresence, the class of experience where the environment is mediated by a technology". An arbitrary presence measure can be used, if they result in a relation $y_2(y_1)$. See 2.1.3 for a description of the different measurement methods for presence.
In this elaboration two measures are used to assess the experienced presence. Firstly, a common presence questionnaire, which is the dominating method to evaluate presence systems. Secondly, a new psychophysical method that maps presence against a multimodal conflict. A multimodal conflict is a well defined parameter in psychophysics and signal detection theory. Several resource parameters, e.g. haptic data reduction rate, cause multimodal conflicts on the HSI. Hence, a multimodal conflict is a generic parameter to represent changes in resource savings. In the following the psychophysical measure is defined.

Definition of the Psychophysical Presence Measure

The introduced presence measure operationalizes the factor 'consistence of multimodal information displayed by the human system interface'. This means, that the presence system is evaluated for its capability to veridically intermediate the target environment. The measure introduced serves as an example how the objectives 'performance' y_2 and 'resource savings' y_1 can be comprised into a mathematical relation. Several authors (e.g. [10; 15; 52]) emphasized the importance of this factor. The measure is based on the assessment of multimodal perceptual conflicts c produced by a multimodal HSI (See Subsection 2.2.1 for more information). A perceptual conflict is a perceivable difference between realistic information about a certain environmental state, which sources from different modalities (redundant information). The operators' perceptual system is only partly able to generate a realistic, coherent estimate about the environmental property by the obtained information. Redundant information about an environmental property is displayed consistently, if the operator can integrate the sensed information (See Subsection 2.2.1 *Perception of Redundant and Complementary Information* for detailed information.). According to signal detection theory the conflict that can be discriminated with a performance $d' = 1$ (corresponding to correct identification in 69% by an unbiased observer) is defined to be the just noticeable difference (JND) in psychophysics (See Subsection 2.2.1 for detailed information). In the current measure the conflict c is caused by a certain amount of resource savings. It is assumed that conflict and resource savings are related by unity, consequently the objective value for resource savings y_1 is defined by the conflict c

$$y_1 := c. \tag{3.1}$$

As stated above it is assumed that perception of presence decreases, with increasing conflict. Hence, the presence measure can be formulated by

$$y_2 = f_p(y_1) = 1 - r(y_1). \tag{3.2}$$

The function $r \in \mathbb{R}$ describes the degradation of presence caused by the conflict. It is related to the type of psychophysical measurement used to measure the extent of the

perceived bimodal conflict. If a rating method is used to assess the conflict, then f should be the mean rating of the extent of the bimodal conflict. It is zero if the conflict cannot be perceived. By definition it cannot be larger that one.

$$0 \leq r(y_1) \leq 1. \tag{3.3}$$

The relative bimodal conflict c induced by the human system interface is described by

$$y_1 = \frac{|\mu_1 - \mu_2|}{\mu_2}, \tag{3.4}$$

where μ denotes the mean value of the normally distributed percepts and the indices denote the different modalities. This equation is also called *Weber Fraction* in psychophysics. Ideal presence is experienced if the objective value is unity $y_2 = 1$. No presence is experienced if the objective value is zero $y_2 = 0$.

Implications for VR and Telepresence

The proposed measure can be applied to assess presence generated by arbitrary presence systems. In both cases the measure quantifies the quality of the presence-mediating technology. According to this measure an ideal presence system is a system that generates no perceptual conflicts between the multimodal information. For a VR-System, that means that the HSI produces no perceptual conflicts. For a telepresence, system that means that HSI, COM, and TO do not induce distortions leading to a perceptual conflict.
These statements raise the question to what extent the measure assures that the generated reality by the HSI equals the target reality (VR or remote environment). Under the assumption that at least one modality involved in the feedback signal carries veridical information about the environmental property of the target environment then, in the ideal case, the measure assures the veridical equality between displayed environment and target environment as well.

3.1.2 Experiment

An experiment is conducted to define the Pareto-frontier $\mathbb{P}$ of a presence system. A new presence measurement method, assessment of bimodal conflicts by magnitude estimation, is used concurrently to an existing questionnaire method to define the optimal alternatives. Experiments are conducted on a visual-haptic HSI connected to a VR that represents a compliant cube. Results show that both measures could successfully identify a Pareto-frontier. Furthermore, both measures provided nearly equal results for the Pareto-frontier $\mathbb{P}$.

Method

Presence system: A human system interface is used that provides visual and haptic feedback at high accuracy. The visual subsystem consists of a TFT screen mounted in the

line of sight of the hand showing the visual virtual reality. The haptic subsystem consists of two SCARA robots providing two degrees of freedom each. The system interacts with index finger and thumb to allow the exploration of the compliant environment by gripping movements. Workspace is about 80 mm and maximal force is about 35 N. Position information is measured by angle encoders and force is measured by strain gauges. Haptically, the compliant environment is generated using an admittance control scheme. Visually, the environment is presented by a compliant cube with yellow spheres representing the finger tips. During the psychophysical experiments participants were able to insert their answers using a joystick that provides them with the different answer possibilities. The system works under realtime conditions and is programmed by Matlab/Simulink. A detailed description of the HSI including hardware, software, and control structure is provided in Appendix A.1.

Participants: Fifteen participants (8 men, 7 women) of the Technische Universität München as well as the Ludwigs-Maximilians-Universität München took part in this study and were paid for participation. Their average age was 24 years. All participants were right-handed and had normal or corrected to normal vision.

Stimuli: Stimuli were cubes of 80 mm edge length having a certain compliance. They were visually and haptically displayed by the HSI. Visual and haptic compliance information could be varied separately. During the testing, the haptic modality was defined to be the (unaltered) reference modality and amounted to the reference compliance of

$$S_{ref} = 0.85\,\mathrm{mm/N}, \tag{3.5}$$

while the visual modality was set to deviate from the target modality. Conflicts were selected according to the results of [69] and additionally, to cover a relatively broad range above as well as below the perception threshold (amounting ca 30% see [70]). The comparison stimuli were chosen to be

$$S = [0.3; 0.6; 0.8; 1.2; 1.6; 2; 2.4; 2.8] \cdot S_{ref}. \tag{3.6}$$

Design: Perception of bimodal conflict as well as presence had to be tested across the set of all nine bimodal stimuli. To make sure that participants only rated the perceived bimodal conflict or their sensation of presence, both measures were assessed in different blocks. Within one block two sets of all nine stimuli were randomly presented. The order of blocks was randomized across participants introducing the new group factor (further referred to as "group"): Group P started with two blocks rating their presence feeling, group D with four blocks rating the extent of the bimodal conflict as well as the measurement of the JND. Additionally, another control variable was introduced that checked whether a congruent stimulus was presented prior to test session or not (further referred to as "experience").

Experimental procedure: Participants were seated in front of the HSI and grasped the device with their dominant hand. They nearly looked perpendicular at the screen while testing the compliant cube. Each stimulus presentation followed the same basic scheme: As indicated by auditory signals the compliant stimulus was presented for 4s. Subsequently, participants had to response according to the assessment block. Participants were instructed and performed a test session before the experiment started.

Presence assessment by conflict measurement: The block for assessing the perceived conflict started with a short baseline response time measurement. After that four blocks had to be conducted. Participants were instructed to enter their answer through the joystick as fast and as accurate as possible by deciding whether they had perceived a conflict between the visual-haptic information (Yes/No) between the visual and the haptic compliance presentation. Afterward, they rated the extent of the bimodal conflict ranging from "0" (no conflict perceived) to "10" (large intermodal conflict perceived).

Presence assessment by questionnaire: After each stimulus presentation participants rated their feeling of presence according to an item of the Witmer & Singer presence questionnaire (see [21]). The item was chosen to be "How natural did your interaction with the environment seem" (translated into german by [71]). The answer had to be given on a seven-point scale with "1" indicating "very naturalistic" and "7" "not very naturalistic". Both questions (extent of the bimodal conflict, presence) as well as their rating scale were fixed above the screen. The non-target question was covered. Between both assessment blocks (extent of the bimodal conflict, presence), participants were asked to fill in a questionnaire consisting of the subscale "immersive tendency" depicted from [21] and translated into german by [71].

Data analysis: Firstly, questionnaire data was descriptively analyzed, and the potential influence on the assessed ratings was determined. Secondly, both ratings of the perceived conflict were transformed to range between 0 and 100% in order to be comparable across participants. Afterwards, ratings (presence, conflicts) of presence were averaged across repetitions.
Prior to testing the hypotheses, the influence of the control variables, namely the factors "group" and "experience" (see Section 4.3), have been tested. Therefore, two separate ANOVAs with the repeated measurement factor visual-haptic conflict and both between-participant factors were computed. When a control variable reached statistical significance, it was considered in further tests. The first hypothesis regarding rating extent of the bimodal conflict was tested with a two-factorial ANOVA with repeated measurements (displayed intermodal conflict above the perception threshold of 85%, see [32]). An ANOVA tested the effect of displayed bimodal conflict (ranging from 0 to 280%) on the presence rating (Hypothesis 2). A significant main effect of bimodal conflict was further tested by contrast tests. All effects were corrected for assumed sphericity by Greenhouse Geisser correction, if necessary. Significance level was set to 5%. The standard deviation of the Pareto-frontier was defined to be the mean value of the standard deviations of its sampling

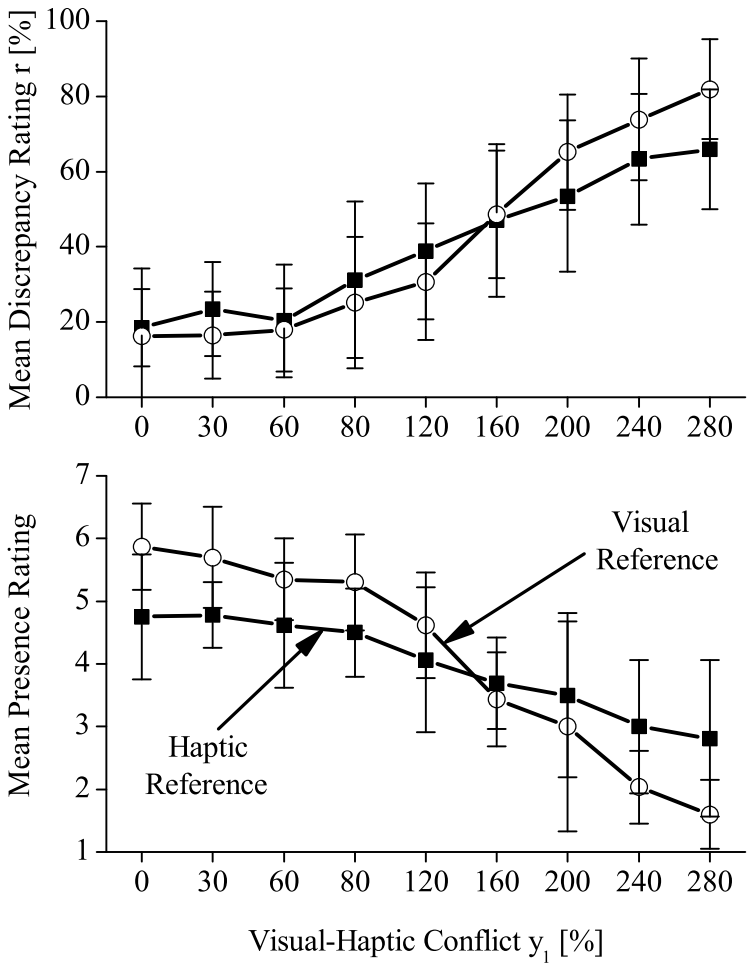


Figure 3.2: Comparision between psychophysical method (upper diagram) and questionnaire method (lower diagram): Mean presence rating decreases and mean rating of extent of the bimodal conflict increases with increasing visual-haptic conflict.

points according to

$$\sigma_{\mathrm{P}} := \frac{1}{n} \sum_{i=1}^{n} \sigma(y_{2_i}), \tag{3.7}$$

Hence, the experimentally derived Pareto-frontier is a statistical quantity characterized by its mean, given by the set $\mathbb{P}$ and its standard deviation σ_{P}.

Results

Questionnaire data: The immersive tendency subscale is comprised of two factors, namely emotional involvement and degree of involvement. Participants' emotional involvement ($\mu = 28.8$, standard deviation $\sigma = 6.1$) as well as degree of involvement ($\mu = 19.3$, $\sigma = 5.6$) did not statistically significantly differ from the German norm sample (see [71]). However, no significant correlation was found between both factors with respect

to immersive tendency and either presence or rating of extent of the bimodal conflict. This indicates that not the personal trait, but the experimental variation accounts for the individual ratings.

Control variables Firstly, influence of control variables, namely influence of order of block (group) and of experiencing of a congruent stimulus on both ratings was tested. The ANOVA revealed no effect of "group" on neither presence rating ($F(1, 11) = 0.19$, $p = 0.675$) nor rating of perceived bimodal conflict ($F(1, 11) = 1.66$, $p = 0.237$). Therefore, order of blocks (factor "group") had not to be considered in further analysis. However, an influence of "experience" with a congruent stimulus prior to the test session could be observed in presence ratings ($F(1, 11) = 11.10$, $p < 0.05$; $\eta^2 = 0.502$) indicating that without presentation presence ratings yields higher scores. Presentation of the congruent stimulus did not generally influence ratings of extent of the bimodal conflict ($F(1, 11) = 2.73$, $p = 0.127$).
Additionally, the test factor showed a significant interaction with displayed visual-haptic conflict ratings (Greenhouse Geisser corrected: $F(2.2, 24.6) = 5.61$, $p < 0.05$; $\eta^2 = 0.338$), but not on presence ratings (Greenhouse Geisser corrected: $F(1.9, 21.5) = 1.99$, $p = 0.162$).

Rating of extent of the bimodal conflict: Displayed perceived visual-haptic conflict above the perception threshold of 85% [32] influenced the ratings (Greenhouse Geisser corrected: $F(2.4, 51.5) = 16.78$, $p < 0.05$; $\eta^2 = 0.563$) and was due to a linear trend ($F(1, 13) = 32.17$, $p < 0.05$; $\eta^2 = 0.712$). As can be seen in Figure 3.2, rating of extent of bimodal conflict, $f(c)$, increased with increasing displayed visual-haptic conflict, c.
Additionally, experiencing a congruent stimulus prior to the test session resulted in an increased mean discrepancy rating ($F(1, 13) = 7.73$, $p < 0.05$, $\eta^2 = 0.373$).

Presence assessment by questionnaire: An ANOVA revealed no overall significant main effect of conflict (Greenhouse Geisser corrected: $F(2.0, 21.5) = 2.35$, $p = 0.120$). Additionally, the a-priori assumption of influence of conflict above or beyond perception threshold was tested. No significant main effect could be found, it visual-haptic conflict is beyond perception threshold (Greenhouse Geisser corrected: $F(1.5, 19.8) = 0.24$, $p = 0.871$). However, a significant main effect of displayed conflict above conflict threshold could be observed ($F(4, 52) = 6.86$, $p < 0.05$; $\eta^2 = 0.345$) which was due to a linear trend ($F(1, 13) = 19.43$, $p < 0.05$; $\eta^2 = 0.599$). As can also be seen in Figure 3.2, presence rating decreased with increasing bimodal conflict.
As has been already reported above, experiencing a congruent stimulus prior to the test session influenced presence ratings across all stimuli ($F(1, 11) = 11.10$, $p < 0.05$; $\eta^2 = 0.502$) revealing lower mean presence ratings.

Presence assessment by conflict measurement: Figure 3.3 shows the result of the introduced presence measure $y_2(y_1)$ as introduced in equation (3.2). Since it is directly related to the perceived conflict, presence is rather high if the conflict is small. Presence is not ideal if the conflict is smaller than the JND ($y_1 < 85\%$). If y_1 is smaller than the JND ($y_1 < 85\%$) the mean value of y_2 is $\mu = 0.75$ at a standard deviation $\sigma = 0.22$. Firstly, this can be explained by the fact that the JND is not an absolute but a statistical

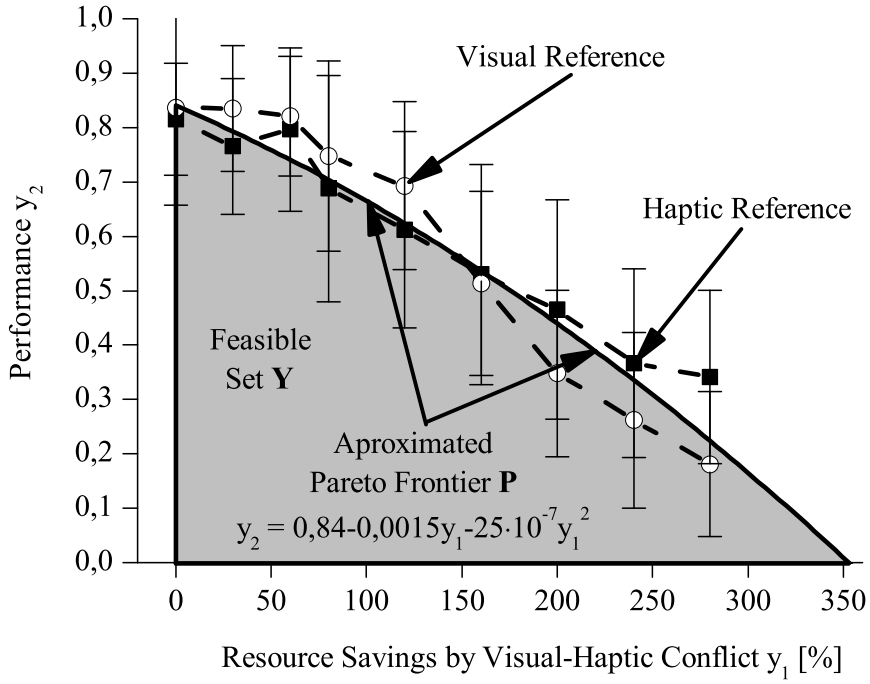


Figure 3.3: Pareto-frontier by psychophysical presence measure: The feasible set $\mathbb{Y}$ is defined by abscissa, ordinate, and Pareto-frontier $\mathbb{P}$. The approximated Pareto-frontier has a convex shape, i.e. the Feasible set $\mathbb{Y}$ is a convex set.

value. Some subjects perceived the conflict also it remained below the JND. Secondly, this can be explained by the general limit of the presence system to veridically display the target environment. For conflicts equal or larger the JND ($y_1 \geq 85\%$) the value of $y_2 = f_p(y_1)$ decreased monotonically. Both measures, the post test rating and the new measure decreased linearly with increasing perceivable conflict. Both trends were statistically significant. The new presence measure showed a higher effect size,η^2. This indicated that operators were more reliable when judging the inconsistent information with the conflict-based presence measure. This might be due to the fact that it was easier for operators to rate a question that pointed directly toward the inconsistent information.
The Pareto-frontier $\mathbb{P}$ can be obtained e.g. by approximating the variant results of the presence ratings. The approximation can easily performed by a polynomial least squares algorithm with $n = 2$. Approximating the experimental outcome yields the Pareto-frontier according to equation (2.50) as

$$\mathbb{P} = \{\mathbf{y} | 0 = -y_2 + 0.84 - 0.0015y_1 - 25 \cdot 10^{-7} y_1 \ ; \quad \mathbf{y} \geq \mathbf{0}\}, \tag{3.8}$$

where the vector y contains the two objective values $\mathbf{y} = (y_1, y_2)$. The Feasible set $\mathbb{Y}$ according to equation (2.49) can be inferred by the Pareto frontier only since the remaining two borders were defined to be abscissa and ordinate.

$$\mathbb{Y} = \{\mathbf{y} | -p(\mathbf{y}) \geq 0 \ , \quad \mathbf{y} \geq \mathbf{0}\}, \tag{3.9}$$

where $p(\mathbf{y}) = -y_2 + 0.84 - 0.0015y_1 - 25 \cdot 10^{-7}y_1$. The approximated Pareto-frontier has a convex shape, i.e. the Feasible set $\mathbb{Y}$ is a convex set according to Definition 2.3.1.
The standard deviation σ_{P} of the Pareto-frontier according to equation (3.7) is

$$\sigma_{\mathrm{P}} = 0.16\,\mathrm{mm/N}. \tag{3.10}$$

Discussion

In this experiment the Pareto-frontier was recorded by two presence measurement methods. The first method was a well-established method based on a questionnaire and the second method was a psychophysically-based magnitude estimation method. The performance decrease was induced artificially.
The outcome shows a monotonically, nearly linearly decreasing Pareto-frontier that could be approximated by a second order polynomial at high precision. However, the assessment shows little reliability ($\sigma_P = 16\%$). This can be explained by the used methods. Questionnaires and magnitude estimation seem not to be able to produce high reliable results, also in presence measurement. A switch to other methods (Method of constant Stimuli, 2AFC-Method) should be considered.

3.2 Derivation of the Utility Function

After the assessment of the Pareto-frontier, the final problem of the decision maker is to choose the best alternative amongst the Pareto-optimal alternatives. This part of multiobjective optimization is called multiobjective decision aid. Based on a decision maker's preferences a utility function is developed that is maximized subject to the Pareto-optimal alternatives. The form of the utility function is strongly influenced by the application scenario. A flexible utility function for the different applications of presence systems is the CES-utility function. Depending on the choice of its parameters it can represent complementary, partly substitutional, or perfect substitutional preference structures arising from three different application scenarios.
In Subsection 3.2.1 the mathematical foundations of a utility functions are laid down. In Subsection 3.2.2 the requirements for utility functions in the context of presence systems are developed. Finally, in Subsection 3.2.3 different utility function based on the CES-utility functions are introduced and linked to the presence system application.

3.2.1 Theory

On the way to a mathematically useful formulation of the decision maker's ophelimity he has to define his *preferences structure* for both objectives. Based upon this preference structure the *utility function* is developed. The relations between preference structure and utility function are important to tailor a function that fits the needs of presence systems applications. Generally, properties of the preference structure translate directly into properties of the utility function.

In the following the concept of preferences is introduced. Thereafter, basic properties of utility functions are explained and methods to set up utility functions are introduced.

Preferences

The concept of *preferences* assumes a choice (real or imagined) between alternatives and the possibility of ranking these alternatives, based on happiness, satisfaction, gratification, enjoyment, ophelimity. More generally, it can be seen as a source of motivation or utility. The concept of preferences builds the formal basis for a mathematical modeling of utility. If a decision maker has a preference ranking of alternatives, and if his choices are made in accordance with these preferences, than the preferences can be said to rationalize his choices. The choices are rational with respect to their utility (happiness, satisfaction, etc.). Hence, the notion of rationality is the underlying concept of a formalization of preferences.
In modern theory of consumer choice it is assumed that a rational decision maker performs choices amongst the alternatives in the feasible objective set $\mathbb{Y}$ and that he chooses the alternatives that are optimum with respect to his preference. Preferences are binary relations on the feasible objective set $\mathbb{Y}$ in a way that the decision maker can compare an alternative $\mathbf{y}^{(1)}$ to an alternative $\mathbf{y}^{(2)}$ so that one and only one of the following holds:

$$\begin{array}{lcll} \mathbf{y}^{(1)} & \text{is indifferent to} & \mathbf{y}^{(2)} & (\text{written } \mathbf{y}^{(1)} \sim \mathbf{y}^{(2)}), \\ \mathbf{y}^{(1)} & \text{is preferred to} & \mathbf{y}^{(2)} & (\text{written } \mathbf{y}^{(1)} \succ \mathbf{y}^{(2)}), \\ \mathbf{y}^{(1)} & \text{is less preferred to} & \mathbf{y}^{(2)} & (\text{written } \mathbf{y}^{(1)} \prec \mathbf{y}^{(2)}). \\ \mathbf{y}^{(1)} & \text{is at least as good as} & \mathbf{y}^{(2)} & (\text{written } \mathbf{y}^{(1)} \succeq \mathbf{y}^{(2)}). \end{array}$$

A preference relation can fulfill properties such as *completeness*, *transitivity*, *reflexivity*, *monoticity*, *convexity*, *continuity*, *non-saturation*, etc. Preference relations with properties transitivity and reflexivity are called weak order or *total preorder* and also called *rational*. It is denoted by $\succeq$. See Subsection 2.3.2 for a more detailed introduction.

Modeling of Utility

A utility function $u : \mathbb{Y} \to \mathbb{R}$ associates a positive, real number $u(\mathbf{y})$ to each alternative $\mathbf{y}$, thereby, representing the decision maker's preferences structure. The form is up to the decision maker. Different decision makers might have different utility functions and it might be difficult to mathematically define a utility function in detail. A utility function can rationalize a decision maker's preferences only if his preferences are complete and transitive. As a basic property most utility functions are *strictly monotonically increasing*. This means, that the preference of the decision maker increases if the value of an objective function increases while all the other objective values remain unchanged. *Continuity* is a sufficient condition for the existence of a utility function. In fact, it guarantees the existence of a continuous utility function. Generally, characteristics on preferences translate into characteristics of the representing utility function. A strictly monotone increasing preference relation $\succeq$ implies that the utility function is strictly monotonically increasing too. A non-saturated preference relation $\succeq$ is represented by a *non-saturated* utility function. This means, all indifference curves to the right always

represent a greater utility than the indifference curves to the left. Convex preferences imply that u is *quasiconcave*. Strict convexity implies *strict quasiconcavity*, respectively. Increasingness, saturation and quasiconcavity are ordinal properties of a utility function. The indifference curves of a convex preference relation $\succeq$ are always bent to the origin. For analytical purposes it is also convenient if u can be assumed to be twice continuously differentiable. See Subsection 2.3.2 for a more detailed introduction.

3.2.2 Utility Requirements for Presence Systems

Pareto-optimal alternatives can be evaluated for a preferred alternative by the help of a utility function (as introduced in Subsection 3.1.1). As applications for presence systems cover a huge scope, from microsurgery to space teleoperation, a single utility function will not match the different preference structures of a human operator generated by different applications. The application scenario has always a major influence on the decision maker's preferences. A class of functions that currently dominates applied research in economical production theory is used as flexible utility function to match the different application-dependent preferences. This class is called constant elasticity of substitution (CES) utility function. It provides freely adjustable features with respect to slope, concavity, and substitution.
In the following, the basic requirements of a utility function in the context of presence systems are described. Thereafter, the CES utility function is explained. In third part, parametrizations of the CES-utility function for different presence systems applications are proposed.

Basic Properties

Basic properties of a utility function for presence systems are *continuity* to evaluate all possible alternatives and *differentiability* to perform mathematical maximization procedures. Since it is obviously better to have more of performance and more resource savings, the function must be *monotonically increasing*, too. Hence, the marginal utility must be positive

$$\frac{\partial u}{\partial y_1} > 0 \quad \text{and} \quad \frac{\partial u}{\partial y_2} > 0. \tag{3.11}$$

In addition to these properties, further properties are necessary to tailor the utility function to the needs of presence system applications. The first property arises from the realization that the utility gain of both objectives should be decreasing with increasing objective value. That means, although the need for performance as well as for resource savings is always positive, the gain of utility is the less the larger the initial objective value is. This can be expressed by *decreasing marginal utility*

$$\frac{\partial^2 u}{\partial y_1^2} < 0 \quad \text{and} \quad \frac{\partial^2 u}{\partial y_2^2} < 0. \tag{3.12}$$

Quasiconcativity (convex preferences) is a feature that assures that well bundled alternatives are preferred over alternatives with unbalanced objective values for performance and resources savings, i.e. averages are preferred over extremes. Quasiconcativity is also important for the solution of the optimization problem as quasiconcave utility functions support global optimal states.
Another important property of the utility function is its inherent possibility to model the *substitution* of both objectives for each other. According to the applications the operator might define the objectives y_1, y_2 to be substitutable (objectives are substitutes), partly substitutable, or complementary (objectives are complements). Measures of the substitution effect are the marginal rate of substitution and the elasticity of substitution. The *marginal rate of substitution* (MRS) is the slope of the indifference curve or the negative, reciprocal relation of the marginal utilities

$$\xi = \frac{dy_2}{dy_1} = -\frac{u_{y_1}}{u_{y_2}}, \tag{3.13}$$

where u_y denotes the partial derivatives (marginal utility).
The *elasticity of substitution* (EoS) is defined to be the quotient of percentaged difference of y_2/y_1 and dy_2/dy_1:

$$\epsilon = \frac{d(y_2/y_1) : (y_2/y_1)}{d(dy_2/dy_1) : (dy_2/dy_1)} \tag{3.14}$$

A utility function that possesses the features demanded above (continuity, differentiability, non-saturation, decreasing marginal utility, quasiconcavity, adjustable substitution) is the constant elasticity of substitution (CES) utility function.

The Constant Elasticity of Substitution (CES) Utility Function

CES-utility functions are defined as

$$u = (ay_1^{\varrho} + by_2^{\varrho})^{1/\varrho}, \qquad a + b = 1,\ \varrho \leq 1, \tag{3.15}$$

where a, b are the distribution parameters, that define the share of utility contributed by the objective values. The parameter ϱ is the substitution parameter that defines the possibilities of substitution between both objectives. The function was introduced in the context of production theory [72]. The elasticity of substitution according to equation (3.14) of the CES-utility function becomes

$$\epsilon = \frac{1}{1-\varrho}. \tag{3.16}$$

As the name suggests it is always constant.
Depending on the substitution parameter ϱ the CES-utility function results in utility functions with different substitution properties:
$\varrho = 1$: The utility function becomes a *linear utility function*

$$u = ay_1 + by_2. \tag{3.17}$$

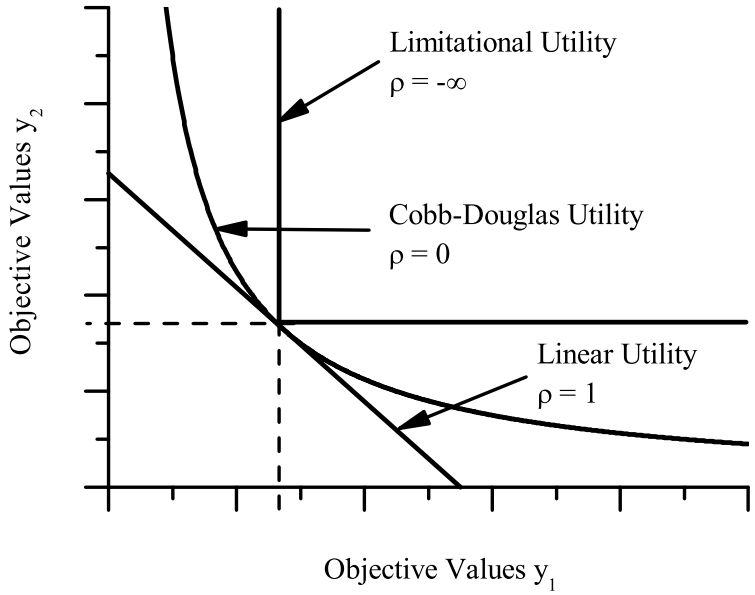


Figure 3.4: Indifference curves of CES-utility functions: Depending on the substitution parameter ϱ the CES-utility function results in either linear, Cobb-Douglas, or limitational utility functions. Indifference curves of linear utility are linearly decreasing representing perfect substitutes. Indifference curves of Cobb-Douglas utility are bent to the origin representing partly substitution. Indifference curves of limitational utility are rectangular defining the objectives as perfect complements.

The MRS is constant $\xi = a/b$, since the indifference curves are linear and the objectives are perfect substitutes with EoS $\epsilon = \infty$.
$\varrho \to 0$: The utility function becomes a *Cobb-Douglas utility function* since both functions have the same indifference curves. The Cobb-Douglas utility function is defined as

$$u = y_1^a y_2^b. \tag{3.18}$$

The MRS does not depend of the absolute objective function values $\xi = ay_2/(1-a)y_1$. The objectives are partly substitutive with unity EoS $\epsilon = 1$, that means, 10% change in the MRS yields a 10% change in the input mix. The Cobb-Douglas function has diminishing marginal utilities according to equation (3.12). Its indifference curves are bend toward the origin, hence, the function is quasiconcave and represents a convex preference structure. Indifference curves of a Cobb-Douglas utility function are displayed in Figure 2.13. The exponents a and $(1-a)$ represent their objectives share of output.
$\varrho \to -\infty$: The utility function becomes a *limitational utility function* (Leontief utility function)

$$u = \min\left\{\frac{y_1}{a}, \frac{y_2}{b}\right\}. \tag{3.19}$$

The MRS is infinite or zero since the indifference curves are rectangular. The objectives are perfect complements and elasticity of substitution is $\epsilon = 0$. The objective that limits utility is called *shortage factor* and the remaining objective is called *excess factor*. Since the indifference curves of a limitational utility function are rectangular only the corner point is important for utility maximization.

An indifference curve of a linear, a Cobb-Douglas, and a limitational utility function is depicted in Figure 3.4.

3.2.3 CES-Utility Functions for Presence Systems

The utility function, which represents the decision maker's preference structure is strongly influenced by the application the presence system is used for. Different applications generate different preferences even if the presence system remains the same. Three different preference structures are identified that are rationalized by three different CES-utility functions.
In the following three different preference structures including utility functions are presented.

Utility Function 1

Maximal performance under constraint resources: In many applications presence systems are limited in resources, e.g. the bandwidth of the COM in space or subaqueous teleoperation systems. Hence, resource savings cannot fall below a certain value and a minimal conflict c_{RS} is perceived. On the same time, highest possible performance is intended in every situation. Objective y_1 cannot be substituted by y_2 if $c < c_{RS}$ (resource constraint) and y_2 is not substitutable by y_1 when $f_p(c) < f_p(c_{RS})$ (maximal performance).
This preference structure can be represented by a limitational utility function as introduced in equation (3.19) and described in the related paragraph. The utility function only depends on the resource constraint and on the parameter a which can be chosen arbitrarily positive. The parameter b then becomes

$$b = \frac{ac_{RS}}{f_p(c_{RS})}. \tag{3.20}$$

Defining parameter a to be $a = 1$ the utility function only depends on the conflict c_{RS} induced by the resource constraint

$$u = \min\{y_1, \frac{y_2}{c_{RS}/f_p(c_{RS})}\}. \tag{3.21}$$

Utility Function 2

Performance and resource savings strongly preferred on equal terms: Another class of applications is characterized by costly resources. Resources are not constraint but the higher the amount of a certain resource used by the presence system, the higher the monetary costs the decision maker has to pay. Examples can be find in space or subaqueous teleoperation again or in quality-of-service enhanced networks like implemented in the Internet protocol IPv 6. But also power consumption could be a costly resource that effects performance. On the same time, the decision maker prefers high performance of the

Preference Structure	CES-Utility Function	Possible Application
non-substitutional, y_1 limited	$\varrho = -\infty, \xi = [\infty, 1], \epsilon = 0$ $u = \min\{y_1, \frac{y_2}{c_{RS}/f_p(c_{RS})}\}$	teleoperation with constraint bandwidth
partly substitutable, convex	$\varrho = 0, \xi = 1, \epsilon = \frac{ay_2}{(1-a)y_1}$, $u = y_1^a y_2^b$	teleoperation with costly bandwidth, QoS-networks
perfect substitutional, $y_1 \sim y_2$	$\varrho = 1, \xi = a/b, \epsilon = \infty$ $u = ay_1 + by_2$	evaluation and training experiments

Table 3.1: CES-utility functions for presence systems: Depending on the the application scenario non-substitutional, partly substitutional, and perfectly substitutional utility functions rationalize the decision maker's preferences for technology-mediated presence.

presence system. Hence, the decision maker prefers well-balanced alternatives to equal-utility alternatives dominated by one objective.
This convex preference structure can be represented by a Cobb-Douglas utility function as introduced in equation (3.18) and described in the related paragraph. The share parameter a represents the relative share of utility endowed by the objective 'performance' while $(1 - a)$ defines the relative share of utility endowed by the objective 'resource savings'. Diminishing marginal utility assures that a change in lower objective values has a higher effect on utility than a change in higher objective values.

Utility Function 3

No objective preferred in particular: The third class of applications is characterized by a decision maker who has no certain preferences for either one of the two objectives but seeks high utility. Hence, objectives are perfect substitutable. Examples for this class are hard to imagine as for presence systems high performance is mandatory and only substitutable if resources are limited or costly. However, in basic performance tests or training application the utility endowed by well-balanced and unbalanced combinations could the same.
The underlying preference structure generated by this class of applications can be represented by a linear utility function. It is introduced in equation (3.17) and described in the related paragraph. The share parameters a and $(1 - a)$ again define the relative utility endowed by the different objectives.

3.3 Solution of the Optimization Problem

The problem to maximize the utility with respect to the non-dominated feasible alternatives is single objective optimization. Since the Pareto-frontier can be represented algebraically, the utility maximization problem is a static optimization problem with equality constraint. The solution can be derived using analytic methods such as the Lagrangian method or numerical methods. The forms of the Pareto-frontier and of the indifference curves are well known. This simplifies the solution process as the existence and charac-

teristic of the solution can be verified graphically. However, since the Pareto-frontier is a statistical quantity with mean and standard deviation the solution of the optimization problem is only the most likely solution. Depending on the variability of the Pareto-frontier and on the utility function other optimal solutions are possible and have to be discussed. Example calculations are conducted using three utility functions introduced before subject to an experimentally derived Pareto-frontier.
Subsection 3.3.1 introduces the utility optimization problem and discusses the problem of an uncertain equality constraint. Section 3.3.2 describes the solution of the utility maximization problem for different preference structures and presents calculations based on experimental results.

3.3.1 Utility Maximization under Measurement Uncertainty

The general solution of the utility optimization problem is to maximize the utility function with respect to all feasible alternatives, i.e. within the feasible set. This problem is termed *utility maximization problem* (UMP) and is a static maximization problem with inequality constraints. It is defined in (2.54)-(2.56) and described in the related paragraph.
However, since the Pareto-optimal alternatives are always at the left border of the feasible set, the UMP can be simplified to the *reduced utility maximization problem* (rUMP). The rUMP is solved by maximizing the utility function subject to the Pareto-frontier reducing the problem to static maximization with respect to a single equality constraint. The rUMP is defined as

$$\max\ u(\mathbf{y}), \tag{3.22}$$
$$\text{s.t. } p(\mathbf{y}) = 0, \tag{3.23}$$

The necessary condition for an optimum of the rUMP can be derived by the Lagrangian equation

$$L(\mathbf{y}, \mu) = u(\mathbf{y}) + \mu p(\mathbf{y}). \tag{3.24}$$

Then, if $\mathbf{y}^*$ is a local maximum the necessary conditions of first order hold:

$$L_{y_1}(\mathbf{y}^*, \mu^*) = 0, \tag{3.25}$$
$$L_{y_2}(\mathbf{y}^*, \mu^*) = 0, \tag{3.26}$$
$$L_{\mu}(\mathbf{y}^*, \mu^*) = 0, \tag{3.27}$$

where L_{y_1} is the partial derivation of the Lagrangian equation with respect to y_1, L_{y_2} the partial derivative with respect to y_1, and L_{y_μ} the partial derivative with respect to μ. The solution is only valid if the elements represent a feasible alternative, hence, are non-negative. Sufficient conditions are not necessary since it is obvious by the problem statement that the optimal alternative depicts maximal, constraint utility (non-saturated increase of the utility function with increasing objective values).
Whether the solution is global or local can be identified according to the shape of the Pareto-frontier and the shape of the indifference curve of the utility function. A global solution is enabled by a convex Pareto-frontier and a strictly concave indifference curve or by a concave indifference curve and a strictly convex Pareto-frontier. A global or local solution is enabled by a convex Pareto-frontier and a concave indifference curve or

by a non-convex Pareto-frontier and a convex indifference curve. What kind of solution exists can be verified by mathematically analyzing the form of the Pareto-frontier and of the indifference curve, or, more simply, by looking at the graphical representation of the rUMP. An example is given in Figure 2.14 with a global solution.
Due to the finite reliability of the Pareto-frontier, their final form is not stiff but only defined statistically within a two-dimensional tube around $p(\mathbf{y}) = 0$. The mean Pareto-frontier $\mathbb{P}$ is only the most likely set of the efficient alternatives. Consequently, the solution $\mathbf{y}^*$, of the rUMP is only the most likely solution. Changes in slope, offset, or convexity characteristics lead to other, less likely, but possible solutions. All possible solutions form a two dimensional set $\mathbb{Y}^*$ around the most likely solution $\mathbf{y}^*$. The expansion and the topology of this set is defined by the form of the Pareto-frontier, its variability, and by the utility function. The faster Pareto-frontier and indifference curve distance from each other when moving away from the optimal alternative $\mathbf{y}^*$, (At $\mathbf{y}^*$, Pareto-frontier and maximal utility indifference curve are tangent to each other having same slope.) the smaller the set of possible optimal alternatives $\mathbb{Y}^*$. Hence, the more L-shaped the utility function, i.e. the higher the EoS (see equation (3.14) for a definition and Subsection 2.3.2 for a detailed description) the smaller the expansion of $\mathbb{Y}^*$. Furthermore, the less the variability of the Pareto-frontier the smaller the set of possible optimal alternatives. An illustration is given in Figure 3.5.

3.3.2 Utility Maximization for Different Preference Structures

The solution of the rUMP (3.22), (3.23) is strongly influenced by the utility function which encodes the preferences of the decision maker amongst his objectives 'performance' and 'resource savings'. Depending on the EoS of the utility function the set of possible optimal solutions $\mathbb{Y}^*$ varies, since the Pareto-frontier is a statistical quantity.
In the following the solution is characterized with respect to the preference structures proposed for presence systems in Subsection 3.2.3. The Pareto-frontier is described by a second order polynomial $p = -y_2 + k_0 + k_1 y_1 + k_2 y_1^2$ and experimental data is taken from Subsection 3.2.3.

Maximization of Utility Function 1

Maximal performance under constraint resources: In case of the limitational utility function given in equation (3.21) the rUMP becomes

$$\max \left[\min\left(y_1, \frac{y_2}{c_{RS}/f_p(c_{RS})}\right)\right], \tag{3.28}$$

$$\text{s.t. } p(y_1, y_2) = 0. \tag{3.29}$$

The optimal alternative is only depending on the perceptual conflict $y_{1,RS} := c_{RS}$ induced by the lower limit of the resource savings objective. All alternatives left of this resource savings constraint are not feasible and, therefore cannot become an optimal alternative. Since the objective y_2 'performance' is, according to the preferences, always maximal,

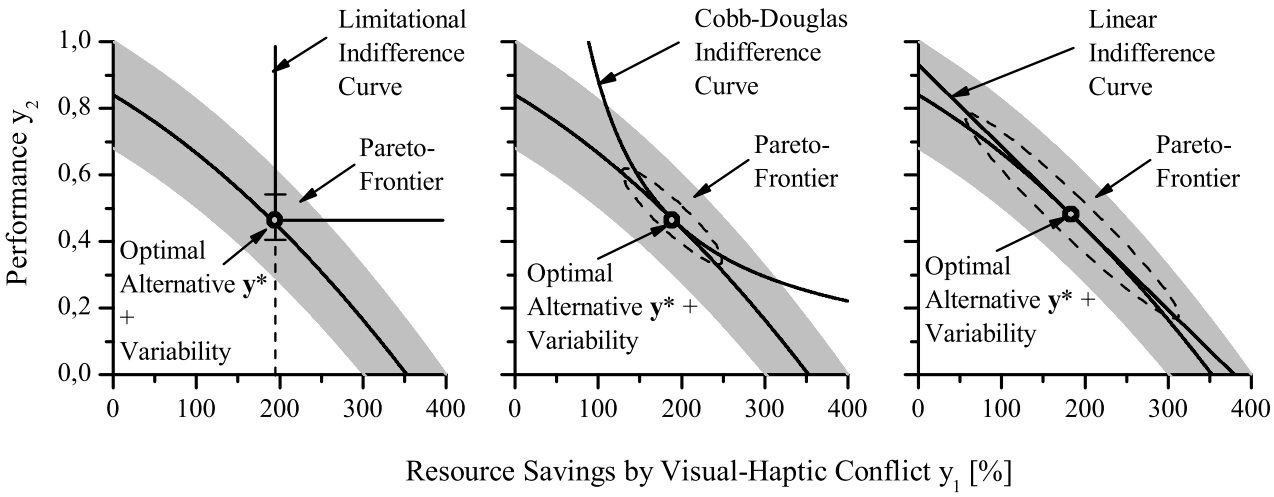


Figure 3.5: Utility optimization for presence systems at different preferences: Since the Pareto-frontier is a statistical quantity the optimal alternative $\mathbf{y}^*$ is only the most likely objective bundle satisfying maximum utility. The likelihood structure of optimal alternatives can be described by level sets centered around $\mathbf{y}^*$. Depending on the variability of the Pareto-frontier and on the utility function the expansion of the level sets differs: The higher the standard deviation of the Pareto-frontier and the larger the Elasticity of Substitution of the decision maker the the larger the expansion of the level sets, i.e. the less reliable the optimal alternative. The diagrams show the results of utility maximization subject to the experimentally recorded Pareto-frontier of Subsection 3.1.2 under the three different preference structures described in Subsection 3.2.3.

the most likely optimal alternative $\mathbf{y}^*$ is located in the corner of the indifference curve. Hence, the solution of the rUMP can be obtained by evaluating the expression of the Pareto-frontier at $y_1 = c_{RS}$. Due to the non-smooth characteristic of the indifference curves a solution by the Lagrangian equation fails. For the Pareto-frontier recorded in the experiment presented in subsection 3.1.2, assuming a perceptual conflict $y_1 := 195$, the most likely alternative is located at

$$\mathbf{y}^* = [195, 0.46]. \tag{3.30}$$

Optimization was performed on the mean Pareto frontier described by equation (3.8). The utility endowed by the most likely optimal alternative is $u = 0.46$. The solution is depicted in the left diagram of Figure 3.5. Because of the statistical character of the Pareto-frontier the less likely optimal alternatives are located on the vertical limb of the indifference curve and also at its lower extension. Utility remains the same if the Pareto-frontier is shifted upward. The utility decreases if the Pareto-frontier is shifted downward (indicated by the dashed line in the left diagram if Figure 3.5.
If no resource constraint is given and, therefore, no lower resource savings limit, the utility function rationalizes a preference structure that seeks maximal performance only.

Maximization of Utility Function 2

Performance and resource savings strongly preferred on equal terms: In case of the Cobb-Douglas utility function the most likely optimal alternative is located at the point where Pareto-frontier and an indifference curve tangent each other. Assuming the mean Pareto-frontier is expressed by a second order polynomial as introduced before, the rUMP with Cobb-Douglas utility is defined as

$$\max\ y_1^a y_2^b, \tag{3.31}$$
$$\text{s.t. } p(y_1, y_2) = 0. \tag{3.32}$$

The Lagrangian equation becomes

$$L = y_1^a y_2^b + \mu(-y_2 + k_0 + k_1 y_1 + k_2 y_1^2) \tag{3.33}$$

and the first order necessary conditions according to (3.25)-(3.27) become

$$L_{y_1}(\mathbf{y}^*, \mu^*) = a y_1^{a-1} y_2^b + \mu(k_1 + 2k_2 y_1) = 0, \tag{3.34}$$
$$L_{y_2}(\mathbf{y}^*, \mu^*) = b y_1^a y_2^{b-1} - \mu = 0, \tag{3.35}$$
$$L_{\mu}(\mathbf{y}^*, \mu^*) = -y_2 + k_0 + k_1 y_1 + k_2 y_1^2 = 0. \tag{3.36}$$

Dissolving this equation system yields a quadratic equation, which contains the solution for y_1^*

$$k_0 + (k_1 + \frac{bk_1}{a})y_1 + (k_2 + \frac{2bk_2}{a})y_1 = 0. \tag{3.37}$$

Negative solutions y_1 can be neglected since they are not contained within the feasible set $\mathbb{Y}$. The remaining objective function value y_2^* can be calculated by evaluating the Pareto-frontier at y_1^*.
For the Pareto-frontier recorded in the experiment presented in Subsection 3.1.2, the solution of the rUMP ($a = 0.5, b = 0.5, k_0 = 0.84, k_1 = -0.0015, k_2 = -25 \cdot 10^{-7}$) becomes

$$\mathbf{y}^* = [189.87, 0.46]. \tag{3.38}$$

Optimization was performed on the mean Pareto frontier described by equation (3.8). The utility endowed by the most likely optimal alternative is $u = 9.35$. The solution is depicted in the center diagram of Figure 3.5.

Maximization of Utility Function 3

No objective preferred in particular: The solution of the rUMP in case of the linear utility function (3.17) is similar to solving the problem with Cobb-Douglas utility. With linear utility the rUMP becomes

$$\max\ a y_1 + b y_2, \tag{3.39}$$
$$\text{s.t. } p(y_1, y_2) = 0. \tag{3.40}$$

The solution by the first order conditions based on the Lagrangian equation

$$L = a y_1 + b y_2 + \mu(-y_2 + k_0 + k_1 y_1 + k_2 y_1^2) \tag{3.41}$$

becomes

$$y_1^* = -\frac{ak_1 + b}{2ak_2}. \tag{3.42}$$

The remaining objective function value y_2^* can be calculated by evaluating the Pareto-frontier at y_1^*.
For the Pareto-frontier recorded in the experiment presented in subsection 3.1.2, the solution of the rUMP ($a = 1, b = 0.00245, k_0 = 0.84, k_1 = -0.0015, k_2 = -25 \cdot 10^{-7}$) becomes

$$\mathbf{y}^* = [190.07, 0.46]. \tag{3.43}$$

Optimization was performed on the mean Pareto frontier described by equation (3.8). The utility endowed by this optimal alternative is $u = 190.07$. The solution is depicted in the right diagram of Figure 3.5.

3.4 Conclusion

Summary

In this chapter a new method to evaluate mediated-presence (performance) against used system resources (resource savings) of a presence system is presented. The method is embedded within the mathematical theory of multiobjective optimization and can be divided into two parts.
The first part deals with the identification of the non-dominated alternatives, i.e. combinations of a certain performance and a certain degree of resource savings that represent feasible working points of the presence system. This is an optimization problem with two objectives. It is proposed that the non-dominated alternatives are identified by presence measurement or psychophysical methods. According to the theory the set of non-dominated alternatives is called Pareto-frontier. An experiment is conducted assessing the Pareto-frontier using methods of both classes. Thereby, a new psychophysical method to assess perceived presence is proposed concurrently.
The second part deals with the derivation and maximization of an utility function in order to identify the most preferred alternatives among the technical efficient, i.e. the Pareto-optimal alternatives. Based on a structured analysis different preference structures are identified especially suitable for presence systems. Utility functions are developed that rationalize the different preference structures and give them a mathematical form. Solutions of the optimization problem subject to a deterministic model of a Pareto-frontier are presented. The solutions are also discussed when being subject to an uncertain Pareto-frontier, which will be the outcome of an experimental assessment. Calculations based on an experimentally recorded Pareto-frontier are presented, too.

Scientific Contribution

The method opens a new way to the understanding, the analysis, and the evaluation of technology-mediated presence and presence systems. Similar results from literature are

not known. The method provides the following benefits:
1.) The former problem of evaluation methods being too specialized evaluating presence systems solely by measuring maximum achievable presence, is improved by extending the evaluation process to an additional dimension. The technology that enables as well as hinders veridical experience of presence is taken into account and the presence achieved is seen as a trade-off between the system's capabilities and the system's resources. Thereby, a much more holistic description of a presence system is possible represented by the Pareto-frontier. The form of the Pareto-frontier provides more information about the system's capabilities than the former evaluation quantity, i.e. the highest possible presence that could be achieved under ideal circumstances measured by non-standardized methods. The Pareto-frontier allows for a more detailed description of 'improvement' of the presence-mediating technology:

Definition 3.4.1 *(**Improvement of mediating-technology**) Improvement of a presence system's technology is achieved if, ceteris paribus, one or more alternatives of the improved system dominate the Pareto-optimal alternatives of the non-improved presence system.*

Based on this definition, definitions of local and global improvement can be derived

Definition 3.4.2 *(**Global and local improvement of mediating-technology**) Global improvement occurs if all alternatives of the improved system's Pareto frontier dominate the Pareto-frontier of the non-improved system. Local improvement occurs if only some alternatives of the improved system's Pareto frontier dominate the Pareto-optimal alternatives of the non-improved presence system, respectively.*

Finally, conditional improvement terms the gain of improvement in some regions and the loss of improvement in other regions.

Definition 3.4.3 *(**Conditional improvement of mediating-technology**) Conditional improvement has occurred if improvement has occurred and one or more alternatives of the improved system's Pareto-frontier are dominated by alternatives of the Pareto-frontier of the non-improved system.*

Thereby, improvement can be generated by any development of the subsystems of the mediating technology (HSI, COM, TO).
2.) The former problem of evaluation methods being to general to take certain application scenarios into account is solved by this method as it can be freely customized to different application scenarios of presence systems. Utility functions rationalize different preference structures generated by different applications. Arbitrary preference structures for any kind of application can be implemented, if these preference structures can be defined as binary relations within order theory. Here, three different utility functions are derived for three basic application scenarios. Thereby, mainly the elasticity of substitution was adjusted to represent preference structures generated by space teleoperation or surgery procedures, by presence application with costly resources, and by test scenarios or psychophysical evaluation scenarios of presence systems. The solutions were calculated

and the impact of the different preference structures were discussed. Depending on the preference structure the character of the solution differed tremendously. The solutions gave plausible implementation advice for different application scenarios. The method also unveils shortcomings of current evaluation methods. As depicted in Figure 3.5 any utility optimization is superfluous, if the variability of the method to measure the Pareto-frontier is too large. Any evaluation statement about performance, improvement, or utility must be based on a reliable Pareto-frontier having low standard deviation. As can be seen at the outcome of the experiment conducted in Subsection 3.1.2 questionnaires or magnitude estimation to measure presence are not suitable for the assessment of a Pareto-frontier and, therefore, their application is doubtful, especially if the presence system is not parametrized based on a standardization or at least a clear structure.
3.) The method can serve different intentions. Firstly, it provides clear means for the understanding and analysis of presence and technology within presence systems and their applications. As the relation between presence and resources are modeled using a neat mathematical concept, multiobjective optimization, understanding of the processes within presence systems are facilitated. Secondly, the method can be used purposely to parametrize the technology of a presence system. Especially, when the technology should be used in an application that is characterized by costly resources, the method provides clear advice which working point has to be chosen. However, the quality of the advice crucially depends on the reliability of the Pareto-frontier and on the mathematical formulation of the decision maker's preferences. Thirdly, the method can be used as basic method for an evaluation center to test the impact of arbitrary algorithms to be implemented in the technology within presence systems. Based on a precisely recorded Pareto-frontier of a presence system, which is especially used for the evaluation, the algorithm to be tested can be analyzed under different utility functions. Reliable statements about the improvement (global, local, or conditional) caused by this algorithm can be conducted for the encoded application scenarios.

Open Problems

Firstly, the method can be extended to incorporate resource savings of multiple resources in generally two ways. One way is to apply the presence measurement method to record the Pareto-frontier on a system influenced by savings on multiple resources. However, that would not provide advice how to parametrize the different technical resources, since the effect of multiple resources is observed through one parameter, the performance degeneration measured by the presence measurement method, only. The second way is to record a Pareto-frontier for each resource independently. The result would be a Pareto-frontier with a dimension > 2, since every resource under consideration adds another dimension. To identify the preferred alternative on this hyper plane, a utility function with the same dimension is necessary.
Secondly, performance measurement methods have to be identified that provide a high reliability, i.e. low variability of the Pareto-frontier. The experimental assessment described in Subsection 3.1.2 unveiled that the methods questionnaire and magnitude estimation provided a Pareto-frontier with a reliability ($sd = 0.16\,\mathrm{mm/N}$, see equation (3.10)) that is too low for further statements. Methods that provide a higher reliability are methods of psy-

chophysics, namely the Method of Constant Stimuli and the Two-Alternative Force-Choice Task (see Subsection 2.2.2). The drawback of these measures are that their implementation and execution is laborious and could only be done in a well-established experimental test-bed.
Thirdly, the method should be applied using a existing and costly/constraint resources. In the experiment described in Subsection 3.1.2 the conflict measured was induced artificially and not by savings of a resource used to drive the presence system. An proposition of a resource are the impact of bandwidth savings by the reduction algorithms for haptic data that are introduced in Chapter 6.

4 Visual-Haptic Perception of Compliance: Explorative Studies

If you want realism,
look at the world around you;...

R. Dorfman, *The Price System* (1964)

Mechanical environments are mainly perceived by processing visual and haptic information. As mass and damper, compliance is an atomic part of every mechanical environment. It is mathematically described by *Hooke's law* as the division of position through force. Knowledge about the perception of compliance explains us how we explore our daily environments. But also for research on presence systems, knowledge about the perception of compliance is of crucial importance as it is intended to veridically resemble a mechanical target environment. In contrary to literature on the perception of mechanical signals, such as force and position-based information, literature on the perception of compliant environments is sparse. Even more sparse is literature on the bimodal perception of compliant environments as most publications are constraint to the haptic modality. Hence, the perception of compliant environments has in store a huge number of unsolved questions e.g. about temporal perception, crossmodal perception, or about fusion and combination of redundant and non-redundant information.
In this chapter three different studies are presented in which we analyzed the discrimination performance of a human operator when perceiving artificially rendered compliances. Firstly, human's performance was assessed when perceiving visual-haptic stimuli with respect to the temporal presentation. The problem was analyzed by an iterative staircase method and resulted in a clear answer: Humans are more precise when visual and haptic information is presented sequentially. Based on the outcome of this study the second study aimed at identifying the general performance of human's visual-haptic compliance perception using a matching task. As main results we found that the JND of compliance perception is around $25 - 35\%$ and that the visual modality might not contribute to this performance. Because of the unclear result about the role of the visual modality, the third study directly dealt with the visual-haptic fusion of compliance information using a 2AFC-task. One of the results of this study supported the previous result as we found that visual-haptic fusion seems not to occur in compliance perception. The studies had been partly conducted in [73–77]. The studies have been published on different conferences [70; 78; 79]. Based on the results of these studies, in Chapter 5 a theoretical model of compliance perception was developed and experimentally evaluated.
The remainder is divided in four Sections. The three different studies are presented in Section 4.1, Section 4.2, and in Section 4.3. A final conclusion is presented in Section 4.4.

4.1 Concurrent vs. Sequential Assessment

A human system interface (HSI) enables a human operator to perceive and act in virtual or remote environments (see Figure 1.1). Perception and manipulation capabilities can be increased by feeding multimodal (e.g. visual and haptic) information of the target environment back (for an overview see [80]). Haptic feedback is very sensitive since mechanical energy is exchanged over command and feedback signals. This closed feedback loop is susceptible to different kinds of disturbances and can even become instable (e.g. [18; 19]). Therefore, at the operator's site incongruences between visual and haptic information about the target environment can occur. However, humans can perceive even incongruent bimodal information without any conflict. Measurement of this perceptional process requires an HSI with high accuracy and extensive experiments using psychophysical procedures. Perception of bimodal mechanical information is analyzed here at the example of compliance information. The study was published to a large extent in [78].
It is known that information of more than one modality is integrated to form a coherent percept (e.g. [81; 82]). Precondition of integration is spatial as well as temporal congruence of information (e.g. [81; 82]). However, if an intermodal conflict is below threshold of perception, integration still takes place (e.g. [83]), even though only one modality has been attended (e.g. [84]). In this context, Marks introduces the terms *stimulus* and *perceptual congruence* (see in [83], pp. 85-105): Stimulus congruence denotes that there are no differences in the physical stimulus parameters. Perceptual congruence is a psychological construct: Even without physical congruence discrepancy in visual-haptic information remains to some extent unnoticed by the observer. Because a qualitative difference of the integrated percept depending on physical or perceptual congruent stimulus presentation has been observed, different mechanisms may be involved [85; 86]. Attentional processes might influence integration and the extent of discrepancy, respectively (see e.g. [82; 87]). One branch of integration theory states that information of two or more modalities are combined by differently weighting them (e.g. [88; 89]). The relative contribution of each sense depends either on the appropriateness (e.g. [90]), the effectiveness (e.g. [91]) or the reliability (e.g. [92; 93]) of each modality, or on the direction of focusing attention (e.g. [94]). A wealth of research into visual-haptic integration exists and has concentrated either on size (e.g. [92; 95; 96]), shape (e.g. [97–99]) or texture perception (e.g. [100–102]), as well as on visual influence on proprioceptive localization (e.g. [103–105]), respectively. Bimodal perception of compliance information has evoked only few studies yet [106; 107]. Most of the research on integration indicates visual dominance over the haptic modality (for a review see [83; 90]), especially in spatial properties (e.g. [91]). Moreover, some factors, such as age (e.g. [108]), response modality (e.g. [109]), instruction (e.g. [110]) or noise (e.g. [92]), have been found to reduce visual influence. Some studies even show tactile dominance over vision (e.g. [100; 101; 111]) or at least in some tasks (e.g. [112]). Visual dominance therefore is no general phenomenon and depends on additional task relevant factors. Bimodal perception of compliance information has evoked only few studies yet [106; 107].
The contribution of this study is to identify the the just-noticeable difference (JND) (see e.g. [113]) when perceiving object compliance through a visual-haptic HSI. The relative JND is defined as

$$\text{JND} = \frac{|S - S_{ref}|}{S_{ref}}, \tag{4.1}$$

where S is the compliance stimulus[1]. Compliance is the combination of force f and position x information and can be expressed by Hooke's law

$$S = \frac{x}{f}. \tag{4.2}$$

Three hypotheses were tested. *Hypothesis 1*: Attending separately to information of two modalities arising from one source might introduce a high bias: The human perceptual system tries to integrate even conflicting information to provide a coherent percept, especially if information is derived from one source (see [81; 82]). Therefore, concurrent comparison of visual and haptic information should result in reduced detection performance (further referred to as 'method A'). On the other hand, a low discrepancy threshold should result when attending to the object as a whole, and hence to compare a visually presented object to an haptically presented object sequentially (further referred to as 'method B'). *Hypothesis 2*: Most of the research on visual-haptic integration reports visual capture in intermodal conflict situations (see above). It is therefore expected that visual dominance should occur: The detection of intermodal discrepancies in object compliance should be impaired when the visual modality remains unaltered (as the target or reference modality) and hence the haptic modality varies. *Hypothesis 3*: As has been shown in different studies (see above), visual dominance seems not to be a general phenomenon and to depend on task relevant factors. It is expected that modality dominance is not constant over the whole stimulus range of object compliance: Low compliant objects provide scant visual compliance information, whereas high compliant objects might be easier perceived when relying on visual information. All three hypotheses were tested regarding influence of assessment method, reference modality (i.e. unaltered target modality), and reference compliance with respect to the detection threshold.
In Subsection 4.1.1 the method is explained. The results are described in Subsection 4.1.2. A discussion is presented Subsection 4.1.3.

4.1.1 Method

Presence System

A human system interface is used that provides visual and haptic feedback at high accuracy. The visual subsystem consists of a TFT screen mounted in the line of sight of the hand showing the visual virtual reality. The haptic subsystem consists of two SCARA robots providing two degrees of freedom each. The system interacts with index finger and thumb to allow the exploration of the compliant environment by gripping movements. Workspace is about 80 mm and maximal force is about 35 N. Position information is measured by angle encoders and force is measured by strain gauges. Haptically, the compliant environment is generated using an admittance control scheme. Visually, the environment is presented by a compliant cube with yellow spheres representing the finger tips. During the psychophysical experiments participants were able to insert their answers using a joystick that provides them with the different answer possibilities. The system works under realtime conditions

[1] A capital S is used to indicate that the stimulus is a combination of two other stimuli, i.e. a relation, a mapping from one stimulus to another.

and is programmed by Matlab/Simulink. A detailed description of the HSI including hardware, software, and control structure is provided in Appendix A.1.

Participants

Thirty-two (32) students of the Technische Universität München and the Universität der Bundeswehr München took part in this study and were paid for participation. Half of the participants were assigned to group A (method A), the other half to group B (method B). Due to missing values, two (group A) and five participants (group B) had to be excluded from further analysis. The average age of participants amounted to 25 years (group A) and 26 years (group B). Eleven (11) men and 3 women (group A) and 6 men and 5 women (group B) participated. All of them were right-handed and had normal or corrected to normal vision.

Stimuli

Seven (7) reference compliances were selected covering a broad range feasible by the HSI. Reference compliance amounts to

$$S_{ref} = [0.2; 0.4; 0.5; 0.8; 1.4; 2.5; 4.9]\,\mathrm{mm/N}. \tag{4.3}$$

Additionally, intermodal discrepancy should be assessed with either the visual or the haptic modality remaining unchanged and therefore being the reference modality: Reference compliance of the unchanged modality was one of the seven values, whereas compliance of the comparison or non-target modality was varied according to the procedure.

Procedure

An adaptive staircase method was used to assess performance of participants (see Subsection 2.2.2 for a detailed explanation). Thereupon, two ways of assessment were defined: Method A demanded comparing visual and haptic information *concurrently* within one trial and to decide whether sensory information deviates from each other, while in method B participants had to *sequentially* compare a congruent bimodal stimulus with an incongruent bimodal stimulus. With both methods a total of 14 threshold values had to be assessed, namely each reference stimulus (7) and target modality (2) combination. In order to reduce overall testing time for participants, method of assessment was chosen to be a between-participants-variable and testing of the 14 experimental conditions was divided into two sessions with seven randomly chosen stimuli. These stimuli were selected with the following restrictions: Neither the same reference compliance nor the same reference modality were presented in succession.
Because the ability to be drawn into a book, film or VE, better known as *immersive tendency* (see [114]), has been known to play an important role in designing human-machine-interaction, discrepancy of bimodal information might be mediated by this personal factor. Therefore, an additional 12-item questionnaire was included to control for that variable ([114], translated by [71]): Immersive tendency was assessed by answering

12 items building the two factors, *tendency to get emotionally involved* and *degree of involvement* (see [71]).
An additional group specific question was included in the demographical questionnaire: Group A rated under which reference (unchanged) modality they felt easier to perform the task, whereas group B rated the sensory information they mostly had relied on during test sessions.
Participants were seated in front of the HSI with their dominant hand grasping the device and while looking nearly perpendicular at the screen. They were carefully instructed according to their group membership to which they were randomly assigned. A training period had to be completed prior to each test session. Afterward, seven experimental conditions (one test session) were randomly presented. Participants explored the stimulus depending on their group membership and responded by a joystick.
The start and end of each trial was signaled by a sound. Duration of stimulus presentation depended again on group membership: Group A compared visual and haptic information for 4 s with an intertrial-interval amounting to 4 s. Group B tested the stimulus compliance for 2 s with an interstimulus-interval of 2 s and an intertrial-interval of 4 s. Masking of environmental noise was regarded to be not necessary due to the HSI making no disturbing noise, which might influence the participants' responses.
At the end of the second test session participants filled in questionnaires assessing their demographical data, their experience during testing (additional group specific question) as well as the immersive tendency questionnaire.
Two different procedures of measuring the perceptual threshold according to the group variable had been used (method A, B). Both methods assessed the relative just noticeable difference JND. The psychophysical procedure to derive the JND was an adaptive staircase method targeting the 50% performance level. Initial stimulus and step size were adopted according to group membership.

Group A: Participants were instructed to concurrently compare within one trial information from both modalities given the reference modality which was announced by the experimental instructor prior to measurement. Therefore, the experimental task was to make *crossmodal* comparisons and to choose between two response alternatives, namely 'difference' vs. 'no-difference'. In case of haptic reference the visual modality had to be adjusted to match the haptic modality. Since only position measurement is possible by vision, the perceptual task was to match the visual position to the haptic position encountered while exploring the compliant object

$$\text{Method A, haptic ref.:}\quad x_v \overset{\text{adjust to}}{\rightarrow} S = \frac{x_o}{f_o^e}, \tag{4.4}$$

where x_v denoted the position information obtained by the visual modality and x_o denotes the information measured by the haptic modality as used before. In case of visual reference the haptic modality had to be adjusted to match the visual modality. Since participants had to match compliance information (measured by the haptic modality) to a position (measured by the visual modality) they had to filter out the haptic position x_o from the haptic compliance estimate to perform the matching

$$\text{Method A, visual ref.:}\quad S = \frac{x_o}{f_o^e} \overset{\text{adjust to}}{\rightarrow} x_v, \tag{4.5}$$

Threshold measurement by the staircase procedure started with the comparison stimulus

$$S_0 = 2S_{ref}, \tag{4.6}$$

while step size was adapted according to

$$S_1 = S_0 - xS_{ref}. \tag{4.7}$$

After the third turning point step size was reduced from $x_1 = 0.1$ to $x_2 = 0.03$. Congruent stimuli were interspersed with a probability of 10%. After having reached the threshold ten times, threshold measurement ended: The difference threshold was computed as the mean of the limit cycle consisting of the last, unchanged turning points. The principle of the adaptive staircase method is depicted in Fig. 2.9.

Group B: Participants explored two objects sequentially one of which characterized by a discrepancy and the other being the congruent one

$$\text{Method B: } S \overset{\text{compare with}}{\leftrightarrow} \hat{S}, \tag{4.8}$$

where S denotes the congruently displayed compliance and $\hat{S}$ the incongruently displayed compliance. Reference modality was not announced by the experimental instructor. The participants' task was to decide whether the second stimulus felt more or less compliant than the first or whether there was a difference between the first and second stimulus. Threshold measurement started with the comparison stimulus yielding a discrepancy of

$$S_0 = S_{ref} + 0.8S_{ref}, \tag{4.9}$$

while step size was varied according to equation (4.7). Until the third transition point has been reached, step size amounted to $x_1 = 0.1$ and was then reduced to $x_2 = 0.03$. Congruent comparison stimuli were interspersed with a probability of 5%. After having reached the threshold six times, the sequence ended. The difference threshold was defined as the mean of the limit cycle consisting of the last, unchanged turning points.

4.1.2 Results

Immersive Tendency

Participants rated their immersive tendency on a 7-point-scale building the two factors *emotional involvement* and *degree of involvement* which were computed for each participant. Group A showed an average emotional involvement of 23.3 (standard deviation $sd = 5.8$) and an average degree of involvement of 25.6 ($sd = 6.9$), whereas mean emotional involvement amounted to 16.8 ($sd = 4.4$) and mean degree of involvement to 24.6 ($sd = 8.7$) in group B. All values did not statistically significant differ from those reported by Scheuchenpflug [71], indicating that the participants are a good sample of population.
In order to find out whether the two groups (A, B) differed from one another regarding the immersive tendency, because there seemed to be a difference in emotional involvement

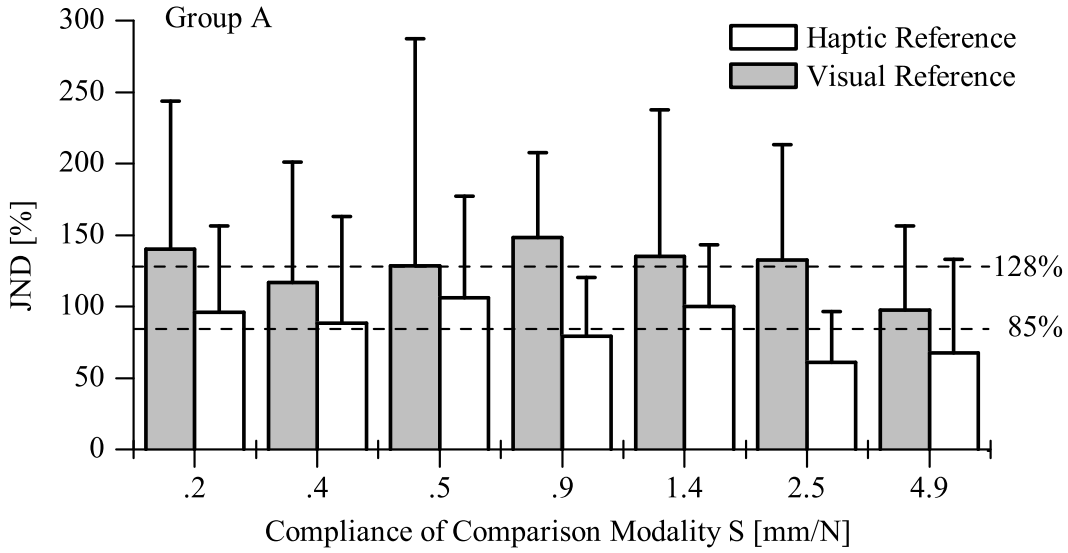


Figure 4.1: Results of Group A: When visual and haptic compliance information of one object were compared concurrently the JND is higher if the visual modality is the reference modality. Error bars indicate one standard deviation.

at least descriptively, a t-test for independent groups was computed. No difference in degree of involvement could be found ($t(21) = 0.3, p = 0.8$). However, both groups differed statistically significantly in their emotional involvement ($t(21) = 3.0, p < 0.05$): Group A rated to be higher emotionally involved than group B. However, only 29% of variance could be explained by this effect which therefore can be neglected.

Group A: No correlation between emotional involvement and performance could be observed. Only two variables when the haptic modality was the (unaltered) target modality showed a statistically significant (significance-level of 5%) correlation with degree of involvement: Reference compliance of 2.45 mm/N (Spearman ϱ=+0.7) and of 0.42 mm/N (ϱ=−0.8). A positive correlation indicates a higher JND along with a higher degree of involvement, whereas a negative correlation indicates better performance (reduced JND) with a higher degree of involvement.

Group B: Additionally, emotional involvement had no influence on performance, whereas degree of involvement statistically significantly influenced perception threshold: With the visual modality being the target modality and a reference compliance of 0.42 mm/N a negative correlation could be observed ($\varrho = -0.7$).

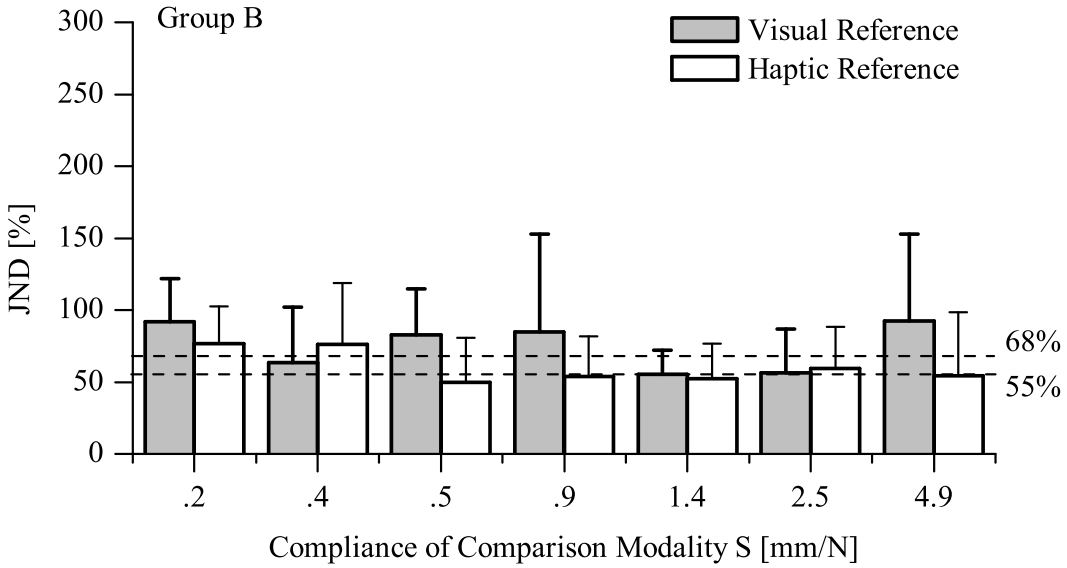


Figure 4.2: Results of Group B: When visual and haptic compliance information of one object were compared sequentially the JND was higher in only some stimuli when the visual modality is the reference modality. In general, JNDs were much lower compared to the concurrent presentation (Group A). Error bars indicate one standard deviation.

Group Specific Questions

Participants answered an additional question according to the modality which facilitated the given task (group A) and according to the modality participants mostly attended to (group B). To determine whether there was an influence on performance or a relation to immersive tendency separate correlation analyses for both groups (A, B) were computed.

Group A: Participants answered performing the task was easier when the reference modality was either the haptic (n=2) or the visual (n=4) modality or both together (n=8). There was neither a correlation with immersive tendency nor performance.

Group B: Participants reported that they primarily attended the haptic modality (n=5), the visual modality (n=3), both modalities without preference (n=2). Rating of the attended modality affected statistically significantly (significance-level of 5%) performance when the visual modality was the (unaltered) target modality: Reference compliance of 0.5 mm/N ($\varrho = +0.7$) and 0.42 mm/N ($\varrho = +0.8$) indicating higher performance when attending to the haptic modality, medium performance for the visual and lowest performance when attending to both modalities. However, there was no correlation with immersive tendency.

Descriptive Analysis

The JND was computed for each experimental condition within both groups. The mean performance is presented in Table 4.1.

Group A: As can be seen in Figure 4.1, reference modality affects performance: When the visual modality remains unchanged according to equation (4.5), the average JND is higher than in the case of the haptic reference according to equation (4.4), i.e. the task described by equation (4.5) is more difficult than the task described by equation (4.4). This indicates visual dominance, i.e. it is more difficult to filter the haptic position information to match the visual reference than to adjust the visual information to a filtered haptic position reference (visual modality measured position only).

Group B: Group B showed overall higher performance than group A (see Figure 4.2): JND is lower and standard deviation are smaller indicating that group B had less difficulties performing the detection task sequentially according to equation (4.8). Again, there seems to be an influence of reference modality on JND, but only in some reference stimuli. Most of the participants who had to be excluded were unable to perform the tasks when the visual modality remained unaltered and thus more likely when attending to the visual modality.

Testing Hypotheses

In order to determine the influence of reference compliance (0.22 to 4.88 mm/N) and reference modality (visual, haptic) depending on assessment method (A, B) on the ability to detect discrepancies in intermodal information, a $7 \times 2 \times 2$ analysis of variance (ANOVA) with repeated measurements and method as between-participants-variable was computed (significance level of 5%).
Threshold was significantly different in both groups ($F(1, 23) = 9.5, p < 0.05$; partial $\eta^2 = 0.29$): Group B showed higher sensitivity, i.e. detected smaller discrepancies in intermodal information. Additionally, reference modality influenced the JND: Performance was higher when the haptic modality was reference modality and therefore the visual modality was changed during the testing ($F(1, 23) = 8.6, p < 0.05$; partial $\eta^2 = 0.27$). No interaction of target modality and reference compliance could be observed ($F(6, 138) = 0.7, p = 0.6$). No other effects reached statistical significance.
However, effect size (partial η^2) of both main effects is very low and, as can be seen comparing Figure 4.1 to Figure 4.2, the above reported influence of reference modality seems primarily due to performance of group A. Therefore, a 7×2 ANOVA with repeated measurements was computed for each group.
Again, main effect of modality was statistically significant in group A ($F(1, 13) = 10.3, p < 0.5$); the effect now accounted for 44% of the variance. However, no influence of reference modality on performance could be observed in group B ($F(1, 13) = 1.2, p = 0.3$). The only other though negligible effect that reached significance was the interaction between target modality and reference compliance ($F(6, 60) = 2.5, p < 0.05$; partial $\eta^2 = 0.20$) indicating

Reference (Target) Modality	Comparison	
	Concurrently (Group A)	Sequentially (Group B)
Vision	JND= 128%	JND= 55%
Haptic	JND= 85%	JND= 55 to 68%

Table 4.1: Summary of the results: When comparing compliances reference modality seems only to play a role when comparisons are performed concurrently. Sequentially comparing compliances yields better performance and is independent of the reference modality.

higher JND when reference compliance is $0.5, 0.85$ and 4.88 mm/N and the visual modality is the reference.

4.1.3 Discussion

Difference thresholds in visual-haptic compliance information were assessed for different experimental conditions. Participants either had to concurrently compare visual with haptic compliance information (method A) or to sequentially compare two compliant objects displayed visual-haptically (method B). The chosen method affected the detection performance, as was expected (*Hypothesis 1*). As can be seen in Table 4.1, concurrent comparisons yield low performance (around 128% to 85%), whereas sequential comparisons yield performances between 55% to 68%. Similar results are obtained by Srinivasan, Beauregard & Brock who showed that participants' ability to identify the less compliant of two easiliy distinguishable compliant stimuli decreased as the ratio between visual and haptic discrepant compliance information increased to around 0.5 [107].
The different JNDs reflect that detecting visual-haptic conflicts is very difficult as long as no congruent comparison is available. The perceptual system may integrate the information in order to provide a coherent percept (e.g. [81; 82]). On the other hand, conflicts can be more easily detected when comparing to congruent information. As expected (*Hypothesis 2*), reference modality influenced discrimination performance: When varying haptic information (visual modality is unchanged), the JND is higher than with the haptic modality being the reference (see also [107]). This indicates that participants relied more on visual information when performing the discrimination task. Although this visual dominance can be observed for both groups (main effect), especially performance of group A contributes to this effect since this group had to perform an additional filtering to extract the position information from the haptic percept. In group A, 128% difference between the visual comparison and the haptic reference information are necessary in order to be detected, whereas 85% intermodal difference can be detected when the haptic modality remains unchanged (see Table 4.1). This indicates that the filtering of position information is more demanding if the information that has to be matched has to be filtered than if the reference information has to be filtered.
Visual dominance depended on reference compliance in group B, as expected (*Hypothesis 3*). However, effect size is rather low. Additionally, a positive correlation between

directing attention and discrepancy threshold could be observed in some reference stimuli: Performance decreased when participants attended to the 'wrong' modality, i.e. vision. Moreover, the analysis of missing values revealed that expecting the 'wrong' modality resulted in a non-convergence of the iterative psychophysical method. The cost of attending the wrong modality has already been shown to decrease performance (e.g. [94; 115]). Whether the influence of attention accounts for the observed result has not been systematically addressed in this study. Further experiments have to clarify, whether this effect might account for this interrelation.

4.2 Sequential Crossmodal Matching

Mechanical environments can be perceived by combining force information f and position-based information (i.e. position x, velocity, acceleration). Considering visual-haptic perception of compliance, expressed by Hooke's law

$$S = \frac{x}{f}, \tag{4.10}$$

position-based information can be detected by both modalities while force information can only be detected by the haptic modality. The obtained information has to be mathematically processed to obtain an estimate of the explored compliance. An analysis of the underlying perceptual process of information integration will contribute not only to a psychological understanding of compliance perception but also to the design and control of human system interfaces used to access artificial environments. A sound analysis requires a visual-haptic human-system-interface (HSI) with high accuracy in displaying mechanical environments and extensive experiments using psychophysical procedures. The study was published to a large extent in [70].

Information derived by different senses has to be integrated into a single percept of the manipulated object. Most research in this area has concentrated on the intersensory integration of a single object attribute (e.g. [92; 116–118]), e.g. position (see e.g. [92]). In addition, intrasensory integration, i.e. within one modality, has also been addressed (e.g. [119–121]): Research indicates qualitative differences when integrating intra- or intersensory information (e.g. [117]). However, studies concerning more complex variables such as compliance have, as yet, rarely been undertaken (see e.g. [107; 122–126]). When a person explores a compliant object, the haptic system provides information about arm and finger displacement along with signals as to force (kinesthesis), as well as information about the indentation of the fingertip (cutaneous or tactile information) (e.g. [127; 128]). The visual system adds information about the finger positions over time (see Figure 4.3.2). These inputs give rise to a percept of the object's compliance. When there is redundant information from both modalities, i.e. both arise from the same physical event, integration presumably occurs (e.g. [81]). However, the way in which haptic and visual information are combined to determine compliance, and whether additional cognitive factors influence integration, is not as yet known. Tan and colleagues found evidence that people tended to rely on force cues to discriminate levels of compliance, but were affected by position cues as well [122]. They computed the relative just-noticeable difference (JND),

$$\text{JND} = \frac{|S - S_{ref}|}{S_{ref}}, \tag{4.11}$$

when people discriminated compliance with varying pinch force over a constant displacement and when people discriminated force with varying pinch displacement. These JND were lower than when compliance was discriminated by applying force over randomly varying displacements. The increased JND seems to result from the perceptual system computing the perceived compliance when both force and position signals vary (see [122]). While more than $20 - 30\%$ differences in compliance appear necessary for discrimination to be successful (e.g. [122–126]), differences in finger distance of around 5% (e.g. [92; 129; 130]) and 8-10% difference in force information can be discriminated (e.g. [122; 123; 131]). On the other hand, visual perception of object compliance might primarily be based on visual position information, because there exist no visual receptors to decode visual force information (see Figure 4.3.2). Visual discrimination ability for position information has primarily been addressed by comparing length or size of objects: Deviations of around 3% can be detected (e.g. [132–134]). Similar results have been found when people discriminate line length (e.g. [135]). Furthermore, an influence of line orientation on the accuracy of discrimination has repeatedly been shown (e.g. [133–135]): A difference of approximately 10% between lengths of vertically oriented lines can be visually perceived (e.g. [135]). Force information, and thus object compliance, cannot be directly derived by the visual system. However, an estimate of an object's compliance could be obtained even by mere visual observation and thus involve expectancies about visual deformation of a compliant object ([136], see also e.g [137]). Some preliminary results on visual-haptic perception of compliance have been reported, showing that easily discriminable stimuli became harder to discriminate with decreasing visual reliability [78; 107]. Multisensory perception has often been reported to result in a more reliable percept than unimodal [92]. However, intermodal discrepancy can remain unnoticed (e.g. [78]) and result in a single, altered percept [85]. In a task where people were to compare two multi-modal (visual and haptic) stimuli, one with congruent compliance cues and another where one modality was discrepant, the detectability of the discrepant information was found to depend on the modality. Specifically, detection was lower when the visual modality was held constant across the stimuli and haptic cues were made discrepant, than in the reverse situation (for further details see [78]). In addition, the method of threshold assessment also influenced the overall magnitude of the detectable intermodal discrepancy. On this basis, it appears that a threshold assessment method which allows the participants directly to match perceived compliance could offer valuable clues to intermodal interaction. The present study was designed with this goal. Participants directly matched compliant stimuli signaled by vision, haptics, or both. On each trial, a standard stimulus was presented, and the participant adjusted a comparison stimulus to match it in compliance.

The contribution of this study is three-fold. First, it addressed the role of active exploration in a visual position discrimination task (i.e. of actively indenting a visually displayed cube without receiving force or relevant haptic position information). Based on the results by Tan et al showing increased thresholds when position varies [122], information about the process of indenting, when added to visual information about indentation, should add noise, resulting in an increase of the discrimination threshold (*Hypothesis 1*). Furthermore, we expected that when compliant cubes had to be discriminated by static indentation without active exploration, participants would rely on static position information, and therefore the value obtained should be comparable to the JND observed in visual position perception (see [135]). Secondly, the study investigated intermodal interactions, by having subjects adjust stimuli rendered haptically and visually. On these trials, the comparison stimulus

matched the standard stimulus in one modality (called the reference modality), and the task was to adjust the second modality of the comparison stimulus so that the stimuli matched completely. According to reported results on intermodal discrepancy thresholds (see [78]) and the phenomenon of visual dominance (e.g. [91]) an influence of reference modality is also expected with the present matching paradigm (*Hypothesis 2*): Smaller errors should occur with the visual modality matching and the haptic modality being the reference (which remains unaltered during the visual matching) than with the reverse. A third issue is how the thresholds obtained in the bimodal condition compare to unimodal thresholds. Keeping in mind that integration often occurs by combining information from more than one sense by weighting them e.g. according to their reliability (see [92]), a difference between unimodal and bimodal assessed threshold is expected (*Hypothesis 3*). Specifically, the threshold for a given comparison modality is expected to be greater in the bimodal condition than the unimodal, if the presence of the unaltered reference modality in the bimodal condition adds irrelevant cues to the judgment.
In Subsection 4.2.1 the method is explained. The results are described in Subsection 4.2.2. A discussion is presented Subsection 4.2.3.

4.2.1 Method

Presence System

A human system interface is used that provides visual and haptic feedback at high accuracy. The visual subsystem consists of a TFT screen mounted in the line of sight of the hand showing the visual virtual reality. The haptic subsystem consists of two SCARA robots providing two degrees of freedom each. The system interacts with index finger and thumb to allow the exploration of the compliant environment by gripping movements. Workspace is about 80 mm and maximal force is about 35 N. Position information is measured by angle encoders and force is measured by strain gauges. Haptically, the compliant environment is generated using an admittance control scheme. Visually, the environment is presented by a compliant cube with yellow spheres representing the finger tips. Within psychophysical experiments participants are able to adjust the compliance of the comparison stimulus with a potentiometer: A 360 degree turn of this rotary knob corresponds to a 50% change of compliance. The matched comparison compliance was recorded when the participant ended the trial by commanding the joystick. The system works under realtime conditions and is programmed by Matlab/Simulink. A detailed description of the HSI including hardware, software, and control structure is provided in Appendix A.1.

Participants

Twenty-three (23) students of the Technische Universität München and the Ludwigs-Maximilian-Universität München took part in this study and were paid for participation. Three (3) participants had to be excluded from further analysis because of missing data. The remaining 20 students (15 women and 5 men) were 24 years on average. All of them were right-handed and had normal or corrected to normal vision.

Stimuli

Cubes of 80 mm edge length with a standard compliance amounting to

$$S_{ref} = 0.85\,\mathrm{mm/N} \tag{4.12}$$

were chosen to represent the compliant stimuli and displayed by the HSI. Additional dummy cubes with a different standard compliance (see below) were included in testing, in order to prevent response perseveration. Virtual cubes were either displayed unimodally, i.e. visually or haptically, or bimodally, i.e. visual and haptically. No visual cues were given during the unimodal haptic presentation as well as no haptic cues during the unimodal visual presentation. An additional haptic standard compliance of 0.5 mm/N was chosen to represent the compliant cube of the dummy trials. Unimodal visual stimuli were presented in two ways, with and without active exploration. In the active condition, participants were allowed to indent the visual cube by moving the grasp device while no haptic force feedback as well as no haptic position information was given: The grippers could be closed, causing the cube to be visually indented to the pre-defined standard indentation depth (Va). In the passive condition, visual discrimination was done with watching the static indented cube (Vp). Additionally, a visual standard indentation depth of 13 mm was selected for the dummy trials.
Due to the method of threshold assessment there were two bimodal conditions: Either the haptic or the visual modality provided a basis for comparison while the other modality served as a constant reference. Therefore 5 modality conditions were realized: unimodal haptic (H), unimodal visual active (Va) as well as unimodal static passive (Vp) and the two bimodal conditions, visual reference with haptic comparisons (vH) as well as haptic reference with visual comparisons (hV).

Procedure

The method of adjustment was used to assess participant's performance (see Subsection 2.2.2 for a detailed explanation). Thresholds of each reference modality (H, Va, Vp, vH, hV) were assessed from above (down-series, i.e., the comparison started detectably above the standard) and below the threshold (up-series). Each series was repeated 5 times. Thresholds were assessed in three modality-specific blocks: Unimodal haptic (H), unimodal visual (Va, Vp) as well as bimodal (vH, hV) stimulus presentation. Five dummy trials were included in each block; order of series or modality condition within one block was randomized. Order of blocks was counterbalanced across participants using latin squares. All 65 threshold assessments had to be completed by each participant.

Unimodal Matching: As the psychophysical method of assessing the unimodal threshold, the method of adjustment was chosen, also called method of average error (see e.g. [113]): Participants tested a standard stimulus and were instructed to adjust the compliance of the comparison stimulus with a rotary knob until it matched the compliance of the standard stimulus. In order to minimize any tendency for stereotypic responses, each of the ten repeated adjustments started from variable start levels (80, 75, 70, 65, and 60%) and participants adjusted the compliance by rotating the knob in

one direction. The trial ended when indicated by the participant and the adjusted value was recorded. From the matching data, two measures are extracted. One is the mean of the matched comparisons, the PSE. The difference between the PSE and the standard, or the constant error, measures bias in responding. The second measure is the variability around the PSE, as measured by the population SD of the matched comparisons. As a measure of dispersion around the mean, this can be treated as a measure of the difference threshold. When normalized by the standard and multiplied by 100, it is treated here as a just noticeable difference or JND.
The visual task was either to match the visual indentation while being allowed to actively indent the visual displayed cube or to match the visually indented cube without active exploration. Although participants were instructed to adjust the visually perceived compliance the task was a visual position matching task with or without active exploration.

Bimodal Matching: The procedure of assessing the bimodal thresholds was adapted from the unimodal one. The participant was given a bimodal standard stimulus, which had congruent visual and haptic cues, and a comparison stimulus with a reference modality that matched the standard and an adjustable modality that was clearly discrepant. His or her task was to adjust the discrepant modality with the rotary knob until it matched the compliance of the standard. Again, each threshold assessment was repeated ten times, five times from above the standard compliance and five times from below while start levels of the adjusting modality were variable (160, 155, 150, 145, and 140%). Bimodal trials ended with the participant indicating that the adjusted level matched that of the perceived congruent one. The bimodal PSE and its standard deviation SD (threshold) were computed as described above.
Participants were seated in front of the HSI with their dominant hand grasping the device and looking nearly perpendicular at the screen. They were carefully instructed during a short training. Each trial started with testing the standard stimulus (unimodal matching) or the congruent bimodal stimulus (bimodal matching); switching between standard and comparison was possible. They were allowed to re-adjust the selected comparison stimulus, although they were instructed to decide as accurately as possible and by matching the standard stimulus with the least adjustments possible. In order to prevent visual or haptic matching, the robot arms were set to the start position and exploring the standard or comparison was only possible after a delay of 2 s. The delay between standard and comparison mode amounted to 1s (interstimulus-interval). The intertrial-interval lasted for 2 s.
After having completed all three blocks, participants were asked to fill in a questionnaire assessing their demographic data as well as their immersive tendency (subscale of a presence questionnaire by [114], translated by [71]) to control for any personal factors that might eventually influence the data.

Data Analysis

PSE and difference threshold (JND) (Standard deviation of the PSE) were computed. Results were descriptively analyzed. In order to test hypothesis 1, a t-test with dependent groups (exploration method active or passive) was computed on the dependent variables

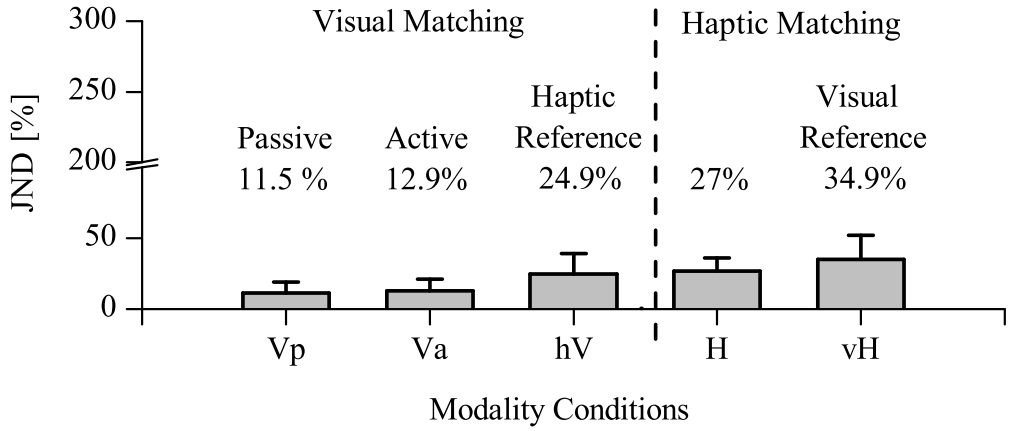


Figure 4.3: Results in discrimination performance: The largest JND was observed when participants had to haptically match a visual reference (vH). Haptically matching a haptic reference yielded higher performance (H). Visual matching a reference compliance (hV) yielded similar performance. In general, results were much smaller than in the experiment of Section 4.1 (Figures 4.2 and 4.1 are identically scaled for easy comparison) indicating a method effect. Error bars indicate one standard deviation.

(SD around the PSE as well as PSE) using the visual unimodal data only. Afterwards, hypothesis 2 was tested with a two-factorial ANOVA with repeated measurements on the two factors 'comparison modality' (active vision, haptics) and 'number of modalities' (uni-, bimodal).

4.2.2 Results

Immersive Tendency

Participants rated their immersive tendency on a 7-point-scale for each of two factors, *emotional involvement* and *degree of involvement*, which were computed for each participant. Average emotional involvement was 27.7 (standard deviation sd = 6.5) and average degree of involvement 17.7 (sd = 5.5); these values did not statistically significantly differ from those reported by Scheuchenpflug [71], indicating that the participants are a sample of a comparable population. No correlation between emotional involvement and the SD around the PSE could be observed. In order to determine the influence of preferred modality, participants answered an additional question concerning their preference for either the visual, the haptic or both modalities during the bimodal threshold assessment. Performing the task was rated to be easier when either the haptic (n=2), the visual (n=13) modality was changed or without preference (n=5). Neither immersive tendency nor performance (JND) was correlated with the individual preference rating.

Descriptive Analysis

Haptic matching: The PSE varied little across conditions, corresponding to an average constant error of CE, as normalized relative to the standard.

Visual matching: PSE and JND of visual indentation were assessed with active exploration or by mere visual observation. As can be seen from Figure 4.3 there is essentially no difference in PSE or JND between the two conditions. The JND amounted to 11.5% with passive and 12.9% with active exploration.

Haptic and bimodal matching: There are only small differences between the PSE depending on modality condition. However, differences in thresholds can be observed (see Figure 4.3): Bimodal matching results in an increase of the standard deviation compared to unimodal matching; furthermore, haptic unimodal matching results in a higher SD than visual. The JND amounted to 27.0% with unimodal haptic matching and increased to 34.9% with additional visual information (vH). The addition of haptic information resulted in an increase of the normalized standard deviation of active visual matching to 24.9% (hV).

Testing Hypotheses

As has already been descriptively observed, the difference in SD between active and static passive matching of visual indentation tested with a t-test was not statistically significant ($t(19) = 0.73, p = 0.475$). This indicates that active exploration has no effect on either PSE or JND.
A two-factor ANOVA tested the influence of 'comparison modality' (visual, haptic) and 'number of modalities' (unimodal, bimodal). The visual active condition was selected to represent the unimodal visual results. The main effect of 'comparison modality' was statistically significant ($F(1, 19) = 44.89, p < 0.05$; partial $\eta^2 = 0.703$). Haptically, adjusting the comparison stimulus yielded a higher threshold than visual matching. Additionally, 'number of modalities' significantly influenced the threshold ($F(1, 19) = 7.00, p < 0.05$; partial $\eta^2 = 0.269$) indicating an increase in threshold in the bimodal matching tasks. The interaction term turned out not to be statistically significant ($F(1, 19) = 1.18, p = 0.291$).

4.2.3 Discussion

Using the psychophysical method of adjustment, individuals' matching of compliance (and visual indentation) discrimination was measured when participants explored cubes haptically or visually, in unimodal and bimodal conditions. The effect of actively vs. passively perceiving a visually displayed compliant cube was also assessed; participants either tested the cube by visually indenting it or only by observing the (static) indentation depth. A difference between these testing methods was expected (*Hypothesis 1*): Static passive exploration should result in a JND comparable to those observed in visual position dis-

crimination (approximately 10% according to [135]); active exploration could increase this value due to added noise. The data showed, however, that about 11.5% deviation when passively matching the static standard indentation depth or compliance could be detected, and no statistical difference in the threshold obtained by active matching was found.
Visual dominance was expected. Therefore, when matching visual compliance with added haptic information, the threshold should be smaller than when displaying additional haptic information while adjusting the visual modality (*Hypothesis 2*). Also expected was an increase in threshold when matching the bimodal standard compared to matching the unimodal one, due to noise from the additional unaltered modality (*Hypothesis 3*). The results showed that haptic matching led to a higher threshold, and this effect did not differ, according to whether matching was unimodal or whether congruent visual information was also present. In fact, the presently obtained JND is similar to that reported in the literature for haptic matching of non-virtual stimuli [107; 122–126] and for visual position discrimination [135].
In addition, the present bimodal thresholds are relatively low compared to the results reported in [78]. This indicates again a difference of assessment method on the minimal detectable intermodal threshold (see also [113]).
Integration of visual as well as haptic information when exploring an objects' compliance is demanding. As has been shown for haptic perception of compliance, force and position information have to be combined and result in an increase of the JND [122]. On the other hand, only position information can be sensed by the visual system: We found that visual exploration method, i.e. static passive matching vs. active visually indenting, did not influence performance in terms of SD or JND. The observed JND is comparable to those reported for position discrimination in [135], amounting to approximately 12%. Additionally, we found that visual matching of indentation was superior in both unimodal and bimodal conditions; presumably this is attributable to not being forced to compute a compliance estimate as in haptic mactcing tasks (see [122]). Furthermore, bimodal matching resulted in an increase of thresholds compared to unimodal matching: When one modality remains constant while another is compared, the threshold increases relative to matching on the comparison modality alone. The size of this increase is comparable, whether the comparison modality is vision or haptic. This indicates that the additional alternate modality - although congruent - adds noise, and the added noise does not depend on the modality being matched.

4.3 Different Compliances and Multimodal Fusion

Mechanical environments are mainly perceived by processing position-based information (position, velocity, acceleration) and force information. However, humans do not have distinct modalities for each of these kinds of information. Considering visual-haptic perception, position-based information can be detected by both modalities while force information can only be detected by the haptic (kinesthetic) modality. Furthermore, the obtained information has to be computationally processed to obtain an estimate of the explored mechanical environment. An analysis of the underlying fusion process will contribute not only to psychophysical knowledge but also to the design and control of human system interfaces used to access artificial environments. A sound analysis requires a visual-haptic

HSI with high accuracy in displaying mechanical environments and extensive experiments using psychophysical procedures. The study was published to a large extent in [79].
Information from different senses is integrated to form a coherent perception of our environment. Several models of integrating multisensory information have been developed and can be subdivided into the following three model types (e.g. [88; 138]): A) Sensory penetration (one modality influences the other during unimodal stimulus processing), B) feedforward convergence and C) feedforward convergence with feedback-loop. Model B of integration is the most popular one (e.g. [88; 89; 138]): Information from different senses is separately processed and converges to a coherent percept at higher processing levels (e.g. [81]). However, modalities do not necessarily contribute equally to the bimodal percept: Visual dominance has often been reported, such that the visual modality captures e.g. the haptic modality (see [139]; for a review see [83; 90]). Different approaches argue that the relative contribution of each sense depends on modality appropriateness (e.g. [90]), effectiveness (e.g. [91]) or on the direction of attention (e.g. [94]). Another approach states that information from different senses are fused in such a way that the percept is the best possible estimate (e.g. [93]). Ernst & Banks point out that under a maximum-likelihood model this estimation depends on the reliability of each modality: The modality with the highest reliability contributes most to the bimodal percept. On the other hand, if the reliability of one modality is reduced, its relative contribution to the bimodal percept decreases (e.g. [92]). Considerable research on integration based on the maximum-likelihood theory has already been done [92; 116–118]. Most research on bimodal integration has concentrated on one sensory signal, e.g. position [92]. Additional studies have addressed the question of integrating information within a modality (e.g. [119–121]) and have also shown differences between integration within and between senses (e.g. [117]). However, investigating the integration of more complex variables that induce observers to integrate bimodal signals of different dimensions has not been undertaken yet. Haptic perception of compliance has already been shown to require the combination of force and position cues. This results in a loss of sensitivity (see e.g. [122–126]).
The contribution of this study is an analysis of visual-haptic compliance integration. According to *Hooke's law* compliance is expressed as the combination of position x and position information f

$$S = \frac{x}{f}. \tag{4.13}$$

Since position information is redundant (perceived by the visual and the haptic modality) the perception of compliance could include fusion of redundant information (see Subsection 2.2.1 for a mathematical derivation of the maximum likelihood estimation to fuse redundant information). As a basis of our investigation it was assumed that humans independently identify the presented compliance visually and haptically. In a second step they integrate the visual and the haptic compliance identification result to a single percept. Since humans do not have visual force sensors we assumed that people inferred force cues from the cube itself or from a force prior. We used 2AFC-tasks to record the psychometric functions of visual, haptic, and visual-haptic compliance perception. A test of whether integration of visually and haptically estimated compliance obeyed the maximum-likelihood model [92] was conducted. The results showed that participants had difficulties in identifying compliance from pure visual cues. The reliability of compliance perception decreased with increasing compliance. In our study we could not confirm the maximum-likelihood model.

In Subsection 4.3.1 the method is explained and embedded into the theory of bimodal fusion. The results are described in Subsection 4.2.2. A discussion is presented Subsection 4.2.3.

4.3.1 Method

Presence System

A human system interface is used that provides visual and haptic feedback at high accuracy. The visual subsystem consists of a TFT screen mounted in the line of sight of the hand showing the visual virtual reality. The haptic subsystem consists of two SCARA robots providing two degrees of freedom each. The system interacts with index finger and thumb to allow the exploration of the compliant environment by gripping movements. Workspace is about 80 mm and maximal force is about 35 N. Position information is measured by angle encoders and force is measured by strain gauges. Haptically, the compliant environment is generated using an admittance control scheme. Visually, the environment is presented by a compliant cube with yellow spheres representing the finger tips. During psychophysical experiments participants were able to insert their answers using a joystick that provides them with the different answer possibilities. The system works under realtime conditions and is programmed by Matlab/Simulink. A detailed description of the HSI including hardware, software, and control structure is provided in Appendix A.1.

Participants

Fifty-eight (54) students of the Technische Universität München and the Ludwigs-Maximilian-Universität, München took part in this study and were paid for participation. Participants were randomly assigned to one of the experimental groups according to modality in which stimuli were explored. Thirteen (13) participants had to be excluded from further analysis, because no psychometric function could be fitted to the data; most of theses participants had been assigned to group V (n=8). Eleven persons (6 men, 5 women) with an average age of 24 years explored the stimuli without vision (group H), and 11 participants (5 men, 6 women) with an average age of 26 years explored without haptic feedback (group V). The other two groups tested the cube bimodally: Twelve (12) persons (6 men, 6 women) with an average age of 24 years tested congruent stimuli (group HV), whereas 11 participants (4 men, 7 woman) with an average age of 25 years explored incongruent stimuli (group HVS). All participants were right-handed and had normal or corrected to normal vision.

Stimuli

All stimuli were virtual rendered compliant cubes of 80 mm edge length having a certain compliance s. They were either unimodally, i.e. visually or haptically, or bimodally, i.e. visual-haptically, displayed by the HSI. The bimodal stimuli were presented congruently, i.e. haptic and visual information were equal or as conflicting stimuli, i.e. haptic and

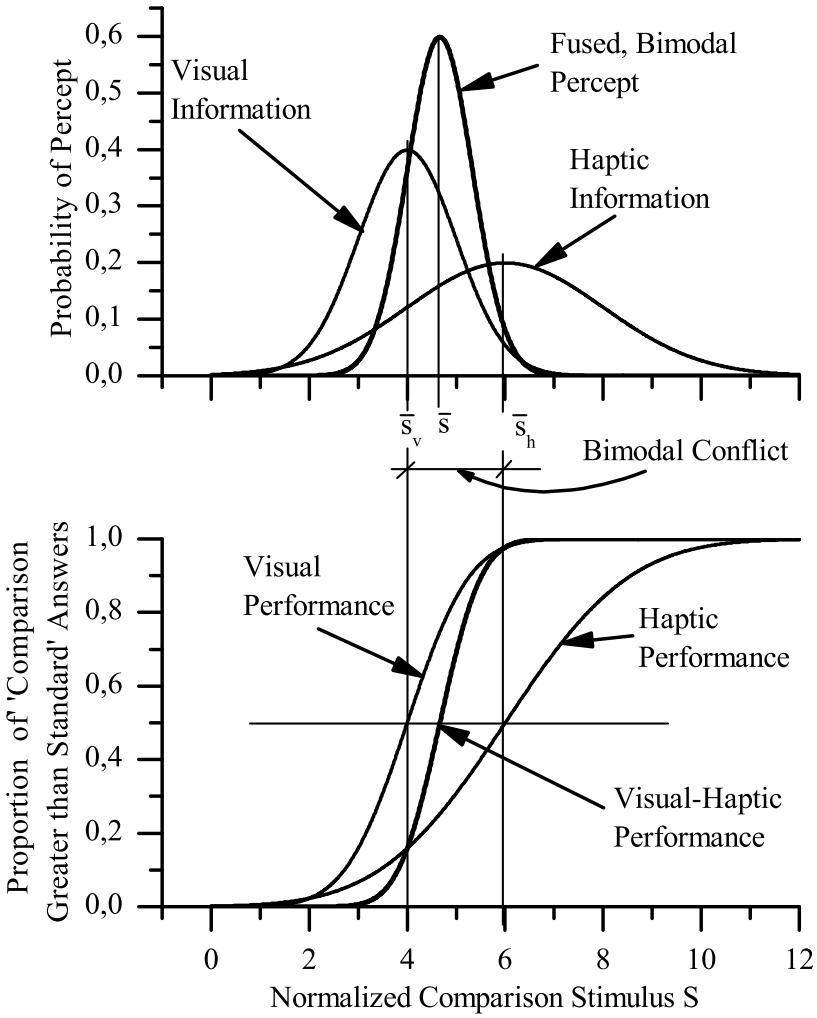


Figure 4.4: Perception of bimodal information: Minimizing the expectation value of the quadratic estimation error yields a combined percept with optimal variance (upper diagram). Characteristics of estimates can be obtained by differentiating the psychometric function that represent observer performance in stimulus discrimination (lower diagram).

visual information differed according to equation (4.19). No visual feedback was given during the unimodal haptic presentation nor was haptic feedback given in unimodal visual presentation (except the possibility to move the grasp device in order to move the visual fingertips).
Five standard stimuli were selected. The values were

$$S_{ref} = [0.2; 0.4; 0.9; 1.4; 2.1]\,\mathrm{mm/N}. \tag{4.14}$$

Seven comparison stimuli were selected to be linearly distributed above the standard stimuli. The comparison stimulus with the lowest compliance was the same as the standard stimulus. Since the detection threshold is known to be around 30% (see [70] for detailed information), the most compliant comparison stimulus was twice the standard value to assure that nearly all participants would detect the difference. The reason why the standard

stimuli were not centered in the middle of the comparison stimuli is due to the fact that it is difficult to haptically display relatively compliant environments accuratly by a human system interface, especially when the standard stimuli are less compliant themselves, which was the case here.
Additionally, bimodal conflicting stimuli were generated using a reference modality, which defined the standard stimulus. With the haptic modality as reference the (haptic, visual) combinations were $S = (0.2, 0.4)$, $S = (0.9, 1.6)$, $(S = 1.4, 2.1)$ in [mm/N]. With the visual modality as reference the (haptic, visual) combinations were $(S = 0.4, 0.2)$, $(S = 1.3, 0.9)$, $(S = 2.1, 1.4)$ in [mm/N].

Design and Procedure

Design: In case of visual-haptic perception the single modality estimates S_v, S_h (subscripts denote modalities) are fused to an integrated percept $\hat{S}$. The variable S represents the compliance of the object, i.e. the stimulus value. Maximal exploitation of the sensed redundant information can be computed using concepts of mathematical optimization. Thereby, a cost function is minimized with respect to the estimation error of the final percept $\hat{S}$. It is assumed that the fusion mechanism optimally combines the sensed information according to equation (2.13). As the single modality estimates, the fused estimate is assumed to be normally distributed (see upper diagram in Figure 4.4). The mean value of the optimally fused percept $\hat{S}$ is a weighted average of the mean values μ_v, μ_h of the single modality estimates

$$\mu^*_{vh} = \underbrace{\frac{\sigma_h^2}{\sigma_v^2 + \sigma_h^2}}_{k_v} \mu_v + \underbrace{\frac{\sigma_v^2}{\sigma_v^2 + \sigma_h^2}}_{k_h} \mu_h. \tag{4.15}$$

The standard deviation of the optimally fused percept $\hat{S}$ deviation is given by

$$\sigma^*_{vh} = \sqrt{\frac{\sigma_v^2 \sigma_h^2}{\sigma_v^2 + \sigma_h^2}}. \tag{4.16}$$

As depicted in the upper diagram of Figure 4.4 the combined percept yields a more reliable estimate represented by the highly peaked distribution. Furthermore, the mean value of the combined percept is a weighted average of the means of the single modality estimate. For complete derivation of the optimal solution and the Kalman filter refer to [26].
Assuming a Gaussian characteristic, distributions of single and bimodal estimates can be recorded using e.g. a two-alternative-forced-choice task (2AFC) (e.g. [113]). The result, a Gaussian psychometric function, expresses the observer's performance with respect to a categorical judgment (see lower diagram in Figure 4.4). In this type of task, the observer has to compare a stimulus S to a standard stimulus S_{ref}. If she/he has to decide for one of two alternatives (e.g. *Was the second stimulus more compliant than the first stimulus?*) the task is called *two-alternative-forced-choice* task (2AFC). See Subsection 2.2.2 for a detailed description of the method. The upper bound of the psychometric function (normally 1.0) represents the observer's performance with a comparison stimulus arbitrarily greater than the standard value, i.e. the observer answers that the comparison is greater on 100%

of the trials. The lower bound of this function (usually 0.0) represents the probability of answering that the comparison stimulus is less than the standard stimulus for an arbitarily lesser value. The psychometric function normally has the shape of a sigmoid and can be modeled by a cumulative Gaussian distribution. Hence, the characteristics of the perceptual estimate, i.e. the mean μ and the standard deviation σ, can be empirically estimated by taking the stimulus value and the slope at proportion 0.5. The slope at proportion 0.5 corresponds to $1/(\sigma\sqrt{2\pi})$. The inverse value of the quadrat of the standard deviation σ is also called *reliability* of the perceptual estimate.
The mean value μ of the distribution fit to the data is called *point of subjective equality* (PSE) as it defines the value where 50% of the participants decide for the first response alternative and 50% for the second:

$$\text{PSE} := \mu. \tag{4.17}$$

The constant error (CE) is defined as the difference between the PSE and the standard stimulus:

$$\text{CE} := \text{PSE} - S_{ref}. \tag{4.18}$$

A CE is only possible to compute when perceiving information from a single modality or when bimodal information is congruent, i.e. has no conflict.
The bimodal conflict is defined by the difference between the means of the single modality estimates

$$c := |\mu_v - \mu_h|. \tag{4.19}$$

If the conflict induced by the sensory estimates of the two modalities is too large the perceptual system might not be able to fuse incoming information. In this case the conflict is perceived because the final estimate will not result in a single coherent percept [92].
The relative difference between optimal PSE and empirically estimated PSE is defined by:

$$e_{\text{PSE}} := \left| \frac{\text{PSE}^*_{vh} - \text{PSE}_{vh}}{\text{PSE}^*_{vh}} \right|. \tag{4.20}$$

Respectively, the difference of the slopes is defined by

$$e_\sigma := \left| \frac{\sigma^*_{vh} - \sigma_{vh}}{\sigma^*_{vh}} \right|. \tag{4.21}$$

In both equations the subscript VH refers to an perceptual estimate from a bimodal (visual-haptic) stimulus condition.
The psychometric function for discriminating compliance was assessed within the following modality conditions: Unimodal haptic (H), unimodal visual (V), bimodal congruent and bimodal incongruent stimulus presentation with one modality being the reference modality (VH). Because testing time is relatively long using this method, modality was chosen to be a between-participants variable. Each of the above described five standard stimuli was recorded using the 2AFC-tasks. Standard compliance was a within-participants variable.

Experimental Paradigm: One trial consists of the sequential presentation of two stimuli: the standard and the comparison stimulus; the sequencing of standard stimulus and comparison stimulus differed randomly. Duration of each stimulus presentation was 2 s with an inter-stimulus interval of 2 s and an inter-trial interval of 4 s. Discrimination performance of each standard stimulus was assessed during one block, within which, each combination of this standard and the seven comparison stimuli was randomly presented 8 times.
The task was a 2AFC-task. That is, subjects had to compare both sequentially presented stimuli and to decide whether the second stimulus was more compliant than the first one. Because only stimuli that were more compliant than the standard stimulus had been chosen for this experiment and position of the standard stimulus was randomly varied the resulting psychometric function represents the proportion of correctly detected differences (further referred to as 'proportion correct'). In our design, this is equivalent to the proportion of stimuli rated 'more compliant' than the standard.
Participants were seated in front of the HSI with their dominant hand grasping the device and with direction of gaze essentially perpendicular to the screen. They were carefully instructed according to the group to which they were randomly assigned. A short training had to be completed, before the test session started.
At the end of the experiment, participants were asked to fill in questionnaires assessing their demographical data. Additionally, because the ability to be drawn into a book, film or VE, better known as *immersive tendency* (see [114]), has been assumed to contribute to performance within virtual environments. An additional 12-item questionnaire was included to control for that variable ([114], translated by [71]): Immersive tendency was assessed by answering 12 items building the two factors, *tendency to get emotionally involved* and *degree of involvement* (see [71]).

Data Analysis

For each participant and each standard stimulus a psychometric function was fitted using *psignifit* software, version 2.5.6, described in [29] and Matlab/Simulink. The following parameters were computed by fitting the psychometric function to the experimental data: Means (PSE_h, PSE_v, PSE_{vh}) and standard deviations (σ_h, σ_v, σ_{vh}) of the perceptual estimate were computed for each standard stimulus by taking stimulus level and slope at proportion 0.5. Additionally, the CE was computed.
Whether standard compliance affected discrimination performance and whether visual-haptic integration occurs were tested separately. The first question was tested with the CE and standard deviation of the PSE (which will further be referred to as η) and additionally with the slope of the psychometric function, σ.
The second question was tested by first estimating the optimal mean (PSE^*_{vh}) as well its standard deviation (σ^*_{vh}) according to equation (4.15) and equation (4.16). Results were descriptively analyzed and compared to the experimentally derived bimodal parameters (PSE_{vh} (congruent ans incongruent), σ_{vh} (congruent and incongruent)).

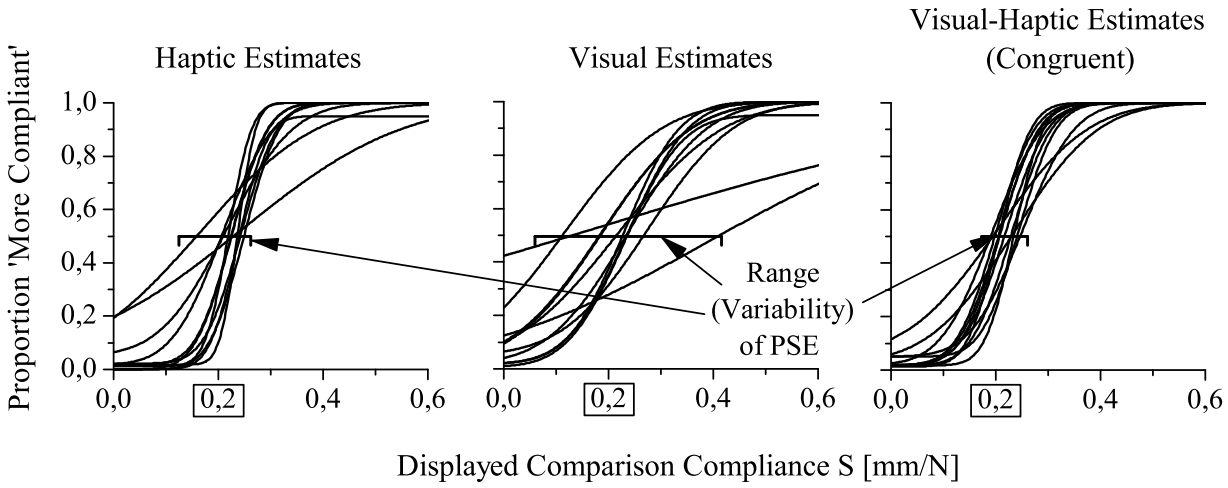


Figure 4.5: Unimodal and bimodal compliance estimates (referenced to the standard compliance $0.2\,\mathrm{N/mm}$): Haptic estimates and visual-haptic estimates show nearly identical slopes and variabilities. Visual estimates show higher variabilities and a lower slopes. (Due to the experimental design values smaller than the standard compliance were extrapolated by using the symmetric characteristic of the sigmoid.)

4.3.2 Results

Immersive Tendency

Both factors, emotional involvement and degree of involvement, were computed for each participant. Rating of emotional involvement ranged between 26.4 and 29.9, and degree of involvement between 17.5 and 21.0 within all 4 groups. These values did not statistically significantly differ from those reported by Scheuchenpflug [71], indicating that the participants are a good sample of the general population. No correlation between emotional involvement and individual thresholds and their standard deviation could be observed. This indicates that no immersive tendency influenced participants.
Participants of both bimodal groups rated which modality they attended to: Most attended to the haptic (group VH: n=9, group VHS: n=7), the visual (group VH: n=2, group VHS: n=1) or both modalities (group VH: n=1, group VHS: n=3).

Perception of Different Compliances

The first result reveals that participants had difficulty in discriminating compliance which was only presented visually: As can be seen from Figure 4.5 the variability of the CE, is small in group H and group VH, but larger in group V. This can also be seen from Figure 4.6 which shows the PSE and the standard deviation of the PSE, η, plotted against the standard compliance: Although the CE (difference between PSE and unity line) is small for nearly all standard stimuli, η from the visual estimate is higher for nearly all standard

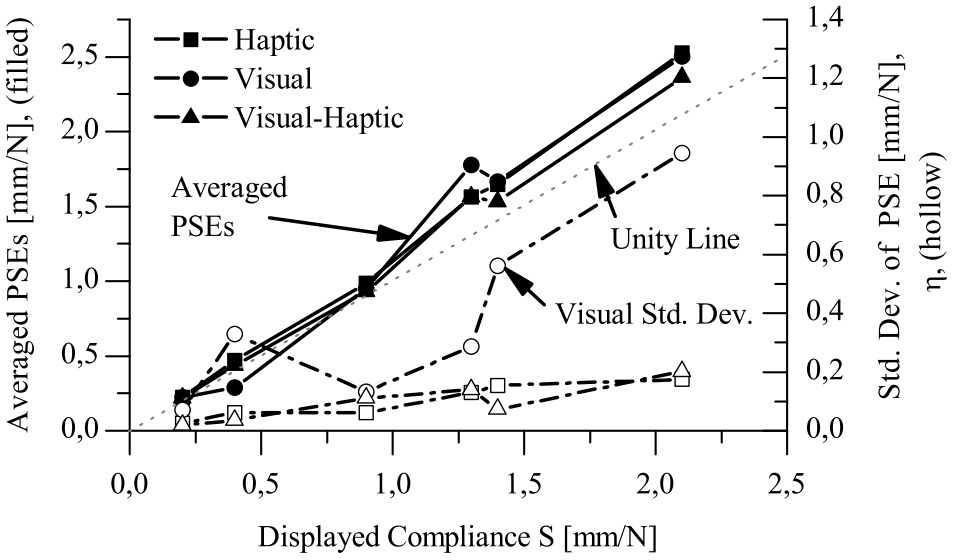


Figure 4.6: Unimodal and bimodal compliance estimates: PSEs (averaged across subjects, solid line) show low constant error independent of displayed compliance and modality. Standard deviation of PSEs (dashed line) across subjects increases with higher compliance. The visual standard deviation, in particular, is highly variable indicating the difficulty when estimating compliance visually.

stimuli. Consequently, the mean CE across all standard compliances in group H amounted to 0.2 N/mm (standard deviation $\eta = 0.1$ N/mm), 0.17 N/mm ($\eta = 0.08$ N/mm) in group VH, and in group V 0.24 N/mm ($\eta = 0.4$ N/mm).
Additionally, as has been reported above, several participants had difficulty in performing the task and had to be excluded form further analysis.
The second result shows that participants' ability to generally estimate compliances decreases with increasing compliance. It was tested whether the CE changed depending on the standard compliance (separately for all groups). With increasing standard compliance, there was an increase in CE in group H (Greenhouse Geisser corrected: $F(2.7, 26.8) = 26.64, p < 0.05$; partial $\eta^2 = 0.727$) as well as in group VH (Greenhouse Geisser corrected: $F(3.0, 32.6) = 20.18, p < 0.05$; partial $\eta^2 = 0.647$), but not in group V (Greenhouse Geisser corrected: $F(1.8, 18.4) = 0.98, p = 0.389$). Descriptively, this can be seen in the second part of Figure 4.6 relating the standard deviation of the PSEs, η, to the displayed compliance. It can be seen that all curves show a positive trend. Especially, η of the visual estimates showed a strong increase with higher standard compliances.
The consistently positive value of the CE is due to the fact that comparison compliance was always greater than the standard compliance.
The third result confirms the second result showing that participants' estimate of compliance became less reliable with decreasing compliance. As depicted in Figure 4.7, the slope of the psychometric function decreases with increasing standard compliance. That means the standard deviation σ of the perceptual estimate increases with larger compliances. Whether this influence is statistically significant was tested separately for all modalities. A statistically significant influence of standard compliance was found in group H (Greenhouse

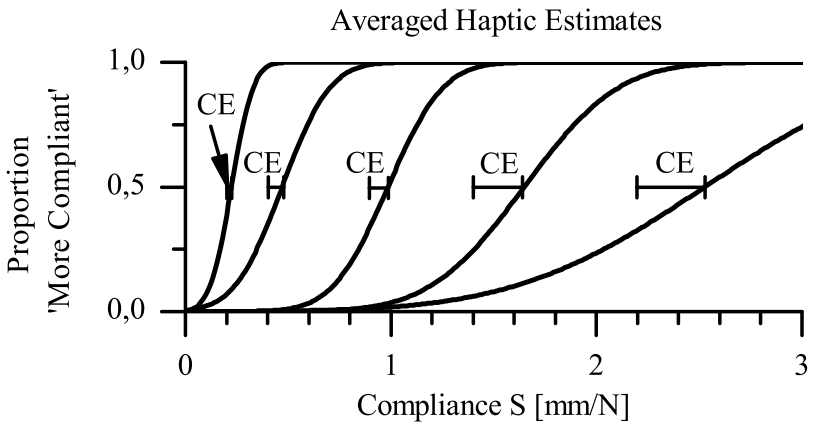


Figure 4.7: Psychometric functions of different compliance estimates (standard compliances indicated by dotted-lines): Haptic estimation of compliance became less reliable with increasing standard compliance. Hence, haptic perception of compliance fulfills the Weber-fraction assumption (constant relative JND). The same results were discovered for the bimodal groups. A comparable increase in standard deviation was not observed in the visual groups, where performance was low for all compliance values. The consistently positive value of the CE was due to the fact that participants were only presented comparison stimuli more compliant than the standard stimulus.

Geisser corrected: $F(1.4, 14.3) = 8.11, p < 0.05$; partial $\eta^2 = 0.448$), in the bimodal congruent presentation group VH (Greenhouse Geisser corrected: $F(2.4, 26.4) = 41.80, p < 0.05$; partial $\eta^2 = 0.792$), and in the bimodal conflict group VHS (Greenhouse Geisser corrected: $F(1.2, 11.6) = 10.43, p < 0.05$; partial $\eta^2 = 0.510$). In each case there was an increase in σ with increasing compliance. No influence of compliance was found to affect the σ in group V (Greenhouse Geisser corrected: $F(1.4, 14.3) = 8.11, p = 0.309$).

Fusion of Compliance Information

Evidence of perceptual fusion is obtained if the PSE of the combined percept (PSE_{vh}) is situated between the single modality PSEs. Furthermore, the reliability of the combined percept should be higher resulting in a smaller standard deviation σ_{vh} compared to the single modality parameters σ_h, σ_v.
The fourth result is that the data of this experiment did not confirm that the PSE of the combined bimodal percept was an average of single-modality percepts. In particular, there seemed to be no optimally weighted fusion of uniformly distributed information as predicted by the theory. It can be seen in Table 4.2 that the PSEs of the combined percepts were close to the modality that presented the less compliant information. (Optimal PSEs (PSE^*_{vh}) and relative error between optimal and estimated PSEs (e_{PSE}) were calculated according to equations (4.15),(4.20).) As indicated by the bold values in Table 4.2 participants seemed to be limited by the modality that presented the less compliant information when perceiving bimodal information. The other modality does not seem to add information since the combined percept cannot be denoted as being 'between' the PSEs of the

Standard		Estimates			Calculated	
S_h	S_v	PSE_h	PSE_v	PSE_{vh}	PSE^*_{vh}	$\mathrm{e}_{\mathrm{PSE}}$
A) congruent standard stimulus [mm/N]						
0.2	0.2	0.2	0.2	0.2	0.2	2.4%
0.9	0.9	1.0	1.0	0.9	1.0	5.7%
1.4	1.4	1.6	1.7	1.5	1.6	7.2%
B) conflicting standard stimulus - haptic reference [mm/N]						
0.2	0.4	**0.2**	0.3	**0.2**	0.2	4.6%
0.9	1.6	**1.0**	1.8	**1.0**	1.1	11.9%
1.4	2.1	**1.6**	2.5	**1.6**	1.9	17.4%
C) conflicting standard stimulus - visual reference [mm/N]						
0.4	0.2	0.5	**0.2**	**0.2**	0.3	32.1%
1.3	0.9	1.6	**1.0**	**1.1**	1.1	3.7%
2.1	1.4	2.5	**1.7**	**1.7**	2.1	23.0%

Table 4.2: Experimentally-estimated and theoretically-calculated, optimal PSEs (PSE^*_{vh} and $\mathrm{e}_{\mathrm{PSE}}$ calculated according to equations (4.15),(4.20); values rounded): PSEs of the combined percept (PSE_{vh}) were similar to the PSEs of the single-modality percepts (PSE_h, PSE_v) that relayed the more compliant information (connection emphasized by bold values). Optimal PSEs according to the maximum-likelihood model differed up to 32.2% from the empirically estimated PSEs of the combined percept (PSE_{vh}).

two modalities. Consequently, the PSEs of the incongruent estimates were smaller than the optimal values. Differences were between 4.6% and 32.1% (mean 15.5%). The PSEs of the congruent estimates did not differ notably from the optimal values, but this does not indicate fusion since optimal PSE and single-modality PSEs are the same in the congruent case.

The fifth result confirms the results above based upon the standard deviation of the perceptual estimates. The standard deviations σ that describe the inverse reliability of the estimates are listed in Table 4.3. (Optimal σ's (σ^*_{vh}) and relative error between optimal and estimated σ's (e_σ) were calculated according to equations (4.16),(4.21).) In the congruent case, participants' reliabilities of the bimodal percept seemed to be closely tied to the reliabilities of the haptic modality (which was reported in the first result to be more reliable than the visual modality when presenting compliance information). In the incongruent case participants' reliabilities of the bimodal percept seemed to be tied to the reliabilities of the modality that presented the less compliant information: In all cases where the reliabilities of the single modality percept differ more than 0.1 mm/N participants' combined estimate was close or equal to the reliabilities of the smaller single-modal percept. Additionally, combined estimates for congruent and incongruent standard stimuli (σ_{vh}) were larger than the values σ^*_{vh} based on the maximum likelihood theory. According to maximum-likelihood model the slope σ of the combined percept has to be smaller than the smallest slope of the two single-modality percepts indicating that information is added by both modalities and reliability increases. However, in this study, the slopes of the combined estimates σ_{vh} are not smaller than the smallest slope of single modality estimates (only in one case). This indicates that there seemed to be no fusion of the bimodal information.

Stimuli		Estimates			Calculated			
s_h	s_v	σ_h	σ_v	σ_{vh}	σ^*_{vh}	e_σ	k_h	k_v
A) congruent standard stimulus [mm/N]								
0.2	0.2	**0.1**	0.2	**0.1**	0.1	7.4%	0.8	0.2
0.9	0.9	**0.2**	0.5	**0.2**	0.2	6.7%	0.8	0.2
1.4	1.4	**0.4**	0.8	**0.4**	0.3	28.2%	0.8	0.2
B) conflicting standard stimulus - haptic reference [mm/N]								
0.2	0.4	**0.1**	0.4	**0.1**	0.1	28.2%	1.0	0.0
0.9	1.6	**0.2**	0.6	**0.3**	0.2	27.2%	0.9	0.1
1.4	2.1	0.4	0.5	0.5	0.3	82.4%	0.6	0.4
C) conflicting standard stimulus - visual reference [mm/N]								
0.4	0.2	0.2	0.2	0.1	0.1	33.4%	0.5	0.5
1.3	0.9	0.7	**0.5**	**0.5**	0.4	33.4%	0.3	0.7
2.1	1.4	0.7	0.8	0.9	0.6	71.1%	0.5	0.5

Table 4.3: Experimentally-estimated and theoretically-calculated slopes (σ^*_{vh} and e_σ calculated according to equations (4.16),(4.21); values rounded): In the congruent case, participants' reliabilities of the bimodal percept seemed to be closely tied to the reliabilities of the haptic modality. In the incongruent case participants' reliabilities of the bimodal percept seemed to be tied to the reliabilities of the modality that presented the less compliant information. All relations are marked bold. Optimal σ's according to the maximum-likelihood model differed up to 82.4% (mean 32.3%) from the empirically estimated σ's of the combined percept (σ_{vh}).

4.3.3 Discussion

To summarize the results, we found that discrimination performance increases with decreasing compliance. Furthermore, the combined percept of visual and haptic compliances could not be predicted by a maximum-likelihood estimation of uniformly distributed information. Rather, participants' percept was close to the modality that presented the less compliant stimulus. These points are expanded on in the following.

Discrimination of compliance with our visual-haptic human system interface was found to produce reliable psychometric functions. The estimate of the mean and standard deviation of the perceptual estimate of compliance indicated that people were less able to discriminate compliance, and perceived compliance became more variable, as the baseline level of compliance increased. This was true for stimuli presented haptically and bimodally; performance with purely visual stimuli was relatively poor at all compliance levels. Thus overall, these results indicate that haptic cues to compliance, as simulated by our interface, are not only effective but necessary.

The more accurate performance in detecting rigid (less compliant) environments than detecting soft (compliant) environments might be explained by the fact that exploration of rigid environments leads to a smaller change in finger position than is the case for soft environments. Consequently, the position estimate in compliant environments might be more reliable, with the result that the visual and haptic compliance estimates are more reliable as well.

Our experiments not only tested the efficacy of the interface for conveying compliance,

but they also provided tests of the hypothesis that visual and haptic cues would be fused. Contradicting this hypothesis, the results indicated that fusion did not occur, and in particular, the combined percept of visual and haptic compliances could not be predicted by maximum-likelihood of uniformly distributed compliance information. Instead, participants' percept was close to the modality that presented the less compliant stimulus.
The PSEs of the incongruent estimates were smaller than the values predicted from optimal (i.e., maximum-likelihood) integration, instead lying close to the values of the modality presenting the more compliant level of the stimulus. The estimated reliabilities of the percept were greater than optimal, and particularly for haptic stimuli, tended to be closer to the haptic value (which was the less compliant).
The results of the bimodal conditions, in short, clearly violate optimal fusion as described by the maximum-likelihood model. Instead, performance appears to be limited by the discriminability and reliability of the less compliant component of an incongruent bimodal stimulus. Why should this occur? One answer to this question can be derived from our data, which clearly indicate that the lower the compliance being perceive, the more discriminable and reliable the perceptual estimate. If two independent stimulus estimates are produced for a discrepant stimulus, with some signal of discriminability and reliability, a higher-order processor may use the better estimate in making the discrimination judgment. This would lead the lower-compliance stimulus level to dominate judgments, as we observe.
Further research is needed to test these hypotheses, but our results clearly indicate a departure from optimal fusion and hence advance our knowledge of compliance judgments using a bimodal interface.

4.4 Conclusion

Summary

Three explorative studies were presented about visual-haptic compliance perception. All studies assessed estimation performance in terms of the difference threshold. However, results were complementary rather than redundant as different kinds of difference thresholds were targeted and different methods were used to obtain the results.
In the first study, Section 4.1, concurrent vs. sequential estimation performance was measured. The focus was on the crossmodal difference threshold, i.e. one modality was held as reference and the other modality had to be matched by the participant as close as possible. An adaptive staircase method was implemented to adjust the target modality toward the reference modality. As main result, participants' performance was lower when they had to perform the matching concurrently. Matching haptic to visual compliance yielded JND = 128% and visual to haptic matching yielded JND = 85%. Performance rised in case of sequential matchings, to a JND = 65% when haptic had be matched to vision and JND = 55% otherwise.
The second study, Section 4.2, focused on sequential crossmodal matching. In contrast to the first study the method of adjustment was used and the participants could actively ad-

just the reference modality. Unimodal matching as well as bimodal matching was analyzed and passive vs. active exploration. The main result indicated that participants are more precise when actively adjusting the conflict using the method of adjustment as if passively using the staircase method (first study). Matching haptic to visual compliance yielded JND = 34.9% and visual to haptic matching yielded JND = 24.9%.
The third study, Section 4.3, concentrated on the fusion of visual and haptic information when perceiving object compliance. Psychometric functions were recorded using 2AFC-tasks. Congruent and incongruent stimuli were displayed to analyze whether and how participants integrated bimodal information to a coherent percept. The main results indicated that participants had difficulties in estimating compliance that was only presented visually. Furthermore, discrimination performance decreased with increasing compliance. No evidence of bimodal fusion could be secured.

Scientific Contribution

As detailed discussions of the scientifical outcome are presented locally this part focuses on results that emerge from comparisons between the different studies.
1.) The first general result becomes salient when comparing the results of the first study with the results of the second study. Crossmodal comparisons of compliance information seem to be more difficult if haptic is matched to vision than if vision is matched to haptic information. As different methods were used in both studies to obtain the result, the result seems to hold independent of the method. Hence, the direction of the matching is a very robust feature of visual haptic compliance perception. See also Table 4.4.
2.) The second general result also emerges between the first and the second study. The JNDs obtained by staircase method are significantly higher than the JNDs obtained by method of adjustment. Using the staircase method, participants had to compare two stimuli after a limited, short exposure. Using the method of adjustment, participants were able to actively go back and forth with the stimulus to be matched for a longer exploration phase. The different procedures account for the different performances. Hence, the JND is a method-dependent statistics of human estimation performance. See also Table 4.4.
3.) As verified in the third study visual compliance estimation is hardly possible. This is due to the circumstance that humans lack a visual force sensor. As for the perception of compliance the combination of position and force is mandatory, the perception has to fail if either force or position information is missing. Although this is an obvious result, it did not came out in the first two studies. In the crossmodal matching tasks observers were able to come up with a reliable matching of haptic to visual compliance. In the unimodal non-crossmodal tasks of the second study participants showed even higher performance than in a haptic matching task. However, what participants did in both cases was actually position matching. In the crossmodal tasks, when a haptic stimulus had to be matched to a visual compliance stimulus, participants performed position matching in presence of force information that was not useful and hence, had to be filtered out. The presence of disturbing information accounts for the low performances. In the visual compliance matching tasks of the second study no distracting information was present and performance was very high (11.5%, 12.9%), comparable to position discrimination performance known from literature. Besides the result that visual compliance perception is hardly possible, one has to admit that, at least, a percept with a small reliability (high JND) was possible with

Crossmodal Matching Performance		
Reference (Target) Modality	**Method**	
	Staircase	**Adjustment**
Vision	JND= 65%	JND= 35%
Haptic	JND= 55%	JND= 25%

Table 4.4: Comparison between first and second study: The direction of crossmodal matching is a robust feature of visual-haptic compliance perception since both methods recorded a higher performance when visual stimuli had to be matched to haptic stimuli. On the other hand, the JND is a method-dependent feature of visual-haptic compliance perception, since the JNDs measured by the method of adjustment are significantly smaller for both conditions.

the used presence system. Since humans do not a have a visual force sensor, participants obviously inferred the visual force information from the way the compliant object was indented or they directly inferred the compliance by the object's color or surface.

Open Problems

As detailed discussions of the open problems are presented locally this part focuses on problems that emerged from inconsistent results between the conducted studies and on problems that emerged in all studies again.
In contrast to the first and second study, in the third study the performance of compliance perception was identified to be depending on the compliance value. Estimation performance decreased with increasing compliance. Why this did not came out in the in the studies before, especially in the first study remains unsolved. It is suggested to perform an additional study to clarify this characteristics of compliance perception.
In all studies the influence of the visual modality could not be clearly identified. Especially in the last study, which deliberately targeted the fusion taking place in visual-haptic compliance estimation, no clear result about the role of the visual information could be obtained. However, the clear description provided will help to setup more specific experiments.
Another problem is strongly related to the latter problem. No theoretical model could be set up describing visual-haptic compliance perception in mathematical detail. The results of the three studies show that human visual-haptic perception is far from being easy to analyze. Results are task-dependent and method-dependent. A theoretical model that would predict all these results would be the final step in visual-haptic compliance perception. However, by now, not even a simple model exists that is able to predict a small fragment of the realm of compliance perception. Even more, the interdisciplinary character make the analysis of compliance perception hard labor. The human system interface has to be very accurate to control the stimulus according to the needs of the experimenter. Secondly, the psychophysical method has to be implemented soundly. Methodical mistakes

might be covered up in the data analysis and therefore leading to contradictory results. Experiments are exhaustive for the participants and the results might be skewed due to fatigue or non-concentration. However, the explorative studies presented in this chapter give several starting points for the development of theoretical, mathematical models of compliance perception. In the next chapter a theoretical model will be analyzed that aims to clarify the processing of force and velocity information within visual-haptic compliance perception.

5 Visual-Haptic Perception of Compliance: A Theoretic Model

...if you want understanding,
look at theories.

R. Dorfman, *The Price System* (1964)

The mechanical properties of an environment (like compliance/stiffness, density/damping, or center of mass/inertia) are perceived during active manipulation primarily by processing position-based information (position, velocity, acceleration) and force cues. Humans do not have specialized sensory channels for the two types of information. Position-based cues are provided by both visual and haptic modalities, for example, by the visible displacement of the fingers pressing into a surface and by kinesthetic cues to the change in joint angle. Force cues are primarily haptic, in the form of kinesthetically sensed resistance, although visual cues to force may also be available, for example, by visible surface deformation. In [79] evidence was given for a compliance estimate based entirely on vision; hence, it seems important to consider visually cued force. In short, multi-sensory cues to both position and force appear to be available to obtain a final estimate of the environment's mechanical properties like surface compliance. A model that accurately captures the perceptual process by which these cues are utilized will contribute not only to psychophysical knowledge, but also to the design and control of human system interfaces used to access virtual or remote environments.

Developing such a model requires an environment in which mechanical interactions can be precisely controlled and psychophysical experiments can be conducted. For this purpose a customized human system interface (HSI) is invaluable.

Psychophysical research into visual-haptic perceptual interactions have often shown that both sources of information influence perceptual outcomes, but with differential impact, e.g. [81; 88; 89; 138]. Various approaches argue that the relative contribution of each channel depends on modality appropriateness [90], effectiveness [91], or on the direction of attention [94]. In [93] it was proposed that redundant information (information from different sources about the same property) can be fused according to a maximum-likelihood estimation. Thereby, the final percept is a weighted average of the single modalities, where weightings are defined by the reliability of the modalities. The outcome is then an optimal fusion of the multimodal information that minimizes variance, i.e. maximizes reliability of the estimated property value. The maximum-likelihood model has been broadly invoked to explain the fusion of redundant information, e.g. [92; 116–121], although it does not apply in every situation [117].

Most research on bimodal perception has focused on *integration* processes estimated using only redundant estimates of the same type of information; for example, how the width of an edge is signaled by simultaneously seeing and pinching it. The estimation of properties that involve the *combination* of different, non-redundant information has generally not

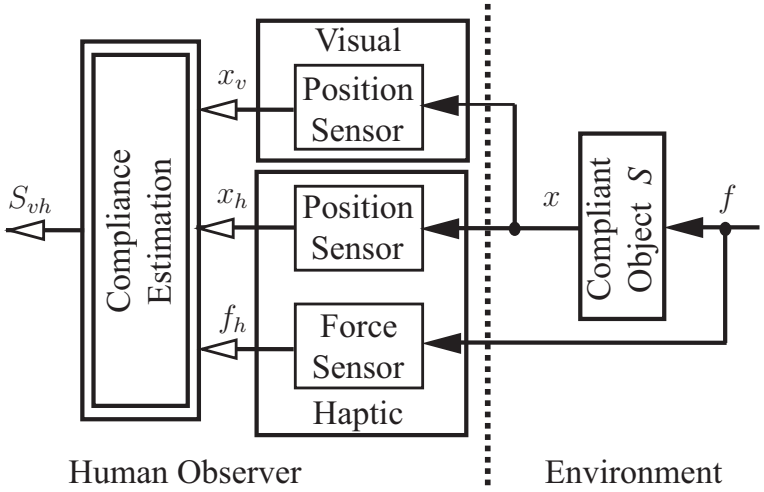


Figure 5.1: Visual-haptic perception of compliance S: Position information x is measured by the visual and the haptic modality. Force information f is only provided by the haptic modality. The final percept S_{vh} can be obtained by combination and fusion processes. (Filled arrows depict physical interaction, hollow arrows depict signal processing.)

been addressed in detail. Perceptual estimation of compliance is one of the latter cases. According to *Hooke's law*, compliance S is expressed as the division of position x by force f

$$S = \frac{x}{f}. \tag{5.1}$$

Hence, perception of compliance requires processing the combination of both force and position information; the two variables are non-redundant.

Various studies have investigated people's ability to perceive compliance when only haptic information is available (e.g. [124; 140–143]). These generally conclude that performance in estimating compliance is considerably worse than in estimating position or force from haptic cues. Just noticeable differences (JNDs) ranging from 15% to 35% have been reported for unimodal estimation of compliance. In these experiments, real compliant objects, like rubber specimens, or virtual objects rendered using a haptic HSI, are manipulated by participants in order to generate force and position estimates that are combined to produce an estimate of compliance. Studies of this type have tended to preclude vision; multimodal studies, which involve the visual modality as well as the haptic one, are rare. The beneficial influence of tactile (cutaneous) information in kinesthetic-tactile compliance perception was analyzed in [144].

Note that whereas force and position data offer non-redundant compliance cues, according to the equation above, it is still the case that in principle, the haptic and visual modalities could offer redundant cues to compliance. This will particularly be true when vision carries information about force as well as position. The first study that analyzed visual-haptic compliance estimation reported a significant influence of visual information on the final percept [145]. However, other studies suggest that the conjoint processing of visual and haptic cues to compliance is relatively difficult. In [146] crossmodal matching performance was analyzed, and it was concluded that matching a stimulus in which compliance was signaled simultaneously by vision and haptics was more difficult than matching on each modality in sequence. In [70] sequential comparisons were used to conclude that there

was no advantage in matching stimulus compliance when vision was added to haptic cues, relative to haptic cues alone. The role of the redundant information added by vision was further investigated in a succeeding study, which directly targeted the issue of whether visual and haptic cues to compliance were integrated according to the maximum-likelihood model [79]. Results showed a significant influence of visual information; however, it could not be confirmed that this influence was based on the fusion of bimodal information.
The contribution of this Chapter is two-fold. First, we offer a theoretical process model of visual-haptic compliance perception (double-lined system in Fig. 5.1). Mathematically, the final visual-haptic compliance percept can be modeled as a serial connection of a combination process and an integration process. Second, we describe the results of two psychophysical experiments that were conducted to determine the actual structure of the model. The first experiment tested whether combination processes occur within-modal or crossmodal, i.e. before or after a possible integration process. The second experiment further refined the model by indicating the role of the visual information. Both experiments used the method of constant stimuli to construct psychometric functions. The results clearly indicate that humans rely on the haptic compliance estimate. Visual information about compliance is not integrated with haptic information, although it consistently biases participants' judgments of compliance.
In Subsection 5.1 the process model is described in it two basic entities. In Subsection 5.2 we describe the first experiment, where we compare the two models by analyzing whether visual-haptic fusion occurs before or after the combination of force and position information. In Subsection 5.3, describe the second experiment in which we further investigate the role of the visual information when compliance is signaled both visually and haptically. An overall Conclusion is presented in Subsection 5.4

5.1 Models for Visual-Haptic Perception of Compliance

The combination and fusion of visual and haptic information for the perception of objects' compliance can be mathematically described in two ways, as depicted in Fig. 5.2.
Process Model 1 assumes that in a first step, each modality provides a compliance estimate by combining non-redundant position information x and force information f. (As noted above, there is evidence that force information is cued at least to some extent by vision [79].) In a second step, the visual compliance estimate S_v and the haptic compliance estimate S_h are fused to determine final percept S_{vh}.
Process Model 2 assumes that, in a first step, redundant position information is fused to produce a composite position estimate x_{vh} and redundant force information is fused to produce a force estimate f_{vh}. In a second step position and force estimates are combined to create the final a percept of compliance S_{vh}.
Estimation from independent, redundant (postion-, force-, compliance) sources, called *integration*, can basically happen by three different ways. The single modality estimates are assumed Gaussian and denoted by the following random variables

$$\begin{aligned} X_v &\sim N(\mu_v, \sigma_v), \\ X_h &\sim N(\mu_h, \sigma_h), \end{aligned} \tag{5.2}$$

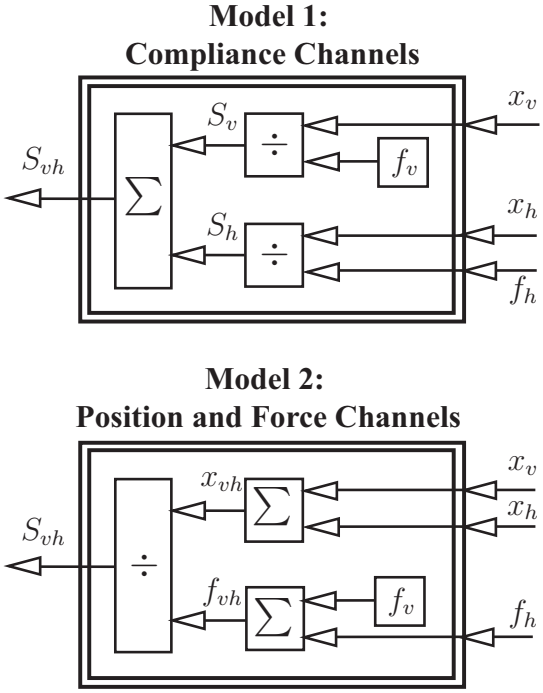


Figure 5.2: Visual-haptic compliance perception: The models assume that non-redundant information about position and force is *combined* to estimate compliance, and that any redundant parameter estimates are *integrated*. The perceptual process can then mathematically be described in two ways. In Process Model 1 the integration process (indicated by $\sum$) is preceded by the combination process (indicated by $\div$). In Process Model 2 integration processes are succeeded by a combination process.

where μ is the mean value and σ the standard deviation. The final percept is captured by the random variable

$$X_{vh} \sim f(\mu_{vh}, \sigma_{vh}). \tag{5.3}$$

It has not necessarily to be Gaussian-distributed.
The reliability of an estimate or a percept is

$$r = \frac{1}{\sigma^2}, \tag{5.4}$$

Fusion of multimodal information occurred if the reliability of the multimodal estimate is larger than the reliability of the most reliable single modality estimate. *Confusion* occurred if the reliability of the multimodal estimate is lower.
The trivial way to deal with two sources of redundant information is to disregard one source and concentrate on the other. Hence, the perceptual system concentrates on most reliable modality and the bimodal percept is equal to the estimate of this modality

$$\boxed{X_{vh} = X_x,} \tag{5.5}$$

where x is either h or v representing the most reliable modality. The integration process can capture neither fusion nor confusion.
The second basic way to integrate redundant information is to generate a weighted average of the different random variables defined by the different estimates. According to this

process the perceptual system would weight and convolute the different distributions. For a bimodal percept the mathematical process is described by

$$\boxed{\begin{array}{lcl} X_{vh} & = & w_v X_v + w_h X_h, \\ w_v + w_h & = & 1. \end{array}} \tag{5.6}$$

The mean of the joint distribution is

$$\mu_{vh} = w_v \mu_v + w_h \mu_h. \tag{5.7}$$

And the variance becomes

$$\sigma_{vh}^2 = (w_v \sigma_v)^2 + (w_h \sigma_h)^2. \tag{5.8}$$

This process can capture no fusion, fusion and confusion. Calculating the weights w_v, w_h to minimize the variance according to

$$\frac{\mathrm{d}\sigma_{vh}^2}{\mathrm{d}w_v} = 0 \tag{5.9}$$

yields a final estimate with minimal variance, i.e. maximal reliability

$$\sigma_{vh}^2 = \frac{\sigma_v^2 \sigma_h^2}{\sigma_v^2 + \sigma_h^2}. \tag{5.10}$$

Optimal fusion is called maximum-likelihood-estimation (MLE) and was confirmed in e.g. [27] for the fusion of redundant position information. The resulting distribution from an optimal fusion process is depicted in Fig. 5.3. See also Subsection 2.2.1 for more information.

A third basic way to integrate redundant information is to randomly draw with a certain probability from either one of the available sources. Hence, the distribution of the final percept is equal to the the weighted sum of the distribution values of the two modality estimates

$$\boxed{\begin{array}{lcl} X_{vh} \sim f & = & w_v N(\mu_v, \sigma_v) + w_v N(\mu_h, \sigma_h), \\ w_v + w_h & = & 1. \end{array}} \tag{5.11}$$

The mean of the joint distribution is again a weighted average

$$\mu_{vh} = w_v \mu_v + w_h \mu_h. \tag{5.12}$$

The variance is based on the variances of the single modality estimates and on the conflict between the two estimates, i.e on the difference between the mean values

$$\sigma_{vh}^2 = w_v \sigma_v^2 + w_h \sigma_h^2 + w_h w_h (\mu_v - \mu_h)^2 \tag{5.13}$$

The process can capture no fusion and confusion. The resulting distribution is depicted in Fig. 5.3.

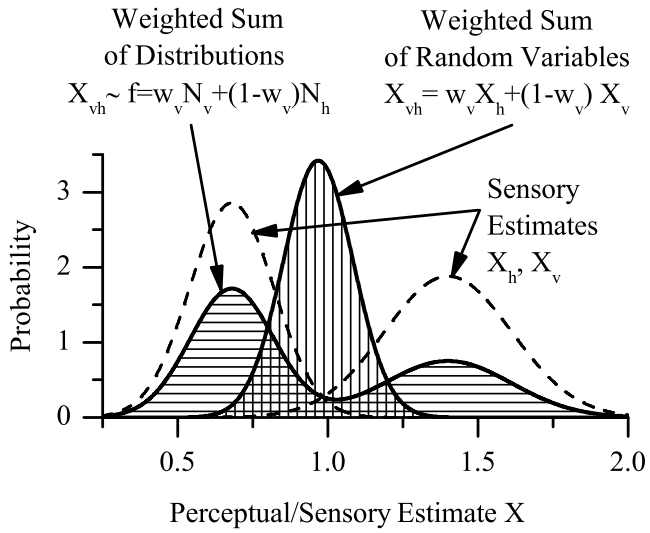


Figure 5.3: Non-trivial integration processes of two sensory estimates: The weighted sum of random variables leads to a Gaussian shape since the estimates are assumed Gaussian, too. Thereby, the perceptual system integrates the two estimates by performing a convolution of the underlying weighted distributions. The weighted sum of distributions does not yield a Gaussian distribution. Thereby, the perceptual system draws with a certain probability of the two estimates. Both processes lead to the same mean value.

5.2 Experiment: First Integration, then Combination?

5.2.1 Hypotheses

To determine which of these process models applies for visual-haptic compliance perception an experiment was conducted with two conditions. In the first condition, called active, participants made a perceptual decision about the compliance of a virtual cube that was explored by pushing the right thumb into its surface. The resulting active compliance information I_a was (where subscripts v, h index the modalities and x, f indicate position and force)

$$I_a = (x_v, x_h, f_v, f_h). \tag{5.14}$$

In the second condition, called resistive, participants made a perceptual decision about the compliance of a virtual cube that was pushed against the right thumb by the robot; participants were told to hold the thumb as stable as possible. The resulting resistive information I_r was then

$$I_r = (x_v, f_h). \tag{5.15}$$

If Process Model 1 applies, and perceived compliance is obtained first by unimodal estimates that are then fused, a compliance percept is only possible for the active condition (5.14) but impossible for the resistive condition (5.15) since neither a haptic-only estimate S_h nor a visual-only estimate S_v is possible. On the other hand, if Process Model 2 applies a reliable compliance percept is possible for both the active (5.14) and the resistive condition (5.15), since the perceptual system combines force and position information ir-

respective of the modality they come from.
Since it was difficult to fully eliminate x_h and f_v in the experiment, a percept was possible in both conditions for both models but presumably differed in the reliability. Hence, we predicted the JND of the percepts, which inversely correlates with the reliability. The hypotheses for the experiment were then as follows.
If Process Model 1 holds, the JND in condition (5.15) should be larger than in condition (5.14)

$$\mathrm{JND}_r \overset{!}{>} \mathrm{JND}_a \tag{5.16}$$

If Model 2 hold the JNDs for both conditions should not differ much.

$$\mathrm{JND}_r \overset{!}{\approx} \mathrm{JND}_a \tag{5.17}$$

The latter prediction results from the fact that the additional information (x_h, f_v) provided in the active condition (5.14) is much less reliable than the information (x_v, f_h) provided in both conditions. Hence, integrating the additional information should not lead to a significant increase of the reliability in the active condition compared to the resistive condition (5.15).

5.2.2 Method

Presence System

A human system interface is used that provides visual and haptic feedback at high accuracy. The visual subsystem consists of a TFT screen mounted in the line of sight of the hand showing the visual virtual reality. The haptic subsystem consists of two SCARA robots providing two degrees of freedom each. The system interacts with index finger and thumb to allow the exploration of the compliant environment by gripping movements. Workspace is about 80 mm and maximal force is about 35 N. Position information is measured by angle encoders and force is measured by strain gauges. Haptically, the compliant environment is generated using an admittance control scheme. Visually, the environment is presented by a compliant cube with yellow spheres representing the finger tips. During psychophysical experiments participants were able to insert their answers using a joystick that provides them with the different answer possibilities. The system works under real-time conditions and is programmed by Matlab/Simulink. A detailed description of the HSI including hardware, software, and control structure is provided in Appendix A.1.

Participants

Thirty (30) students of the Technische Universität München took part in this study and were paid for participation. Data collected from the first ten (10) students was used for test sessions only and to adjust the HSI; those data are not reported here. The results were generated with participants eleven (11) to 30. Participants had an average age of 24 years. All participants were naïve to the purpose of the study, but most of them were experienced psychophysical observers. All participants were run individually and completed the experiment in a two-hour sessions including training and pauses.

Written, informed consent was obtained from the participants prior to their inclusion in the study and the rights of the participants were protected according to the 1964 Declaration of Helsinki.

Stimuli

All stimuli were virtual rendered compliant cubes of 80 mm edge length. The compliance of the standard stimulus was

$$S_{ref} = 1.6\,\text{mm/N}. \tag{5.18}$$

Eight (8) congruent bimodal comparison stimuli were implemented distributed among the standard stimulus

$$S = [0.7, 1, 1.2, 1.4, 1.8, 2, 2.2, 2.4]\,\text{mm/N}. \tag{5.19}$$

In the *active* condition stimuli were provided according to equation (5.14) (denoted by the subscript "a"). Thereby, participants could actively indent the virtual cube by gripping with index finger and thumb, while the index finger was fixed and the indentation was only performed with the thumb (with an excursion of about 2.5cm). On the visual display participants saw their finger representations (yellow spheres) indenting a virtual cube.
In the *resistive* condition, stimuli were provided according to (5.15) (denoted by the subscript "r"). Thereby, participants were told to maintain their thumb at a constant fixed position (index finger was again fixed) as the robot pressed into the cube, rendering the commanded compliance. The cube trajectory was a sine-wave with randomized amplitude (ranging from .5 and 2.5 cm) and period of 1.0s. On the visual display participants saw the cube pressing/moving into their thumb representation (yellow sphere). An additional control condition was implemented in which only haptic stimuli were explored, using the active protocol but with no visual depiction (denoted by the subscript "h").

Procedure

The psychometric function for discriminating compliance was assessed by the method of constant stimuli. A 2-conditions within-subject design was used. The active and restistive conditions, were presented in different blocks that consisted of 8 comparisons and 10 repetitions. The sequence for each of the blocks was randomly chosen for each participant. Participants were seated in front of the HSI, looking at the screen and grasping the device with their dominant hand. After being instructed, participants could were trained using a set of stimuli which were not repeated in the experiment. Training ended when subjects felt confident they understood the task. Hence, $3 \times 19 = 57$ psychometric functions had to be recorded.
One trial consists of the sequential presentation of two stimuli: a standard and a comparison stimulus. Participants explored each stimulus as directed, active or resistive, to assess their compliance of the two stimuli and reported which of the two appeared more compliant. The duration of each stimulus presentation was 2 s with an inter-stimulus interval of 2 s. The next trial began 2 s after the answer was inserted by the joystick.

Data Analysis

Psychometric functions were fitted using "psignifit" software, version 2.5.6, described in [29] and Matlab/Simulink. The following parameters were computed by fitting the psychometric function to the experimental data to describe the final percept $\hat{S}$: The point of subjective equality for each condition (PSE_a, PSE_r, PSE_h) was computed as the stimulus level at proportion 0.5

$$\text{PSE} = F_\Psi^{-1}(0.5) \tag{5.20}$$

where F_Ψ denoted the sigmoid-shaped psychometric function (cumulative probability).
The just noticeable difference for each condition (JND_a, JND_r, JND_h) was computed by taking the weighted stimulus difference between PSE and proportion 0.84

$$\text{JND} = \frac{\sqrt{2}(F_\Psi^{-1}(0.84) - \text{PSE})}{S_{ref}}, \tag{5.21}$$

where the weighting $\sqrt{2}$ represented the performance correction in reminder tasks (method of constant stimuli) to indicate a perceptual sensitivity of $d' = 1$. See [3] p.180/181 for a detailed description of the reminder paradigm[1].

5.2.3 Results

Discriminating compliance with our visual-haptic virtual reality was found to produce reliable psychometric functions.
A paired-sample t-test performed on the JNDs obtained in the two conditions indicates that the JND in the resistive condition is significantly larger than in the active condition ($t(18) = 3.32, p = 0.001$ one-tailed). For the active condition (5.14) the JND_a resulted in a mean $\mu = 28.91\%$ and a standard deviation $\sigma = 11.96\%$. For the resistive condition (5.15) the JND_r resulted in a mean $\mu = 82.67\%$ and a standard deviation $\sigma = 71.56\%$. The mean JND in the resistive conditions is 2.86 times larger than the JND in the active condition, indicating a large, significant increase of the JND (i.e. a decrease of reliability) when information is constrained to visual position information x_v and haptic force information f_h.
The data and the fitted psychometric functions of one subject are depicted in Fig. 5.4.

5.2.4 Discussion

In two conditions, participants estimated the compliance of a cube either by indenting the cube surface actively, or by holding their finger at the same position while the simulated cube was moved back and forth. In the first condition, information to estimate compliance

[1] The performance correction was conducted because participants were assumed to have used a *differencing strategy* when comparing between standard and comparison stimulus. No performance correction would be necessary if participants used a *decisionally separable strategy*, which assumes that participants essentially memorized the standard stimulus. However, since compliance is a dynamical stimulus comprised of force and position information, we assumed a complexity such that participants did not ignore the standard stimulus but processed it anew on each trial.

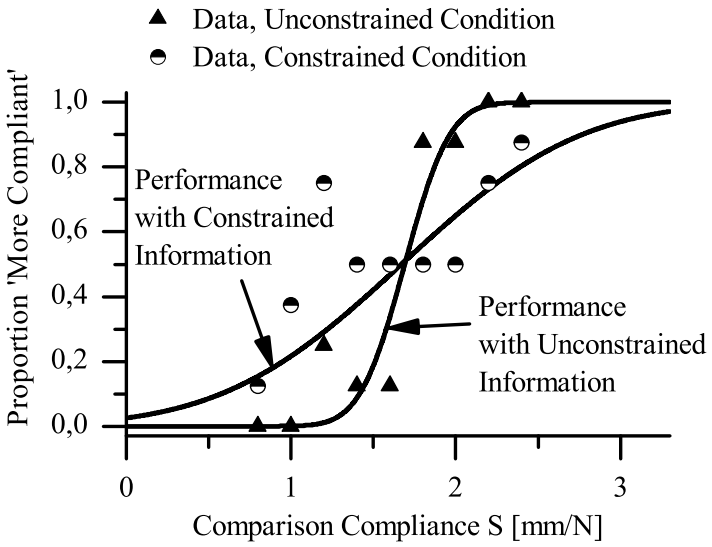


Figure 5.4: Perception of visual-haptic compliance with active and resistive information (experimental data and psychometrics function of participant No. 12): With active visual-haptic information, discrimination performance was better than if only x_v and f_h were provided. On average the JND increased from 28.75% to 82.5% if information was resistive only. (Data points represent the proportion of 10 trials each.)

is available through the haptic and the visual modalities for both position and force. In the second condition, in contrast, participants could estimate the displacement of the cube only by using the visual modality and the force applied in the indentation using the haptic modality. The data collected in this experiment disconfirms hypothesis (5.17), indicating that humans perceive compliance from visual and haptic information according to Process Model 1 (see Fig. 5.2). In this model, perception of objects' compliance is obtained by combining non-redundant position and force information independently for each sensory modality. The different compliance estimates obtained in this way for each modality are fused subsequently. This way of combining information allows a reliable compliance estimate only when information in each modality comprises position and force information, as it is the case in the active condition described above.

It is worth noting as well that of the two models, Process Model 1 can more easily incorporate other information about compliance per se (e.g., priors or top-down knowledge), since it does not require cross-modal integration in terms of position and force.

The results collected in this experiment are similar to the ones reported by Srinivasan and LaMotte using spring cells [144], which investigate tactual perception of compliance (vision was always precluded). In different conditions, participant's hand was either stationary or moving and tactile stimuli were applied on the finger's skin in order to judge compliance. From these data, the authors concludes that tactile information is itself was insufficient for judging the compliance of spring cells. This study extends Srinivasan's result for different reasons. Here, the information about compliance was obtained via force estimates using the kinesthetic sense. Tactile information was not relevant since finger-pad indentations were such that the tactual stimulation was constant and equal in the two conditions. Also, the participant's finger was not glued or supported by a plate as used in Srinivasan's passive

condition. Participants were told only to hold their finger at a fixed position and had to actively apply forces in order to so. However, despite the pre-training, participants could not completely immobilize their finger, because they could not compensate entirely for the force produced by the simulated oscillation of the cube. Their finger movement was very little (they were informed by a beep if they moved their finger more than $\pm 5\,\mathrm{mm}$ from x and the trial was not recorded), but it might have been sufficient to provide information about position from the haptic modality (albeit with minimal reliability). This artifactual information might explain why subjects could provide at least a rough compliance estimate in the resistive condition.

This experiment confirmed that combination processes happen before a possible integration process. The integration process to a final percept $X_{vh} := S_{vh}$ is analyzed in the following.

5.3 Experiment: Is Visual Information Fusing or Confusing?

5.3.1 Hypotheses

The result of the first experiment indicated that compliance perception is performed according to Process Model 1; that is, integration of bimodal information takes place only after the combination of force and finger position to produce an estimate of compliance. This leads to the question of whether there is a general limitation on the integration of visual and haptic information, such that redundant estimates - in this case, of compliance – from vision and touch can be combined by some weighted scheme, but non-redundant parameters - in this case, position and force – cannot be fused across the two modalities.
Experiment 2 was designed to test Process Model 1 in more detail. Specifically, it manipulated visual cues to position and force, holding haptic cues constant, and used psychometric evaluation to determine how those manipulations affected the final compliance estimate. According to Model 1, the visual channel should use the distorted cues to arrive at a compliance estimate, which is then fused with the estimate from touch.
In this experiment we modified the visual position information relative to the haptic by four (4) ratios $(2, 1, 0.5, 0$, used as subscripts in the following equations) producing the following conditions:

$$\begin{aligned} x_{v,2} &= 2x_h, \\ x_{v,1} &= 1x_h, \\ x_{v,0.5} &= 0.5x_h, \\ x_{v,0} &= 0x_h. \end{aligned} \tag{5.22}$$

The magnification of the participants' finger movements ($x_{v,2}$) leads to a higher compliance S_v and damping the visual position ($x_{v,0.5}$) leads to smaller visual compliance. The coefficient of zero corresponds to no visible displacement. Furthermore, we attempted to manipulate the visual cues to force, on the assumption that people may infer forces based upon the manner in which the object is indented and/or by its surface. Accordingly, to portray visual force we used cylinders instead of spheres to represent the indentation tool, and varied their size and surface appearance. In the visual VR we implemented a metal

tool (subscript "M") and a wood tool (subscript "W") to indent the cube in the visual VR, as depicted in Fig. 5.5. The conditions are related according to

$$f_{v,W} \quad < \quad f_{v,M}. \tag{5.23}$$

Using the wood tool is expected to yield a higher compliance estimate than using the metal tool, assuming that it implies less force is used to reach a given indentation.
Specific hypotheses about the changes in the mean and the variability of the responses were formulated in the context of a set of models, as described blow. Here we describe more general hypotheses. For any level of visual cueing, integration of S_v with S_h (confusional, suboptimal or optimal according to the maximum likelihood model) is indicated by an averaged PSE

$$\mathrm{PSE}_x = w_h \mathrm{PSE}_h + w_v \mathrm{PSE}_v, \tag{5.24}$$

where x represents either condition $S0.5$,$S1$, or $S2$ (condition $S0$ was a unimodal condition). Assuming that the mean estimate for the visual cue will shift with the distorted visual cue, and further assuming constant weights, this means that the final compliance percepts should shift in the direction of the visual distortion. That is, the mean unimodal compliance estimate should not shift for the congruent vision case, but should shift downward for $\mathrm{PSE}_{0.5}$ and upward for PSE_2. Hence, the hypotheses for integration, based on the PSEs can be denoted as

$$\mathrm{PSE}_{0.5} \overset{!}{<} \mathrm{PSE}_0 \overset{!}{=} \mathrm{PSE}_1 \overset{!}{<} \mathrm{PSE}_2. \tag{5.25}$$

5.3.2 Method

Presence System

See the first experiment for details.

Participants

Twenty-three (23) students of the Technische Universität München took part in this study and were paid for participation. The data from the first three (3) participants was used to adjust the HSI and those data are reported here. The results were generated with participants four (4) to 23. Participants had an average age of 24 years. All participants were right-handed and had normal or corrected to normal vision.

Stimuli

All stimuli were virtual rendered compliant cubes with a length of 80 mm. Participants were able to squeeze the virtual cube displayed by the HSI actively with both fingers. The compliance of the standard stimulus was

$$S_{ref} = 0.7\,\mathrm{mm/N}. \tag{5.26}$$

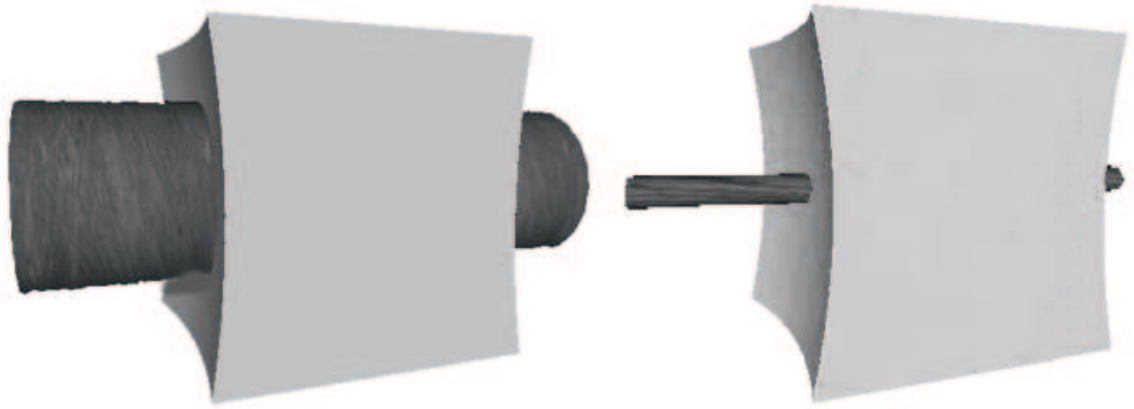


Figure 5.5: Visual force information: To generate different visual forces f_v, different indentation tools were used. The metal tool had a large diameter and a metal-colored surface. The wood tool had a small diameter and wooden-colored surface.

This was considerably lower than in the first experiment to accommodate the expected increase in PSE due to the visual distortions. The visual compliances of the standard stimulus S_v was defined by equations (5.22).
Furthermore, in the visual virtual reality, the fingers were represented not by yellow spheres as in Experiment 1, but by two cylinders. The cylinders were differently shaped and colored as described above. In the first configuration the cylinders had a small diameter (1 cm) and the surface resembled wood (denoted by the subscript "W"). In the second configuration the cylinders had a large diameter (5 cm) and the surface resembled metal (denoted by the subscript "M"). The conditions are depicted in Fig. 5.5.
Eight (8) visually-happtically congruent bimodal comparison stimuli were distributed with the standard stimulus at the center of the distribution

$$S = [0.14, 0.22, 0.36, 0.5, 1, 1.15, 1.3, 1.44]\,\mathrm{mm/N}. \tag{5.27}$$

Procedure

As in the first experiment the method of Constant Stimuli was used to assess performance. Eight (8) conditions were tested, constituting the combination of four (4) levels of the deviation of the visual compliance as defined by equations (5.22), and two (2) levels intended to influence the visual force f_v as defined by (5.23). Each of 19 participants was tested in all condition. Hence $4 \times 2 \times 19 = 152$ psychometric functions had to be recorded.
One trial consists of the sequential presentation of two stimuli: the standard and the comparison stimulus. Duration of each stimulus presentation was 2 s with an inter-stimulus interval of 2 s. The next trial began 2 s after the answer was inserted by the joystick. Participants could elect to have a break whenever they wanted by not inserting the answer. Each condition was presented in one block consisting of the eight (8) comparison stimuli presented in random order ten (10) times. Hence, 80 trials were necessary to obtain a psychometric function. The sequence of the blocks was randomly chosen for each participant. Participants were seated in front of the HSI with their dominant hand grasping the device and with direction of gaze essentially perpendicular to the screen. They were carefully

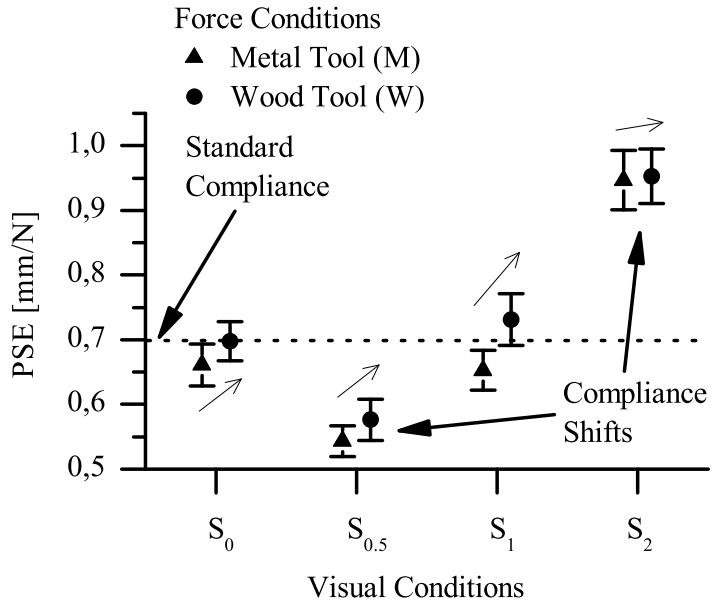


Figure 5.6: Change of PSEs by visual position and force information: The bimodal percept shifted to lower compliance when visual indentation was only half of haptic indentation (condition S_{05}) and to higher compliance when visual indentation was twice the haptic indentation (condition S_2). Percepts that were derived using the visual metal tool tended to be less compliant than percepts generated using a visual wood tool. This was consistent with the idea that the metal tool generated larger visual force information than the wood tool, but these differences were not significant.

instructed according to the group to which they were randomly assigned. A short training was completed before the session started.

Data Analysis

See the first experiment for details.

5.3.3 Results

As in the first experiment, reliable psychometric functions could be fitted to the data in all conditions. From these functions, two parameters were extracted, the PSE and the JND, as described above. The PSE_0 had a mean of $\mu = 0.68\,\mathrm{mm/N}$ and a standard deviation of $\sigma = 0.14\,\mathrm{mm/N}$. The PSE_2 had a mean of $\mu = 0.95\,\mathrm{mm/N}$ and a standard deviation of $\sigma = 0.20\,\mathrm{mm/N}$. The PSE_1 had a mean of $\mu = 0.69\,\mathrm{mm/N}$ and a standard deviation of $\sigma = 0.16\,\mathrm{mm/N}$. The $\mathrm{PSE}_{0.5}$ had a mean of $\mu = 0.56\,\mathrm{mm/N}$ and a standard deviation of $\sigma = 0.12\,\mathrm{mm/N}$.

The changes of the PSEs induced by the visual position information are depicted in Fig. 5.6. Two-way repeated-measures ANOVA performed on the PSE obtained from the individual

fit of subjects' responses in the three conditions that contained visual information showed a significant effect of visual position information ($F(2, 38) = 44.96$, $p < 0.001$). The effect of the force information was not significant ($F(1, 19) = 2.29$, n.s.). This first result confirmed that visual position information influenced the PSE of the bimodal percept as if fusion had occurred. Three Bonferroni-corrected paired-sample t-test were used to compare the unimodal compliance percept S_0 with the three visual conditions, according to the general hypothesis above. Results followed the predictions: Perceived compliance in the visual condition S_2 was higher than in S_0 according to hypotheses (5.25) ($t(19) = 2.57$, $p < 0.016$ one-tailed). In condition S_1 no difference between the bimodal and the unimodal percept was found ($t(19) = 0.88$, n.s). In the condition $S_{0.5}$ the final compliance percept was lower than the unimodal compliance percept ($t(19) = 4.75$, $p < 0.016$ one-tailed).
The experimental outcome for the PSEs in the visual conditions ($S0.5$, $S1$, $S2$) is captured by the mathematical processes described by equation (5.6) and (5.11). The bimodal PSE is a weighted average of the single modality means. The data of conditions $S0.5$, $S2$ clearly rejects the mathematical process described by equation (5.5). The weightings w_v, w_h were calculated according to equation (5.7) or (5.12) using the averaged unimodal and bimodal PSEs ($\text{PSE}_0 = 0.68$, $\text{PSE}_{0.5} = 0.56$, $\text{PSE}_2 = 0.95$) as well as the commanded visual compliances $S_{v,0.5} = 0.35$, $S_{v,2} = 1.4$. For condition $S_{0.5}$ they were calculated to be $w_v = 0.636$, $w_h = 0.364$. For condition S_2 they were calculated to be $w_v = 0.625$, $w_h = 0.375$. Hence, participants used the same weightings for both conditions, which are approximately

$$w_h = 0.63, \qquad w_v = 0.37. \tag{5.28}$$

It is mathematically not possible to calculate the weights from the PSEs in the congruent condition since all combinations would lead to the same result.
The second factor in the ANOVA above indicates that visual force information did not significantly influence the PSEs of the bimodal percepts. However, a trend, which could be observed in all conditions, followed our prediction that the metal tool is associated with larger visual forces. Changes of the PSEs induced by the visual force information are depicted in Fig. 5.6.
The changes of the JNDs induced by the visual position information are depicted in Fig. 5.6. The same ANOVA reported above for the PSEs was repeated on the JNDs. It showed a significant effect of visual position information ($F(2, 38) = 10.45$, $p < 0.001$), whereas the effect of the force information did not reach significance ($F(1, 19) = 0.41$, n.s.). The changes of the JNDs induced by the visual force information are depicted in Fig. 5.6. The finding that the incongruent visual conditions $S_{0.5}$ and S_2 showed a higher JND than the unimodal condition S_0 is clearly inconsistent with the prediction from the maximum-likelihood model that the bimodal inputs will reduce the input. We used a paired t-test to assess the model's prediction that the congruent visual condition S_1 condition would produce a lower JND than the unimodal condition. The advantage for the bimodal condition was significant ($t(19) = 1.85$, $p = .04$, one-tailed).
Two Bonferroni-corrected paired-sample t-tests were used to compare the JNDs obtained in the incongruent visual conditions with the unimodal condition S_0. These tests indicated that in the visual condition S_2 the bimodal JND was greater than the unimodal S_0 ($t(19) = 2.46$, $p < 0.032$ two-tailed); the other comparison of unimodal vs. $S_{0.5}$ was not significant, ($t(19) = 0.19$, n.s.). Finally, in pursuit of the models described below, we tested for differences between JNDs in the incongruent bimodal conditions; the JND for $S_{0.5}$ was

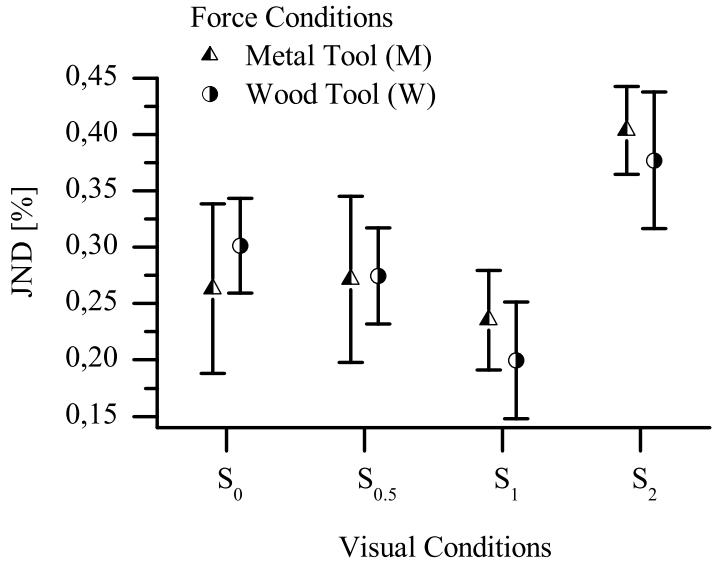


Figure 5.7: Change of JNDs by visual position and force information: The JNDs of the bimodal percepts were not consistently smaller than the unimodal JND. (JNDs are relative to the standard compliance.)

significantly less than for S_2 ($(t(19) = 2.78$, $p = .012$, two-tailed).
The remaining comparisons across visual conditions showed no significant differences: for unimodal vs. S_1 ($t(19) = 0.83$, n.s.), and unimodal vs. $S_{0.5}$ ($t(19) = 0.19$, n.s.).

5.3.4 Discussion

The influence of visual compliance information to the final compliance percept was analyzed in an experiment with conditions varying the visual compliance by changing visual position and visual force.
Visual position was varied by doubling and by halving the indentation of the fingers. According to our hypotheses we found a significant effect for the bimodal compliance percept. If visual position was doubled, participants reported a softer object, i.e. the PSE of the perceived compliance increased compared to the PSE of the unimodal percept. If visual position information halved, participants reported a stiffer object, i.e. the PSE of the perceived compliance decreased. If visual position information was added congruently, participants reported the same compliance PSE as in the unimodal case. Hence, it can be stated that visual compliance information biases the bimodal compliance percept.
Visual force was varied by outfitting the visual VR with either wooden or metal tools. The wooden tools should provide a lower visual force than the metal tools. Therefore, a stiffer cube should be reported, when using the metal cubes. However, the influence of the visual force information on the PSE of the bimodal percept did not become significant, although we could isolate a quasi-significant effect. Hence, it could be stated that the visual force information very likely biases the bimodal compliance percept.
Our result supports the results by Srinivasan [145], who also found that visual compliance

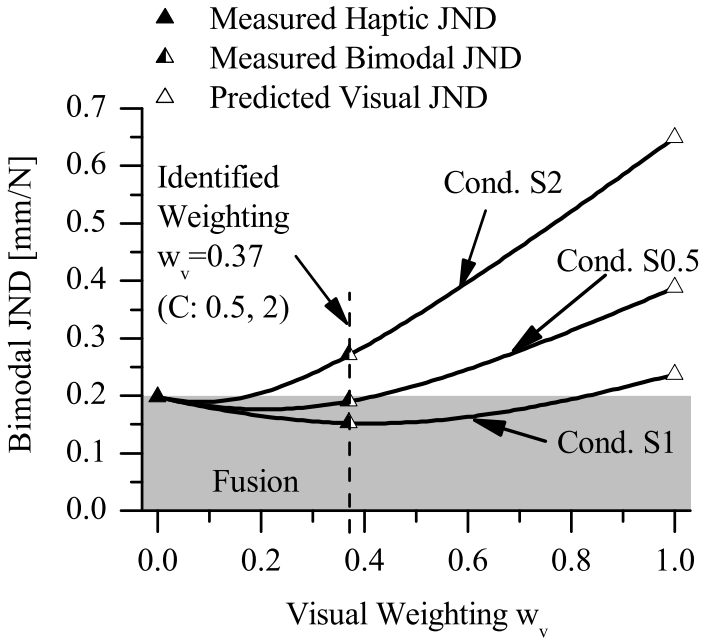


Figure 5.8: Visual-haptic perception of compliance assuming a summation of two random variables defined by the visual and the haptic compliance estimate according to equation 5.6: The weightings were found to be the same for all three conditions. Fusion has occurred if the bimodal JND is smaller than the lower unimodal JND, i.e. the haptic JND (gray-shaded area). For the congruent condition S_1, the bimodal JND is close to the optimal value (minimum of the depicted curves). For the condition $S_{0.5}$, it is nearly optimal and close to the unimodal haptic estimate. The bimodal JND is larger than the unimodal estimate in condition S_2. The predicted visual JNDs increase non-monotonically violating the assumption of a constant Weber fraction, but they increase proportional to the conflict.

information biases the visual-haptic percept when squeezing helix-formed springs displayed by a visual-haptic VR. Furthermore, our results support the studies presented in [70; 79] that also reported an distracting influence of visual compliance information.
The influence of the bimodal PSE by the different conditions can be modeled by the two non-trivial integration processes explained in Section 5.1 which predict the bimodal PSE to be a weighted average of the unimodal PSEs. The weightings were found to be stiff for the different conditions, equation(5.28). This is a very strong result confirming that the perceptual system non-trivially integrates the bimodal information. However, that does not necessarily mean that the perceptual system always fuses the bimodal information, since fusion, no fusion, and even confusion are still possible, depending on the underlying integration process.
More information to identify the underlying mathematical process comes from the JNDs. The relation of the bimodal and unimodal JNDs, assuming that the perceptual system integrates the bimodal information according to a weighted summation of two random variables (equations (5.6)), is depicted in Figure 5.8. The predicted visual JNDs increase non-monotonically ($\text{JND}_{v,0.5} = 0.39\text{mm/N}$, $\text{JND}_{v,1} = 0.24\text{mm/N}$, $\text{JND}_{v,2} = 0.65\text{mm/N}$). This violates the assumption of a constant Weber fraction. However, the absolute visual

JNDs increase nearly proportional to the conflict. Therefore, the visual JNDs can also be predicted based on the JND of the congruent condition according to

$$\mathrm{JND}_{v,c} = (1 + c)\mathrm{JND}_{v,1}, \tag{5.29}$$

where c is the factor defining the bimodal conflict, which is $c = 0.5$ in condition $S_{0.5}$ and $c = 2$ in condition S_2. Then, the predictions are $\mathrm{JND}_{v,0.5} = 0.36\mathrm{mm/N}$ for the visual JND of condition $S_{0.5}$ and $\mathrm{JND}_{v,2} = 0.71\mathrm{mm/N}$ for condition S_2, respectively. This close relation might result from a perceptual process that is performed right after the combination process but before the integration process. That means, the reliability of the visual estimate is corrected ex post depending on the mean conflict between haptic and visual compliance information.
Latter result is supported by the prediction of the visual JNDs based on the remaining integration process (equations (5.11)), which assumes that the perceptual system draws randomly from both estimates. Assuming this process the visual JNDs for all conditions result in complex values. That means, the process over-predicts the bimodal JNDs. Hence, out of the three basic mathematical integration models explained in Section 5.1, only the second integration process (equation (5.6)) enables a psychophysically satisfying interpretation of our experimental results. The remaining integration processes either fail to predict the PSEs (equation (5.5)) or fail to predict real-world JNDs (equation (5.11)).
Since the second integration process predicts optimal fusion (bimodal JND is smallest possible JND) in the congruent condition, it could finally be stated that visual information is fused with haptic information. However, the perceptual fusion process is not robust. It is transformed into no fusion or even confusion by an additional perceptual process if a conflict between visual and haptic compliance information is perceived.

5.4 Conclusion

Summary

In this chapter a theoretic model of visual-haptic compliance perception was elaborated. Two experiments were conducted to develop the model successively. The first experiment unveiled that reliable compliance estimates are only possible when information in each modality comprises position and force information. The combination process takes place before a possible fusion process of redundant information. The second experiment unveiled the role of visual compliance information within the visual-haptic percept. Visual compliance information biases the final, bimodal percept. Information is only fused if the bimodal estimates are congruent. If a conflict is perceived, an additional perceptual process decreases the reliability of the visual estimate and turns the integration process to no fusion or even confusion.

Scientific Contribution

1.) The role of combination and integration processes could be clarified for the first time. As reported in the first experiment the combination of position and force information takes

place within modalities but not between. Information of the other modality is not fused before a first modality-specific compliance estimate is built.
2.) The role of visual compliance information could be clarified in detail for the first time. As the second experiment unveiled, visual information does bias the haptic compliance percept. If the visual compliance is higher than the haptic compliance, the final compliance percept is higher than the haptic compliance. If the visual compliance is smaller than the haptic compliance, than the final compliance is also smaller than the haptic compliance. Although, this appears like a fusion process we could clarify that fusion is only the valid integration process if the bimodal information is congruent. Based on the results a perceptual process is postulated that takes conflicting information into account and decreases the reliability.
3.) Weber's law seems also to hold for the perception of compliance. Comparing the bimodal congruent percepts in both experiments we observed nearly equal JNDs for both reference compliances. Since the reference compliances were rather different (0.7 and 1.6 mm/N) we can assume that the Weber fraction at least applies in the range defined by these stimuli.

Open Problems

The presented research provides a theoretical model for visual-haptic compliance estimate but some questions still remain to be treated.
Firstly, it is not clear to which extent visual force information can be inferred by the observer. By our method using two different squeezing tools we could only isolate a quasi-significant effect. However, since subjects have at least an unreliable visual compliance percept (as reported in the third study of Chapter 4) it can be deduced that visual force information must exist (Despite humans have no visual force sensor.). Further experiments to identify the character of the visual force information would be helpful.
Secondly, the postulated additional perceptual process has to be analyzed in more detail. Simulations have to be conducted to understand all implication of this process. An interesting question is how the variance of a single estimate can be increased without changing the mean value of the bimodal estimate. This might be done by additional noise elements or dynamical systems.
Thirdly, the present study does only cover the perception of a single type of mechanical environment, i.e. compliance. It is not clear if the results apply also for inertia or density. Further studies are necessary.

6 Haptic Data Compression with Perceptual Coding

When you blow it, babe
You got to blow it right
Oh baby, if you fake it, mama
Baby, fake with all your might
When you fake it, mama
Please fake it right

Led Zeppelin, *For Your Life* (*Presence*, 1976)

Presence systems (as introduced in Section 2.1) that enable realistic, immersive experiences of complex tasks are usually equipped with bimanual, highly dexterous facilities to perform and display kinesthetic-haptic stimuli. In such a system HSI an TO are equipped with haptic devices that have several degrees of freedom (> 20) each using highly accurate sensors (> 14 bit) and actuators controlled at high sample rates (> 1 kHz). Hence, for the synchronization of HSI and TO a large amount of data has to be exchanged over the COM. However, bandwidth is generally limited or costly motivating the need for effective compression algorithms for haptic information.
In bilateral, kinesthetic-haptic (by now 'haptic') telepresence systems, force and position-based signals are commanded from the human operator to the TO and sensed information from the remote environment is reflected and displayed by the HSI. Thereby, a control loop is closed between HSI and TO. Main objectives in the control system design are stability and transparency. Ideal transparency means that the operator does not perceive the presence-mediating technology (HSI, COM, TO) when experiencing the target environment through distal attribution or proximal stimulation (see Subsection 2.1.2 for clear information on the different modes of mediated presence). The key challenges associated with the loop closed over the COM are delay and limited bandwidth. A COM affected by delay can lead to instability of the overall system if not treated by appropriate measures. These measures, e.g. passivity, reduce transparency to an extent positively correlated with the height of the delay [47; 49; 147].
While communication delay is frequently treated, few researchers consider communication resource limitations. Compression algorithms compress or discard information to reduce the amount of data to be stored, processed, or transmitted. The lossless compression scheme proposed in [148] results in a trade-off between compression efficiency and delay required for compression. Differential pulse code modulation (DPCM) together with a fixed rate quantization has been proposed in [149]. Adaptive DPCM together with Huffman coding has been considered in [150]. General considerations about haptic lossy compression methods and perceptual performance have been presented in [151].
The only approach that retains passivity of the COM in delayed telepresence is *deadband control* introduced in [152]. Therein, compression is achieved by setting up a deadband,

which defines a deviation from a sample value. Each submitted sample defines a deadband. If the next sample is within the deadband, it will be discarded. If it exceeds the deadband, it will be submitted. Passivity is achieved by assuring that the reconstructed velocity and force samples have equal or less power than their initial correlates. This is achieved by a passive hold-last-sample algorithm for instantaneous telepresence [64]. A passive hold-last sample for delayed telepresence was introduced in [152]. The implementation can be deployed in telepresence systems with arbitrarily constant communication delay. Experiments revealed data compression rates up to 87% without decreasing transparency. The main drawback of the algorithm is its inability to maintain an upper bandwidth limitation. If force or velocity signals are subject to fast changes the compression savings can decrease to zero. Different extensions of deadband control have been proposed such as incorporation of perceptual Kalman filtering [153], additional use of prediction algorithms [154], and application to three-dimensional data [155]. However, all extensions suffer from not being passive and therefore possibly causing instability of the overall system.
In this chapter passive haptic compression methods are proposed that provide high compression rates and additionally enable constant data compression rates. Based on a classification for haptic lossy data compression (LDC) sufficient passivity conditions are proposed for interpolative, extrapolative, and direct compression methods. The algorithms themselves can be arbitrarily complex. Interpolative and extrapolative LDC-algorithms are implemented and their impact is psychophysically analyzed. As a result, the implementations provided transparent compression up to a compression rate of $10:1$ (90% data rate savings). The contributions had been partly elaborated in [66; 156; 157] and have been published to a large extent in [65].
The remainder is organized as follows: In Section 6.1 design goals and the classification is presented. In Section 6.2 the passivity criteria are introduced and explained. Algorithms are derived in Section 6.3. A psychophysical analysis and a parametrization method are presented in Section 6.4. Finally, a conclusion is drawn in Section 6.5.

6.1 Design Goals and Classification

Haptic lossy data compression algorithms are deployed within the COM of a presence system. Their goal is to discard information not perceived by the human operator. As delayed communication can lead to instability of the presence system, the COM is designed passive. Hence, only compression methods are considered that preserve the passivity of the COM. Therefore, the algorithms are designed in Scattering domain. The performance of the compression methods can be evaluated in two dimensions. Firstly, the compression method has to achieve high data compression rates. Secondly, the compression should be not perceivable by the human operator. A classification is proposed that structures the different algorithm in frequency-based, interpolative, extrapolative, and direct algorithms. In Subsection 6.1.1 the basics of passive compression are laid down. Subsection 6.1.2 introduces the data compression ratio and Subsection 6.1.3 introduces the notion of transparency. The classification is presented in Subsection 6.1.4.

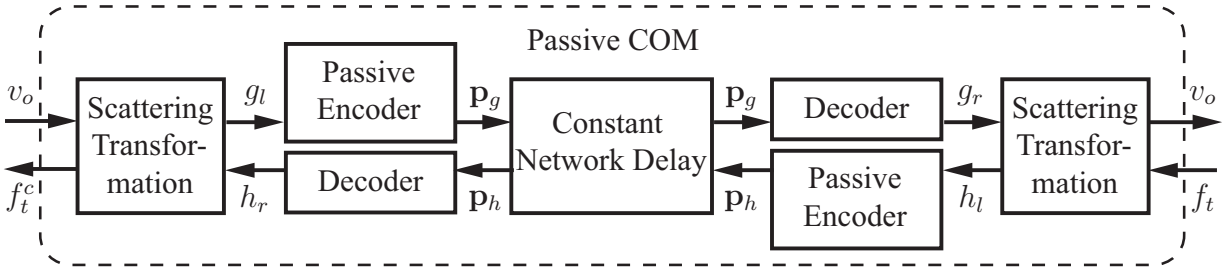


Figure 6.1: Structure of the COM for passive compression: Encoder and decoder are applied in Scattering domain, which assures passivity for arbitrary constant delays. Furthermore, encoder and decoder have to be passive themselves.

6.1.1 Passivity for Haptic Compression Algorithms

In contrast to audio and video compression, haptic compression algorithms are applied within an energy-exchanging closed control loop between operator and environment. The dynamics of the compression algorithm effects the dynamics of the overall system and computations have to be real-time. To stabilize bilateral haptic presence systems the passivity paradigm is widely deployed. Therefore, the presence system is divided into subsystems that were designed passive. See Subsection 2.4.2 for a detailed introduction and Figure 2.17 for an illustration. A passive system does not produce energy. A reachable dynamical system with zero initially stored energy is passive if the *integral passivity condition* (2.68) holds

$$\int_0^t P_{in}\, \mathrm{d}\tau \geq 0, \qquad \forall t \geq 0, \tag{6.1}$$

where P_{in} is the power input of the system. For the COM of a presence system in velocity-force architecture the input power is defined as the scalar product of force and velocity

$$P_{in} = \dot{x}_o f_t^c - \dot{x}_o^c f_t, \tag{6.2}$$

where $\dot{x}_o$ is the velocity commanded by the operator and $\dot{x}_o^c$ is the commanded velocity on teleoperator side and f_t is the force reflected by the teleoperator and f_t^c is the reflected force on operator side. The power entering the system is counted positive and the power leaving the system is counted negative.
The COM of a presence system, especially in telepresence architecture, is very likely to be afflicted with communication delay (due to delay in the communication networks like the Internet, etc.). Therefore, passivity measures for the COM are mandatory. Usually the Scattering transformation is applied to passivate the COM in presence of constant communication delay mapping power variables (velocity, force) into wave variables

$$\begin{aligned} g_l &= \frac{b\dot{x}_o + f_t^c}{\sqrt{2b}}, \quad & h_l &= \frac{b\dot{x}_o - f_t^c}{\sqrt{2b}}, \\ g_r &= \frac{b\dot{x}_t^c + f_t^e}{\sqrt{2b}}, \quad & h_r &= \frac{b\dot{x}_t^c - f_t^e}{\sqrt{2b}}. \end{aligned} \tag{6.3}$$

Thereby, g_l, $h_r \in \mathbb{R}$ denote the incident wave and h_r, $h_l \in \mathbb{R}$ denote the reflected wave (also called *wave reflections*). The parameter b (wave impedance) is a positive constant that can be chosen arbitrarily. The transformation is *bijective*, i.e. unique and invertible. Hence, no information is lost or gained by encoding power variables into wave variables or wave variables into power variables. The passivated COM is depicted in Figure 6.1.
Applying the Scattering transformation (2.71) to equation (2.73) yields the power input in Scattering domain

$$P_{in} = \frac{1}{2}(g_l^2 - g_r^2 + h_r^2 - h_l^2), \tag{6.4}$$

where the index indicates the wave variables on the right and on the left hand side. Using condition 6.1 latter equation can be divided into a passivity condition for the incident wave

$$\int_0^t g_l^2(\tau)\,\mathrm{d}\tau \geq \int_0^t g_r^2(\tau)\,\mathrm{d}\tau \tag{6.5}$$

and a passivity condition for the reflected wave

$$\int_0^t h_r^2(\tau)\,\mathrm{d}\tau \geq \int_0^t h_l^2(\tau)\,\mathrm{d}\tau \tag{6.6}$$

illustrating that waves carry their own power (unit of measurement $\sqrt{W}$). Hence, waves remain passive if they do not accumulate energy when passing the forward path (incident wave g) or the reflection path (wave reflections h).
For the passivation of the COM is mandatory in haptic telepresence, only LDC-methods are considered that preserve passivity of the COM. Hence, the compression algorithms introduced in the following are designed for the Scattering domain and have to be passive themselves. The dynamics of a lossy compression scheme can be divided into phase characteristics and amplitude characteristics. The phase of the signal is influenced by the encoder and the decoder as well as by the procedure of compression. Every stage adds a delay between original data and reconstructed signal. The amplitude of the signal is changed due to the rejection of information by the lossy reduction. The passivation of the delay caused by the LDC-algorithm is superfluous since waves remain passive for constant delays. The passivation of the amplitude changes caused by the lossy compression has to be assured by limiting the energy of the decoded wave to obey the passivity conditions in wave domain (6.5), (6.6).

6.1.2 Data Compression Ratio

The data compression ratio is defined by the ratio of the original data rate [bit/s] to the reduced data rate yielding

$$\mathrm{CR} = \frac{\text{Uncompressed Data}}{\text{Compressed Data}}. \tag{6.7}$$

It is usually denoted as an explicit ratio, e.g. $5:1$ (read "five to one"). High compression ratios are desirable CR$>>1$. Data rate savings are expressed by

$$\mathrm{DRS} = \frac{\mathrm{CR} - 1}{\mathrm{CR}} \tag{6.8}$$

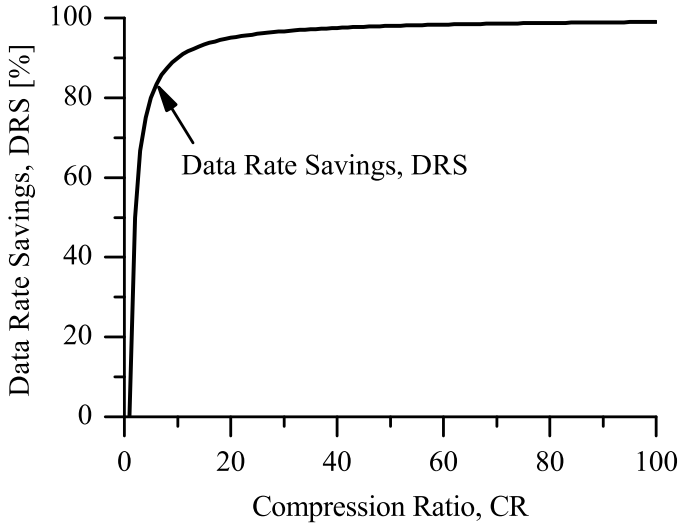


Figure 6.2: Data rate savings: Marginal data rate savings decrease when the compression ratio is increasing linearly indicating that each additionally saved sample yields less relative data rate savings.

and usually denoted in [%]. The relation between CR and DRS is depicted in Figure 6.2 and reveals that every additionally saved sample yields less relative data rate savings.
Lossless compression of digitized data such as video and audio preserves all the information, but can rarely do much better than 2:1 compression because of the intrinsic entropy of the data. In contrast, lossy compression or lossy reduction can achieve much higher compression ratios at the cost of a decrease in quality. Compression artifacts from loss of important information decrease transparency and have to be avoided by perceptual coding.

6.1.3 Perceptual Transparency

A presence system is called *transparent* if the human operator can 'look through' the technology directly perceiving the target environment. For haptic presence systems criteria like (2.1) and (2.2) give advice whether a certain system is transparent or not in terms of a dynamical residual. Residuals are caused by all involved subsystems. Kinematical and dynamical restrictions are imposed by the robotical devices (HSI, TO). Communication delay, noise, or bandwidth constraints are caused by the COM. In telepresence applications especially the communication delay T severely effects transparency due to wave reflections. In data compression the residuals are called *artifacts*. Hence, an LDC-algorithm is called *transparent* if the human operator does not perceive an artifact caused by the compression algorithm. Artifacts have two origins. Firstly, they can be caused by the loss of information leading to reconstruction failures (*approximation artifacts*). Secondly, artifacts can be caused by the delay introduced by the compression algorithm (*phase artifacts*). According to [63; 158] the dynamical residuals caused by delay depend on the dynamics of the target

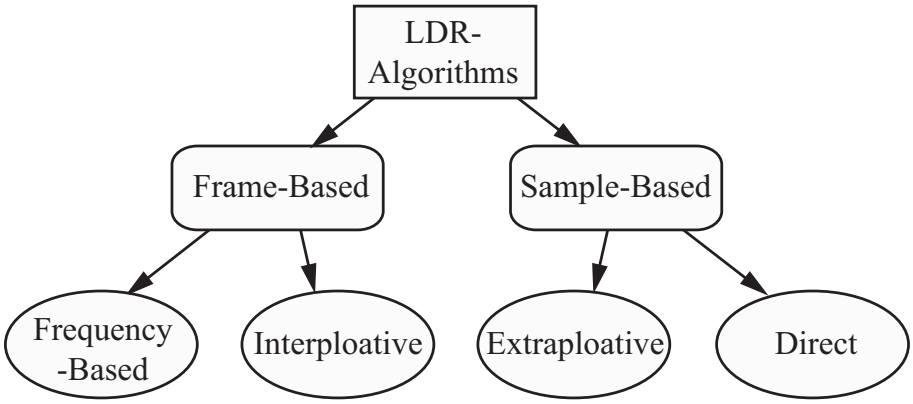


Figure 6.3: Classification of compression algorithms: Frame-based algorithms insert a delay due to storing samples into a frame. They can further be divided into frequency-based and interpolative algorithms. Sample-based algorithms insert no delay and can be further be divided into extrapolative and direct algorithms.

environment. If the environment represents free space the displayed impedance becomes

$$Z_d \approx G_f bTs. \tag{6.9}$$

That means, in free space the force-controlled HSI, represented by the transfer function G_f, renders a mass proportional the amount of the delay T. If the environment shows a rigid wall (infinite stiffness) the displayed impedance becomes

$$Z_d \approx G_f \frac{b}{Ts} \tag{6.10}$$

meaning a rigid wall is displayed as a compliant wall with compliance increasing with the delay T.
Although the transparency criteria (2.1), (2.2) give advice whether a certain compression effects transparency or not, the human operator has to be taken into account and psychophysical experiments have to be conducted to determine whether artifacts are perceivable or not. It is proposed that transparency of an LDC-algorithm is analyzed with respect to its psychophysical detectability. The statistics governing psychophysical signal detectability is called *empirical threshold* and can be recorded by the psychometric function (see Subsection 2.2.2). Henceforth, an LDC-algorithm is called *psychophysically transparent* if a human operator cannot detect its artifacts, i.e. the operator's detection performance is below the empirical threshold.

6.1.4 Classification

Data compression schemes can be divided into lossless and lossy compression. *Lossless compression* schemes exploit statistical redundancies to achieve the reduction and the decoded data equals the original data. On the other hand, *lossy compression* schemes are methods that retrieve data that is different from the original data but close enough to be useful in some way. Only lossy compression methods are considered in the following.

Since especially phase artifacts decrease transparency a classification is proposed that divides the methods in methods that store a certain part of the incoming wave to apply the compression and methods that apply the compression without inducing a delay. These different classes are called *frame-based* and *sample-based*. Frame-based schemes can further be divided into schemes that perform a mathematical transformation to apply the compression (e.g. frequency- or wavelet transformation). This class is called *frequency-based* compression. Frame-based compression schemes can also be divided into schemes that perform the compression directly on the signal frame. This class is called *interpolative*. Sample-based schemes can further be divided into schemes that perform an extrapolation. This class is called *extrapolative*. Eventually, sample-based schemes can be divided into schemes that apply the compression on each sample directly. These schemes are called *direct*. The classification is depicted in Figure 6.3.

6.2 Compression Strategies

Passive haptic compression algorithms introduce phase and reconstruction artifacts. Phase artifacts, i.e. delay, severely decrease transparency. However, the longer the delay the higher the data rate savings. Frequency-based, interpolative, extrapolative and direct compression strategies differently implement this trade-off. Conditions to passivate the reconstruction artifact are proposed with exception of the frequency-based strategy, which is analyzed to be inappropriate for passive haptic compression.
The compression strategies are presented in Subsection 6.2.1 - Subsection 6.2.4.

6.2.1 Frequency-Based Compression

Frequency-based compression strategies are characterized by storing a certain part of the incoming signal into a frame. Afterward the signal frame is processed by an integral transformation (frequency transformation, wavelet transformation, etc.). The compression is performed in the new domain and eventually, the encoder performs the back-transformation to the original domain. Nearly all lossy coding algorithms for acoustic and visual data use frequency-based compression methods.
In case of passive haptic compression the strategy suffers from two peculiarities. Firstly, sophisticated psychophysical models for haptic information to be effectively applied by the strategy, do not exist (in contrast to acoustics or vision), especially not for wave variables[1]. Secondly, the strategy is assumed to use large frame-lengths to efficiently apply the encoder. For example, the minimal signal length to effectively apply a frequency transformation is at least 500 samples. If the system sampling rate is 1 kHz then the delay introduced just by storing the signal frame would be 0.5 s. However, a delay of even 0.5 s extremely deteriorates transparency of the presence system (see [159] for a comprehensive treatment) by decreasing the signal-to-noise ratio of the wave reflections.

[1]For example, the concept of JND is not feasible for haptic information in Scattering domain. Imagine a movement from free space (incident wave $g_1 = b\dot{x}_1 + f_1$ with $\dot{x}_1$-high and $f_1 - low$) into a rigid wall ($g_2 = b\dot{x}_2 + f_2$ with $\dot{x}_2$-low and f_2-high). For a certain wave impedance b it could happen that $g_1 = g_2$ and the difference would be falsely identified as 'not noticeable' although being highly salient instead.

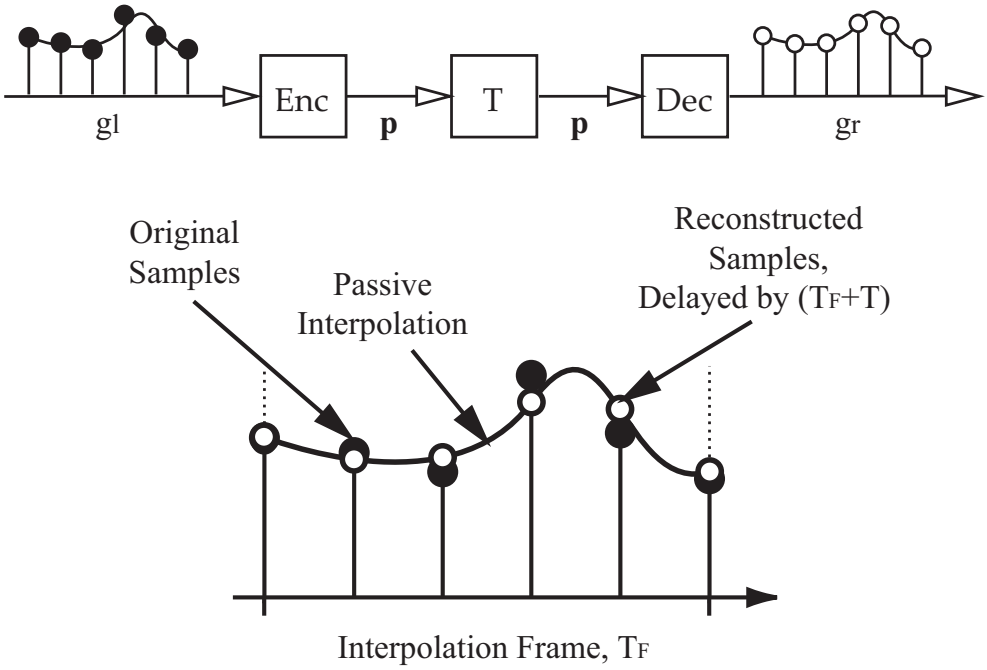


Figure 6.4: Principle of interpolative compression strategies: The interpolated signal has less energy as its original correlates.

As result, frequency-based strategies are considered as inappropriate for passive haptic compression.

6.2.2 Interpolative Compression

The interpolative compression strategy approximates the incoming signal within an *interpolation frame*, without performing a transformation. The encoder works as follows: k_F samples are accumulated to a frame, an approximation algorithm, e.g. a spline interpolation, is applied, and the resulting parameter vector $\mathbf{p}$ is transmitted over the network. The decoder reconstructs the signal using the parameter vector $\mathbf{p}$. The compression principle is depicted in Figure 6.4.
The delay introduced by encoding and decoding the signal, i.e. the computational time to perform the compression, depends on the algorithm but shall be short or can even be neglected. The delay introduced by the interpolation frame amounts to

$$T_F = \frac{k_F}{f_s}, \tag{6.11}$$

where f_s is the sampling frequency. According to conditions (6.5), (6.6) passivity can be assured by forcing the interpolated wave less or equal powerful than the original wave. Hence, the *passivity criterion for interpolative compression strategies*[2] is

$$\int_{t=t_j}^{t_j+T_F} g_l(t)^2 \mathrm{d}t \geq \int_{t_j+T_F+T}^{t_j+2T_F+T} g_r(t)^2 \mathrm{d}t \tag{6.12}$$

[2]The proposed passivity criteria are denoted for the incident wave only. The conditions for the reflected wave are straightforward due to the similarity of the conditions (6.5) and (6.6).

with t_j the starting time of frame j.
The compression ratio is

$$\mathrm{CR} = \frac{k_F}{\dim(\mathbf{p})}. \tag{6.13}$$

Transparency is decreased by phase artifacts and reconstruction artifacts: The delay T_F inherent in the strategy and the signal error resulting from the approximation expressed by the dimension of the parameter vector $\dim(\mathbf{p})$. Compression is only achieved if

$$\dim(\mathbf{p}) < k_F \tag{6.14}$$

This expression unveils the trade-off that to be made when compressing haptic data with passive interpolative compression strategies. The larger the frame length k_F the more data can be saved, i.e. the higher the compression ratio, but the lower the transparency due to induced delay T_F (phase artifacts).
The advantages of the interpolative compression are:

1. A constant, freely adjustable data rate, i.e , the instantaneous compression ratio is equal to the average compression for all times. Hence, any communication bandwidth limits can be satisfied.
2. Arbitrary algorithms are possible as long condition (6.12) is satisfied.

6.2.3 Extrapolative Compression

The extrapolative strategy estimates future samples to a certain extent, called *estimation horizon*. The encoder works as follows: k_{EH} samples are estimated and a signal is constructed based on certain assumptions resulting in the parameter vector $\mathbf{p}$ transmitted over the network and reconstructed by the decoder. Every k_{EH} samples an estimation of the next k_{EH} samples is performed. The duration of the extrapolation horizon amounts to

$$T_{EH} = \frac{k_{EH}}{f_s}, \tag{6.15}$$

The compression principle is depicted in Figure 6.5.
The delay introduced by encoder and decoder depends on the algorithm but shall be short or can even be neglected. The delay introduced by the extrapolation procedure depends on the energy difference between original signal and estimated signal. According to conditions (6.5), (6.6) passivity of the amplitude change can be assured by forcing the extrapolative wave less or equal powerful than the difference between original wave and the preceding extrapolations starting from the beginning. Hence, the *passivity criterion for extrapolative compression schemes* becomes

$$\int_0^{t_j} (g_l^2 - g_r^2)\mathrm{d}t \geq \int_{t_j+T}^{t_j+T_{EH}+T} g_r^2 \mathrm{d}t, \tag{6.16}$$

with t_j representing the time when a new estimation is performed and T_{EH} as the length of the estimation horizon.
The compression ratio is

$$\mathrm{CR} = \frac{k_{EH}}{\dim(\mathbf{p})}. \tag{6.17}$$

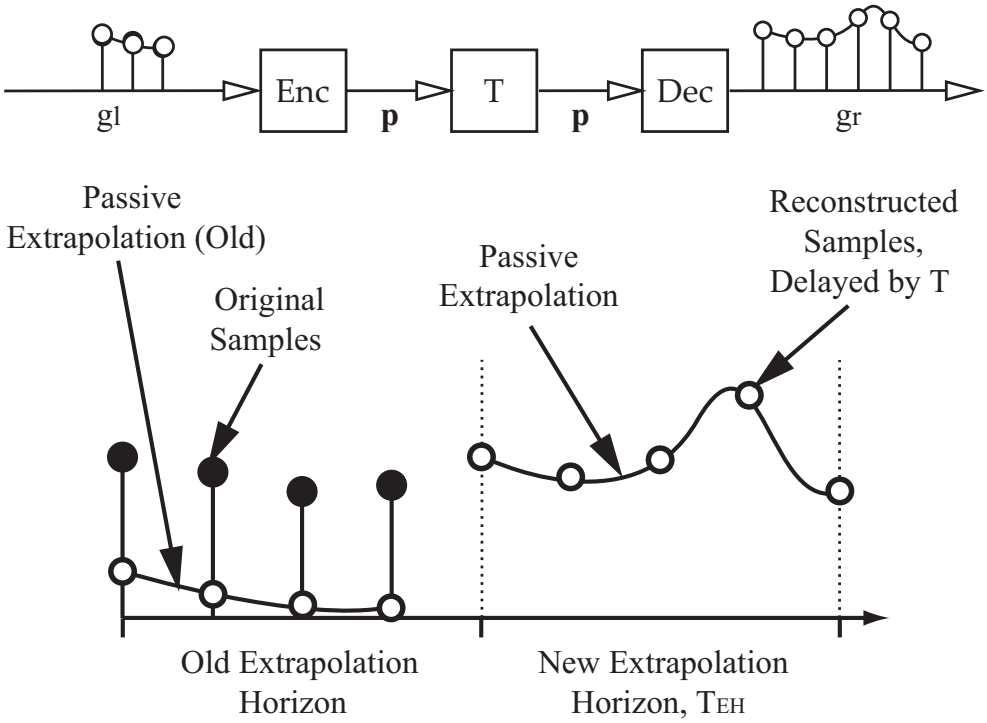


Figure 6.5: Principle of extrapolative compression strategies: The signal within the extrapolation horizon is estimated using the residual energy of the difference between precedent extrapolations and its real correlates.

Transparency is influenced only by the signal error resulting from the extrapolation. It is straightforward to show that compression is only achieved if

$$\dim(\mathbf{p}) < k_{EH} \tag{6.18}$$

The advantages of the extrapolative compression are:

1. A constant, freely adjustable data rate. Hence, any communication bandwidth limits can be satisfied.
2. No strategy-inherent delay. Hence, no phase artifacts will deteriorate transparency.
3. Arbitrary algorithms are possible as long condition (6.16) is satisfied.

6.2.4 Direct Compression

Direct compression algorithms are based on a strategy such that the compression is applied to each sample directly without using frames or estimation horizons.
The delay introduced by encoder and decoder depends on the algorithm but shall be short or can even be neglected. As it is a sample-based scheme, the delay introduced by the procedure is zero and no phase artifacts will deteriorate the transparency. The passivation of the amplitude change has to assure that the absolute value of the decoded sample has to be decreased or left unchanged compared to its original correlate $g_l(t)$

$$|g_l(t)| \geq |g_r(t+T)|\,. \tag{6.19}$$

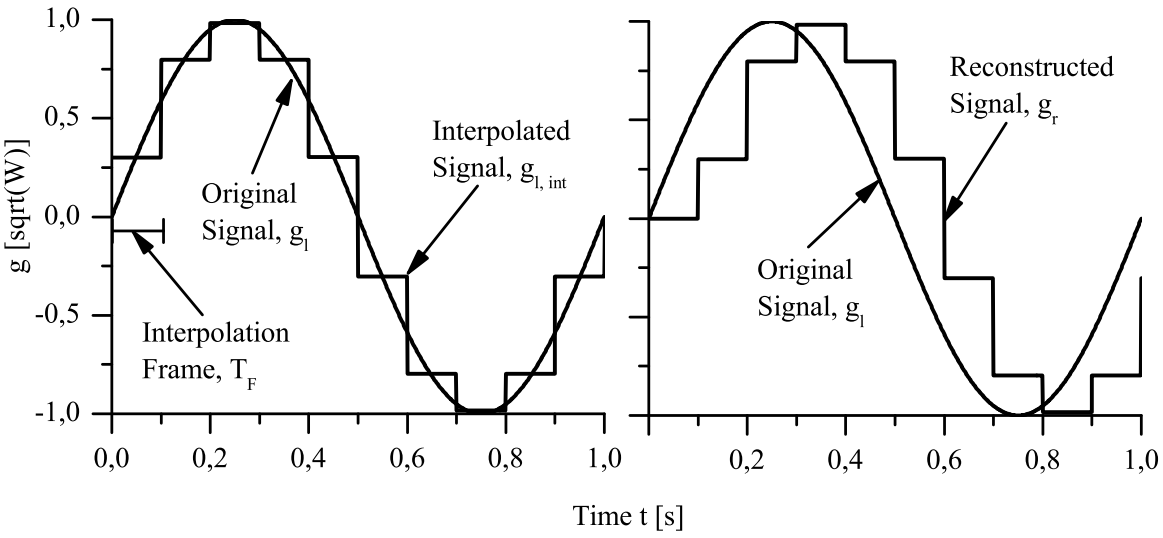


Figure 6.6: Operation mode of iDS (simulated at $T_F = 0.1\,\mathsf{s}$, $f_s = 1\,\mathsf{kHz}$): An interpolation is calculated as the mean of a wave signal in an interpolation frame (right diagram). The interpolation causes the delay T_F, hence, the reconstructed signal (left diagram) is delayed by T_F. (Network delay T=0.)

The advantage of the strategy is due to its direct character: No delays are induced.
An implementation of the direct compression scheme is deadband control and was proposed, together with the passivity condition, in [152] (read the Introduction for a short discussion). Dead band control reaches average data rate savings of 87%. However, the due to the compression principle no deterministic data rate can be defined and therfore given data rate limits cannot be maintained.

6.3 Algorithms

Based on the passivity criteria (6.12), (6.16), (6.19) arbitrary algorithms that passively compress haptic signals in wave domain can be implemented. Two different algorithms are explained. Passive interpolative downsampling uses the interpolative compression strategy. Passive extrapolative downsampling is based on the extrapolative compression strategy.
In Subsection 6.3.1 the interpolative algorithm is explained. In Subsection 6.3.2 the extrapolative algorithm is explained.

6.3.1 Interpolative Downsampling

The interpolative compression method proposed here is called *passive interpolative down sampling* (iDS). The main idea is to interpolate the original values within an interpolation

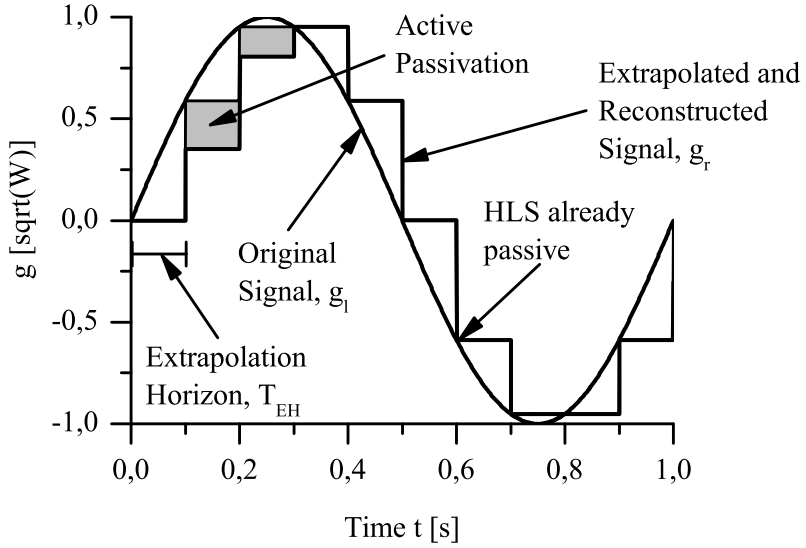


Figure 6.7: Operation mode of eDS (simulated at $T_F = 0.1\,\text{s}$, $f_s = 1\,\text{kHz}$): A new extrapolation horizon is calculated based on the energy difference between old extrapolations and its real correlates. If the last original sample before a new extrapolation satisfies the passivity criterion, it is taken for the extrapolation (HLS). Otherwise this sample is reduced in energy such that it satisfies condition (6.16). (Network delay is T=0.)

frame by their mean value

$$g_{l,int}(t) = \frac{1}{T_F} \int_{t_j}^{t_j+T_F} g_l(t) \quad \forall t = [t_j, t_j + T_F]. \tag{6.20}$$

The parameter vector contains a single element $\mathbf{p} = g_{l,int}$, the mean value over the frame length. Naturally, the mean value has minimal possible energy during the interpolation frame and, therefore, satisfies the passivity condition (6.12). The compression ratio is determined as

$$\text{CR}_{\text{iDS}} = k_F : 1. \tag{6.21}$$

Simulation results for interpolation frames $k_F = 100$ at a sampling frequency $f_s = 1\,\text{kHz}$ yielding $T_F = 0.1\,\text{s}$ are shown in Figure 6.6. Since the original samples are replaced by their mean, the compression ratio is CR= 100 : 1 and the data rate savings are 99%. Depending on the length of the interpolation frame higher frequencies will be filtered out. The compression algorithm has lowpass-filter characteristics. However, primarily, a transparency decrease is caused by the phase artifacts, induced by framing the signal.

6.3.2 Extrapolative Downsampling

An implementation using the extrapolative compression scheme is *passive extrapolative down sampling* (eDS). The main idea is to extrapolate within the extrapolation horizon T_{EH} by a single value. The extrapolation is performed in two modes. If the the most

recent value measured, already satisfies the passivity criterion for extrapolative compression schemes (6.16) than it is taken as extrapolation throughout the extrapolation horizon. This mode is known as hold-last-sample (HLS). If the HLS-value does not satisfy the passivity criterion, than a value reduced in energy is computed such that (6.16) is satisfied. The equations are

$$g_{l,ext}(t) = \left\{ \begin{array}{ll} g_l(t-1) & \text{if (6.16) holds,} \\ g_l & \text{such that (6.16) holds.} \end{array} \right\} \quad \forall t = [t_j, t_j + T_{EH}] \tag{6.22}$$

The parameter vector contains one element $\mathbf{p} = g_{l,ext}$. The compression ratio is

$$\mathrm{CR}_{\mathrm{eDS}} = k_{EH} : 1. \tag{6.23}$$

Simulation results for an extrapolation horizon of $k_{EH} = 100$ at a sampling frequency $f_s = 1\,\mathrm{kHz}$ yielding $T_{EH} = 0.1\,\mathrm{s}$ are shown in Figure 6.7. Since the extrapolation for each results in a single value, the compression ratio is CR= 100 : 1 and the data rate savings are 99%.
A new extrapolation value for the actual extrapolation horizon is calculated based on the energy difference between old extrapolation and its real correlate. As can be seen in Figure 6.7, the reconstructed signal is only retarded if not enough energy is available for the extrapolation to predict the original signal by HLS. Depending on the length of the estimation horizon higher frequencies will be filtered out. The algorithm has lowpass-filter characteristics. The performance of the eDS solely depends on the quality of the extrapolation.

6.4 Psychophysical Evaluation and Parametrization

The application of the proposed interpolative and extrapolative algorithms (iDS, eDS) causes phase and reconstruction artifacts, which have complex dynamical characteristics dependent on the target environment rendered. The higher the compression ratio the larger the deterioration of transparency due to the artifacts. (See Subsection 6.1.3 for details on transparency.)
Two general questions were answered by this extensive experimental study.

Q1: Which algorithm (iDS, eDS) yields higher performance?

Q2: Secondly, what is the highest compression ratio still leading to perceptual transparency of the compression algorithm?

Therefore, the two algorithms were applied to a telepresence system. Trials were performed at seven different compression ratios and in three different environments.
Based on the outcome of the simulation two hypotheses were tested.

H1: 'The performance is lower when telepresence is performed in a rigid environment than in a soft environment.' Since artifacts deteriorate transparency by decreasing the compliance of an environment (refer to equation (6.10)), the lowest detection threshold should be measured in an infinite stiff environment. In a softer environment the additional compliance increase by the artifacts should concealed.

H2: 'eDS provides higher performance than the iDS.' Due to the phase artifacts caused by the frame delay T_F of interpolative methods, iDS might reveal a lower detection threshold than eDS.

The study is performed using a 2AFC-task experiment and a high performance telepresence system. The information enables the human operator to deliberatively adjust the parameter of the LDC-algorithms to match his preferences for the presence system used in a certain application.
In Subsection 6.4.2 the method is explained in detail. Subsection 6.4.2 the results are presented. Finally, a discussion is conducted in Subsection 6.4.2.

6.4.1 Method

Presence System

A telepresence system is used that provides haptic feedback at high accuracy. The system consists of two identical linear actuators. One is used as HSI and equipped with a knob for the human operator to grasp. The other one is used as TO and its end effector that interacts with the remote environment, which represented by different springs. The springs were directly connected to the end effector of the TO and fixed to a support plate on the other side.
During psychophysical experiments participants were able to insert their answers using a joystick that provides them with the different answer possibilities. The system works under real-time conditions and is programmed by Matlab/Simulink. A detailed description of the HSI including hardware, software, and control structure is provided in Appendix A.2.

Participants

Twelve (12) students of the Technische Universität München participated in this study. All participants were right-handed. Participants were paid for participation. All participants delivered consistent results.

Stimuli

The stimuli were generated by the human operator actively by exploring the different remote environments by the telepresence system. Thereby, different parametrizations of the compression algorithms yielded differently transparent operation of the presence system. The interpolative algorithm iDS was parametrized by framelengths

$$T_F = [1; 2; 4; 7; 10; 15; 20]\,\text{ms} \tag{6.24}$$

($f_s = 1\,\text{kHz}$) resulting in compression ratios

$$\text{CR} = [1:1; 2:1; 4:1; 7:1; 10:1; 15:1; 20:1]. \tag{6.25}$$

The extrapolative algorithm eDS was parametrized by estimation horizons

$$T_{EH} = [1; 2; 4; 7; 10; 15; 20]\,\text{ms} \tag{6.26}$$

($f_s = 1\,\text{kHz}$) resulting in compression ratios

$$\text{CR} = [1:1; 2:1; 4:1; 7:1; 10:1; 15:1; 20:1]. \tag{6.27}$$

Three remote environments are used. Wood is used as a stiff environment with a compliance of $S = 0\,[\text{mm/N}]$. A spiral spring is used as soft environment with a compliance of $S = 13\,[\text{mm/N}]$. Three spiral springs are used to generate a medium-stiff environment $S = 14\,[\text{mm/N}]$.

Procedure

One trial consists of the sequential presentation of two stimuli: the standard and the comparison stimulus; the sequencing of standard stimulus and comparison stimulus differed randomly. Duration of each stimulus presentation was 2 s with an inter-stimulus interval of 2 s and an inter-trial interval of 4 s. The standard stimulus was always generated by the telepresence system with no compression algorithm applied (CR= 1 : 1). The comparison stimuli were based on the different parametizations (see above). Detection performance was assessed during one block for each compression scheme and each environment. Every combination of standard and comparison stimuli was randomly presented 10 times. Psychometric function were recorded for each algorithm-environment combination, i.e. 6 psychometric functions for each participant. The detection threshold (2.40) and its variability (2.42) ware computed.
The task was a 2AFC-detection task. Subjects had to compare both sequentially presented stimuli and to decide whether the second stimulus was more compliant than the first one. Proportion of correctly detected LDR-algorithm (further referred to as 'proportion correct') was recorded.
Participants were seated in front of the HSI with their right hand grasping the HSI. They were carefully instructed. A short training had to be completed, before the test session started. At the end of the experiment, participants were asked to fill in questionnaires assessing their demographical data.

Data Analysis

For each participant 6 psychometric functions were fitted using "psignifit" software, version 2.5.6, described in [29] and Matlab/Simulink. Hence, 72 psychometric functions were fitted in total. The detection threshold was considered as the CR that leads to the 75%-correct value of the psychometric function[3]

$$\text{CR}_{\text{DT}} = F_{\Psi}^{-1}(0.75), \tag{6.28}$$

which starts at proportion 0.5 (*guess rate*), for imperceivable stimuli, and settles at proportion 1.0, for arbitrary large stimuli.

[3]Since a 2AFC-task was used the performance criterium 'proportion correct' could be considered bias-independet. Participants most likely adopted a symmetric decision rule.

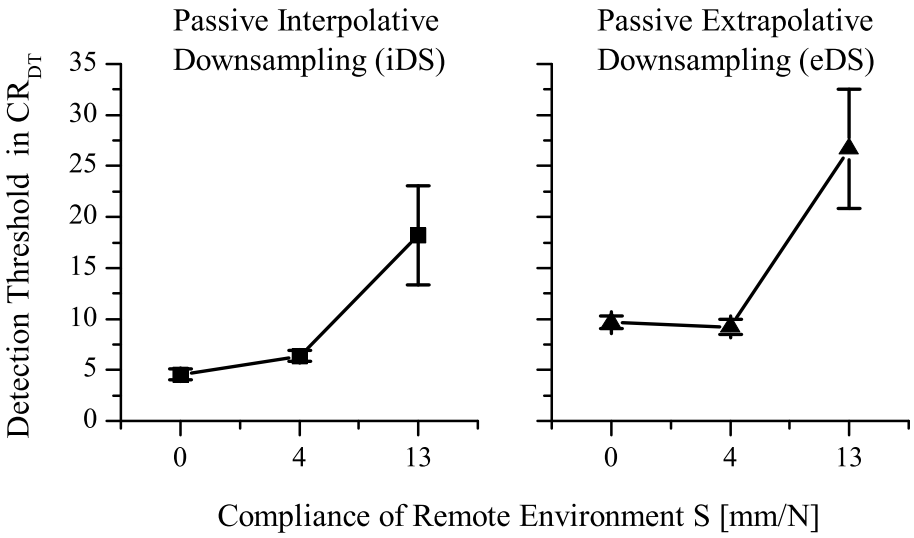


Figure 6.8: Detection threshold CR_{DT} of compression algorithms with standard error: Participants showed highest detection performance in the stiff environment, i.e. in stiff environments the compression algorithm has the highest impact on transparency. Furthermore, participants showed higher detection performance when detecting the interpolative algorithm, i.e. iDS had a higher impact on transparency than eDS. Based on this result the perceptual transparent compression ratio was defined to be $CR^*_{iDS} \approx 5:1$ and $CR^*_{eDS} \approx 10:1$.

6.4.2 Results

As a first result it could be stated that participants showed lowest detection performance in the soft environment. As depicted in Figure 6.8 applying iDS the stiff environment (0 mm/N) was detected at an average compression ratio of CR=4.56 : 1. eDS was detected at an average compression ratio of CR=9.67 : 1. Both values differed statistically significant for a paired t-test resulted in $t(22) = 6.17, p = 0.00003$. The detection threshold increased significantly for the interpolative algorithm when participants had to explore the medium compliant remote environment (4 mm/N) to CR=6.37 : 1. For the extrapolative algorithm the detection threshold did not differ significantly from the eDS-threshold in the stiff environment: CR=9.33 : 1. Also in the medium environment eDS was superior, for a paired t-test denoted a significance of $t(22) = 3.21, p = 0.0049$. An increase of detection threshold for both algorithms was observed for the soft environment with CR=18.21 : 1 (iDS) and 26.71 : 1 (eDS), however, the mean values did not differ significantly from each other due to high standard deviations.
The reliability of the detection thresholds depicted in Figure 6.9 supports this statement. The standard deviations (which are positively correlated to the reliability) increase with increasing compliance indicating that participants were more secure to give an estimate for the stiff environment than in the softer environments. Hence, H1 could not be falsified. (A 12×3-ANOVA revealed that the results in the different remote environments differed statistically significant. $F(2, 33) = 8.76, p < 0.05$ for the DT of eDS and $F(2, 33) =$

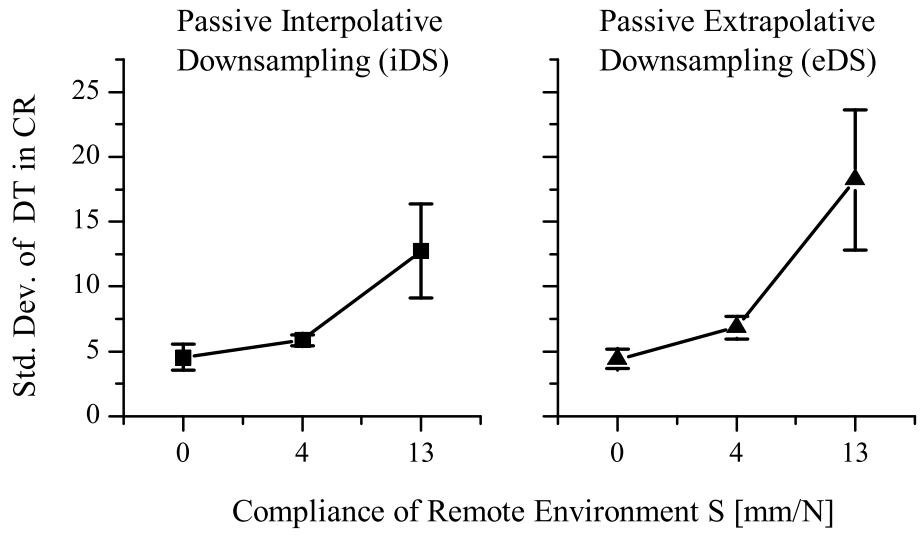


Figure 6.9: Reliability of detection thresholds of compression algorithms with standard error: Participants' reliability in detecting the compression algorithms were nearly equal, decreasing with higher compliances.

$5.29, p < 0.05$ for its variability. $F(2, 33) = 6.81, p < 0.05$ for the DT of iDS and $F(2, 33) = 4.10, p < 0.05$ for its variance.)
The second result was already mentioned in the first result but has to be expanded. eDS is less detectable than iDS. The detection thresholds in all three environments are higher using eDS instead of iDS. Since for eDS values did not differ significantly for the stiff and the medium environment the detection threshold was defined to be the average of both thresholds. Hence, the detection threshold for eDS was CR= 9.5 and thereupon the optimal configuration for eDS was defined to be

$$\mathrm{CR}^*_{\mathrm{eDS}} \approx 10:1 \quad (\mathrm{DRS}_{\mathrm{eDS}} = 90\%) \tag{6.29}$$

The minimal average detection threshold for iDS was CR= 4.56, hence the optimal configuration for iDS is defined to be

$$\mathrm{CR}^*_{\mathrm{iDS}} \approx 5:1 \quad (\mathrm{DRS}_{\mathrm{iDS}} = 80\%). \tag{6.30}$$

As already mentioned in the first result the results were highly statistically significant ($t(22) = 6.17, p = 0.00003$). As the stiff environment was the environment were the artifacts were most salient (first result) and the compression ratios stated above are the highest compression ratio that lead to a perceptually transparent compression, the performances (6.29), (6.30) can be considered as the overall performance of the algorithms. H2 could not be falsified.
As a third result the mean psychometric functions depicted in Figure 6.10 could serve as a parametrization guide. Therefore, the psychometric functions of the 2AFC-tasks have to be corrected for guessing by rescaling them from proportion $0.5 \ldots 1.0$ to $0 \ldots 1.0$. The original scale starts from proportion 0.5 called *guess rate*, which is the proportion participants will settle to if they cannot recognize the stimulus and, therefore, start guessing

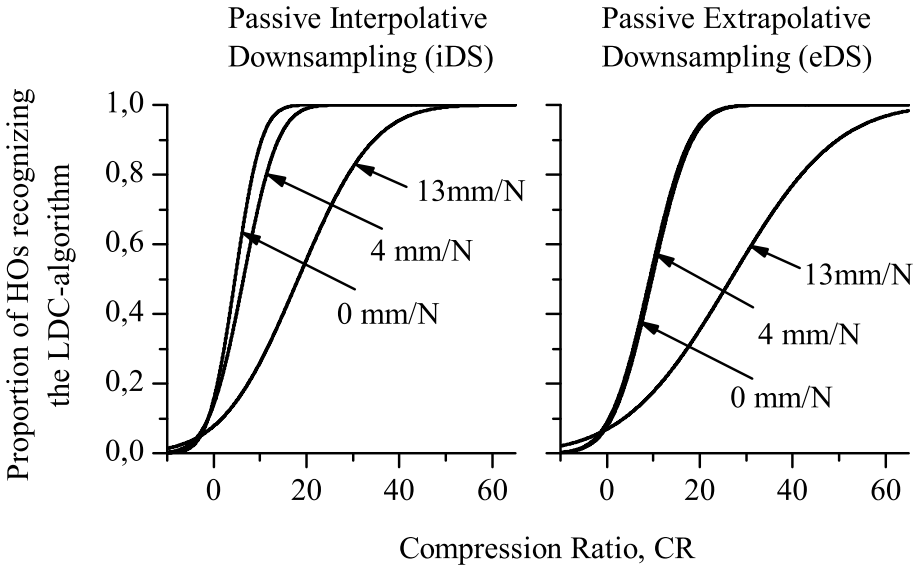


Figure 6.10: Perceptual transparency of the compression algorithm: Based on the experimental results the performance of each algorithm can be estimated for different environments using the averaged, rescaled psychometric functions. The higher the compression ratio, the higher the proportion of the algorithm detected. The extrapolative algorithm shows higher performance indicated by a higher compression ratio at proportion 0.5.

between the two alternatives. The functions give clear advice how to parametrize each of the algorithms. The proportion of human operators being able to detected the compression algorithm is given depending on the compression ratio. Thereby, the functions also give information about the detection of compression ratios which were not part of the experiment, by using the psychophysical knowledge that underlying principle of detection performance has a Gaussian or approximately Gaussian shape.

6.4.3 Discussion

The study assessed the performance of human operators that perceived remote environments through a telepresence system with passive haptic compression schemes applied in command and reflection path. Two algorithms were analyzed, passive interpolative downsampling (iDS) and passive extrapolative downsampling (eDS). A high performance telepresence system was used. Detection thresholds and their variability were assessed by recoding psychometric functions (proportion 'correct') via a 2AFC-Task.
The results gave reliable information to which proportion compression algorithms at different compression ratios could be detected. eDS provided a deterministic perceptually transparent compression with data rate savings of $\mathrm{DRS}_{\mathrm{eDS}} \approx 90\%$ in the worst case, i.e. in the infinitely stiff environment were operators were most sensitive. iDS provided deterministic data rate savings of $\mathrm{DRS}_{\mathrm{iDS}} \approx 80\%$ in the same conditon. Therefore, eDS can be considered superior to iDS (Q1) and perceptual transparent data rate savings are 90%

(Q2). The averaged psychometric functions for each environment give detailed information how to parametrize the compression algorithms, if data rate savings are aspired that exceed the perceptually transparent paremetrizations.

6.5 Conclusion

Imperceivable, deterministic network traffic reduction of 90% is achieved by applying the proposed passive, explorative compression algorithm for haptic information.

Summary

In this chapter new lossy methods to compress haptic data as exchanged in telepresence systems are proposed.
The first part deals with the analysis of the needs of haptic lossy data compression (LDC). Passivity, high compression ratio, and perceptual transparency were set as the goals of the upcoming algorithms. To achieve passivity, the algorithms were designed in Scattering domains and had to be passive themselves. It was proposed to define perceptual transparency by the empirical threshold, a statistics from signal detection theory widely used in psychophysics, which defines the smallest detectable impact of a certain stimulus. A classification was developed based on the possibly inserted delay that structured the compression methods into frequency-based, interpolative, extrapolative, and direct methods.
The second part deals with the development of different passivity criteria. Except for the frequency-based method, which was analyzed to be inappropriate for haptic data compression, passivity criteria were proposed for interpolative, extrapolative, and direct compression methods. The proposed criteria allow arbitrary implementations of compression algorithms such that they will not affect the stability of the overall system.
In the third part two compression algorithms are developed using the interpolative and the extrapolative compression strategy. The interpolative compression algorithm is called passive interpolative downsampling (iDS), the extrapolative strategy is called passive extrapolative downsampling (eDS). Simulations were conducted to illustrate the different operation modes.
The last part presents an evaluation of the proposed algorithms. An extensive psychophysical study was conducted recording psychophysical functions of twelve (12) subjects by 2AFC-tasks.

Scientific Contribution

The proposed methods open new ways to haptic lossy data compression. Following benefits are provided:
1.) The proposed classification gives first advice about the performance of an algorithm with respect to its perceptual transparency and its data rate savings. Algorithms were divided into frequency-based, interpolative, extrapolative, and direct algorithms. Frequency-based and interpolative algorithms store the original data into a frame thereby inducing

a delay of the frame-length. Therefore, they were termed 'frame-based' methods. Their compression potential is high since all information is present about the data to be compressed. However, the inherent delay produces phase artifacts that deteriorate termed 'sample-based'. Their compression potential is lower than of frame-based algorithms since they do only have indirect information from the past samples about the data to compress. But since they cause no static delay phase artifacts are smaller and perceptual transparency is more likely achieved.
2.) The proposed passivity criteria were sufficient and necessary for passive signal-processing algorithms in Scattering domain. Passivity criteria were proposed for the interpolative, extrapolative, and direct compression (or signal processing). The criteria allow the design of arbitrary algorithms, which were passive as long as they satisfy one of these conditions.
3.) Two algorithms were implemented. The interpolative algorithm (iDS) performs a interpolation of an adjustable part of the signal (interpolation frame) by its mean value. The extrapolative algorithm used a modified hold-last-sample strategy to forecast the original signal within the extrapolation horizon. If the last sample satisfies the passivity condition for extrapolative algorithms, it was used for the next extrapolation. If it does not satisfy the condition, it was reduced in energy until fulfills the condition. Both algorithms provided a constant, freely adjustable compression ratio, i.e. the instantaneous compression ratio is equal to the average compression ratio for all times. Hence, any hard communication bandwidth constraints can be satisfied.
4.) The psychophysical evaluation clearly indicated that eDS was superior to iDS. The extrapolative algorithm achieved 90% data rate savings without perceivably imparing transparency, the interpolative algorithm achieved only 80%. This was predicted by the classification. Hence, eDS provided the same compression performance than deadband control (see Introduction). However, deadband control only provides an average compression ratio (instantaneous compression ratio can drop to zero if signal changes exceed the deadband in succession) and therefore cannot be deployed if hard bandwidth constraints exist. The averaged psychometric functions give clear advice how to parametrize the algorithms for different environments if higher data rate savings were aspired. The reason why the human operator did not perceive the algorithms when parametrized perceptually transparent can only be surmised. Most probable, the transparency deteriorations caused by motor dynamics and flexibility of the presence system covered the transparency deteroration by the compression algorithm. Since a high fidelity presence system was used with low flexibities and low mass, it is assumed that the perceptual transparent configurations are even higher for less perfromant presence systems.

Open Problems

Since the proposed passivity criteria are generic and only two algorithms have been implemented improvements could be achieved by tackling the following problems.
Enhanced estimation capabilities of the compression algorithms could increase the perceptually transparent compression ratio. eDS and iDS only generate first order approximations of the original signal. The quality of the estimation could be improved using a higher order approximation. On the same time more information of the signal should be used to parametrize the approximation. E.g. the iDS could be improved using a second or third

order interpolation. Thereby, more information would be used about the original signal, however, the compression ratio would decrease. To achieve perceptually transparent results psychophysical knowledge should be incorporated. JND-information about force and velocity should be incorporated into the approximations.
The dynamical impact of the compression algorithms should be analyzed in more detail. Compression algorithms induce phase artifacts (delay) and reconstruction artifacts that both deteriorate transparency. The impact phase artifacts in scattering domain is already analyzed. It is known that the target environment of the presence systems influences the dynamical impact of the artifacts. The impact is modeled in case the environment represents free space and an infinite stiff wall (see Subsection 6.1.3. However, for more complex environments no dynamical models of the deterioration exist (in [158] basic models are provided that could serve as starting point for a more detailed analysis) The impact of the reconstruction artifacts also deserves a closer look. Due to the downsampling character of the proposed algorithms higher frequencies are impaired. However, quantitative results have to be generated to give clear advice about the dynamical properties of the overall presence system.

7 Conclusion

7.1 Results in a Nutshell

Perception-Based Utility Optimization

In Chapter 3 a new method is proposed for the evaluation of presence in presence systems. A high degree of presence is conceived as an objective that has to be maximized amongst other objectives. Objectives that are conflicting with a high degree of presence are efforts and resource savings, i.e. display resolution, stability margin, etc. Multiobjective optimization theory is used for the mathematical structure of the problem to maximize the conflicting objectives. It is proposed that the non-dominated alternatives in the objective space are identified by certain psychological or psychophysical methods. Three different utility functions are proposed that represent different preference structures. The utility functions are customized for applications presence systems. Eventually, the optimal combinations of presence and effort/resource savings are identified based on the solution of the optimization problem. The solution of the optimization problem is discussed with respect to the uncertainty induced by the experimental procedure. An example using a visual-haptic presence system is provided. The results were published to a large extent in [63; 68; 160].

Compliance Perception

In Chapter 4 three explorative studies are presented analyzing an operator's visual-haptic perception of compliant objects. The objects are artificially rendered cubes by a visual-haptic presence system. Crossmodal difference thresholds were assessed in the first two studies. Crossmodal thresholds were found to be significantly higher than unimodal or bimodal difference thresholds. Values range from JND= 128% to 85% for crossmodal comparisons and 35% to 27% in the unimodal or bimodal case. Crossmodal difference thresholds of compliance perception have never been assessed before. A method effect was found and analyzed between the adaptive staircase method and the method of adjustment, with the latter providing more accurate results. In the third study fusion of redundant information was analyzed. The result indicated that fusion between haptic and visual compliance estimates did not occur. Furthermore, it was found that the absolute JND increases with increasing compliance. Hence, the relative JND for compliance perception remains nearly constant and it can be stated that the Weber fraction also applies when perceiving compliances. The results were published to a large extent in [70; 78; 79].
In Chapter 5 a new process model was developed for visual-haptic compliance perception, based on the results of the explorative studies presented in Chapter 4. The model was verified by two experiments. The first experiment unveiled that combination processes of

non-redundant position and force information take place before possible fusion processes. The second experiment unveiled that visual compliance information is fused to the haptic compliance information. However, this fusion process is not robust, since an additional perceptual process decreases reliability if conflicting information is perceived. This process can turn fusion into no fusion or confusion. The final PSE is biased by the visual compliance. Eventually, it could be stated for the first time, that visual-haptic compliance perception is based on the intermodal combination of haptic information and that visual and haptic information are fused optimally.

Haptic Data Compression

In Chapter 6 comprehensive research on lossy haptic data compression with perceptual coding is presented. A new classification is proposed that structures compression methods especially suited for haptic data. Based on this classification passivity criteria are proposed that allow the implementation of arbitrary compression algorithms not affecting the stability of the overall presence system. The passivity conditions are formulated in the Scattering domain, which is a transformation that passivates constant communication delays of force and velocity information. Two new algorithms are proposed using the interpolative and the extrapolative compression strategy. The interpolative algorithm achieves the reduction using a passive average of the original signal. The extrapolative algorithm achieves the reduction using a passive hold last sample strategy. Both algorithms are parametrized perceptually transparent by an extensive psychophysical evaluation. The interpolative algorithm reaches constant data rate savings of 80% without being detected by the human operator. The extrapolative algorithm reaches constant data rate savings of 90%. This represents today's highest performance in haptic data compression. Further parametrization advice is provided if perceptual transparency is not necessary. The results were published to a large extent in [49; 65].

7.2 Outlook

This thesis deals with visual-haptic presence systems and provides new results for presence measurement, compliance perception, and haptic data compression. Thereby, a consequent interdisciplinary approach was chosen to tailor the results to the human operator. The rigorous and close combination of engineering methods (optimization theory, control theory) and psychological methods (psychophysics, statistics) is new for research on presence systems. As the different, highly innovative results state the methodology is efficient. Future directions for interdisciplinary research on multimodal or haptic presence systems should contain the following issues.
1.) Method effects have to be analyzed in more detail. As shown in Chapter 4 the discrimination performance represented by the JND can vary according to the assessment method and the procedure. As especially haptic perception is still a young research field, a detailed categorization of the different psychophysical methods does not exist. As a consequence, experimental results vary tremendously and are hardly suitable for the comparison of the performance of human-robot interaction. More research should be spent on the effects of

the single methods subject to the assessment of haptic perception. A detailed description of what the methods measure should be provided in an understandable context for psychologists and engineers. Even a standardization should be pursued defining a certain assessment method and a certain procedure to evaluate human-machine interaction in presence systems on common ground.
2.) Models for haptic perception should be developed in more detail. Only two models exist for haptic perception. The first model is the assumption of a constant JND, this model is called *Weber Fraction*. The Weber Fraction applies to a large range of many stimuli, but is only a coarse description. The second model is the maximum-likelihood estimation according to [27], which describes the fusion of bimodal information. It has been successfully applied to predict visual-haptic perception, but cannot be deployed to describe haptic stimuli alone. For the perception of visual and acoustic stimuli, models exist (psychovisual models, psychoacoustic models) that replicate the perception of those stimuli in much higher resolution. These models are successfully applied to different implementations, e.g. lossy visual and acoustic compression algorithms. The need for high resolution models for haptic perception is obvious. Since haptic human-robot interaction covers a large range, from interaction with small robots to interaction with large, powerful devices, models are needed that predict the perceptual performance of a human operator in more detail.
3.) Kinesthetic-tactile perception should be analyzed in more detail. Haptic perception involves the perception of kinesthetic, tactile, and thermal stimuli. In this theses haptic stimuli were constrained to kinesthetic haptic stimuli only. However, especially the interplay of kinesthetic and tactile perception is crucial for human's perception of mechanical environments. Hence, since the utmost intention of haptic displays is the veridical display of a certain environment, the concurrent display of kinesthetic and tactile stimuli should be aspired. Therefore, the impact of combination and fusion processes of kinesthetic and tactile stimuli should be analyzed.
4.) Control algorithms should be developed especially to incorporate psychohaptic models. The performance of haptic displays is very sensitive to the used control structure. Haptic displays can even become instable. Hence, the veridical display of haptic stimuli is only possible if controllers exist that use information about the perceptual capabilities of the human operator. In Chapter 6 a method is proposed how data compression algorithm can be tuned perceptually transparent by an psychophysical evaluation. Although, this method has proven effective, it would be more desirable to incorporate a pychohaptic model directly into the compression algorithm. This would make a psychophysical evaluation superfluous and would provide more flexibility for the control systems engineer, since she/he can define the trade-off between stability and performance more directly.
5.) Hardware should be developed especially to generate dexterous kinesthetic and tactile stimuli for the human hands. Haptic perception focuses to a large extend on the human hand. Exploration and manipulation tasks are utmost performed by bimanual movements. Also for the analysis of haptic human perception, procedures that involve human hands are preferably used. However, the current hardware available for manipulations only provides limited dexterity. Furthermore, tactile stimuli are mostly excluded. Hence, the quality of technology-mediated experiences in presence systems will be significantly increased, if displays for kinesthetic and tactile stimuli reach a similar fidelity as visual or acoustic displays already do.

Appendix Presence Systems

A.1 Presence System 1: Visual-Haptic VR-System with Admittance Control

A visual-haptic presence system in VR structure for basic research on perception is described. The main purpose of the system is to render a mechanical characteristic S of a virtual cube. The system is composed of a TFT-screen and two SCARA-robots assembled in a rack to display a virtual environment to a sitting operator. The kinematical configuration of the SCARA-robots allows for gripping movements. The control structure is an admittance control scheme, which is suited to render non-rigid environments.
In Subsection A.1.1 the used hardware and software are described. In Subsection A.1.2 the dynamical structure and the control concept are presented.

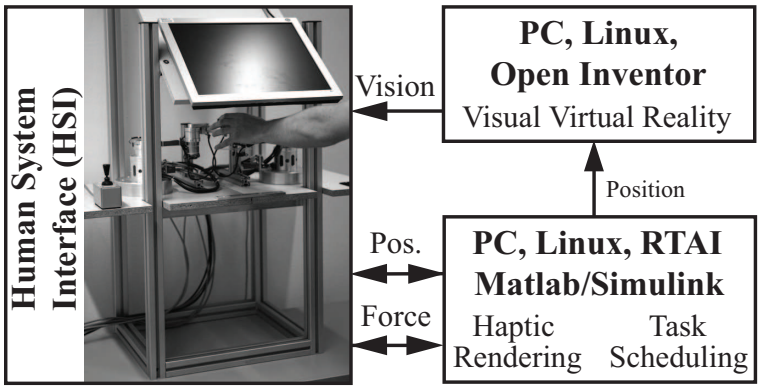


Figure A.1: Presence system 1: The system consists of a HSI and real-time processing unit. Visual and haptic information is exchanged and positions and forces are measured.

A.1.1 Hardware and Software

Haptic information is exchanged via a haptic interface comprised of two self-made SCARA robots providing a single degree of freedom each. See [161; 162] for a detailed description of the hardware development. The system interacts with index finger and thumb enabling the operator to perform gripping movements. High fidelity components like *Maxon* motors and *Harmonic Drive* gears enable best possible control. Workspace is about 80 mm and maximum force is about 50 N. Position information is measured by angle encoders and force is sensed by strain gauges attached to both robot links. Visual information is provided by a TFT screen. Thereby, the compliant environment is represented by a gray cube squeezed by two orange spheres (on opposed cube sides) representing finger positions. See Figure A.1

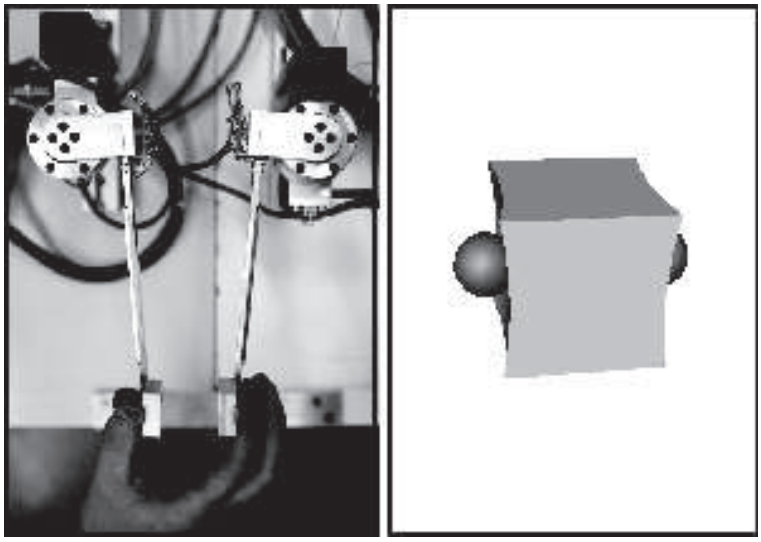


Figure A.2: Haptic and visual feedback: The haptic feedback renders a compliant cube to be explored by thumb and index finger. In the visual feedback fingers are replaced by orange spheres.

for an illustration of the overall system and Figure A.2 for close ups on the two types of feedback. The TFT screen is slanted by 40° and mounted in the line of sight to the hand enabling participants to look at the display as if there were looking at their hand [1].
The system is connected to a PC running *RTAI*-RealTime Application Interface for Linux. SCARA sensor signals are recorded by a *Sensoray 626* DAQ-Card providing 16 bit sensing resolution. Signal processing algorithms are implemented as *Matlab/Simulink* models with real-time code generated automatically. The system operates at 1 kHz sampling frequency. Measured positions are transferred to a second PC running the visual VR programmed in *Open Inventor*.

A.1.2 Dynamics and Control

The identical robots of the HSI are controlled independently using the same admittance control scheme (see Figure A.3 for kinematical configuration). In the following, the concept is explained using a single robot system without loss of generality. Kinematical transformations (forward kinematics, inverse kinematics) mapping torques T to forces f are omitted for simplicity.
For dynamics consider a mechanical robot with a single translational degree-of-freedom. The dynamical equation is given by

$$M_h\ddot{x}_h + D_h\dot{x}_h + K_h x_h + n_h = g_h - f_o^e, \tag{A.1}$$

where M_h, D_h, K_h denote mass, damping, and stiffness of the HSI and $n_h \in \mathbb{R}$ denotes the nonlinear dynamics of the HSI. Robot force $g_h \in \mathbb{R}$ depends on motor torque T and on

[1] The tool transformation should have no influence on the dynamics of the gripping movement, if participants are given a learning phase (e.g. see [163])

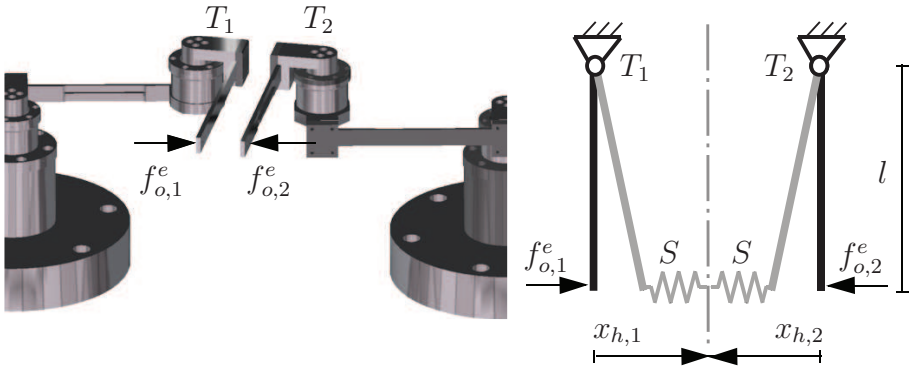


Figure A.3: Kinematical structure of the haptic display: Two SCARA robots haptically render compliant cubes for gripping movements.

link length l, respectively. The force f_o^e results from the interaction between operator and HSI and is measured by the force sensors. The position of the end effector is denoted by x_h. Input-output linearization [164] and force compensation is achieved by commanding

$$g_h = f_h^m + n_h + f_e^o. \tag{A.2}$$

The resulting linear dynamics are

$$M_h\ddot{x}_h + D_h\dot{x}_h + K_h x_h = f_h^m, \tag{A.3}$$

where f_h^m is the new motor force of the linearized HSI. Force compensation by adding the measured force f_e^o directly on the control output is feasible if only low to medium forces are applied and tho motor output is not exceeded.
A PID controller, $C : U \to M$, realizes the control signal f_h^m according to the position difference of robot and stimulus compliance

$$f_h^m = C[x_s - x_h], \tag{A.4}$$

where the square brackets indicate that C contains differential and integral operations. The HSI directly interacts with the human operator. The velocity of the HSI and the velocity of the operator's fingers are opposite, hence

$$x_h = -x_o. \tag{A.5}$$

The dynamics of the robot interacting actively with the human operator is described by

$$f_o = M_o\ddot{x}_o + K_o x_h + f_o^m, \tag{A.6}$$

where M_o, K_o denote mass and flexibility of the operator's fingers and is f_o^m is the force actively exerted by the human operator impeded by the force f_o that mediates the stimulus compliance (VR displayed to the operator).
The dynamics of the stimulus compliance is described by the admittance $S : U \to M$ (force input, position output), which represents compliance according to Hooke's law

$$x_s = Sf_o^e, \tag{A.7}$$

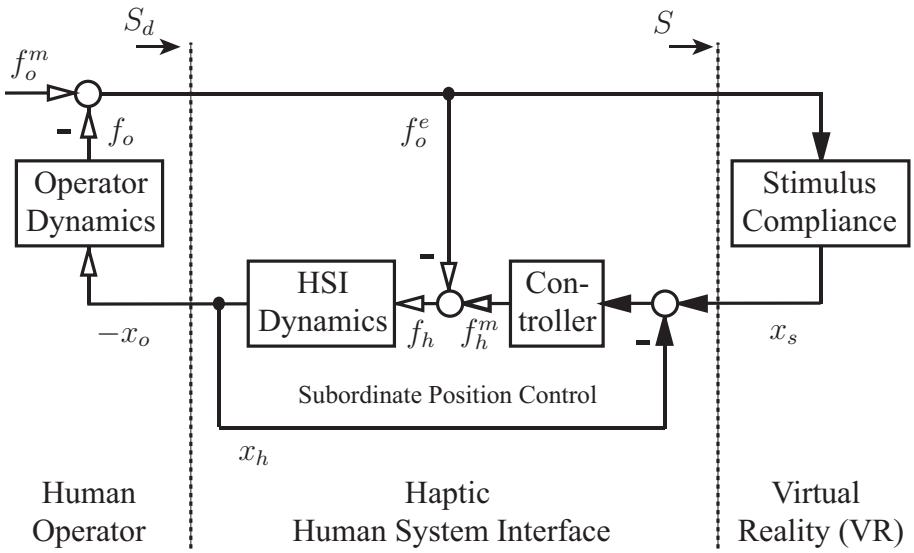


Figure A.4: Control of presence system 1: Different compliances are rendered using a high-fidelity robot driven by admittance control. The displayed stimuli compliances S_d show nearly no differences to the commanded stimuli compliances S. (Hollow arrows represent physical interaction, filled arrows represent signal processing.)

where S [mm/N] is the compliance whose perception is addressed in this study. The control concept employing inner position control driven by a VR with force reference is called *admittance control* explained in Section 2.4. It is best suitable for rendering non-rigid environments like compliant environments. Minimum compliance (= maximum stiffness) that can be rendered is $S = 0.14\,\mathrm{mm/N}$.
A block diagram of the human operator interacting with the admittance-controlled HSI is depicted in Fig. A.4. Hollow arrows depict physical interactions, filled arrows are used for signal interactions. All subsystems are considered linear(ized) and time-invariant. The fidelity of the VR depends on dynamics and control of the HSI. The robot is light weighted, dynamics of the motor current control are negligible, and velocities are small (i.e. friction effects negligible). Consequently, the transparency of the system can be assumed as nearly ideal and the displayed dynamics S_d can be considered equal to the dynamics of the VR

$$S_d = S. \tag{A.8}$$

A.2 Presence System 2: Haptic Telepresence System with Passivated COM

A haptic presence system in telepresence structure for basic research on perception and control is described. The system is composed of two linear actuators that enable fast responses, high forces, and friction compensation to render a broad bandwidth of remote environments from high stiffnesses to free space. The control structure provides force control of the HSI and velocity control of the TO (velocity-force architecture). It also features a passivation of the COM for constant delays.

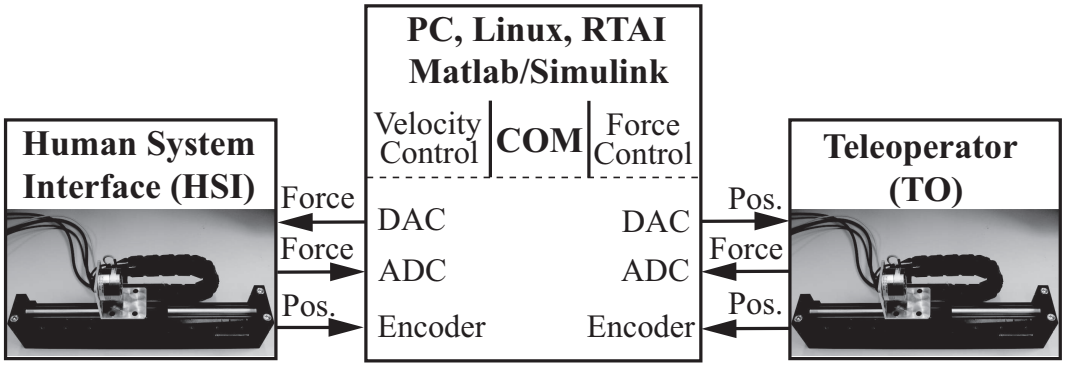


Figure A.5: Presence system 2: The telepresence system consists of a haptic HSI, a haptic TO, and a real-time processing unit that emulates the COM. Positions and forces are measured and exchanged. The COM is passivated for constant delays.

In Subsection A.2.1 the used hardware and software are described. In Subsection A.2.2 the dynamical structure and the control concept is presented.

A.2.1 Hardware and Software

The two identical linear actuators are mounted lengthwise next to each other on a table. One actuator is used as HSI and equipped with a knob for the human operator to grasp. The other one is used as TO and its end effector interacts with the remote environment. Remote environments with arbitrary impedances can be attached to the end effector of the TO.
The system is connected to a PC running *RTAI*-RealTime Application Interface for Linux. SCARA sensor signals are recorded by a *Sensoray 626* DAQ-Card providing 16 bit sensing resolution. Signal processing algorithms are implemented as *Matlab/Simulink* models with real-time code generated automatically. The system operates at 1 kHz sampling frequency.

A.2.2 Dynamics and Control

Due to the simple linear kinematical construction of the actuators the models for the HSI and the TO can be restricted to one translational degree-of-freedom.

The Human Operator and the HSI

Consider the dynamics of a human operator simplified as a linear system having one translational degree-of-freedom

$$-M_o\ddot{x}_h - K_o x_h = f_o^m - f_o^e, \tag{A.9}$$

where M_o, K_o denote mass and stiffness of the operators arm, f_o^m denotes the muscle torque, and f_o^e the external torque. Hence, the force exerted on the human arm is

$$f_o = M_o\ddot{x}_h + K_o x_h. \tag{A.10}$$

Consider furthermore the HSI as mechanical robot with one translational degree-of-freedom. It is assumed to be stiff and it is actuated by an electrical motor. Motor and gear are modeled as uniform bodies having their center of mass on the translation axis. Force and velocity sensors have zero weight and ideal responses.
The movement direction of the human operator's arm is opposite to the movement direction of the HSI

$$v_h = -v_o. \tag{A.11}$$

Hence, the dynamical model of the HSI driven by the human operator is given by

$$M_h\ddot{x}_h + D_h\dot{x}_h + K_h x_h + n_h = g_h - f_h^e, \tag{A.12}$$

where M_h, D_h, K_h denote mass, damping, and stiffness of the HSI and n_h comprises non-linear and linear forces based on gravity and friction. The parameter g_h represents the force generated by the motor of the HSI. The parameter f_h^e represents the force exerted on the human operator and is measured by a sensor mounted on the link.
Input-output-linearization (computed-torque method) according to [164] is conducted by commanding a force

$$g_h = f_h^m + n_h, \tag{A.13}$$

where f_h^m is the new force generated by the linearized HSI. The linearized dynamics of the HSI is then described by

$$M_h\ddot{x}_h + D_h\dot{x}_h + K_h x_h = f_h^m - f_h^e. \tag{A.14}$$

Consequently, the force moving the HSI becomes

$$f_h = M_h\ddot{x}_h + D_h\dot{x}_h + K_h x_h. \tag{A.15}$$

The motor of the HSI imprints the force reflected from the remote environment passing the COM, f_e^c. Force control C_f assures that the human operator feels the reflected force despite the dynamics of the HSI,

$$f_h^m = C_f[f_e^c - f_h^e], \tag{A.16}$$

where the square brackets indicate that C_f contains differential and integral operations of PID-control.

The COM

Exchanged command and feedback signals afflicted with communication delay can lead to instability of the overall system. The passivity paradigm is applied to passivate the COM and to prevent the system going instable by constant communication delay.
By definition a passive system does not generate energy. An observable single-input-single-output system with no initially stored energy is passive, if the input power exceeds the output power for all times

$$\int_0^t p_{in}(\tau)\mathrm{d}\tau \geq 0, \qquad \forall t \geq 0. \tag{A.17}$$

Where p_{in} represents the netto instantaneous power input. Power is defined as scalar product of effort (e.g. force) and flow (e.g. velocity) variables and therefore

$$p_{in} = \dot{x}_o f_t^c - \dot{x}_t^c f_t^e \tag{A.18}$$

is the power entering the COM-two-port.
In case of constant communication delays, the active COM can be passivated using the scattering transformation explained in Section 2.4. The bijective transformation of effort and flow variables into scattering variables is given by

$$\begin{aligned} g_l &= \frac{b\dot{x}_o + f_t^c}{\sqrt{2b}}, \quad & h_l &= \frac{b\dot{x}_o - f_t^c}{\sqrt{2b}}, \\ g_r &= \frac{b\dot{x}_t^c + f_t^e}{\sqrt{2b}}, \quad & h_r &= \frac{b\dot{x}_t^c - f_t^e}{\sqrt{2b}}. \end{aligned} \tag{A.19}$$

Where $g_l,\ h_r \in \mathbb{R}$ denote the incident waves and $g_r,\ h_l \in \mathbb{R}$ denote the reflected waves. The parameter $b > 0$ represents the wave impedance and can be chosen arbitrarily. The passive dynamics of the COM disturbed by constant delay T are finally given by

$$\begin{aligned} g_r(t) &= g_l(t-T), \\ h_l(t) &= h_r(t-T). \end{aligned} \tag{A.20}$$

The TO and the Remote Environment

Consider a mechanical robot, the TO, in same configuration as the HSI, replacing the human operator in the remote environment. Consequently, the dynamical equation is given by

$$M_t\ddot{x}_e + D_t\dot{x}_e + K_t x_e + n_t = g_t - f_t^e \tag{A.21}$$

where M_h, D_h, K_h and n_t denote again the linear dynamics and the nonlinearities of the robot. $g_t \in \mathbb{R}$ and $f_e \in \mathbb{R}$ denote the motor force and the external force, respectively. Input-output linearization [164] is achieved by commanding

$$g_t = f_t^m + n_t, \tag{A.22}$$

The resulting linear dynamics are

$$M_t\ddot{x}_e + D_t\dot{x}_e + K_t x_e = f_t^m - f_t^e, \tag{A.23}$$

where f_t^m is the new motor force of the linearized TO. A velocity controller, $C_v : U \to M$, realizes the command signal $\dot{x}_o^c$ according to

$$f_t^m = C_v[\dot{x}_o^c - \dot{x}_e]. \tag{A.24}$$

The squared brackets denote the differential and integral operations by the controler. The TO is serially connected to the remote environment described by M_h, D_h, K_h and by definition the velocity of the TO and the velocity of the remote environment are opposite

$$\dot{x}_\mathrm{e} = -\dot{x}_\mathrm{t}. \tag{A.25}$$

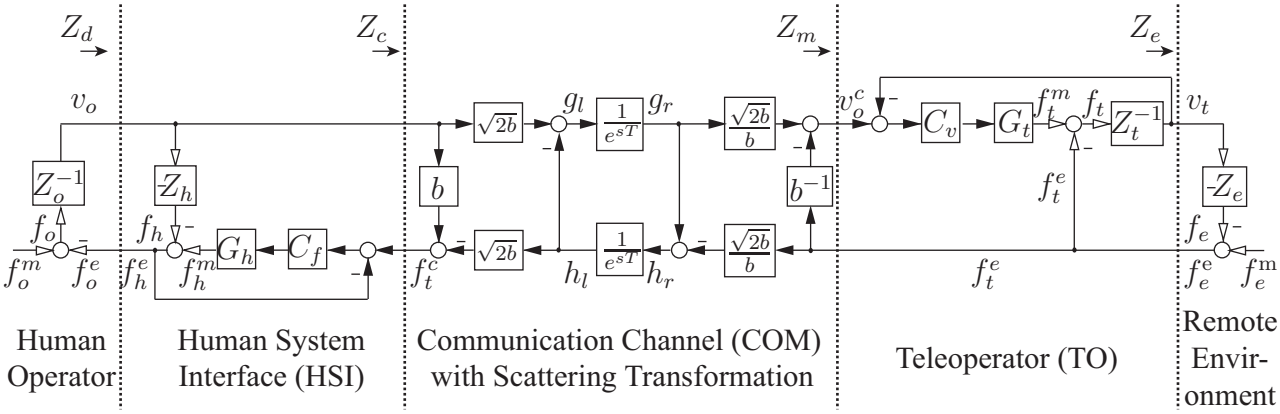


Figure A.6: Control of presence system 2: Arbitrary remote environments were rendered using high-fidelity robots connected in a velocity-force architecture. The COM is passivated for constant communication delays by the Scattering transformation (Hollow arrows represent physical interaction, filled arrows represent signal processing.)

The passive dynamics of the environment are described by

$$f_\mathrm{e} = M_e\ddot{x}_e + D_e\dot{x}_e + K_e x_e. \tag{A.26}$$

Finally, the external force sensed by the TO and reflected to the human operator is

$$f_t^e = f_e^e, \tag{A.27}$$

since sensor dynamics are omitted.

The Overall System and Related Work

A block diagram of the overall telepresence system is depicted in Figure A.6. The dynamics of the human operator are comprised to Z_o, the dynamics of the HSI are coprised to Z_h, the dynamics of the TO are comprised to Z_t, and the dynamics of the remote environment are comprised to Z_e. The variables represent the impedances of the mechanical systems, Z^{-1} denotes the admittance respectively. The systems G_h and G_t denote the dynamics of the actuators of HSI and TO and assumed to have unity dynamics. Hollow arrows depict physical interactions, filled arrows are used for signal interactions. All subsystems are considered linear(ized) and time-invariant.
The used telepresence architecture was introduced as *forward flow* in [165]. In [18] it was referred to as *position-force architecture* because position-based information (velocity) is commanded to the TO and force is reflected to the operator. Therein, it was compared to other possible architectures with respect to stability and transparency. Stability under communication delay was also analyzed with and without passified communication delay by wave variables. However, the control of the HSI did not link the reflected force to the force actually felt by the human operator. In [59] the presented architecture was extended by local feedback loops (actually introduced in [60]) to compensate for the dynamics of the HSI and the TO. Furthermore, a force control at operator side relates the reflected force to the force felt by the human operator. Moreover, a lead-lag compensator was introduced

to improve stability and transparency. However, no experiments were conducted. A first analysis of how the dynamics induced by the passified COM influence the dynamics of the telepresence system was conducted in [61]. A more detailed analysis was presented in [62; 63] incorporating psychophysical knowledge of the human operator. Stability of the overall system is hard to prove. We refrained from inserting additional measures to guarantee passivity of the overall system, such as a coordinating force (proposed in [18]) or small gain controllers (proposed in [49]). These measures are conservative and henceforth deteriorate transparency. The passive structure of the COM at least guarantees that no destabilizing effects are induced by the communication delay. The control loops at operator and teleoperator side are parametrized to be stable as well.

Bibliography

[1] R. M. Pirsig, *Zen and the Art of Motorcycle Maintenance: An Inquiry into Values*. William Morrow & Company, 1974.

[2] T. B. Sheridan, "Descartes, Heidegger, Gibson, and God: Toward an eclectic ontology of presence," *Presence*, vol. 8, pp. 551–559, 1999.

[3] N. A. Macmillan and C. D. Creelman, *Detection Theory - A User's Guide*. Lawrence Erlbaum Associates, Inc., 2nd ed., 2005.

[4] L. Newman, "Descartes' epistemology," *Stanford Encyclopedia of Philosophy*, 2005.

[5] M. Heidegger, *Being and Time*. London: SCM Press, 1962.

[6] T. Winograd and F. Flores, *Understanding Computers and Cognition: A New Foundation for Design*. Ablex Publishing Coperation, 1986.

[7] H. L. Dreyfus, *What Computers Still Can't Do - A Critic of Artificial Reason*. MIT Press, 3rd ed., 1992.

[8] P. Zhaorik and R. L. Jenison, "Presence as being-in-the-world," *Presence*, vol. 7, pp. 78–89, 1998.

[9] J. M. Flach and J. G. Holden, "The reality of experience: Gibson's way," *Presence*, vol. 7, pp. 90–95, 1998.

[10] G. Mantovani and G. Riva, ""Real" presence: How different ontologies generate different criteria for presence, telepresence, and virtual presence," *Presence*, vol. 8, pp. 540–550, 1999.

[11] R. Lauria, "In answer to a quasi-ontological argument: On Sheridan's "Toward an eclectic ontology of presence" and Montovani and Riva's "Building a bridge between different scientific communities"," *Presence*, vol. 10, pp. 557–563, 2001.

[12] F. Biocca, "Inserting the presence of mind into a philosophy of presence: A response to Sheridan and Mantovani and Riva," *Presence*, vol. 10, pp. 554–556, 2001.

[13] T. B. Sheridan, "Musings on telepresence and virtual presence," *Presence*, vol. 1, pp. 120–125, 1992.

[14] J. M. Loomis, "Distal attribution and presence," *Presence*, vol. 1, pp. 113–119, 1992.

[15] T. B. Sheridan, "Further musings on the psychophysics of presence," *Presence*, vol. 5, pp. 241–246, 1994.

[16] F. Biocca, "The Cyborg's dilemma: Embodiment in virtual environments," *Journal of Computer-Mediated Communication*, vol. 3, pp. –, 1997.

[17] L. Floridi, "The philosophy of presence: From epistemic failure to successful observation," *Presence*, vol. 14, pp. 656–667, 2005.

[18] D. A. Lawrence, "Stability and transparency in bilateral teleoperation," *IEEE Transactions on Robotics and Automation*, vol. 9, pp. 624–637, 1993.

[19] Y. Yokokohji, "Bilateral control of master-slave manipulators for ideal kinesthetic coupling - formulation and experiment," *IEEE Transactions on Robotics and Automation*, vol. 10, pp. 605–620, 1994.

[20] D. W. Schloerb, "A quantitative measure of telepresence," *Presence*, vol. 4, pp. 64–80, 1995.

[21] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionaire," *Presence*, vol. 7, pp. 225–240, 1998.

[22] IJsselsteijn, de Ridder, and F. Avons, "Presence: Concept, determinants and measurement," in *Proceedings of the SPIE*, 2000.

[23] M. Usoh, E. Catena, S. Arman, and M. Slater, "Presence questionaires in reality," *Presence*, vol. 9, pp. 497–503, 2000.

[24] M. C. Cavosoglu, A. Sherman, and F. Tendick, "Design of bilateral teleoperation controllers for haptic exploration and telemanipulation of soft environments," *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 18, pp. 641–647, 2002.

[25] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*. Wiley, 1966.

[26] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*. John Wiley & Sons, 1997.

[27] M. Ernst and M. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, pp. 429–433, 2002.

[28] S. A. Klein, "Measuring, estimating, and understanding the psychometric function: A commentary," *Perception & Psychophysics*, vol. 63,8, pp. 1421–1455, 2001.

[29] F. A. Wichmann and N. J. Hill, "The psychometric function: I. fitting, sampling, and goodness of fit," *Perception and Psychophysics*, vol. 63, pp. 1293–1313, 2001.

[30] W. Stadler, "A survey of multicriteria optimization or the vector maximum problem, part 1:17-76-1960," *Journal of Optimization Theory and Applications*, vol. 29, pp. 1–52, 1979.

[31] D. Kreps, *A Course in Microeconomic Theory*. Princeton University Press, 1990.

[32] A. Mas-Colell, M. D. Whinston, and J. R. Green, *Microeconomic Theory*. Oxford Univ. Press, 1995.

[33] Y. Sawaragi, *Theory of Multiobjective Optimization*. Academic Press, 1985.

[34] K. Miettinen, *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, 1999.

[35] M. Ehrgott, *Multicriteria Optimization*. Springer, 2005.

[36] V. Pareto, *Manual of Political Economy*. Kelley, USA (Reprint), 1969.

[37] R. L. Keeney and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. John Wiley & Sons, 1976.

[38] C. A. Bana e Costa, ed., *Readings in Multiple Criteria Decision Aid*. Springer, 1990.

[39] T. Hemming, *Multiobjective Decision Making Under Certainty*. PhD thesis, Stockholm : Economic Research Institute, 1978.

[40] M. Papageorgiou, *Optimierung*. R. Oldenbourg, 1991.

[41] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 2003.

[42] K. Sydsaeter and P. Hammond, *Mathmatics for Econmomic Analysis*. Prentice Hall, 1995.

[43] N. Hogan, "Controlling impedance at the man/machine interface," in *IEEE International Conference on Robotics and Automation*, pp. 1626 – 1631, 1989.

[44] M. Überle, *Design, Control, and Evaluation of a Familiy of Kinesthetic Haptic Devices*. PhD thesis, Technische Universität München, 2006.

[45] A. Peer and M. Buss, "A New Admittance Type Haptic Interface for Bimanual Manipulations," *IEEE/ASME Transactions on Mechatronics*, 2008.

[46] C. A. Desoer and M. Vidyasagar, *Feedback systems: Input Output Properties*. Academic Press, 1975.

[47] R. J. Anderson and M. W. Spong, "Bilateral control of teleoperators with time delay," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. vol.34, pp. 494–501, 1989.

[48] G. Niemeyer, *Using Wave Variables in Time Delayed Force Reflecting Teleoperation*. PhD thesis, MIT, Department of Aeronautics and Astronautics, September 1996.

[49] **M. Kuschel**, S. Hirche, and M. Buss, "Communication-induced disturbances in haptic telepresence systems," in *Proceedings of the International Federation of Automatic Control*, (Prague, Czech Republic), 2005.

[50] R. M. Held and N. I. Durlach, "Telepresence," *Presence*, vol. 1, pp. 109–112, 1992.

[51] C. Heeter, "Being there: The subjective experience of presence," *Presence*, vol. 1, pp. 262–271, 1992.

[52] M. Slater, M. Usoh, and A. Steed, "Depth of presence in virtual environments," *Presence*, vol. 3, pp. 130–144, 1994.

[53] G. Mantovani and G. Riva, "Building a bridge between different scientific communities: On sheridan's eclectic ontology of presence"," *Presence*, vol. 10, pp. 537–543, 2001.

[54] M. Slater, "Presence and the sixth sence," *Presence*, vol. 11, pp. 435–439, 2002.

[55] K. M. Lee, "Why presence occurs: Evolutionary psychology, media equation, and presence," *Presence*, vol. 13, pp. 494–505, 2004.

[56] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks, "Physiological measures of presence in virtual environments," in *Presence 2001, International Workshop*, 2001.

[57] T. Schubert, F. Friedmann, and H. Regenbrecht, "The experience of presence: Factor analytic insights," *Presence*, vol. 10, pp. 266–281, 2001.

[58] C. Sas and G. M. P. O'Hare, "Presence equation: An investigation into cognitive factors underlying presence," *Presence*, vol. 12, pp. 523–537, 2003.

[59] J. E. Speich, K. Fite, and M. Goldfarb, "Transparency and stability robustness in two-channel bilateral telemanipulation," *Transactions of the ASME*, vol. 123, pp. 400–4007, 2001.

[60] K. Hashtrudi-Zaad and S. E. Salcudean, "On the use of local force feedback for transparent teleoperation," in *Proceedings of the IEEE International Conference on Robotics & Automation*, (Detroit, USA), 1999.

[61] H. Baier and G. Schmidt, "Transparency and stability of bilateral kinesthetic teleoperation with time-delayed communication," *Journal of Intelligent and Robotic Systems*, vol. 40, pp. 1–12, 2004.

[62] S. Hirche, A. Bauer, and M. Buss, "Transparency of haptic telepresence with constant time delay," in *IEEE Conference on Control and Applications*, (Toronto, Canada), pp. 328–333, 2005.

[63] **M. Kuschel** and M. Buss, "Impedance-based performance analysis of haptic telepresence systems," in *Proceeding of the 2nd International Workshop Human Centered Robotics*, (Munich, Germany), 2006.

[64] Hirche, Hinterseer, Steinbach, and Buss, "Network traffic reduction in haptic telepresence systems by deadband control," in *Proceedings of the International Federation of Automatic Control*, (Prague, Czech Republic), 2005.

[65] **M. Kuschel**, P. Kremer, S. Hirche, and M. Buss, "Lossy data reduction methods in haptic telepresence systems," in *IEEE Proceedings of the International Conference on Robotics and Automation*, (Orlando, USA), 2006.

[66] I. Borutta, "Ein Präsenzmaß für visuell-haptische Telepräsenzsysteme," Sudienarbeit - supervised by **M. Kuschel**, Institute of Automatic Control Engineering, Technische Universität München, 2008.

[67] Z. Yan, "Echtzeit Transparenzschätzung," Diploma thesis - supervised by **M. Kuschel**, Institute of Automatic Control Engineering, Technische Universität München, 2008.

[68] **M. Kuschel**, F. Freyberger, B. Färber, and M. Buss, "A presence measure for virtual reality and telepresence based on multimodal conflicts," in *Proceedings of the ISPR Presence 2007. The 10th Annual International Workshop on Presence*, (Barcelona, Spain), 2007.

[69] M. A. Heller, J. A. Calcaterra, S. L. Green, and L. Brown, "Intersensory conflict between vision and touch: the response modality dominates when precise, attention-riveting judgments are required.," *Percept Psychophys*, vol. 61, pp. 1384–1398, October 1999.

[70] F. Freyberger, **M. Kuschel**, R. L. Klatzky, B. Färber, and M. Buss, "Visual-haptic perception of compliance: Direct matching of visual and haptic information," in *Proceedings of the IEEE International Workshop on Haptic Audio Visual Environments and their Applications (HAVE)*, (Ontario, Canada), 2007. (**Best Paper Award**).

[71] R. Scheuchenpflug, "Measuring presence in virtual environments," in *HCI International 2001*, (New Orleans, USA), 2001.

[72] K. Arrow, H. Chenery, B. Minhas, and R. Solow, "Capital-labor substitution and economic efficiency," *Review of Economics and Statistics*, vol. 43, pp. 225–250, 1961.

[73] J. Faber, "Verbesserung der Darstellungsqualität von Mensch-Maschine Schnittstellen durch crossmodale Wechselwirkungen," Sudienarbeit - supervised by **M. Kuschel**, Institute of Automatic Control Engineering, Technische Universität München, 2006.

[74] M. Werner, "Identifikation dynamischer Prozesse in der menschlichen Wahrnehmung," Diploma thesis - supervised by **M. Kuschel**, Institute of Automatic Control Engineering, Technische Universität München, 2006.

[75] N. Gattinger, "Analyse der Wahrnehmung von Steifigkeit in visuell-haptischen Telepräsenzsystemen," Sudienarbeit - supervised by **M. Kuschel**, Institute of Automatic Control Engineering, Technische Universität München, 2007.

[76] A. Schmid, "Minimierung von Unsicherheiten in der Hand-Auge Koordination," Sudienarbeit - supervised by **M. Kuschel**, Institute of Automatic Control Engineering, Technische Universität München, 2007.

[77] Z. Mingxiang, "Admittanzregelung mit psychophysischen Parametern für ein haptisch-visuelles Telepräsenzsystem," Sudienarbeit - supervised by **M. Kuschel**, Institute of Automatic Control Engineering, Technische Universität München, 2008.

[78] F. Freyberger, **M. Kuschel**, B. Färber, and M. Buss, "Perception of compliant environments through a visual-haptic human system interface," in *Proceedings of the IEEE VR Workshop On Haptic and Tactile Perception of Deformable Objects (HAPTEX)*, (Hannover, Germany), 2007.

[79] **M. Kuschel**, F. Freyberger, R. L. Klatzky, B. Färber, and M. Buss, "Visual haptic perception of compliance: Fusion of visual and haptic information," in *Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems*, (Reno, USA), 2008.

[80] M. Buss, "Mechatronics and control issues in multi-modal telepresence," in *IEEE International Conference on Mechatronics and Robotics, MechRob2004*, 2004.

[81] B. E. Stein and M. A. Meredith, *The merging of the senses*. Cambridge: MIT Press, 1993.

[82] C. Spence and J. Driver, *Crossmodal space and crossmodal attention*. Oxford: University Press, 2004.

[83] G. A. Calvert, C. Spence, and B. E. Stein, *The handbook of multisensory processes*. Cambridge: MIT Press, 2004.

[84] D. W. Massaro, *Perspectives on perception and action (pp. 273-299)*, ch. Information-processing theory and strong inference: A paradigm for psychological inquiry. Hillsdale: Lawrence Erlbaum Associates, 1987.

[85] L. E. Marks, *The handbook of multisensory processes*, ch. Cross-modal interactions in speeded classification, pp. 85–105. Cambridge: MIT Press, 2004.

[86] S. Soto-Faraco, J. Lyons, M. Gazzaniga, C. Spence, and A. Kingstone, "The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities.," *Cognitive Brain Research*, vol. 14, pp. 139–146, 2002.

[87] W. R. Garner, *Handbook of Perception, Vol. 2, Psychological Judgement and Measurement (pp. 23-59)*, ch. Attention: The processing of multiple sources of information. New York: Academic Press, 1974.

[88] J. Driver and C. Spence, "Multisensory perception: Beyond modularity and convergence.," *Current Biology*, vol. 10(13), pp. 731–735, 2000.

[89] S. Shimojo and L. Shams, "Sensory modalities are not separate modaltities: Plasticity and interactions.," *Current Opinion in Neurobiology*, vol. 11, pp. 505–509, 2001.

[90] R. B. Welch and D. H. Warren, *Handbook of Perception and Human Performance: Sensory Processes and Perception, Volume 1*, ch. Intersensory Interactions. New York: J. Wiley & Sons, 1986.

[91] G. A. Calvert, M. J. Brammer, and S. D. Iversen, "Crossmodal identification.," *Trends in Cognitive Sciences*, vol. 2(7), pp. 247–253, 1998.

[92] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion.," *Nature*, vol. 415, pp. 429–433, 2002.

[93] M. O. Ernst and H. H. Bülthoff, "Merging the senses into a robust percept," *Trends in Cognitive Sciences*, vol. 8(4), pp. 162–169, 2004.

[94] S. Guest and C. Spence, "What role does multisensory integration play in the visuotactile pereption of texture?," *International Journal of Psychophysiology*, vol. 50, pp. 63–80, 2003.

[95] I. Rock and C. S. Harris, "Vision and touch," *Scientific American*, vol. 216, pp. 96–104, 1967.

[96] J. A. S. Kinney and S. M. Luria, "Conflicting visual and tactual-kinesthetic stimulation.," *Perception & Psychophysics*, vol. 8(3), pp. 189–192, 1970.

[97] I. Rock and J. Victor, "Vision and touch: An experimentally created conflict between the twosenses.," *Science*, vol. 143, pp. 594–596, 1964.

[98] M. A. Heller, "Haptic dominance in form perception with blurred vision," *Perception*, vol. 12, pp. 607–613, 1983.

[99] M. A. Heller, "Haptic dominance in form perception: Vision versus proprioception," *Perception*, vol. 21, pp. 655–660, 1992.

[100] S. J. Lederman and S. G. Abbott, "Texture perception: Studies of intersensory organization using a discrepancy paradigm, and visual versus tactile psychophysics.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 7(4), pp. 902–915, 1981.

[101] S. Guest and C. Spence, "Tactile dominance in speeded discrimination of textures.," *Experimental Brain Research*, vol. 150(2), pp. 201–207, 2003.

[102] M. A. Heller, "Visual and tactual texture perception: Intersensory cooperation," *Perception & Psychophysics*, vol. 31(4), pp. 339–344, 1982.

[103] D. H. Warren and H. L. Pick, "Intermodality relations in localization in blind and sighted people.," *Perception & Psychophysics*, vol. 8(6), pp. 430–432, 1970.

[104] H. L. Pick and D. H. Warren, "Sensory conflict in judgements of spatial direction," *Percpetion & Psychophysics*, vol. 6(4), pp. 203–205, 1969.

[105] R. Over, "An experimentally induced conflict between vision and propriozeption.," *British Journal of Psychology*, vol. 57, pp. 335–341, 1966.

[106] W. C. Wu, C. Basdogan, and M. A. Srinivasan, "Visual, haptic, and bimodal perception of size and stiffness in virtual environments," in *Proceedings of the ASME Dynamic Systems and Control Division - 1999, DSC-Vol. 67 (pp. 19-26)*, 1999.

[107] M. A. Srinivasan, G. L. Beauregard, and D. L. Brock, "The impact of visual information on the haptic perception of stiffness in virtual environments," in *Proceedings of the ASME Dynamic Systems and Control Division - 1996, DSC-Vol. 58 (pp. 555-559)*, 1996.

[108] G. F. Misceo, W. A. Hershberger, and R. L. Mancini, "Haptic estimates of discordant visual-haptic size vary developmentally," *Perception & Psychophysics*, vol. 61(4), pp. 608–614, 1999.

[109] M. A. Heller, J. A. Calcaterra, S. L.Green, and L. Brown, "Intersensory conflict between vision and touch: The response modality dominates, when precise, attention-riveting judgments are required," *Perception & Psychophysics*, vol. 61(7), pp. 1384–1396, 1999.

[110] S. Lederman, G. Thorne, and B. Jones, "Perception of texture by vision and touch: Multidimensionality and intersensory integration," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 12(2), pp. 169–180, 1986.

[111] M. O. Ernst, M. S. Banks, and H. H. Bülthoff, "Touch can change visual slant perception," *Nature Neuroscience*, vol. 3 (1), pp. 69–73, 2000.

[112] S. M. Fishkin, V. Pishkin, and M. L. Stahl, "Factors involved in visual capture," *Perceptual and Motor Skills*, vol. 40, pp. 427–434, 1975.

[113] G. A. Gescheider, *Psychophysics - Method and Theory.* Hillsdale: J. Wiley & Sons, 1976.

[114] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire.," *Presence: Teleoperators and Virtual Environments.*, vol. 7(3), pp. 225–240, 1998.

[115] C. Spence, M. E. R. Nicholls, and J. Driver, "The cost of expecting events in the wrong sensory modality," *Perception & Psychophysics*, vol. 63(2), pp. 330–336, 2001.

[116] J.-P. Bresciani, F. Dammeier, and M. O. Ernst, "Vision and touch are automatically integrated for the perception of sequences of events," *Journal of Vision*, vol. 6, pp. 554–564, 2006.

[117] J. M. Hillis, M. O. Ernst, M. S. Banks, and M. S. Landy, "Combining sensory information: Mandatory fusion within, but not netween, senses," *Science*, vol. 298, pp. 1627–1630, 2002.

[118] W. J. Adams, E. W. Graf, and M. O. Ernst, "Experience can change the 'light-from-above' prior," *Nature Neuroscience*, vol. 7(10), pp. 1057–1058, 2004.

[119] J. M. Hillis, S. J. Watt, M. S. Landy, and M. S. Banks, "Slant from texture and disparity cues: Optimal cue combination," *Journal of Vision*, vol. 4 (13), pp. 1–24, 2004.

[120] K. Drewing and M. O. Ernst, "Integration of force and position cues for shape perception through active touch," *Brain Research*, vol. 1078, pp. 92–100, 2006.

[121] D. C. Knill and J. A. Saunders, "Do humans optimally integrate stereo and texture information for judgements of surface slant?," *Vision Research*, vol. 43, pp. 2539–2558, 2003.

[122] H. Tan, N. Durlach, G. Beauregard, and M. Srinivasan, "Manual discrimination of compliance using active pinch grasp: The roles of force and work cues," *Perception & Psychophysics*, vol. 57(4), pp. 495–510, 1995.

[123] S. Yamakawa, H. Fujimoto, S. Manabe, and Y. Kobyayshi, "The necessary conditions of the scaling ratio in master-slave systems based on human difference limen of force sense," in *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 35, pp. 275–282, 2005.

[124] N. Dhruv and F. Tendick, "Frequency dependence of compliance contrast detection," in *Proccedings of the ASME Dynamic Systems and Control Division*, 2000.

[125] M. K. O'Malley and M. Goldfarb, "The implications of surface stiffness for size identification and perceived surface hardness in haptic interfaces," in *Proceedings of the 2002 IEEE International Conference on Robotics and Automation*, (Washington (DC), USA), 2002.

[126] S. A. Wall and S. A. Brewster, "Scratching the surface: Preliminary investigations of haptic properties for data representation," in *Proceedings of Eurohaptics 2003, Dublin, Ireland*, pp. 330–342, 2003.

[127] F. J. Clark and K. W. Horch, *Handbook of Perception and Human Performance, volume 1*, ch. Kinesthesia. N.Y.: Wiley and Sons, 1986.

[128] P. E. Roland and H. Ladegaard-Pedersen, "A quantitative analysis of sensations of tension and of kinasethesia in man - evidence of a peripherally originating muscular sense and for a sense of effort," *Brain*, vol. 100, pp. 671–692, 1977.

[129] G. B. Evans and E. Howarth, "The effect of grip-tension on tactile-kinaesthetic judgement of width," *Quarterly Journal of Experimental Psychology*, vol. 18, pp. 275–277, 1966.

[130] N. I. Durlach, L. A. Delhorne, A. Wong, W. Y. Ko, W. M. Rabinowitz, and J. Hollerbach, "Manual discrimination and identification of length by the finger-span method," *Perception & Psychophysics*, vol. 46(1), pp. 29–38, 1989.

[131] L. A. Jones, "Matching forces: constant errors and differental thresholds," *Perception*, vol. 18(5), pp. 681–687, 1989.

[132] G. Raffel, "Visual and kinaesthetic judgments of length," *The American Journal of Psychology*, vol. 48, pp. 331–334, 1936.

[133] J. F. Norman, J. T. Todd, V. J.Perotti, and J. S. Tittle, "The visual perception of three-dimensional length," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 22(1), pp. 173–186, 1996.

[134] W. T. Pollock and A. Chapanis, "The apparent length of a line as a function of its inclination," *Quarterly Journal of Experimental Psychology,*, vol. 4, pp. 170–178, 1952.

[135] S. Gephstein and M. S. Banks, "Viewing geometry determines how vision and haptics combine in size percpetion," *Current Biology*, vol. 13, pp. 483–488, 2003.

[136] J. R. Anderson, *Cognitive Psychology and Its Implications*. New York: Freeman, 1995.

[137] R. R. Ellis and S. J. Lederman, "The golf-ball illusion: evidence for top-down processing in weight perception," *Perception*, vol. 27, pp. 193–201, 1998.

[138] D. W. Massaro, "Speechreading: illusion or window into pattern recognition," *Trends in Cognitive Science*, vol. 3(8), pp. 310–317, 1999.

[139] F. Pavani, C. Spence, and J. Driver, "Visual capture of touch: Out-of-the-body experience with rubber gloves," *Psychological Science*, vol. 11(5), pp. 353–359, 2000.

[140] L. A. Jones and I. W. Hunter, "A perceptual analysis of stiffness," *Experimental Brain Research*, vol. 79, pp. 150–156, 1990.

[141] H. Z. Tan, Pang, X. D. Pang, and N. I. Durlach, "Manual resolution of length, force, and compliance," in *Winter Annual Meeting of the American Society of Mechanical Engineers: Advances in Robotics*, 1992.

[142] H. Z. Tan, N. I. Durlach, G. L. Beauregard, and M. A. Srinivasan, "Manual discrimination of compliance using active pinch grasp: The roles of force and work cues," *Perception and Psychophysics*, vol. 57, pp. 495–510, 1995.

[143] R. H. LaMotte, "Softness discrimination with a tool," *Journal of Neurophysiology*, vol. 83, pp. 1777–1786, 2000.

[144] M. A. Srinivasan and R. H. LaMotte, "Tactual discrimination of softness," *Journal of Neurophysiology*, vol. 73, pp. 88–101, 1995.

[145] M. A. Srinivasan, G. L. Beauregard, and D. L. Brock, "The impact of visual information on the haptic perception of stiffness in virtual environments," in *Proceedings of the ASME Dynamics Systems and Control Division*, 1996.

[146] F. Freyberger, **M. Kuschel**, B. Färber, M. Buss, and R. L. Klatzky, "Tilt perception by constant tactile and constant proprioceptive feedback through a human system interface," in *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environments and Teleoperator Systems*, (Tsukuba, Japan), 2007.

[147] G. Niemeyer and J.-J. E. Slotine, "Using wave variables for system analysis and robot control," in *IEEE Conf. on Robotics and Automation*, (Albuquerque, NM), pp. vol. 2, pp. 1619–1625, 1997.

[148] C. Shahabi, A. Ortega, and M. R. Kolahdouzan, "Comparison of Different Haptic Compression Techniques," in *Proceedings of the International Conference on Multimedia and Expo (ICME)*, (Lausanne, Switzerland), 2002.

[149] C. W. Borst, "Predictive Coding for Efficient Host-Device Communication in a Pneumatic Force-Feedback Display," in *Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, (Pisa, Italy), 2005.

[150] A. Ortega and Y. Liu, *Touch in Virtual Environments: Haptics and the Design of Interactive Systems*. Prentice Hall, 2002.

[151] M. H. Zadeh, D. Wang, and E. Kubica, "Perception-based lossy haptic compression considerations for velocity-based interactions," *Multimedia Systems*, vol. 13, pp. 275–282, 2008.

[152] Hirche, Hinterseer, Steinbach, and Buss, "Towards deadband control in networked teleoperation systems," in *Proceedings of the International Federation of Automatic Control*, (Prague, Czech Republic), 2005.

[153] P. Hinterseer, E. Steinbach, and S. Chaudhuri, "Perception-based compression of haptic data streams using kalman filters," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2006.

[154] M. F. Zaeh, S. Clarke, P. Hinterseer, and E. Steinbach, "Telepresence across networks: A combined deadband and prediction approach," in *Proceedings of the Information Visualization*, 2006.

[155] P. Hinterseer, S. Hirche, S. Chaudhuri, E. Steinbach, and M. Buss, "Perception-based data reduction and transmission of haptic data in telepresence and teleaction systems," *IEEE Transactions on Signal Processing*, vol. 56, pp. 588–597, 2008.

[156] P. Kremer, "Signalverarbeitung in passivierten Kommunikationskanälen zum Einsatz in haptischen Telepräsenzsystemen," Diploma thesis - supervised by **M. Kuschel**, Institute of Automatic Control Engineering, Technische Universität München, 2005.

[157] J. Renner, "Implementierung eines fensterbasierten Algorithmus zur Datenreduktion für haptische Telepräsenzsysteme," Sudienarbeit - supervised by **M. Kuschel**, Institute of Automatic Control Engineering, Technische Universität München, 2006.

[158] S. Hirche, *Haptic Telepresence in Packet Switched Communication Networks*. PhD thesis, Technische Universität München, 2005.

[159] S. Hirche and M. Buss, "Transparency of haptic telepresence systems with constant time delay," *at-Automatisierungstechnik*, vol. 54, pp. 51–59, 2006.

[160] R. Aracil, M. Buss, S. Cobos, M. Ferre, S. Hirche, **Martin Kuschel**, and A. Peer, *Advances in Telerobotics*, ch. The Human Role in Telerobotics, pp. 11–24. Springer, STAR series, 2007.

[161] **M. Kuschel**, "Konstruktion eines SCARA-Roboters zur Erforschung von Regelungsstrategien in der Telepräsenz," Sudienarbeit, Technische Universität Berlin, 2004.

[162] **M. Kuschel**, "Performance-orientated control of haptic telepresence systems," Diploma thesis, Institute of Automatic Control Engineering, Technische Universität München, 2004.

[163] E. Brenner and J. B. J. Smeets, "Fast corrections of movements with a computer mouse," *Spatial Vision*, vol. 16, pp. 364–376, 2003.

[164] L. Sciavicco and B. Siciliano, *Modelling and Control of Robot Manipulators*. Springer, 1 ed., 2003.

[165] B. Hannaford, "A design framework for teleoperators with kinesthetic feedback," in *IEEE Transactions on Robotics and Automation*, vol. 5, pp. 426–434, 1989.