Technische Universität München
Fakultät für Elektrotechnik und Informationstechnik
Lehrstuhl für Medientechnik

Data Compression and Quality Evaluation for Haptic Communications

Rahul Gopal Chaudhari, M.Sc.

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. Bernhard U. Seeber
Prüfer der Dissertation:	1.	Univ.-Prof. Dr.-Ing. Eckehard Steinbach
	2.	Univ.-Prof. Dr.-Ing. Sandra Hirche

Die Dissertation wurde am 15.01.2015 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 26.03.2015 angenommen.

To all honest scholarly endeavor.

Abstract

By imparting us the ability to touch and feel virtual or remote environments, haptics introduces physical interactivity into multimodal communication systems. The inclusion of haptic media has been shown to improve task performance, immersiveness, and the overall experience of task execution. While several decades of research have been dedicated to the acquisition, processing, coding, and display of audio-video streams in multimodal systems, similar aspects for haptic streams have started getting addressed only recently.

This thesis presents novel data compression (coding) approaches for haptic signals generated while stroking a textured surface with a mechanical tool. Detailed investigations of the mechanisms of texture signal production and perception are described. Mathematical models for these mechanisms are then used as the basis for the presented data compression approach. The performance of the compression algorithm is comprehensively evaluated both objectively as well as through subjective user studies. These evaluations show that the compression algorithm achieves, for a comparable perceptual quality, a compression ratio two times better than the state-of-the-art.

To obviate the constant need for difficult and time-consuming user studies, this thesis also presents methods for the model-based prediction of haptic perceptual quality. This part of the thesis addresses both wide frequency-band haptic texture signals as well as low-frequency kinesthetic haptic signals. For both cases, high correlations are obtained between the predicted and the actual perceptual quality data.

Kurzfassung

Haptische Systeme ermöglichen das Erfühlen einer virtuellen oder entfernten Umgebung und führen somit physische Interaktivität in multimodale Kommunikationssysteme ein. Haptische Signale steigern die subjektive Interaktionsqualität sowie die Fähigkeit anspruchsvolle feinmotorische Aufgaben auszuführen. Während die Erfassung, Verarbeitung, Kodierung und Darstellung von Audio- und Videoströmen in multimodalen Systemen seit mehreren Jahrzehnten intensiv untersucht wird, sind solche Betrachtungen für die Haptik ein relativ neues Thema.

In der vorliegenden Arbeit wird ein neuartiger Ansatz zur Kompression von haptischen Signalen, die durch das Überstreichen einer texturierten Oberfläche mit einem Stift erzeugt werden, vorgestellt. Die Mechanismen hinter der Erzeugung haptischer Textursignale und deren Wahrnehmung durch den Mensch werden detailliert beschrieben. Mathematische Modelle von diesen Mechanismen bilden die Basis für den präsentierten Komprimierungsansatz. Die Leistungsfähigkeit des Verfahrens wird durch umfangreiche objektive sowie subjektive Qualitätsbewertungen evaluiert. Die Ergebnisse zeigen, dass der Komprimierungsansatz für eine vergleichbare Wahrnehmungsqualität eine Kompressionsrate zwei mal besser als der neueste Stand der Technik erzielt.

Desweiteren, um den ständigen Bedarf an schwierigen und zeitaufwendigen Benutzerstudien zu umgehen, werden Verfahren zur modellbasierten Prädiktion der durch den Mensch wahrgenommenen haptischen Qualität beschrieben. Dieser Teil der Arbeit befasst sich sowohl mit breitbandigen haptischen Texturen sowie niederfrequenten kinästhetischen haptischen Signalen. In beiden Fällen wurde eine hohe Korrelation zwischen den prädizierten und experimentellen Daten erzielt.

Acknowledgements

The work presented in this dissertation was carried out as a member of the research and teaching staff at the Chair of Media Technology (LMT) at the Technische Universität München. My research was supported, in part, by the German Research Foundation (DFG) under the project STE 1093/4 and, in part, by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no. 258941. Many people have contributed to the success of this work in various personal and professional ways. I am sincerely thankful to all of them; I apologize for not being able to explicitly mention each and every one of them here.

I am eternally grateful to Prof. Eckehard Steinbach for the scientific expertise with which he has supervised my research work, and for the kind generosity and trust with which he has supported me all throughout my PhD. Despite his busy schedule, he has always been accessible for discussing new ideas and giving me fresh perspectives. While guiding me in the right direction, he has always left room for letting my own ideas flourish. I have learnt more from him than I can ever repay. I would like to thank Prof. Sandra Hirche for providing important impulses to my research at the beginning of my PhD and for agreeing to become the second examiner. I also thank Prof. Bernhard U. Seeber for chairing the thesis committee.

I would like to extend my sincere thanks to Dr. Julius Kammerl for initiating me into the world of research as my master's thesis advisor, and for the exciting joint work later on model-mediated teleoperation. Very many thanks go to Clemens Schuwerk for his extremely reliable support in my research over the past many years and for the excellent teamwork in the Computational Haptics Laboratory. I would like to thank Burak Cizmeci for contributing significantly to my work through his theoretical insights in signal processing and the joint work on developing the haptic texture codec. I would like to thank Xiao Xu, Fernanda Brandi, and Dr. Giulia Paggetti for collaborating on various interesting technical topics. My sincere appreciation also goes to Dr. Iason Vittorias for the occasional insights into robotic teleoperation systems from the control engineering perspective. I would like to thank Stefan Diewald for co-supervising an excellent bachelor's thesis with me. Special thanks go to all my other colleagues at LMT, especially Damien Schroeder, Anas Al-Nuaimi, Fan Zhang, Christoph Bachhuber, and Robert Huitl for interesting discussions on a variety of topics

ranging from research, technology, travel, art, philosophy, and international developments. I have learnt a lot from all of you.

I would like to extend my heartfelt thanks to Prof. Seungmoon Choi and Prof. Katherine Kuchenbecker for giving me the opportunity to visit their labs at POSTECH and UPenn, respectively. These visits resulted in very fruitful joint research on haptic textures. Special thanks go to Yongjae Yoo for being a caring host for me at POSTECH, and for his reliable scientific support and keen attention to detail. Furthermore, I would like to thank Prof. Subhasis Chaudhuri for the insightful joint work during his visits to our institute, Ercan Altinsoy for the wonderful chapter we jointly contributed to the Quality-of-Experience book, and Verena Nitsch for bringing in her statistical expertise in investigating haptic task performance.

I sincerely appreciate the efforts of all the highly-motivated students with whom I have worked over the years, especially Mojtaba Danaei, Jinjian Wang, Gaganjot Singh, Felix Willgerodt, Amit Sama, and Matti Strese. They contributed immensely to the success of my work. Special thanks also go to Ingrid Jamrath, Dr. Martin Maier and Gabriele Kohl for their quick and reliable administrative support.

Finally, I would like to express my warm gratitude to my beautiful wife Shruti for lighting up my world, my brother Nakul for being consistently zestful and full of life, and my parents and grandparents for believing in my abilities and for supporting me throughout.

Contents

Chapter 1

Introduction

Remarkable technical innovation and tremendous growth in audiovisual communications have improved the Quality-of-Experience and productivity in networked interaction between distant people, e.g. through video conferencing. State-of-the-art commercial teleconferencing solutions already simulate very convincing online environments, where participants experience a high Sense-of-Togetherness (e.g. [Cis14, Hew14]). Sophisticated audiovisual sensing/display devices, efficient audio/video coding standards and high-capacity communication networks have driven this progress.

The ability to *physically* interact with distant objects and humans, a feature that would contribute substantially towards improved task performance and ultimate immersion, is however not yet fully realized. The rapidly rising field of haptics (the science, technology, and applications related to the sense of *touch*) is widely believed to be a big step forward in this direction [Sad07, She92b, SHE+12]. The broad field of haptics can be categorized into roughly three classes [Sri10]:

- Human haptics — referring to the study of human sensing, perception, and manipulation through touch,
- Machine haptics — referring to the design, construction, and evaluation of haptic devices for humans as well as for robots, and
- Computer haptics — referring to the haptic rendering of virtual objects.

Combined with powerful communication technologies, it promises to comprehensively aid and transport not only human communication, but also control capabilities, enabling us to physically transcend the boundaries of space and time. Figure 1.1 shows the block schematic of a multimodal communication system that depicts this vision. As can be seen from the figure, haptic communication is inherently *bidirectional*. That is, humans not only perceive

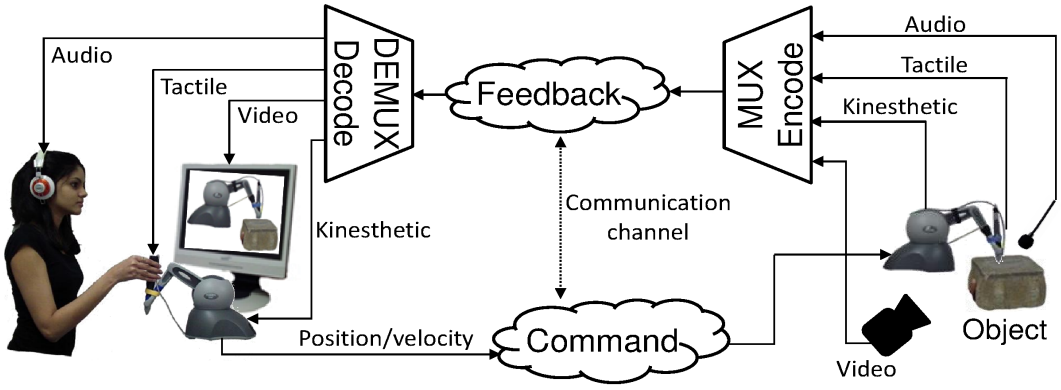


Figure 1.1: A multimodal communication system. The human user can interact physically with the remote object by commanding the movement of the remote robot. At the remote side, audio-visual-haptic signals are encoded, multiplexed together, and transmitted back to the user side. At the user side, the incoming signals are demultiplexed and decoded for presentation to the user. The user can react physically to this feedback, thus closing a global control-loop between herself and the remote environment over the communication channel. Figure reproduced from [CcK+12].

haptic feedback — similar to audio/video — but can also physically act upon the environment. Accordingly, a bidirectional human-centric design and analysis of haptic technology is necessary [ABC+07].

All applications that involve physical interaction with environments that are remote, inaccessible, hazardous, or too big/small in scale for a human can benefit extensively from haptic technology. Some examples include on-orbit servicing for satellites [RLH07], space exploration [ST92], tele-surgery [CCK+12], deep-sea exploration [RCHP07], safety-critical situations [FUS92, KSP+04], tele-manufacturing [FB01], tele-manipulation and tele-assembly [PUB06]. Haptics also plays a key role in virtual simulations of real environments (e.g., collaborative assembly and design [ICG+08]) or fantasy worlds (e.g., games [OSSP12]).

The acquisition, processing, communication, and display of audio and video are traditionally well-established fields [Z08, WOZ02]. But the study of similar aspects for haptics is still at a nascent stage. Although the acquisition and display of haptics have been a subject of intense research [HAC+04, CK13], both these aspects are still far from standardization.

Nevertheless, there is consensus on the various physical properties of objects that need to be captured to realistically reconstruct remote environments for the human user. These properties include hardness, surface texture, shape, etc. To represent these properties, researchers have largely relied on interaction force and torque signals, and vibrotactile signals from surface texture scans. When it comes to communicating these signals over networks, each of them have their own associated challenges. Force and torque (kinesthetic) signals lead to high packet-rates on a network [SHE+12], whereas vibrotactile signals may present large data rates (bits per second), if not high packet-rates [CcK+12].

Until now, the communication of kinesthetic signals has been the prime focus of attention within the field of haptic communications. Highly efficient communication schemes that ascertain system-stability and perceptual transparency have been developed in [HMF+01, SOK02, Bor05, HB04, HHSB05b, KKH06, HB07b, HHC+08a, VKHS09, VRH10, SHE+12, Hin09, Kam12]. The field of vibrotactile communications, on the other hand, has received much less attention [OY13].

This thesis has vibrotactile texture communication as its center of focus. The main contributions made to this field are outlined next. In the following, these contributions are linked to the corresponding chapters/sections of this thesis, where they are presented in further detail.

Major contributions and thesis organization

This thesis presents original contributions towards the efficient delivery of haptic media streams, in particular vibrotactile texture signals, over communication networks. It also investigates perceptual quality aspects of haptic communication systems. The main contributions can be summarized as follows (also see Figure 1.2).

- **Models for the production and perception of vibrotactile texture signals**
  Lossy data compression schemes usually take advantage of the following factors to achieve low bitrates:

  1. modeling a phenomenon or a signal corresponding to this phenomenon with a small set of parameters, and transmitting this (quantized) parameter set, and
  2. modeling human limitations under which this phenomenon or signal is perceived.

  Such models collectively aid in capturing the essence of the signal and "throwing away" redundant or perceptually irrelevant information.

  Section 3.1 addresses point 1 from above for vibrotactile textures. It analyzes, qualitatively, the mechanism behind the production of vibrotactile signals when a mechanical tool is used to scan a textured surface. Furthermore, it motivates the use of a source-filter signal model to capture this mechanism.

  Section 3.2 addresses point 2 from above. It presents novel masking data, which confirm that the well-known perceptual Simultaneous Masking phenomenon [Gel10] also holds true for vibrotactile signals over a wide frequency range. Both these aspects form the basis for the efficient data compression technique proposed in Section 3.3.

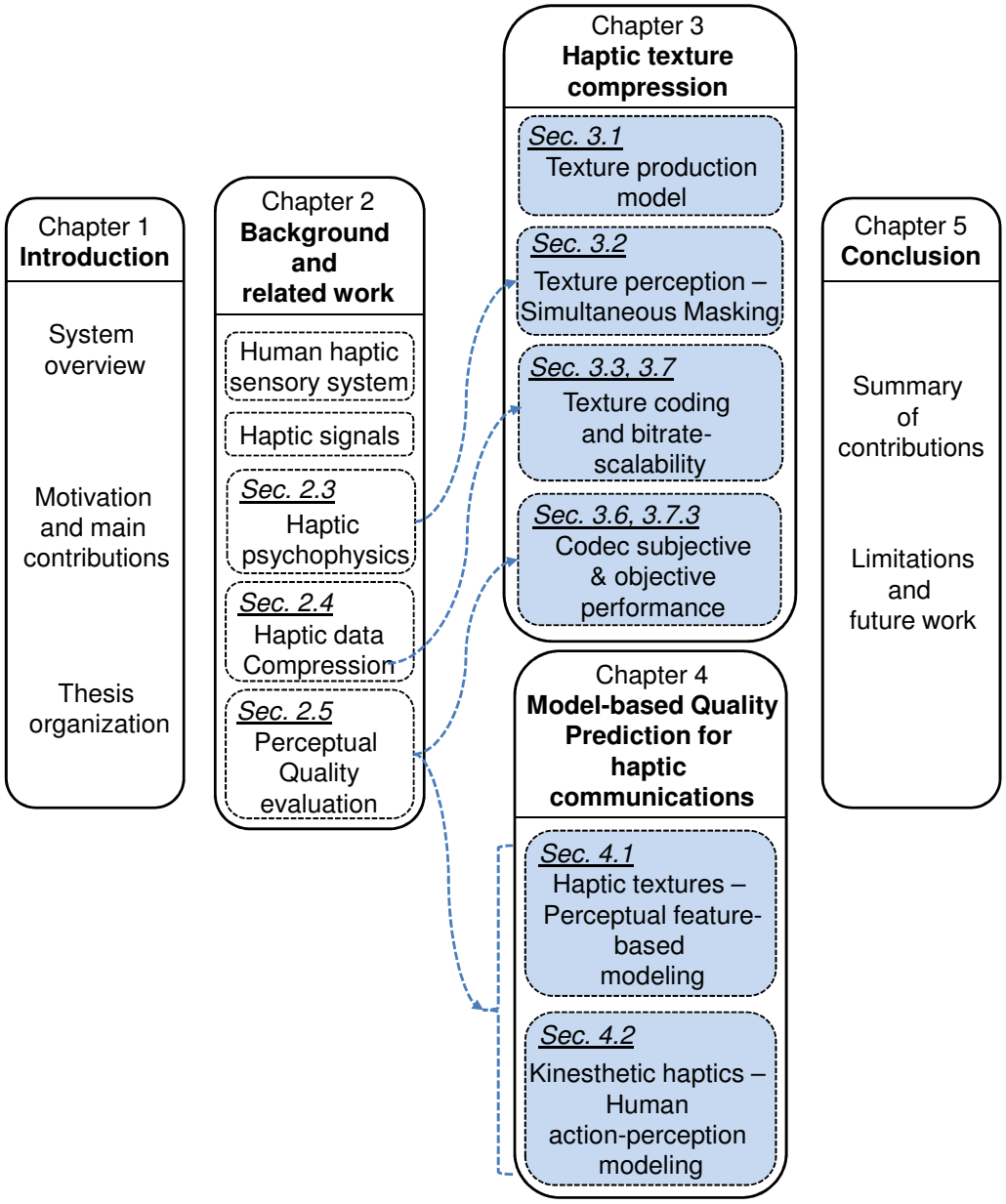


Figure 1.2: Overview of this thesis. The shaded boxes highlight original contributions made in this thesis. Blocks to the left of the arrows denote thesis sections that review related work for a given topic. These sections also identify any gaps in existing research or opportunities for extending it. The arrowheads point to thesis sections that supply novel results to fill the identified gaps or extend the state-of-the-art.

- **Haptic (vibrotactile) texture compression**

  Vibrotactile signals have a bandwidth much wider (upto 1 kHz [BH03]) than kinesthetic signals (0–12 Hz [Jon00]), demanding a sampling rate of at least 2 kHz according to the Nyquist sampling theorem. Moreover, an average human body is covered with 2 m^2 of skin that can be stimulated by distributed tactile stimuli. Combining the above two

factors together, huge data volumes and rates may potentially result, necessitating data compression for cost-efficient transmission and storage.

Section 3.3 presents a novel texture compression scheme that exploits the production and perception models mentioned in the first contribution. In particular, it proposes an encoder based on a Linear Predictive model of the texture signal. The coding noise introduced by this modeling and the subsequent parameter quantization process is judged for imperceptibility using a Simultaneous Masking-based weighting filter.

Sections 3.4 explains the operation of the decoder, and Section 3.5 gives implementation details for the codec. Section 3.6 then analyzes the performance of this codec objectively and subjectively.

The codec should be bitrate-scalable to enable operation in a range of different bitrates under varied network traffic conditions and quality demands. Section 3.7 describes methods for achieving such bitrate-scalability. It also presents objective and subjective evaluation of the performance of the bitrate-scalable codec.

- **Model-based Quality Prediction for haptic communications**
  Human-user studies form an integral part of the evaluation of haptic systems. It is necessary to rigorously test the performance of every new development in haptic sensors, actuators, signal processing and synthesis algorithms from a perceptual perspective. However, such studies are difficult, time-consuming, and expensive. They significantly slow down the progress of research. This issue can be addressed by developing model-based methods to predict haptic perceptual quality.

  Chapter 4 deals with this topic for both vibrotactile texture and kinesthetic signals. Section 4.1 presents a model that predicts perceptual quality based on high-level features for vibrotactile texture signals. Section 4.2 treats quality prediction for kinesthetic signals. In addition to a perception model, it also presents a model for human action to account for the bidirectional closed-loop nature of kinesthetic haptics. This action-perception modeling allows for a fully automatic software simulation of a kinesthetic communication system, enabling an instrumental prediction of perceived quality.

Parts of the work presented in this thesis have been published in various international peer-reviewed scientific journals and conferences [CSH11, CSN$^+$11, CcK$^+$12, CAS14, CSDS15, CYS$^+$15, SPCS14, SHE$^+$12, KCS11, SHK$^+$11, CCX$^+$14, SCS14, SCS12, XKCS11, BCS11, KCS10, CKS09, BCH$^+$11].

Chapter 2

Background and related work

2.1 The human haptic sensory system

The tactual inspection of a physical object leads to a holistic percept of the object. This percept combines the contributions of two distinct senses - the kinesthetic one, and the tactile one. When reconstructing tactual signals in a technical system, this distinction plays a very important role. The technical sensor(s), the signal processing, and the actuator(s) need to be chosen appropriately depending upon which of the above two haptic senses is being targeted for stimulation. The two senses differ functionally as follows.

2.1.1 The kinesthetic sense

The kinesthetic sense notifies us about the static body posture (configuration of the head, torso, limbs, and end-effectors), as well as the dynamic movement of body parts with respect to each other [LL86]. It enables the perception of forces and torques acting on the human body. Thereby, it facilitates the determination of the inertia or stiffness of a touched solid object, or the viscosity of a liquid, based on the resistance to limb movements alone.

It integrates information from: 1) mechanoreceptors located in the muscles (afferent or inward channel), and 2) efferent (outward) information about the body configuration available to the brain [vH54]. The kinesthetic mechanoreceptors include muscle spindles appearing in parallel with the muscle fibers, and Golgi Tendon Organs (GTO) appearing in series with muscle fibers at their connection with the skeleton. The muscle spindles respond to the stretching of muscle, i.e. the change in muscle length. The GTOs, on the other hand, respond to force applied on the muscle (muscle tension) [Jon00].

In mainly kinesthetic tasks, the tactile sense may play the secondary role of indicating a contact. For example, in the primarily kinesthetic task of detecting and turning a door

knob in the dark, the tactile sense just indicates presence of contact through skin-pressure and deformation.

2.1.2 The cutaneous (tactile) sense

In an average adult human, the skin covers almost $2m^2$ of the body surface. Subcutaneous (under-the-skin) tactile mechanoreceptors provide us information about stimuli acting on the surface of the body. Such stimuli may include tapping, stretching, vibration, pressure, indentation, etc. of the skin.

Four kinds of mechanoreceptors exist in the glabrous (hairless) skin (see Figure 2.1) . They are parameterized by the size of their receptive fields (type I - small with well-defined borders, and type II - large with poorly defined borders), and their adaptation rate to a stimulus (slowly or rapidly adapting, i.e. SA or RA). Each mechanoreceptor is responsive to stimuli of a distinct nature [PAF+01]:

- The Merkel cells (SA-I), with their small receptive fields and high spatial resolution, mediate the perception of precise spatial patterns such as Braille dots and sharp edges.
- Meissner corpuscles (RA-I) also have a very high spatial resolution, In fact, they are even more densely packed into the human finger than the Merkel cells. These corpuscles mediate the perception of low-frequency vibrations (30–50 Hz), which may be generated during the relative motion of the finger with respect to a textured surface. The information they provide contributes to the accurate control of grip forces during grasping.
- Pacinian corpuscles (RA-II) are highly sensitive to nanometer-scale skin motion. They are also responsible for the perception of high-frequency vibrations (40 – 1000 Hz), such as the ones generated on the finger by fine surface textures. They adapt to the stimulus more rapidly than the Meissner corpuscles, thus providing information only about

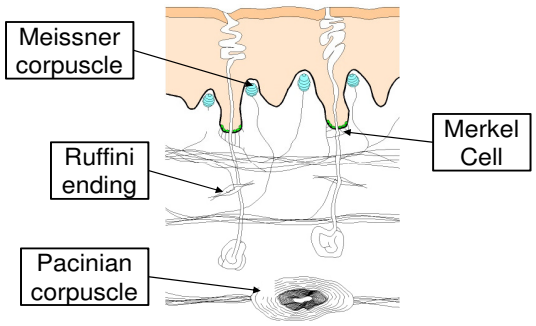


Figure 2.1: Afferent mechanoreceptors of the glabrous skin. Figure adapted from [AYS98], ©1998 IEEE.

dynamic mechanical stimuli. They are believed to be responsible for the perception of an object through a mechanical tool.

- The role of the Ruffini endings (SA-II) is still not as well understood as the other mechanoreceptors. They are believed to respond to the stretching of the skin caused by the movement of digits or limbs.

2.2 Mechanical signals that stimulate human haptic channels

In real-world interactions with objects in the immediate surroundings, humans execute manual tasks and perceive object properties through direct touch. In technical systems however, it is quite challenging to deliver realistic natural or artificial haptic feedback to humans consistent with this real-world experience. The following sections describe which physical signals can be used to represent haptic information, and how they can be sensed and delivered to humans in a technology-mediated way.

2.2.1 Kinesthetic signals

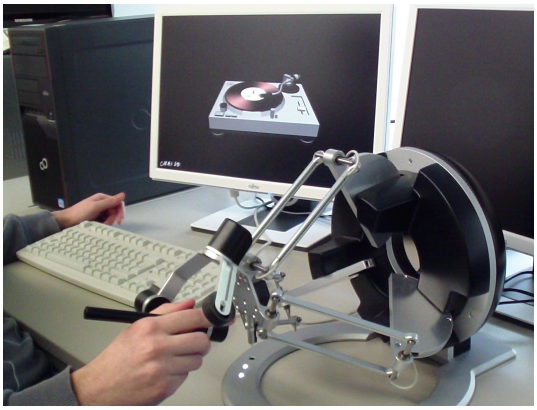
In technical systems, kinesthetic signals can be represented by forces and torques (F/Ts) generated during physical interactions with objects. F/Ts affect the static and dynamic human body state, and can thereby be sensed indirectly through the kinesthetic sensory system.

A haptic device, such as the one in Figures 2.2(a) or 2.2(b), is typically used to sense human motion (position/orientation of the hand, and the corresponding velocities), and present kinesthetic (F/T) feedback in the reverse direction. Haptic devices are characterized by parameters such as shape and size of the workspace, inertia, backdrivability, friction/damping, dynamic range of displayable stiffness, etc. [HA96].

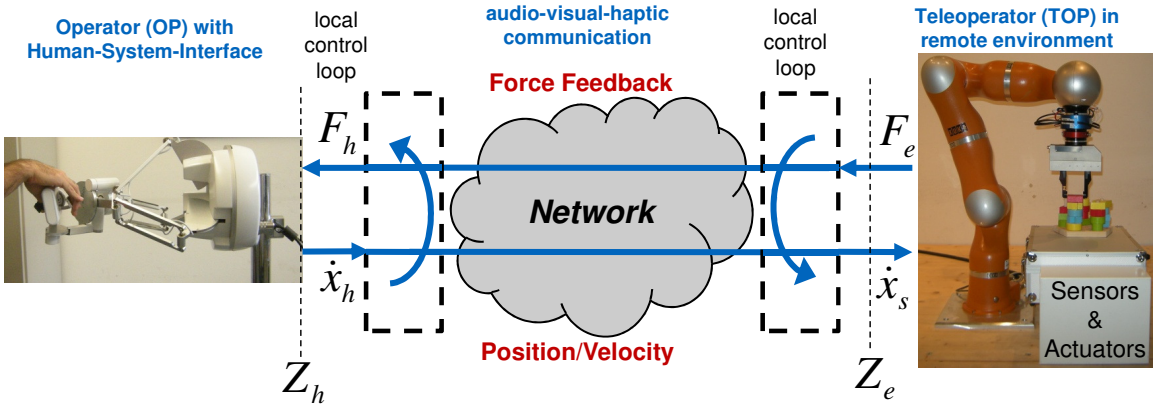
The human motion sensed by the haptic device can be used to control either the motion of a virtual end-effector in a virtual environment (VE) (see Figure 2.2(a)), or that of a real end-effector at the end of a robot-arm (see Figure 2.2(b)). Whenever the controlled end-effector collides with a virtual/real object, feedback F/Ts are either algorithmically computed (in case of a VE), or physically sensed (in case of real-world teleoperation). These F/Ts then drive the haptic device motors to display an impedance to the human hand, which stimulates the human kinesthetic system.

2.2.2 Tactile signals

Tactile mechanical stimuli mainly consist of vibration, pressure, skin-stretch, temperature, etc. Out of these, one of the most prominent — wideband vibration — is discussed more in detail here.



(a)



(b)

Figure 2.2: (a) Haptic interaction with a virtual environment. A standard haptic (force feedback) device, such as the one shown here, reads human motion commands through position sensors and in the reverse direction, displays force to the human hand via motors. (b) A haptic teleoperation system [FS67a]. The human motion $\dot{x}_h$ is sensed at the Human System Interface through the haptic device, and transmitted to the teleoperator robot over a communication network, where it serves as a reference for the robot's motion $\dot{x}_s$. When the robot comes in contact with its environment, forces/torques F_e are sensed and transmitted back to the human operator side, where they are displayed to the user through the haptic device as F_h. Figure reproduced from [CAS14], with kind permission from Springer Science+Business Media.

Vibrations carry information on the texture of surfaces, mechanical system defects or material properties. Wideband vibrations resulting from the scanning of textured surfaces with a mechanical tool can be represented by acceleration signals sensed from the tool [MRRK10]. The spectral response of these vibrations can represent their general feel, whereas their temporal profile can capture important transient events. Texture perception through vibrations is mainly mediated by the Pacinian corpuscles (see Section 2.1.2).

Several researchers have concentrated on the design and construction of vibrotactile transducers for delivering vibrations to the human hand. Electromagnetic shakers (see Figure 2.3), voice-coil actuators (similar in operating principle to shakers, but smaller in size and cheaper),

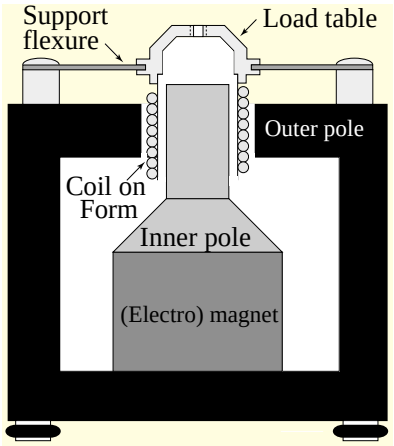


Figure 2.3: Cross-sectional view of an electromagnetic shaker. A permanent magnet or an electromagnet produces a radial magnetic field in a narrow air gap between the inner and outer magnet poles. A conducting coil is suspended in this gap using support flexures. When an AC current is passed through this coil, forces are produced along the coil's axis due to the interaction between the varying magnetic field of the coil and the radial magnetic field. These forces cause the load table to vibrate in proportion to the coil current. Figure based on [LS01], with kind permission from the Sound and Vibration magazine.

eccentric mass motors, piezoceramic actuators, bending wave actuators, surface acoustic wave (SAW) displays, and tactile pattern displays have been used to display tactile feedback. The feedback quality of the actuator is strongly influenced by the bandwidth of the device, frequency response, maximum feedback amplitude, resolution, and latency [Alt12]. Thus, the reproduction quality depends on the ability of a transducer to accurately reproduce a waveform. A detailed discussion of the reproduction characteristics of different actuator technologies can be found in [AM12].

2.3 Human haptic psychophysical limitations

The human sensory-perception system has certain limitations, irrespective of whether the auditory, visual, or the haptic modality is considered. Such limitations are usually captured in terms of:

- the perceptual detection threshold — the smallest perceivable stimulus magnitude,
- the perceptual difference threshold — the smallest perceivable difference in the stimulus magnitude, and
- the time or frequency-domain masking — the raised detection threshold of one stimulus component due to the presence of another one close in time or frequency.

The study of perceptual limitations is critical for the design of technical systems for sensing, reconstructing, and displaying physical stimuli to humans. This is because the knowledge of these limitations gives us the ability to relax the constraints on how precisely and accurately stimuli need to be reconstructed. The following sections discuss some perceptual aspects of kinesthesis and tactility relevant for technical systems.

2.3.1 Kinesthetic perception

The kinesthetic sense is stimulated by low-frequency forces/torques, for which the magnitude difference threshold is a very important perceptual limitation.

2.3.1.1 Weber's law

Ernst Weber, a German physician regarded as the founder of experimental psychology, was among the first to propose a mathematical relationship between the physical intensity of a stimulus and its phenomenologically perceived intensity [Web51b]. He proposed the size of the difference threshold (or Just Noticeable Difference, JND) to be a linear function of stimulus intensity. This phenomenon became known as Weber's Law of JND, and is represented by the following equation:

$$\frac{\Delta I}{I} = \kappa \quad \text{or} \quad \Delta I = \kappa \cdot I, \tag{2.1}$$

where I is the stimulus intensity, ΔI is the so-called Difference Threshold or the Just Noticeable Difference (JND) and κ is a constant called the Weber factor. With some variation, Webers law has been found to apply to every human sense including the haptic sense, over a wide stimulus range [Web51b, TSEC94]. Weber factors for a variety of kinesthetic phenomena are given in [Jon00]: this factor averages 7–10% for forces between 0.5–200N, 15–27% for forces below 0.5 N, 8–22% for stiffness, 14–34% for viscosity, and 21–113% for inertia.

Thus according to the Weber's law, a perceptual deadband exists around the stimulus magnitude, within which stimulus changes go unperceived. Perceptual force limitations captured by the Weber's law have been exploited successfully in several schemes for the efficient transmission of kinesthetic signals over communication networks [HHC$^+$08a, SHK$^+$11, SHE$^+$12].

2.3.1.2 Multi-DoF perceptual deadzone

Human perceptual limitations for multidimensional kinesthetic (force) signals were investigated by Barbagli et al. in [BSH$^+$06], by Tan et al. in [TBS$^+$06], by Pongrac et al. in [PFH$^+$06], and more recently by Kammerl et al. in [KCS10b]. The authors focused on the limitations in perceiving changes in force direction in [BSH$^+$06, TBS$^+$06]. Studies on human limitations in perceiving combined changes in force amplitude and direction were described in [PFH$^+$06]. Force amplitude and direction changes were studied separately in [KCS10b].

These studies collectively proved that the multidimensional extension of Weber's law for force signals is non-isotropic, which means that it is not uniform as a function of the direction of the force vector (see Figure 2.4).

It was found in [KCS10b] that depending upon the direction of force feedback, the Weber factor varied between 2.75% to 6.7%. On the contrary, similar to [BSH+06, TBS+06], force angular thresholds were found to be independent of the direction of force feedback and remained constant at around 1.85°. With these new findings, the work in [KCS10b] achieved an improvement of 26.7% in the packet-rate reduction performance in comparison to a isotropic Weber factor assumption.

Earlier on, the force angular thresholds were also found to be independent of the direction of force feedback in [BSH+06, TBS+06]. Therein the average angular threshold determined was much higher —18.4°, than that from [KCS10b]. It should be noted however that the two results are not directly comparable, since they were obtained under different experiment conditions (see Table 2.1).

Today's picture of haptic perception is still not complete but with more and more investigations being performed, the individual results contribute towards a comprehensive understanding of the underlying mechanisms and limitations that can be exploited in technical systems.

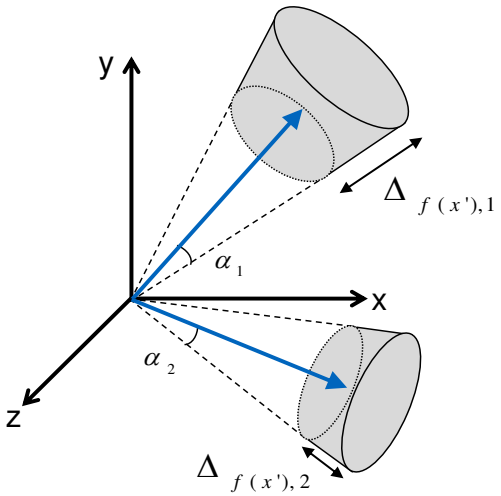


Figure 2.4: Stimulus discrimination zones (gray) have been found to be non-uniform as a function of the direction of the force vector ($\Delta_{f(x'),1} \neq \Delta_{f(x'),2}$ with $\alpha_1 = \alpha_2$, where $\Delta_{f(x')}$ and α denote amplitude and angular discrimination thresholds, respectively) [KCS10b]. Figure reproduced from [CAS14], with kind permission from Springer Science+Business Media.

	[KCS10b]	[BSH+06, TBS+06]
Motivation	To determine perceptual force angular thresholds that can be exploited for efficient networked haptic communication	To provide fundamental psychophysical results on force angular perceptual thresholds
Force feedback	Force was rendered in response to active contact and movement on a spherical surface by subjects.	Subjects passively perceived a fixed force magnitude profile in a test direction, and compared it to the same magnitude profile in the reference direction.
Psychophysical method	Subjects tuned a force-angle quantization parameter until no step-like artifacts could be perceived while moving on the sphere.	A 3-Down 1-Up adaptive staircase method (see [Lev70]) was used with the angle between the test and the reference force vectors as an independent parameter.

Table 2.1: Determination of perceptual thresholds on force direction — a comparison of the experiment conditions in [KCS10b] with those in [BSH+06, TBS+06]

2.3.2 Vibrotactile perception

In addition to the stimulus detection and difference thresholds, the Perceptual Masking phenomenon also plays an important role in tactile perception.

2.3.2.1 Detection and difference thresholds

As mentioned at the beginning of this section, detection thresholds are an important human perceptual limitation. In the related literature, haptic detection thresholds have been studied for different stimulus signals — force, acceleration, or even vibrator position.

Force and position detection thresholds from studies on vibrotactile perception over a frequency range mediated by the Pacinian corpuscles are shown in Figures 2.5 (a) and 2.5 (b), respectively. Both of them are seen to follow U-shaped profiles as a function of stimulus frequency, with peak sensitivity between 100–300 Hz. Furthermore, Figure 2.5 (c) shows difference thresholds for position signals as a function of the sensation level of a 160 Hz sinusoidal stimulus. It is seen that the difference threshold, and hence the Weber factor, is more or less independent of the stimulus intensity.

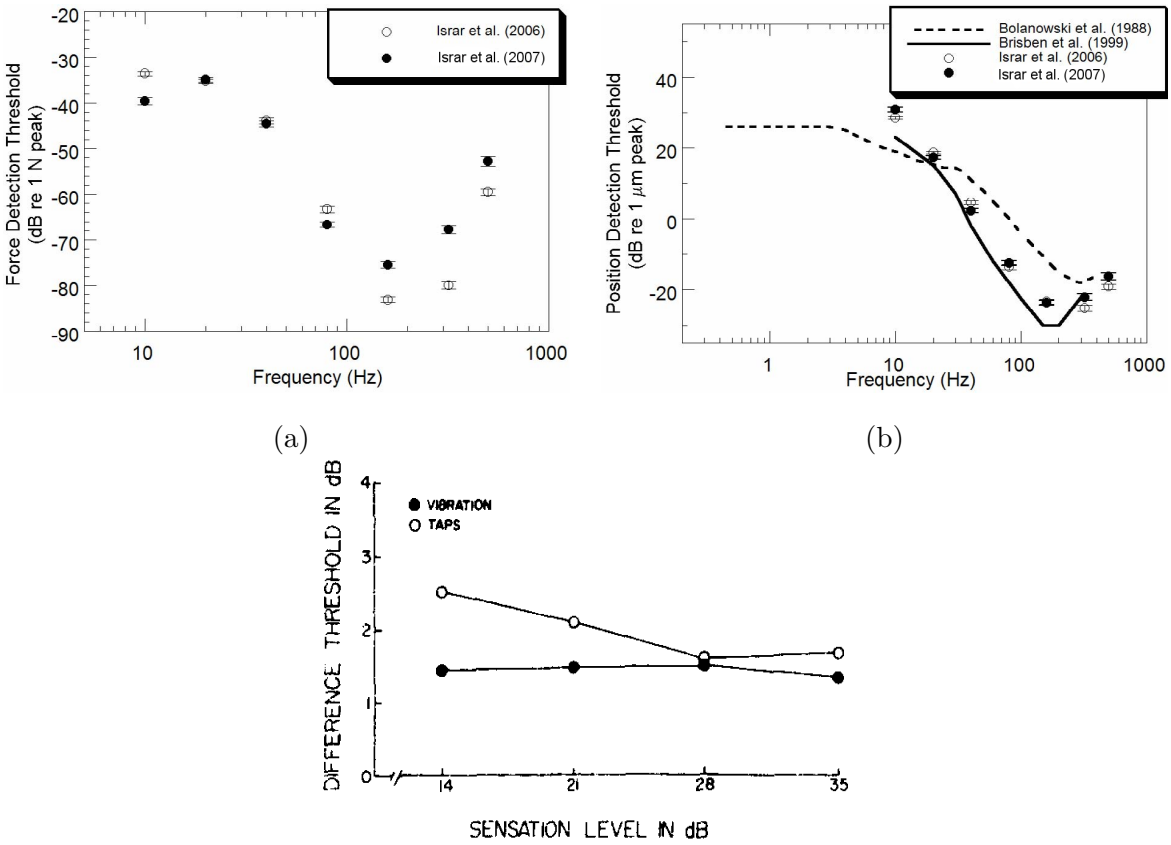


Figure 2.5: (a) Force detection thresholds. Figure reproduced from [ICT07a], ©2007 IEEE, (b) Position detection thresholds. Figure reproduced from [ICT07a], ©2007 IEEE, (c) Position difference thresholds for a 160 Hz sinusoidal stimulus. Figure reproduced from [Cra72], ©1972 Psychonomic Society, Inc.

2.3.2.2 Perceptual masking

It has been observed for all human sensory modalities that the presence of a strong stimulus component masks the perception of a weaker one near to it in space, time, or frequency [FZ56, BO07]. This masking manifests itself by way of increased detection and difference thresholds for the weaker component. Auditory masking has been studied in great detail in the past decades, and has contributed to very successful and efficient audio communication schemes. In the following, masking results for the auditory modality are reviewed first, followed by a survey of the masking studies in vibrotactile perception. It is found that vibrotactile masking has not received all-round attention in the literature, and a number of gaps in the state of scientific knowledge in this area are identified.

Auditory masking models A vast number of studies on auditory masking have been carried out since the 1950s. Psychophysical findings from these studies have been applied to auditory

signal/device design, and to compression of speech and audio signals. For the sake of brevity, this section refrains from a comprehensive review of the auditory masking literature, and focuses on a select few studies which are in essence suitable for application to speech/audio compression techniques.

Comprehensive results on auditory masking have been reported by Feldtkeller and Zwicker in [FZ56]. Figure 2.6 shows simultaneous masking patterns from [FZ56] for pure tones in the range of approximately 4–9kHz, with a narrowband noise centered at 6 kHz as the masker stimulus. It can been seen that the detection thresholds for the pure tones in presence of the masker rise well above the original detection thresholds in quiet. Moreover, it can be seen that the pure tones further away in frequency from the masker are increasingly easier to detect as compared to the ones close to it. This behavior is seen to hold for the entire range of masker noise intensity from 0–60 dB. Similar results were also reported before by Egan and Hake in [EH50] for a masker noise center frequency of about 410 Hz.

In [SAH79], Atal *et al.* proposed a speech coding algorithm exploiting the auditory simultaneous masking for hiding the coding noise below human perceptual thresholds. Despite the availability of previous masking results, further masking studies are reported in [SAH79] with the roles of the masker and maskee reversed as compared to [FZ56] (pure tones as masker stimuli and narrowband noise as maskee). However, the authors made the same qualitative observations about masking as the previous studies.

It should be emphasized here that the masking model deployed in speech coding is qualitative rather than quantitative, and the model parameters are tuned a-posteriori in subjective listening tests. The authors of [SAH79] also found that the model is rather insensitive to the model parameters, and that laborious tuning of these parameters is not required.

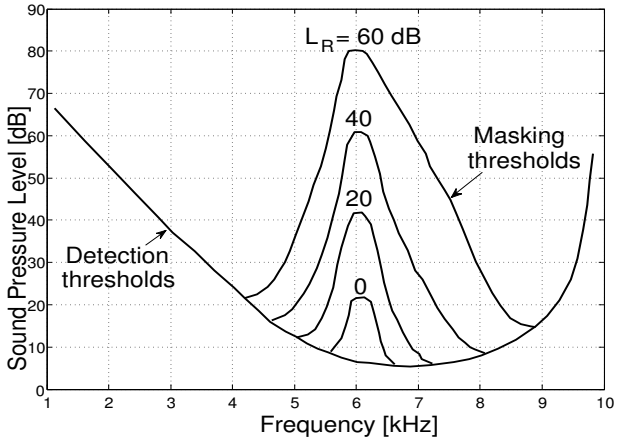


Figure 2.6: Wideband auditory masking characteristics. The increasing peaks represent masking curves for increasing masker loudness (0–60 dB). The pure tone detection thresholds in quiet are shown at the bottom. Figure reproduced from [FZ56], with kind permission from Hirzel Verlag, Stuttgart.

Vibrotactile masking models Vibrotactile perception uses sensory information received by cutaneous mechanoreceptors embedded in the human skin. Of the four kinds of mechanoreceptors, the research presented in this thesis focuses on the perception mediated by Pacinian corpuscles.

Some of the earliest vibrotactile masking results were reported in [GVD82]. Therein, the thenar eminence of the right hand was excited through a 2.9 cm^2 vibrating contactor. A narrowband noise centered at 275 Hz acted as the masker. Pure sinusoidal tones of frequencies 15 (non-Pacinian region), 50, 80, and 300 (all three in the Pacinian region) Hz acted as maskees. Detection threshold shifts for each of the maskees were obtained for masker levels ranging from 0–52 dBSL (sensation-level). It was found that threshold shifts rose in a monotonically increasing manner with increasing masker levels. Maskees located far away in frequency from the masker were found to have lower threshold shifts compared to the ones close to it. Also, from the case of the 15 Hz maskee, it was shown that cross-channel masking did not occur in the duplex (Pacinian and non-Pacinian) model for vibrotactile perception.

In-channel and cross-channel masking situations with both sinusoidal and noise maskers were explored in [HVZ83]. For in-channel masking, for both Pacinian and non-Pacinian regions, threshold shifts for the pure tone maskees increased with increasing masker levels. This was true irrespective of whether the masker was a pure tone or narrowband noise. It was once again shown that signals did not mask each other across channels, corroborating evidence for a critical band like structure in the tactile mechanoreceptor system, similar to that in audition. This conclusion was further consolidated in [MFCJV95b].

The influence of stimulus onset asynchrony (SOA) between the masker and the maskee on threshold shifts was studied in [GSJBV89]. SOA was varied over a range of –1000 to 1500 ms, to generate backward, simultaneous, and forward masking. Again, both sinusoidal as well as narrowband noise signals were used as maskers and in-channel masking was explored for both Pacinian and non-Pacinian channels separately. Higher threshold shifts were observed for simultaneous masking as compared to backward and forward masking.

The studies above have used relatively small localized areas on the human hand (index finger pad or the thenar eminence) to deliver their stimuli directly to the skin in a direction perpendicular to it. The haptic devices used today in teleoperation, however, commonly employ a stylus-like interface attached to a robotic arm, so that the user's movements can be easily mapped to that of the remote robot and appropriate reaction forces can be displayed during contact with remote objects. The human operator usually holds this stylus in a way similar to how one would hold a pen while writing (see Figure 1.1). The rich array of haptic cues like shape, stiffness, friction, texture, etc. that we perceive through the tool in such a setup are displayed to a large area on the hand, tangential to the surface of the skin. To the best of the author's knowledge, no masking data are available for such tool-mediated haptic

display in the literature. Moreover, the range and number of masker as well as maskee frequencies tested in most masking studies is rather limited, hence a wideband masking threshold pattern cannot be derived from these results.

To overcome the limitations mentioned above, new psychophysical experiments are reported in Section 3.2 of this thesis. Table 2.2 shows a detailed comparison of the work presented in Section 3.2 and the most important results available from the literature.

	[GVD82]	[HVZ83]	**Section 3.2 of this thesis**
Mechano-receptors stimulated	Pacinian as well as non-Pacinian	Pacinian as well as non-Pacinian	Only Pacinian, since they are responsible for tool-mediated haptic texture perception
Hand-area stimulated	Thenar eminence of the right hand	Thenar eminence of the right hand	Fingers and side of the hand (pen-like grip)
Vibration actuator used	2.9 cm^2 vibrating contactor	2.9 cm^2 vibrating contactor	Electrodynamic minishaker with a customized stylus
Masker	Narrowband-noise (NBN) centered at 275 Hz	i) Sinusoids at 70, 200, and 400 Hz, ii) NBN centered at 200 Hz	NBN centered at f_c = 120, 200, and 280 Hz
Maskees	Sinusoidal tones at 15, 50, 80, and 300 Hz	Sinusoidal tones collocated with the maskers	$(f_c - 3 \cdot \Delta f)$ to $(f_c + 3 \cdot \Delta f)$ in steps of Δf $(= 10\% f_c)$
Masker levels	0–52 dBSL (sensation-level)	-12 to +34 dBSL (sensation-level)	25 dB above detection-threshold, and only for 280 Hz, also 50 dB above detection threshold
Conclusions	i) No cross-channel masking, ii) masking threshold monotonically increasing with masker level, iii) masking threshold monotonically decreasing function of maskee frequency-difference to masker	Same as i) and ii) on the left	Confirmed conclusions ii) and iii) from [GVD82] for a stylus-like grip, and a wider range of masker and maskee frequencies

Table 2.2: A comparison of the novel Simultaneous Masking study presented in Section 3.2 of this thesis with major related contributions from the literature

2.4 Haptic data communication

When it comes to communicating haptic signals over a network in a multimodal system, kinesthetic and vibrotactile signals come with their own distinct sets of challenges. Table 2.3 shows the main points of difference between the requirements presented by the transmission of these two kinds of signals. The main challenge for kinesthetic (force and torque) signals is the requirement for a high packet-rate and a low-delay in the communication. On the other hand, the main challenge for vibrotactile signals is the requirement for a high bitrate within delay constraints less strict relative to kinesthetic signals.

	Kinesthetic data	**Vibrotactile data**
Signal bandwidth	3–10 Hz, in response to human motion bandwidth, and a haptic environment that can be modeled linearly (e.g. a spring)	40–1000 Hz for Pacinian corpuscle-mediated vibrotactile signals [BH03]
Sampling rate	1 kHz dictated by the control-loop stability requirement and human temporal resolution	2 kHz — Nyquist sampling rate
Delay constraints	Tens of milliseconds can destabilize the global control loop between the human and the environment, without additional control engineering measures [HS06, SHE+12]	40 ms delay is detectable by humans, above 60 ms subjective degradation creeps into the texture perceptual quality [OKST08]
Motivation for data reduction or compression	High packet-rate of 1000 packets/s difficult to maintain, may congest the network [SHK+11]	Distributed multi-channel feedback (e.g., artificial skin) can quickly lead to large data volumes/bitrates. Moreover, output bitrate is shared with high-bitrate audio-video.
Data reduction or compression strategy	Weber's law (see 2.1) used to select only perceptually relevant time-domain samples for transmission	Block-based perceptual linear-predictive signal coding (Section 3.3, [CcK+12])

Table 2.3: A comparison of kinesthetic and vibrotactile data

2.4.1 Packet-rate reduction for kinesthetic data

The Weber's law (see Section 2.3.1.1) forms the basis for almost all of the perceptually motivated haptic data reduction schemes. When a haptic sample is considered to be relevant for transmission, a Perceptual Deadband (PD) is defined using Equation 2.1 for the subsequent haptic samples (see Figure 2.7 for an illustration). As long as the subsequent haptic samples fall within this deadband, they can be dropped from transmission, as the signal-change they constitute is too small to be perceptible.

When a new sample violates the current PD, this sample is considered relevant for transmission. This sample then redefines the PD that is applied for the subsequent samples. Such a Perceptual Deadband Data Reduction (PDDR) scheme allows for signal-adaptive downsampling, since only those samples that constitute perceptually relevant signal changes are sent [HHC+08a].

Mathematically, the size of the applied deadband Δ_i at discrete time i is a function of a deadband parameter k (closely related to Webers JND parameter κ from 2.1), and the signal amplitude of the most recently transmitted haptic sample value h_{i-m}, if the last perception threshold violation took place m samples back in time:

$$\Delta_i^F = \Delta_{i-m}^F = k \cdot |h_{i-m}| \tag{2.2}$$

At the receiver side, the signal needs to be upsampled to a constant rate before display through the haptic device. To this end, a simple hold-last-sample approach has been proposed for data reconstruction in [HHC+08a]. The deadband scheme achieves packet-rate reduction of upto 90% without introducing any noticeable artifacts in the reconstructed signals [HHSB07].

In practical applications, haptic signals are 3D-vectors, corresponding to three Degrees-of-Freedom (DoF) available for motion. Hence, a 3DoF extension of the perceptual deadbands shown in Figure 2.7 is necessary to make sample-transmission decisions. For this purpose, the stimulus discrimination zones (or Perceptual Deadzones) shown in Figure 2.4 are used.

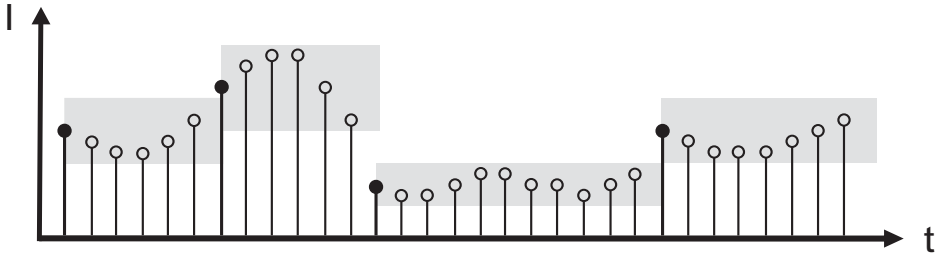


Figure 2.7: Principle of deadband-based data reduction. The thickness of the gray perceptual deadbands is a function of the haptic stimulus intensity I. Black circles represent haptic samples violating the applied perception deadband, and hence need to be sent to the receiver side. Figure reproduced from [SHK+11], ©2011 IEEE.

Whenever new incoming haptic signal vectors fall within the constructed deadzone, the change in signal is expected to be too small to be perceivable. This new haptic vector can therefore be dropped from transmission in order to reduce the transmission packet rate to a minimum. The experiment results in [KCS10b] show that with this extension to 3-DoF, a data reduction performance comparable to that in the 1-DoF case can be achieved.

Figure 2.8 shows an end-to-end block diagram of a kinesthetic communications system. In this diagram, efficient haptic communication is achieved on the forward haptic channel based on PDDR. The human velocity $\dot{x}_h$ is sensed by the Human System Interface (HSI). After undergoing PDDR, it is transmitted to the Teleoperator side, where it may be reconstructed using the hold-last-sample strategy. The reconstructed velocity acts as a reference input to the teleoperator, which further interacts with the environment at a velocity of $\dot{x}_s$. On the feedback haptic channel, the force feedback F_e sensed by the Teleoperator, which may again undergo PDDR before being transmitted to the human side for reconstruction and display.

But, also alternative data reduction approaches, e.g. the contact event-based coding from [KCS11] or the model-mediated approach presented in [Kam12], may be used on the feedback channel. In such approaches, signal models or models for the environment geometry or physics are identified online, and model parameters transmitted to the human side. At the human side, force feedback is rendered fully-locally based on the incoming model parameters. The resulting packet rate over the network is much lower at an average as compared to full 1 kHz haptic transmission.

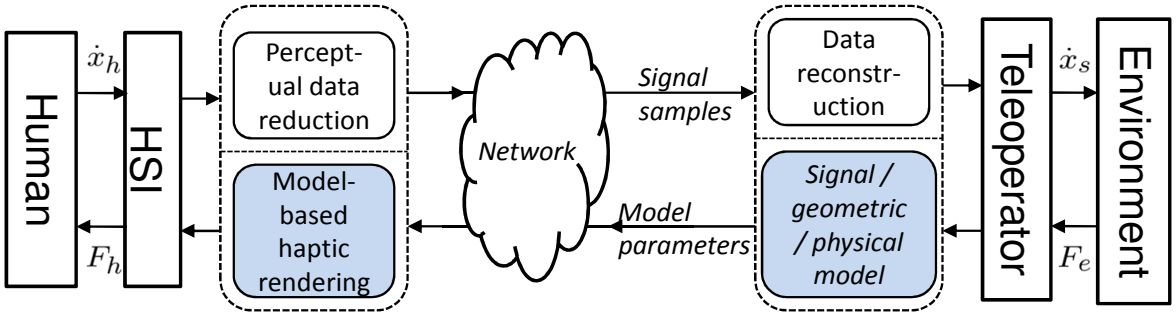


Figure 2.8: Block diagram of a haptic communication system

2.4.2 Compression of vibrotactile texture data

The reduction of kinesthetic data for both offline and online applications has been a very active area of research over the past decade. Various data reduction schemes which differ in their sampling, quantization, and coding strategies have been presented [HMF+01, SKB+01, SOK02, Bor05, HHC+08a, KS08, BS10, SHK+11, Kam12].

On the contrary, the efficient communication of vibrotactile data has received very less attention. To the best of the author's knowledge, the only other work on tactile data com-

pression was presented by Okamoto and Yamada in [OY13]. It describes a frequency-domain texture compression algorithm that exploits human vibrotactile perceptual limitations for compression.

In [OY13], textures were represented as height profiles (the height of a surface point as a function of lateral distance on the object surface), which were captured by an ultra high-resolution laser scanner. The height profiles were transformed to the temporal frequency domain using the Discrete Cosine Transform (DCT) assuming a constant scan velocity (refer to Figure 2.9). The DCT coefficients below human perceptual detection thresholds were then set to zero. The remaining coefficients were quantized with step-sizes determined from perceptual difference thresholds [Cra72]. Finally, this lossy-coded frequency-domain information was transformed back to the time-domain via inverse DCT, and further to the distance-domain assuming the same constant scan velocity as before. The reconstructed height waveform was then used for position-based texture rendering.

The perceptual quality of this algorithm was evaluated in subjective tests, where subjects scanned the height profiles of uncompressed and compressed textures with a haptic device [OY13] and compared them perceptually. It was shown that the proposed algorithm is perceptually transparent (i.e., the coding distortion goes unnoticed) up to a quantization as coarse as 12 levels. The algorithm achieves a very significant compression ratio of 4:1. However, when it comes to realtime haptic teleoperation, there are a couple of limitations to the applicability of this approach. Firstly, the algorithm is offline, meaning that the entire surface height profile is required in advance as an input to the algorithm. The second and more severe limitation is the assumption of a constant (and predetermined) texture scan velocity, necessary to transform the spatial-domain data to the temporal-frequency domain.

Compression was achieved in [OY13] by discarding information based solely on human perception (the *sink* for texture signals) models — namely, perceptual detection and difference thresholds. Section 3.3 of this thesis takes a different approach. In addition to a human perception model, a model of texture signal production (the *source*) with much fewer parameters than the number of texture samples used to build the model is employed for compression. The model parameters, rather than the signal samples themselves, undergo quantization before transmission. The perceptual simultaneous masking phenomenon (see Section 2.3.2.2) is exploited for judging the perceptibility of coding noise introduced by the modeling and quantization processes.

In [OY13], textures were represented by height profiles, whereas this thesis represents them by acceleration signals sensed by Microelectromechanical Systems (MEMS)-based acceleration sensors. MEMS sensors are used since they are small, inexpensive compared to position-sensing systems, and lightweight making them easy to mount on a teleoperator end-effector.

The subjective and objective evaluations show that for a comparable level of subjective

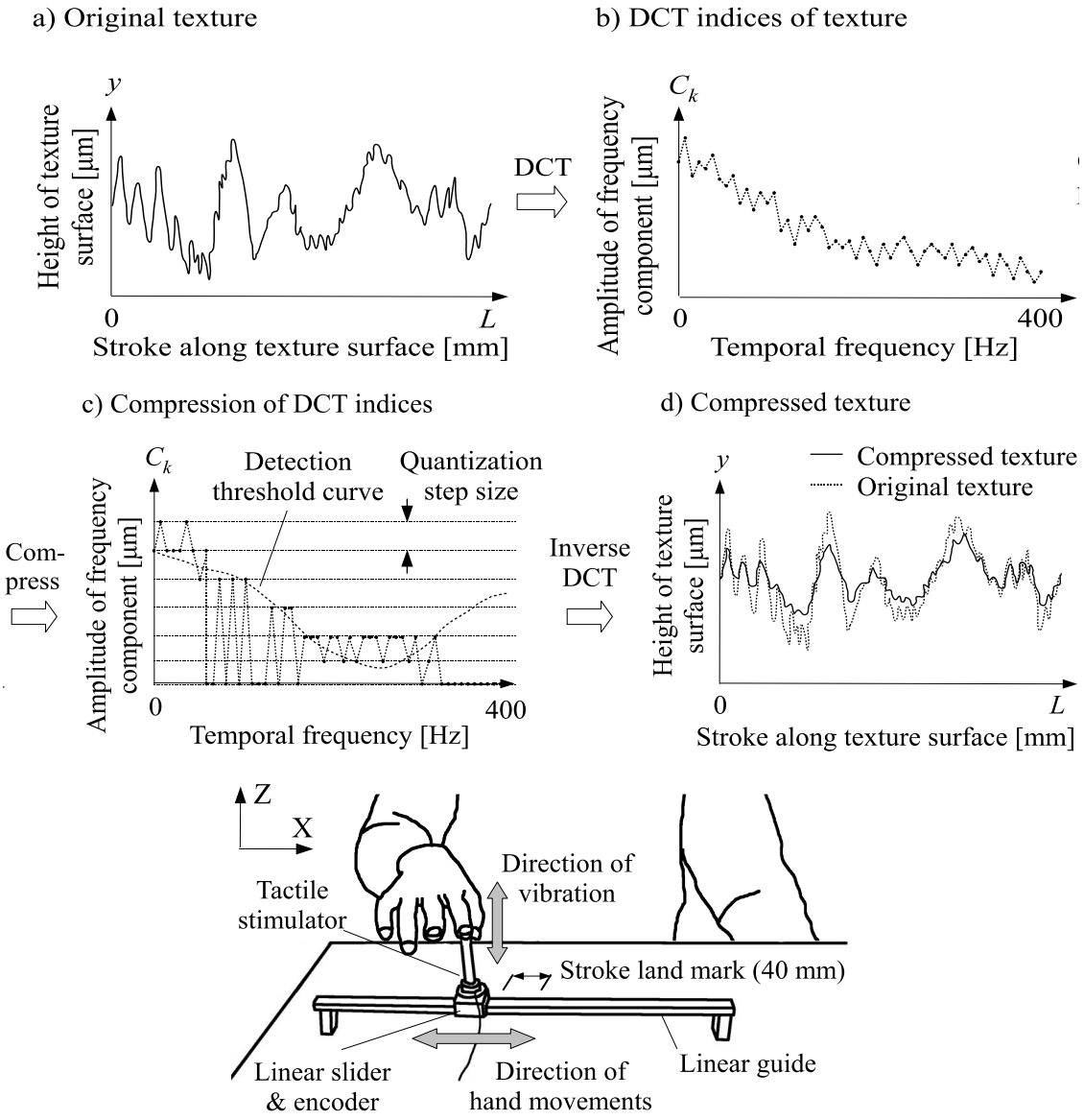


Figure 2.9: Okamoto's texture codec. The tactile actuator that displays the height profiles is shown at the bottom. Figures reproduced from [OY13], ©2010 IEEE.

quality, the compression performance of the algorithm presented in Section 3.3 is significantly better than that of the algorithm in [OY13]. Table 2.4 summarizes the points of comparison between these two algorithms.

2.5 Perceptual quality assessment in haptic communications

Ultimately, a technical (audiovisual-) haptic system would be deemed to have excellent performance, if it facilitates full subjective immersion in a remote/virtual environment. In other words, the system should facilitate the feeling of being "present" in that environment. To

	[OY13]	**Section 3.3,** [CcK+12]
Sensor	Laser-scanner	Micro-electromechanical acceleration sensor
Haptic texture representation	Position signal — height as a function of time	Acceleration signal
Approach	Block-based coding, perceptual thresholding and quantization of Discrete-Cosine-Transform coefficients	Block-based coding, texture production modeling (Linear Prediction), and texture perception modeling (Simultaneous Masking)
Compression ratio	Scalable from 4:1 down to 8:1	Scalable from 8:1 down to 16:1

Table 2.4: A comparison of the state-of-the-art texture codec [OY13] with the codec presented in this thesis

evaluate the Presence phenomenon, special questionnaires have been devised, e.g. in [WS98a]. Such questionnaires inquire for abstract concepts like user control over the system, naturalness of the interaction, the sense of object movement, the consistency with real-world experiences, the anticipation of system response, the ability to concentrate on the task rather than the technical means, etc.

On the other hand, a less ambitious evaluation approach concedes technical limitations, and instead only attempts to evaluate the system performance as perceived by a human user in comparison to the best capability of the system. For example, it may address the question: how much perceptual degradation do communication-induced artifacts introduce in comparison to a perfectly transparent communication link? Such an approach may search for perceptual thresholds within which a user cannot distinguish one system performance level from another.

Inspired by subjective evaluation methods from the audio and video fields, haptic coding algorithms have also been evaluated by rating their perceptual performance based on scales such as the one shown in Table 2.5 [Hin09, Kam12]. There are several shortcomings of this approach. Firstly, the adjectives used in the left column of the table are prone to individual interpretation and can be susceptible to response bias. At best, they should be regarded as categorical responses in the sense that subjects assign stimuli to five different classes according to their own personal criteria. Also, these criteria may be different from subject to subject. Secondly, assigning numerical values to the adjectives is also highly problematic. Since the interpretation of "disturbing" for one subject can be quite different from that for another subject, the same numerical value cannot be assigned to both subject responses. Furthermore,

Description	**Rating**
no difference	100
perceptible, but not disturbing	75
slightly disturbing	50
disturbing	25
strongly disturbing	0

Table 2.5: A rating scale for the perception of communication-induced artifacts in comparison to a transparent communication link

the assumption that the five responses are equally spaced on a numerical scale also cannot be easily supported or verified.

The Magnitude Estimation method from [Ste75, ZG80] avoids some of the above problems. Under this method, subjects assign each stimulus a continuous numerical value proportional to the magnitude of the stimulus they perceive, without the need for any adjectives. The data is processed in a way that normalizes inter-participant differences. But this method also suffers from response bias. Secondly, it may be suitable for rating just stimulus magnitude for brightness, loudness, or tactile intensity. But it cannot be carried over for rating complex communication-induced artifacts in a straightforward manner.

A Signal Detection Theory-based method avoids all of the above problems [MC05]. Herein, subjects are presented with a pair of stimuli and have to make a choice between two alternatives – do the two stimuli feel the "same" or "different". Apart from being very simple for the subjects to perform, this method allows one to separate perceptual sensitivity from response bias.

The methods outlined above entail extensive subjective testing, which is usually costly, difficult, and time-consuming. In haptics, it is especially so, since customized hardware makes it difficult to parallelize tests. Also, since subjects are typically not used to being delivered artificial haptic stimuli, extensive experimenter monitoring is required. These issues slow down the progress of technical developments. To circumvent this problem, mathematical/algorithmic models of human perception may be employed for automatically predicting haptic feedback quality.

2.5.1 Kinesthetic haptics

A technical kinesthetic haptic system is said to be *transparent*, if it enables a human to directly interact with the (remote) real/virtual environment, without noticing the mediating system itself [Min80, She92a]. This goal may be achieved by ensuring that the human-

side and the environment-side positions and forces are equal [YY94] ($x_{HSI} = x_{TO}$, $f_{HSI} = f_e$). Another definition of transparency requires the mechanical impedance displayed to the human to be equal to the environment impedance [Law93] ($Z_{HSI} = Z_e$). Ideal transparency according to the above criteria is not achievable in practice, in particular when lossy data reduction/compression is used or in the presence of time delay on the communication channel. This is since there is a fundamental trade-off between robust stability and transparency [Han89, Law93, HZS01]. Typical performance metrics derived from these transparency criteria include integrals (frequency or time) over position and/or force and/or impedance errors between the operator and the remote side.

A criterion which incorporates human haptic perception limits in the system evaluation was first introduced by Hirche et al. in [Hir05, HB06, HFB+07]. According to this criterion, the haptic system is deemed *perceived transparent* if the difference between the displayed impedance and the environment impedance is within the Just Noticeable Difference (JND) $Z_{displ.} \in [Z_{env} - JND, Z_{env} + JND]$. The effect of control and communication parameters on the perceived impedance was analyzed. Closed form solutions were obtained for certain settings. For example, the relationship between the displayed stiffness, the environment stiffness and the time delay on the communication channel was derived. In addition, this work allows for the optimization of control and communication parameters with respect to stability robustness under the constraint that the system remains perceived transparent.

In [SHPH09], Sagardia *et al.* provided a framework of testing scenarios and measures that help objectively evaluate the quality of haptic rendering algorithms by comparing the response of the algorithms with the expected analytical one.

A comprehensive classification of Quality-of-Experience (QoE) parameters for haptic VE applications was presented by Hamam et al. in [HESG08a, HESG08b, HS13, HSA14]. Both Quality-of-Service (QoS) influences and User-Experience (UX) parameters were considered in this taxonomy. QoS parameters considered were delay, jitter, and packet loss. The relevant UX parameters were classified as: perception-related parameters, quality of haptic rendering, psychological, and physiological parameters.

The perception-related parameters reflected how the user broadly perceived the haptics-based application. The rendering quality jointly considered the rendering of graphics, audio, and haptics. The psychological and physiological parameters captured the subjective and objective user-states respectively. A subset of parameters that represents these classes was selected for further evaluation: media synchronization (QoS parameter), fatigue and user intuitiveness (perception-related), haptic rendering (rendering quality parameter), and degree of immersion (psychological). While working with the same QoS and UX parameters, references [HESG08b, HSA14] differ from [HESG08a, HS13] in the way the subjective data for the above parameter subset was mapped to the overall QoE rating.

In [HESG08a, HS13], a holistic system-level mathematical model for haptic QoE based on weighted linear combinations of QoS and UX parameters was presented and validated:

$$QoE = \zeta \times QoS + (1 - \zeta) \times UX \tag{2.3}$$

where ζ was used to weight QoS parameters versus UX parameters,

$$QoS = \frac{\Sigma_l(\eta_l S_l)}{\Sigma_l(\eta_l)} \tag{2.4}$$

and

$$UX = A\frac{\Sigma_i(\alpha_i P_i)}{\Sigma_i(\alpha_i)} + B\frac{\Sigma_j(\beta_j R_j)}{\Sigma_j(\beta_j)} + C\frac{\Sigma_k(\gamma_k U_k)}{\Sigma_k(\gamma_k)} \tag{2.5}$$

where S_l, P_i, R_j, and U_k represent quality values for QoS measures, perception measures, rendering quality measures, and user-state measures, respectively, and $\eta_l, \alpha_i, \beta_j, \gamma_k$ are weighting factors which depend on the relative quality values within the QoS and user experience categories. Weightings A, B, C were determined empirically. Optimal weighting factors determined in [HS13] led to quality estimates with a high correlation with the subjective ratings (correlation coefficient 0.92, with $p < 0.005$). In comparison to the above approach, in [HESG08b, HSA14], fuzzy membership functions were defined for each of the above parameters, and the user data was mapped to QoE results using fuzzy logic rules.

Different from the previously mentioned approaches in [HB06, BCSS00, HS13, HSA14], Section 4.2 of this thesis takes a signal-based approach to the evaluation of haptic feedback. It focuses on the degree of perceptual quality degradation as the strength of the PDDR data reduction scheme is increased (i.e., as the packet rate is reduced). In the context of a haptic communication system, human action and perception models are presented, which enable the automatic prediction of perceptual quality. A comparison between the most important work in the literature and the work in Section 4.2 is shown in Table 2.6.

2.5.2 Vibrotactile haptics

In [OY11], Okamoto and Yamada presented a metric for quantifying perceptual quality degradation caused by haptic texture data compression. Their metric is based on the spectral model for vibrotactile perception presented by Bensmaïa et al. in [BHY05].

In [BHY05], Bensmaïa et al. hypothesized the presence of frequency-domain Gaussian-shaped "minichannels" mediating vibrotactile perception, and calculated the "activation" generated by a given stimulus in each of these minichannels. The perceptual dissimilarity $D_{S_1S_2}$ between two stimuli S_1 and S_2 was estimated by taking the difference between the activations Z_{S_1} and Z_{S_2} they generated in the minichannels, and by summing up the absolute differences across all minichannels: $D_{S_1S_2} \propto \sum_{f_c} |Z_{S_1}(f_c) - Z_{S_2}(f_c)|$, where f_c's are the center

	System parameters	**Quality parameters**	**Modeling approach**	**Remarks**
[HB06]	Constant communication delay	Transparency (in free-space and in contact)	Exploits the concept of JNDs for stiffness perception to derive wave impedance values for haptic transparency	Free-space: mass increases, in-contact: wall softer than actual
[HESG08a, HS13]	Media synchronization, fatigue and user intuitiveness, haptic rendering, and degree of immersion	Overall QoE rating	Weighted linear combinations of QoS and UX parameters	Correlation coefficient between the subjective rating and the model-based prediction 0.92
[HESG08b, HSA14]	same as above	same as above	Mamdani fuzzy inference system	Correlation coefficient 0.98
[OIFS12]	Network delay	same as above	no model proposed	Position-force control QoE degrades more seriously than position-position control QoE
Section 4.2 of this thesis	Perceptual Deadband data-reduction scheme parameter k	Subjective quality rating	Human action and perception models	Predicted data matches in trend with subjective data, correlation factor: 0.97

Table 2.6: A review of perceptual quality evaluation for kinesthetic haptics in the literature, and its comparison with the approach presented in this thesis

frequencies of 100 minichannels, placed in logarithmic increments in the Pacinian frequency range (40–1000 Hz), and minichannel activation

$$Z_S(f_c) = \sum_f \left(\frac{A_f^2}{T_f^2}\right)^{a_f} \cdot e^{-\frac{(f-f_c)^2}{2(\alpha f_c)^2}},$$

where frequency f lies in (40–1000 Hz), A_f is the signal amplitude and T_f the Pacinian detection threshold both as functions of frequency, and α controls the spread of the Gaussian filter.

The work in [OY11] supplemented the above power-spectral model with an analogous amplitude-spectral model (replacing (A_f^2/T_f^2) with (A_f/T_f)). This extra amplitude-spectral term accounts for Meissner corpuscle-mediated lower frequency vibrations (tens of Hz), which

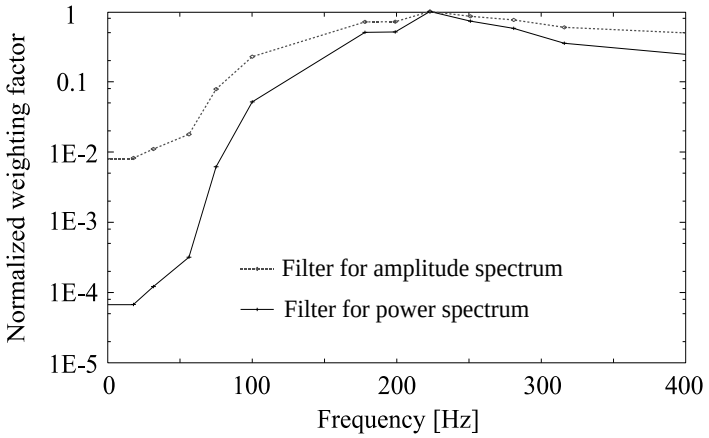


Figure 2.10: Psychophysical weighting filters for vibrotactile stimuli. The filter that weights the power spectrum deemphasizes low frequencies more severely in comparison to the filter that weights the amplitude spectrum. Figure reproduced from [OY11], ©2011 IEEE.

were previously ignored due to the inverse of the square of the (U-shaped) T_f function (see Figure 2.10). With this extension, the algorithm in [OY11] was able to predict perceptual dissimilarity for compressed texture signals with an improved Goodness-of-Fit $R^2 = 0.64$, up from 0.2 for the Bensmaïa model alone.

The above two models capture low-level perceptual properties of texture signals by focusing on one particular aspect of vibrotactile perception — the Pacinian/Meissner detection thresholds. On the other hand, the work presented in Section 4.1 of this thesis uses high-level descriptions of subjective perception — roughness [FL12], coarseness [BH05], brightness [HC10], regularity, time-envelope pattern [PC11], etc. — to predict perceptual data. These "features" are then weighted and combined to best predict perceptual degradation (dissimilarity) data for haptic texture compression. Please refer to Table 2.7 for a comparison of the related literature and this work.

	Haptic features	**Perceptual quality metric (quality degradation w.r.t. reference uncompressed signal)**	**Goodness -of-fit R^2**
[BHY05]	Gaussian-minichannel activation due to signal power weighted by squared detection thresholds	Sum-of-Difference of minichannel activations due to compressed and uncompressed signals	0.2
[OY11]	Linear combination of spectral amplitude and power, weighted by inverse detection thresholds and squared inverse detection thresholds, respectively	Difference of the linear combinations	0.64
Section 4.1 of this thesis	Linear combination of high-level haptic features like roughness, brightness, etc. with low-level psychophysical models [BHY05, OY11]	Same as above	0.84

Table 2.7: A comparison of the state-of-the-art perceptual quality evaluation approaches for haptic textures with the approach presented in this thesis

2.6 Chapter summary

This chapter has covered the background in haptics — physiology, technical haptic systems, haptic psychophysics, data communication, and perceptual quality evaluation — necessary for the reader to understand the motivation for the research presented in the upcoming chapters. The limitations of the state-of-the-art in haptic data compression and perceptual quality evaluation were identified. Furthermore, the chapter also outlined how the work presented in this thesis makes original contributions to address these limitations. Here is a summary of the most important points:

- Most of the research in this thesis deals with haptic textures, represented by wideband acceleration signals, whose perception is mediated by the Pacinian mechanoreceptors in the human hand.

- In technical systems, wideband texture signals can be sensed by an accelerometer, and displayed to the user via an electrodynamic minishaker or a voice-coil actuator.

- Potentially large data volumes and high bitrates motivate the development of efficient compression algorithms for haptic textures.

- Simultaneous masking, an important human perceptual limitation, can be exploited for efficient perceptually-transparent coding of haptic texture signals.
- Automatic model-based perceptual quality prediction is necessary to tackle the challenge of difficult and time-consuming subjects tests, which impede the pace of haptics research.

Chapter 3

Haptic texture compression

This chapter presents one of the main contributions of this thesis — namely, the compression of haptic texture data based on texture production and perception models. Furthermore, the topic of bitrate scalability of the texture codec is also addressed.

3.1 Mechanisms of texture production (source modeling)

As mentioned in Section 2.2.2, stroking a textured surface with a mechanical tool generates wideband vibrotactile signals (represented as acceleration waveforms) that encode the dynamics of the interaction between the hand-tool system and the textured surface. Figure 3.1 shows an illustration of such stroking.

Object surfaces can have coarse features with relatively large physical dimensions, or fine ones with random height profiles. While stroking the surface, the tool-tip usually hits a coarse feature once resulting in a response that can be modeled by a damped sum of sinusoids [GLL11], and then goes on to hit the next feature. On the other hand, the tool-tip has a more prolonged contact with the surface while traveling over fine features, resulting in a continuous random response. Coarse and fine features may also reside on the surface in an interleaved manner. In this case, the response can be thought to be a superposition of the coarse and fine responses.

Such a hand-tool-surface system can be modeled well by linear filters excited with appropriate input. The coarse features can be modeled as large-amplitude pulsed input, whereas the fine features can be modeled as a small-amplitude random noise input. Such modeling of the texture production mechanism is very reminiscent of the modeling of speech production [RS07].

The well-known model for speech production described in [RS07] represents the human vocal tract by an all-pole speech synthesis filter (see Figure 3.2). This speech synthesis filter

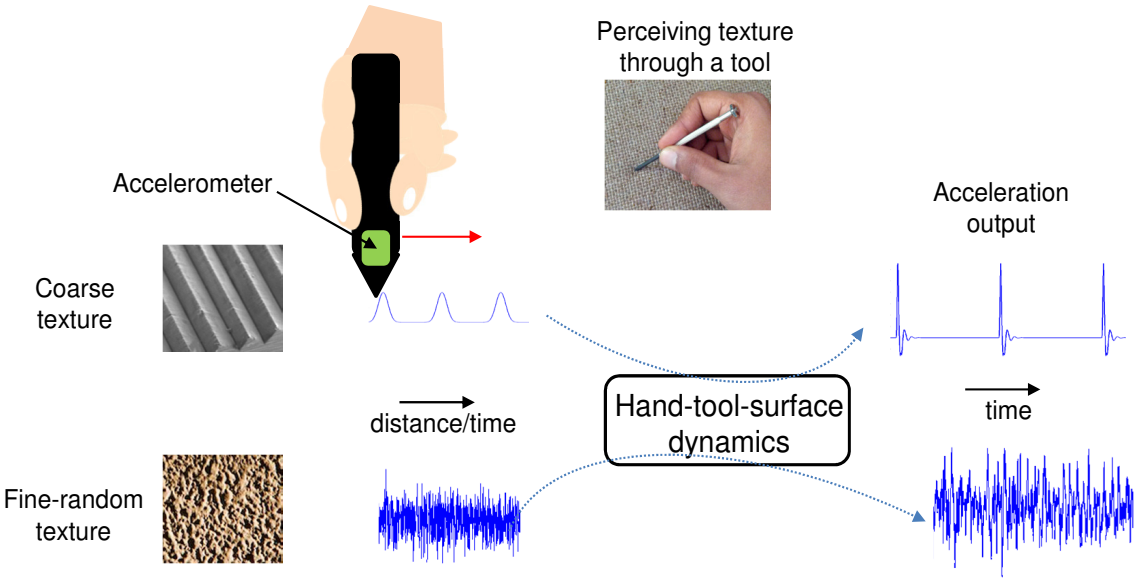


Figure 3.1: Mechanism of texture production. A hand-held tool is used to stroke a textured surface. An acceleration sensor signal mounted on the tool measures the response of the hand-tool system as it hits surface features.

is excited by an air flow originating from the lungs and shaped by the opening/closing of the vocal cords, followed by any intermediate constrictions on its way to the vocal tract. Depending upon the nature of this excitation (quasi-periodic pulses or random noise or a mixture of both), a variety of sounds are produced (voiced sounds or unvoiced sounds or voiced fricatives, respectively).

With the above resemblance, coarse textures can be treated as being analogous to voiced sounds (compare Figure 3.3 (a) to Figure 3.3 (b)), fine textures to unvoiced sounds (compare

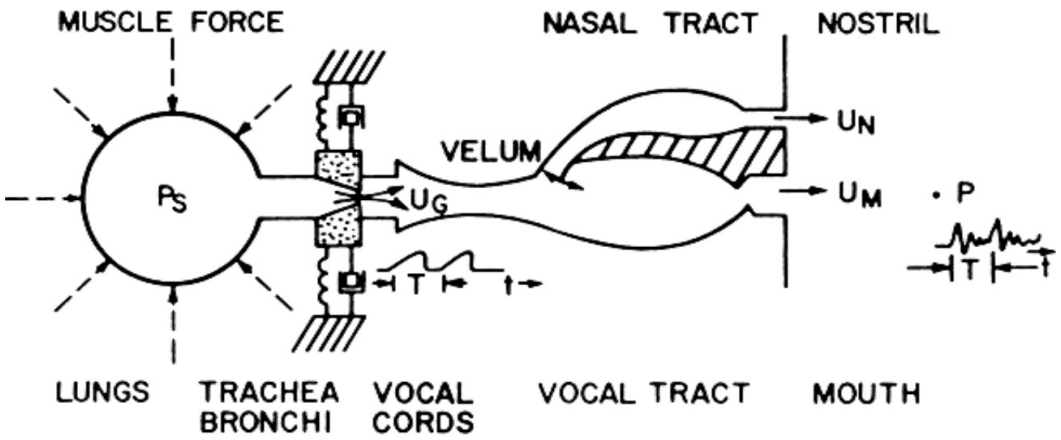
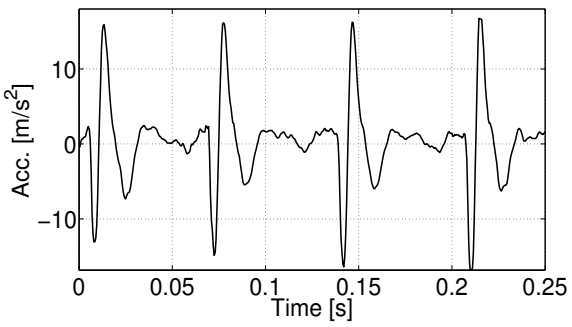
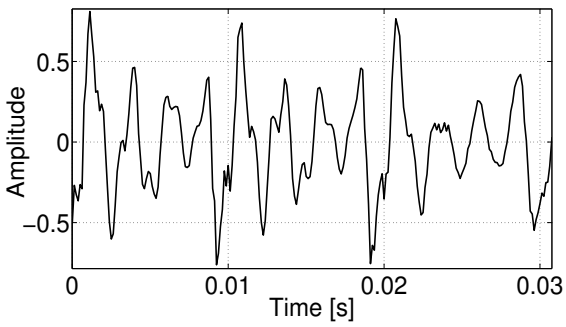


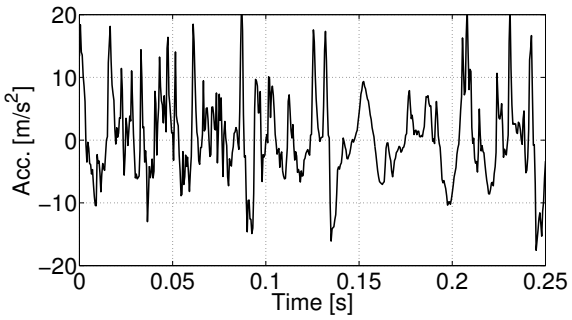
Figure 3.2: Mechanism of speech production. The vocal tract shapes the excitation from the vocal cords to form the sound output at the mouth. Figure reproduced from [FCR+70], ©1970 IEEE.



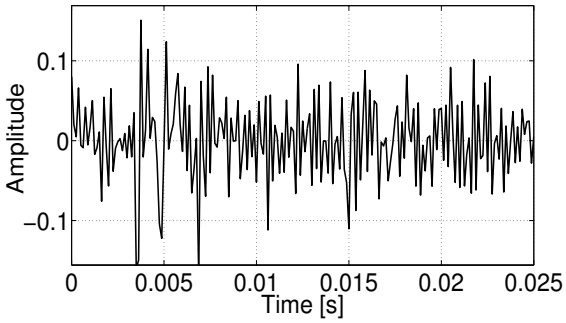
(a) Coarse (here, regularly patterned) texture



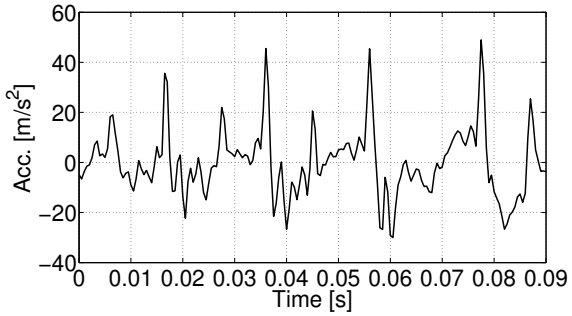
(b) Voiced speech



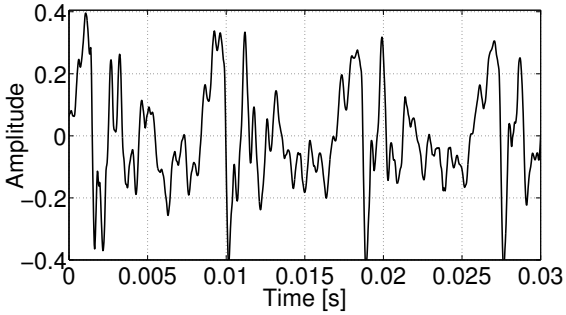
(c) Fine random texture



(d) Unvoiced speech



(e) Mixed texture



(f) Voiced fricative

Figure 3.3: [Left panels] Typical texture signal classes. Coarse (here, regularly patterned or quasi-periodic) texture is shown in (a). The noise-like appearance of fine random textures is shown in (c). A mixture of coarse surface features with fine intermediate features is shown in (e) ("Acc." denotes "Acceleration"). [Right panels] Typical speech signal classes. Quasi-periodicity of voiced speech is visible in (b). Random noise-like appearance of unvoiced speech is shown in (d). A voiced fricative (mixture of quasi-perdiocity and noise) is shown in (f). Figure reproduced from [CcK$^+$12].

Figure 3.3 (c) to Figure 3.3 (d)) and a mixture of coarse and fine textures to voiced fricatives (compare Figure 3.3 (e) to Figure 3.3 (f)). Thus, a very close analogy can be drawn between the production mechanisms of texture signals and speech signals. This provided the inspiration to study the use of speech coding techniques for developing a haptic texture codec. In particular, the Code-Excited Linear Prediction (CELP) coding technique [CT08] was found to be a suitable candidate for haptic texture compression.

The dynamics of hand-tool-surface interactions have been modeled successfully with all-pole Linear Predictive filters in [KRM11]. This past work provides a strong rationale for basing our coding approach on LP models of texture production. Moreover, Analysis-by-Synthesis coders based on Linear Predictive Coding (LPC) are also known to exhibit low bitrates with high subjective quality.

3.2 Perceptual simultaneous masking (sink modeling)

The ultimate sink of the haptic texture signals is the sensory-perceptual apparatus of a human. The distortions introduced in the texture signals over the source–sink chain (compression algorithm, communication channel, reconstruction) will go unnoticed by the human in comparison to the clean original texture signal, if they fall within certain perceptual bounds around the original signal. Therefore, models for dominant perceptual phenomena like Simultaneous Masking form a very critical component of a texture compression algorithm.

3.2.1 Psychophysical experiments

Simultaneous masking is a phenomenon in which the perceptual detection threshold of a weak (maskee) signal increases in the presence of another strong (masker) signal present in the immediate frequency neighborhood of the maskee. This is a fundamental human perceptual limitation that has been exploited successfully by various speech and audio coding schemes in the past, e.g. [SAH79, BS94].

Depending upon the time-offset between the masker and the maskee, three kinds of masking phenomena occur — forward masking, simultaneous masking, and backward masking. The experimental work presented next has simultaneous masking as its focus, while forward and backward masking are left for future work. The objective of the experiments is to investigate if and how the detection thresholds for vibrotactile maskees change in the presence of a masker, and how this change compares to auditory simultaneous masking.

3.2.1.1 Stimuli

The masker and maskee stimuli were displayed to the subjects through a K2004E01 electrodynamic minishaker (The Modal Shop, Inc., USA, see Figure 3.4). To avoid the beats

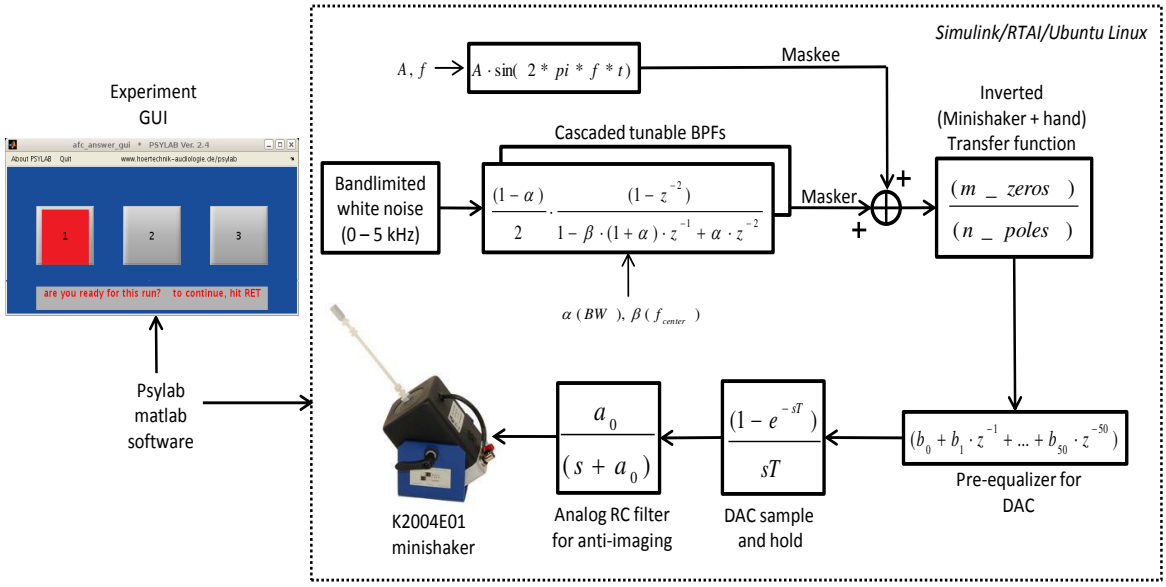


Figure 3.4: Block schematic for generating the masker and maskee stimuli. Figure reproduced from [CSDS15], ©2014 IEEE.

phenomenon, which occurs when two pure tones slightly different in frequency are added together [LKK12], narrowband noise (NBN) was used as the masker and pure tone sinusoids as the maskees.

The frequencies of the stimuli were chosen to be in the range 80–380 Hz, which falls within the sensitivity bandwidth of the pacinian mechanoreceptors. Three NBN center frequencies were spread out over this range — 120 Hz, 200 Hz, and 280 Hz. To generate narrowband noise, the second-order bandpass filter (BPF) given in (3.1) was driven by a white noise generator. The bandwidth of this BPF can be controlled by the parameter α, and its center frequency by β.

$$H_{BP}(z) = \frac{1-\alpha}{2} \cdot \frac{1-z^{-2}}{1-\beta(1+\alpha)z^{-1}+\alpha z^{-2}} \tag{3.1}$$

The three (scaled) BPF responses are illustrated in Figure 3.8 (a). This BPF is symmetric for a center frequency $f_c = Fs/4$, where F_s is the sampling rate. However, it becomes more and more asymmetric as the center frequencies are shifted away from $Fs/4$ on both sides. The effect of this asymmetry on the stimuli was minimized by making two design choices. Firstly, the BPF operated at a sampling rate of 800 Hz. This resulted in a perfectly symmetric BPF at 200 Hz; the other two BPFs at 120 Hz and 280 Hz were also relatively less asymmetric compared to a BPF operating at a higher sampling rate, e.g. 2 kHz. Secondly, the BPF's -3 dB bandwidth was fixed to 40 Hz, and its -40 dB bandwidth was tightened to 150 Hz, in an effort to minimize the asymmetry at 120 Hz and 280 Hz. This resulted in a set of 25 identical cascaded BPFs to generate a given NBN masker. Each of the BPFs in the cascade was controlled by the same values for parameters α and β.

The NBN masker intensity was fixed at 25 dB above the detection threshold. For every masker frequency f_c, the maskee frequency was varied from $(f_c - 3 \cdot \Delta f)$ to $(f_c + 3 \cdot \Delta f)$ in steps of Δf $(= 10\% f_c)$, resulting in 7 maskees per masker. The duration of both the masker and the maskee stimuli was fixed to 1 s, and both of them started and stopped at the same time instant. Both stimuli were windowed with a Hanning window (symmetric 100 ms rise and fall times) to reduce transient effects. Five maskees centered around the 200 Hz masker were also tested for a masker intensity of 40 dB above detection threshold. Figure 3.4 shows the block diagram for generating the masker and maskee stimuli described above.

In the perceptual tests described later in more detail, subjects determined the amplitude at which the maskee was just detected in the presence of the masker. This particular maskee amplitude is the masking threshold for the given masker-maskee combination.

3.2.1.2 Signal processing chain

Inverted minishaker+hand transfer function To ensure that a subject indeed felt stimuli as close to the commanded ones as possible, the non-flat minishaker+hand system response was neutralized as follows. While the subjects held the vibrating handle of the minishaker, the acceleration response at the hand was measured for a chirp input signal to the minishaker. The users maintained a consistent grip force by monitoring visual feedback from a FlexiForce (TekScan Corp., USA) sensor. Matlab functions were used to determine a linear transfer function based on this I/O data (see Figure 3.5):

$$H_{hd}(s) = \frac{3896}{s + 3896} \times \frac{s^2 + 730s + 2.9e04}{0.02128s^2 + 15.56s + 7903} \tag{3.2}$$

The influence of these dynamics was canceled out by incorporating the inverse model into the signal processing chain. Figure 3.5 shows the transfer function estimated for data that was averaged across users. These data were collected two days before the psychophysical experiments began. Before each subject session during the experiments, new hand-device data was collected and if required, the parameters of the above model were re-tuned to fit the new data.

Digital-to-Analog conversion The signal processing chain shown in Figure 3.4 was implemented in the form of a Simulink block diagram. A realtime executable was generated using the realtime interface for the Realtime Application Interface (RTAI)-Linux target operating system running on a standard PC.

The realtime process sends signal samples to the minishaker through the Digital-to-Analog (DAC) converter on a NI PCI-6221 Data Acquistion (DAQ) card. The time-domain sample-and-hold response of the DAC creates a sinc function roll-off in the frequency domain, which extends indefinitely on the frequency axis. This undesirable effect was alleviated to some extent, but not eliminated completely, by operating at an oversampled rate of 8 kHz. Therefore,

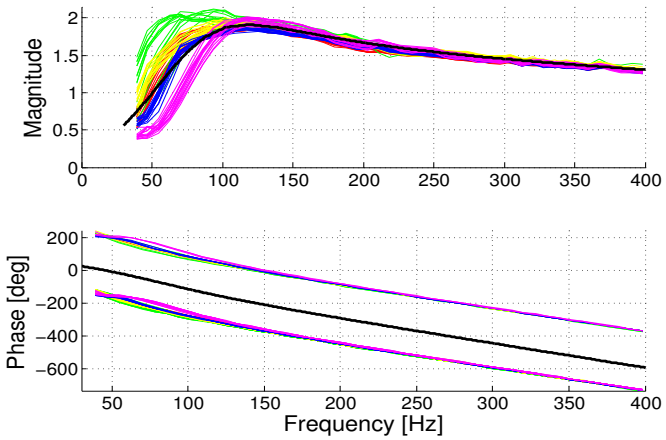


Figure 3.5: Frequency Response Functions (FRF)s plotted for the hand-device input-output data recorded for five users. The thick black curves show the linear transfer function approximation $H_{hd}(s)$ for these FRFs. Figure reproduced from [CSDS15], ©2014 IEEE.

a 50-tap FIR pre-equalizer was used to further cancel out the sinc roll-off and to obtain a flat frequency response in the bandwidth of interest.

3.2.1.3 Subjects and experiment setup

Five subjects participated in the pyschophysical experiments - 1 female, and 4 male. Their ages ranged from 24 to 29, with an average of 26 years. All of them were right handed, four had participated in haptic psychophysical experiments before, while one was inexperienced. Their self-reported experience with haptic devices ranged from "none" to "extensive". None of them reported having any ailments that would affect their sensorimotor performance.

A custom-made stylus-like handle similar to that of the PHANToM Omni (SensAble Technologies, Inc., USA) haptic device was mounted on the K2004E01 minishaker (see Figure 3.6). The subjects were instructed to hold the stylus like a pen in their dominant hand in a standardized 3-finger grip. Their elbow and forearm rested on a wooden plank, which supported a neutral wrist position. When the stylus was held as instructed, the vibrations transmitted through the stylus were mostly tangential to the skin in contact. The subjects wore acoustic noise-canceling headphones (QuietComfort 15, Bose Corp., USA) that played pink noise to mask out auditory cues from the experimental apparatus. They interacted with the experiment GUI through the keyboard.

3.2.1.4 Method

For determining the masking thresholds, a standard psychophysical method was chosen similar to the one used for characterizing the force and position detection thresholds in [ICT07b]. In

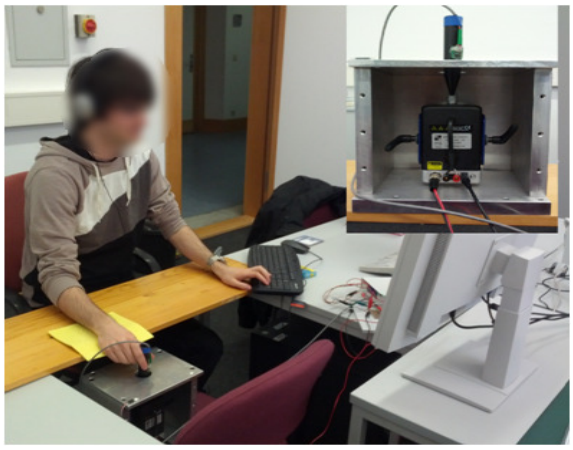


Figure 3.6: Experiment setup. The minishaker device is shown as an inset. Figure reproduced from [CSDS15], ©2014 IEEE.

the literature (see [JT13], for example), this method is referred to as the three-alternative forced-choice (3AFC) 1-up 3-down (1U-3D) adaptive staircase method.

Under this method, on each trial (a given masker-maskee combination) of the experiment, the subject was presented with three consecutive stimulus intervals (3A) with interleaved pauses. The masker was present in each of the three stimulus intervals. However, the maskee was present only in one randomly-selected interval out of the three. The currently active stimulus interval was indicated visually on the GUI. At the end of the trial, the subject had to respond by indicating the interval which she/he perceived to be different from the other two intervals. The next trial started after the subject's response.

The stimulus interval duration was chosen to be 1 s, since haptic adaptation to the stimulus is known to creep in after that. The pause duration was chosen as 600 ms, so that haptic enhancement and summation effects [VG75] were alleviated. Thus, the successive presentation of two stimuli did not affect their individual or collective perception. Haptic enhancement means the increase of the subjective magnitude of the second stimulus due to the presentation of the first one. Haptic summation, on the other hand, is the effect of increment in overall subjective magnitude of two successive stimuli. Each trial thus lasted about 5.2 s, which falls within the constraint of several seconds imposed by the haptic working memory [ASG08].

The step size through which the intensity of the maskee was increased/decreased was initially set to 8 dB for faster convergence towards the threshold. With each reversal of the staircase, it was halved until it reached a minimum of 0.5 dB (see Figure 3.7). The minimum step size of 0.5 dB was found to be a good trade-off between the precision of the masking threshold determined and the size of the experiment. The step-size stayed fixed at this value during the rest of the staircase. A staircase was terminated after three direction reversals (change of intensity from increasing to decreasing) occurred with the minimum step-size. The

masking threshold was calculated to be the mean of all the staircase points across the last three reversals. Every staircase was inspected by the experimenter visually. If the data had not converged, the subject was requested to repeat that staircase.

To make the experiment even more rigorous, each masker-maskee combination was evaluated by the subjects twice — once by tracing an ascending staircase, and once by tracing a descending one (see Figure 3.7). For the ascending staircase, the initial maskee amplitude was chosen to be well below the expected masking threshold level, so that it was not detectable at all at the beginning. On the other hand, for the descending staircase, the initial maskee amplitude was chosen to be well above the expected detection threshold level, so that the maskee was easily detectable at the beginning. The final threshold was calculated by averaging the thresholds obtained from the two staircases. Such averaging is standard practice for staircase methods used in psychophysics [Ges13]. The staircases thus guided the subject towards his perception threshold, at which she/he detected the maskee just as frequently as she/he did not.

Each staircase took about 10 minutes to finish. To avoid fatigue, subjects took a break of 1–2 minutes after every staircase. In all, each subject underwent 3 maskers $\times$ 7 maskees per masker $\times$ 2 staircases per maskee $=$ 42 trials. This amounted to about 7 hours per subject, excluding breaks. The whole experiment was scheduled over two weeks. Each subject performed her/his experiments over two days, 6 disjoint sessions each day (corresponding to 3 maskers, 2 staircases each) with 7 trials (corresponding to 7 maskee tones) per session.

Thresholds obtained with the 1U-3D adaptive method correspond to the 79.4 percentile point on the psychometric function, which is considered to be a reliable level in psychophysics [Lev70]. The higher this percentile, the lower the chance that the subject "guesses" her/his way to the final threshold, so that it is not his real threshold at all. The Matlab-based Psylab li-

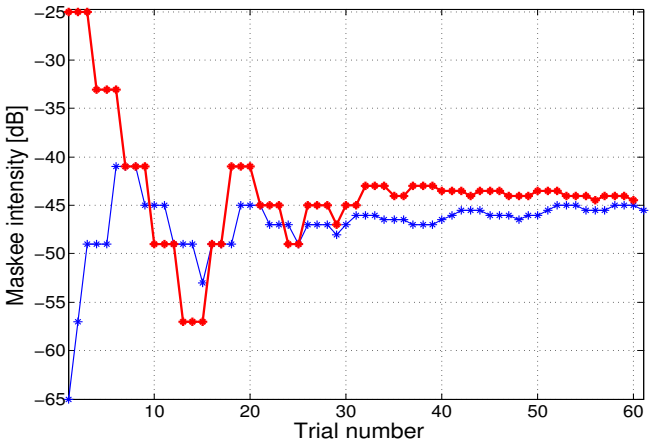


Figure 3.7: Ascending and descending staircases corresponding to a particular masker-maskee combination. Figure reproduced from [CSDS15], ©2014 IEEE.

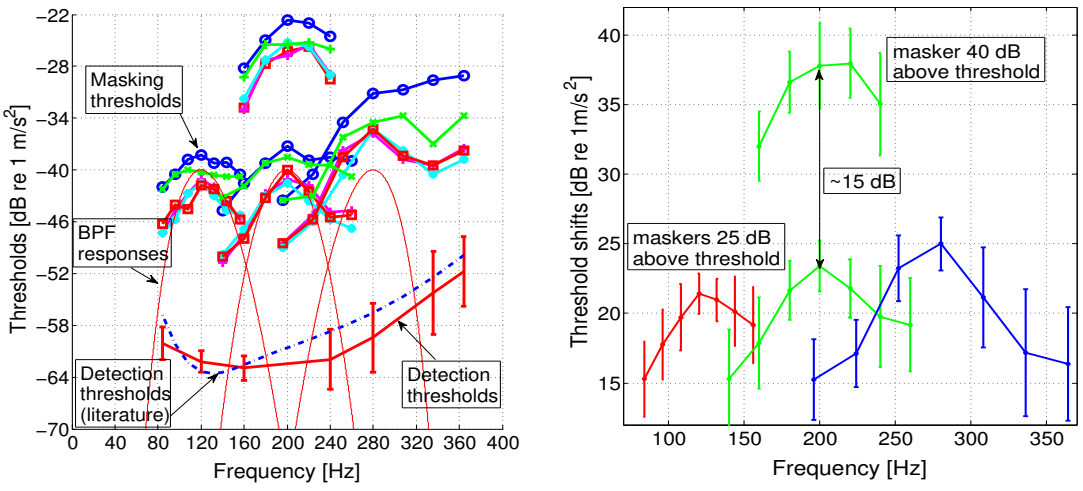
brary [Han06] implements the above staircase procedure, and was used to control the GUI as well as the stimulus parameters during the experiment.

3.2.2 Results

Figure 3.8 (a) shows the masking thresholds in comparison to the corresponding sine-detection thresholds. Figure 3.8 (b) shows the average threshold shifts, whereas Figure 3.9 shows the threshold shifts for individual subjects for each of the maskers. The figures show that the sine-detection thresholds shifted upwards in the presence of the masker for each of the three maskers. The figures also show that the masking thresholds were the highest when the sinusoid's frequency coincided with the NBN masker center frequency, and that they rolled-off gradually on both sides as the sinusoid's frequency went away from the masker center frequency.

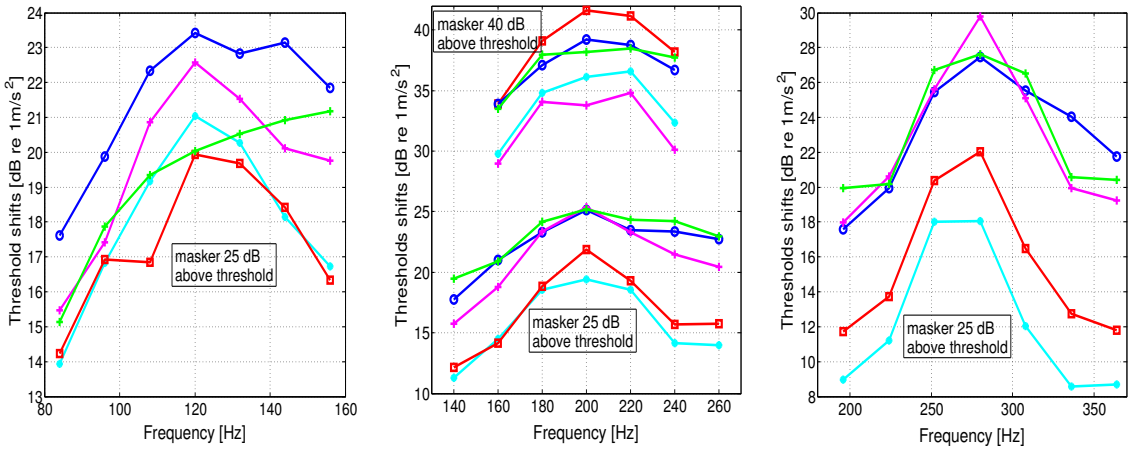
From Figure 3.8(b), it can be seen that for a masker intensity of 25 dB above threshold, the central maskee had an approximately 25 dB threshold shift up for all maskers. For a 200 Hz masker intensity of 40 dB above threshold, the threshold shifts were all approximately 15 dB above the 25 dB results. Thus, the 40 dB threshold shifts were higher proportionally with respect to the masker intensity, while appearing more or less parallel to the 25 dB ones in the frequency range considered.

Figure 3.8 shows that while the masking thresholds corresponding to the 200 Hz masker



(a) (b)

Figure 3.8: Results of the simultaneous masking experiments. (a) Sine-detection thresholds (with 95% confidence intervals) are shown at the bottom of the figure, in comparison with those from [ICT07b]. The upper part shows the masking thresholds for 3 masker frequencies and 5 subjects, (b) Average threshold shifts (= masking thresholds - sine detection thresholds), and the corresponding 95% confidence intervals. Figure reproduced from [CSDS15], ©2014 IEEE.



(a) (b) (c)

Figure 3.9: Threshold shifts for sinusoids around maskers at 120 Hz (a), 200 Hz (b), and 280 Hz (c). Figure reproduced from [CSDS15], ©2014 IEEE.

were more or less symmetric, the ones corresponding to the 120 Hz and 280 Hz maskers are less so. This asymmetry may be attributed to the asymmetric behavior of the band-pass filter that was used to generate the masker stimuli.

These results show that the simultaneous masking phenomenon for vibrotactile perception has a behavior very similar to simultaneous masking for auditory perception. Similar to speech/audio coding, masking can be exploited in the texture codec to hide the coding distortion below human perceptual thresholds. This will be explained in more detail in Section 3.3.5.

3.3 Linear prediction-based texture encoder

As described in Section 3.1, a strong analogy can be drawn between the mechanisms of production of haptic texture signals and speech signals. This provides the motivation for evaluating the adaptability of speech coding techniques for texture coding. The most widely used speech coding techniques are based on linear predictive (LP) modeling of signal segments (or frames), followed by the transmission of the corresponding (quantized) model parameters [CT08, RS07]. This technique, along with an algebraic code approximation of the associated LP filter excitation, was found to adapt well to texture coding. In the following, salient steps in the linear prediction-based coding algorithm are outlined.

3.3.1 Linear Predictive Coding

The sampled signal is subdivided into frames that are processed in succession. As mentioned before, the frame content is modeled by an LP-filter. The LP parameters and the corresponding filter excitation are encoded and sent to the receiver side for decoding and display.

In Linear Predictive Coding (LPC), future samples are predicted on the basis of a linear weighting of N past samples:

$$\hat{x}[n] = \sum_{i=1}^{N} a_i \cdot x[n-i] \tag{3.3}$$

where $\boldsymbol{a} = [a_i]_{i=1:N}$ are the LPC coefficients and N is the order of the LPC filter.

The mean square prediction error is given as:

$$E = \frac{1}{L} \sum_{n=1}^{L} (e[n])^2 \tag{3.4}$$

where L is the number of samples in the frame and $e[n] = x[n] - \hat{x}[n]$. E is minimized by setting the partial derivatives $\left(\frac{\partial E}{\partial a_i}\right)_{i=1:N}$ to 0, leading to a linear system of equations (for zero-mean $x[n]$):

$$\boldsymbol{R}_{N\times N} \cdot \boldsymbol{a}_{N\times 1} = \boldsymbol{r}_{N\times 1} \tag{3.5}$$

where $\boldsymbol{R}$ is the autocorrelation matrix of $x[n]$ and $\boldsymbol{r} = [R(1)...R(N)]$ (first row of $\boldsymbol{R}$). A well-known efficient recursive method, the Levinson-Durbin algorithm [RS07] can be used to solve for $\boldsymbol{a}$. If $\boldsymbol{A}(z) = \left(\sum_{i=1}^{N} a_i \cdot z^{-i}\right)$, the LPC Analysis filter can be written as $H(z) = (1 - A(z))$ and the corresponding synthesis filter as $P(z) = 1/H(z)$ (see Figure 3.10).

The general feel of haptic textures is governed by their spectral signature [KRM11]. In an LP-based algorithm, this feel is represented by the spectral response of the LP synthesis filter. In principle, it can be recreated by simply exciting the synthesis filter with white noise. However, realtime transient temporal events occurring due to surface features and scan velocity changes are equally important. Therefore it is necessary to transmit not only the LP

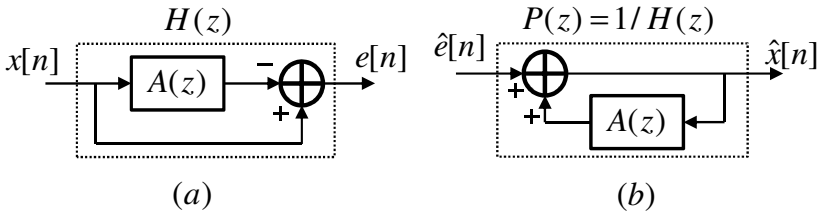


Figure 3.10: $H(z)$ - LPC analysis filter, $P(z)$ - LPC synthesis filter, $x[n]$ - input signal to be predicted, $e[n]$ - residual output of the LPC analysis filter, $\hat{e}[n]$ - an approximation made to $e[n]$ by the encoder, $\hat{x}[n]$ - reconstructed output, i.e., an approximation to $x[n]$ at the decoder output. Figure adapted from [RYK10], ©2010 IEEE.

parameters determined above, but also an encoded version of the excitation parameters that go with them and decide the temporal response of the LP synthesis filter.

As described in Section 3.1, two different kinds of surface features play a role in the haptic interaction of a tool with a textured surface — the coarse features and the fine random ones.

3.3.2 Macro-scale (coarse) surface features

For a tool moving with an almost constant velocity, coarse features produce quasi-periodic acceleration responses over a long time-duration. Coarse features can be modeled by an LP synthesis filter excitation with two parameters — the pitch period T (which corresponds to the time-duration between two coarse features), and the pitch gain β.

A preliminary range for T is first determined by searching for the peaks in the autocorrelation function for a buffer constituting of past and current samples. This is referred to as the "open-loop" search for T. A final refined value for T, and a value for β are then estimated in the so-called "closed-loop" search by minimizing the perceptually-weighted Mean Square Error (MSE), as will be described in Section 3.3.4 (Excitation-parameter search).

3.3.3 Micro-scale (fine random) surface features

Fine random features juxtaposed with the coarse ones, produce random responses. These responses are short-duration relative to the coarse features. Fine random features can be modeled by an LP-synthesis filter excitation with two parameters — an index i into a fixed codebook with an algebraic structure (referred to as "algebraic codebook" hereonwards), and a gain G.

Such a codebook contains codevectors with only a few non-zero pulses. Each of these pulses can have an amplitude of either +1 or –1. The pulses can assume only specific positions in the vector according to an algebraic code. The codebook is very efficiently searchable on account of its algebraic nature, and it does not need to be stored. Similar to the long-term excitation, both these parameters are estimated in a "closed-loop" search by minimizing the perceptually weighted MSE (see below).

3.3.4 Excitation-parameter search

Please refer to the encoder structure in Figure 3.11 for this discussion. Within this structure, branches 1, 2 and 3 successively remove their contributions from the original input texture signal, producing a new target signal for the branch below.

The contribution of Branch 1 is the Zero-Input-Response (ZIR) of the LPC synthesis filter. The ZIR is the response of the filter, only on account of its past memory, without any new excitation (switch s in Figure 3.11 open). This branch facilitates a smooth transition from one frame to the next.

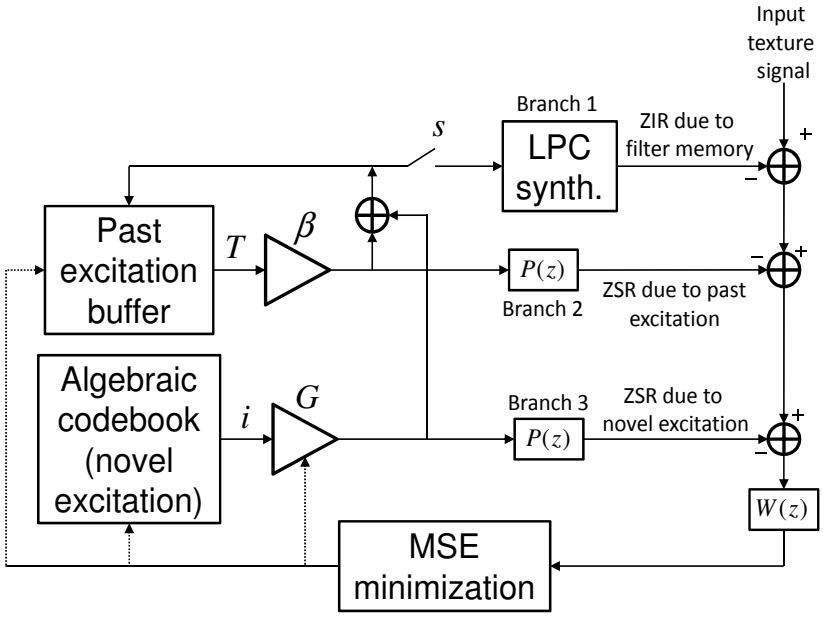


Figure 3.11: Search procedure for excitation parameters at the encoder side. ZIR: zero-input-response, ZSR: zero-state-response, $P(z)$: LP synthesis filter, W(z): perceptual weighting filter (see Section 3.3.5). Figure adapted from [CT08], with kind permission of Springer Science+Business Media and the authors.

The contribution of Branch 2 is the long-term Zero-State-Response (ZSR, the filter response with the filter memory set to zero) that accounts for the quasi-periodicity in the texture signal. Branch 2 searches for the best (corresponding to minimum MSE) past excitation that can be used for the current frame (due to repeatability of responses to coarse features).

The long-term excitation parameters T and β are determined as follows. Once the autocorrelation-based open-loop search gives an initial range for T, a buffer is populated with sections of past excitation corresponding to the initial range of T values. Then, an interative closed-loop search is performed, which yields the best estimates of T and β leading to minimum MSE. Branch 2 removes the contribution of the past excitation corresponding to these final estimates from the target signal generated by Branch 1, producing a new target signal for Branch 3.

The contribution of Branch 3 is the short-term Zero-State-Response. It accounts for the novelty in the current signal frame that could not be "explained away" by the upper branches. Branch 3 also removes its contribution actively by searching for the best algebraic codebook parameters, leading to a minimum perceptual MSE.

After the three branches are executed in an order and all excitation parameters are found, the switch s (in Figure 3.11) closes momentarily to update the filter memory and the past excitation buffer for processing the next frame.

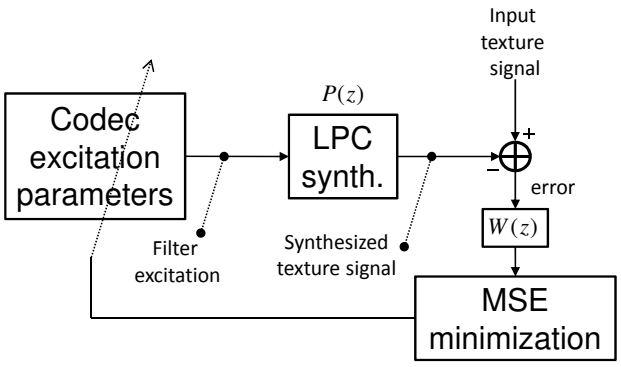


Figure 3.12: The excitation parameters are optimized to obtain minimum perceptual MSE. Figure reproduced from [CcK+12].

3.3.5 Perceptual weighting

The LPC-based coding algorithm is lossy since it *approximates* the spectral response with its poles and the synthesis filter excitation with an algebraic code. Moreover, these codec parameters are vector-quantized before transmission. Thus, the coding process introduces distortion into the encoded signal. To ensure the perceptual transparency of the codec, the coding distortion should be kept as low as possible (ideally below perceptual thresholds). In other words, the difference between the codec input and output signals should be as low as possible (ideally imperceptible). It is therefore necessary to include a perceptual model for texture perception in the coding process.

This is achieved by deploying a perceptual weighting filter $W(z)$ that captures the simultaneous masking phenomenon (see Section 3.2). This filter weights the MSE between the original texture signal and the texture signal reconstructed by the codec. The excitation parameters for the LP filters are searched for by minimizing this perceptually weighted MSE (see Figure 3.12).

The simultaneous masking phenomenon is accounted for in the codec by defining the perceptual weighting filter as a function of the LP synthesis filter response as:

$$W(z) = \frac{P(z/\alpha_2)}{P(z/\alpha_1)} \tag{3.6}$$

where $0 < \alpha_{1,2} < 1$. Thus, the poles of $W(z)$ lie at the same angles as the poles of $P(z)$, but at radii α_2 times those of $P(z)$'s poles. Furthermore, the zeros of $W(z)$ lie at the same angles as the poles of $P(z)$, but at radii α_1 times those of $P(z)$'s poles.

If $\alpha_1 > \alpha_2$, the frequency response of $W(z)$ is like a *controlled* inverse filter of $H(z)$ [RS07] (see Figure 3.13 for an illustration). This nature of the frequency response of $W(z)$ assures that the coding process admits higher noise energy in the frequency regions where the signal energy is high, and low noise energy in the regions where the signal energy is low. Due to simultaneous masking, it is expected that in both these cases, the high/low noise energy will

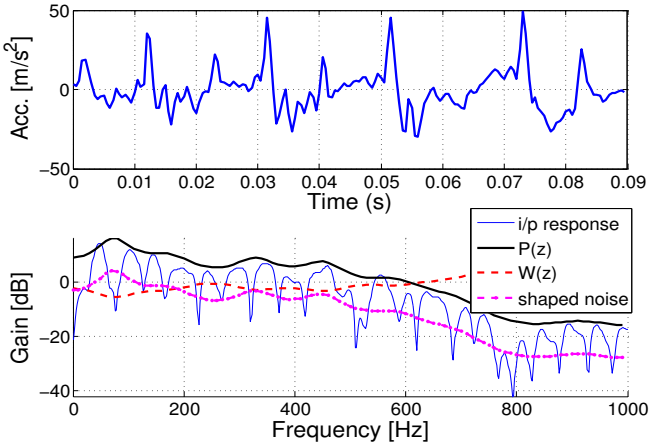


Figure 3.13: [Top] An acceleration signal segment, [Bottom] the frequency response of the LPC synthesis filter $P(z)$ estimated from the segment, and of the corresponding perceptual weighting filter $W(z)$. Figure reproduced from [CSDS15], ©2014 IEEE.

be masked by the corresponding high/low signal energy. Ideally, this would completely hide the coding noise perceptually from the subject. Parameters α_1 and α_2 were tuned to about 0.92 and 0.55 through pilot subjective tests.

3.4 Decoder

At the decoder side (see Figure 3.14), the LP filter and excitation parameters are recovered from the received bit stream. The excitation is then reconstructed by adding the past-excitation and algebraic codebook vectors scaled by their respective gains. It is then passed through the LP synthesis filter to reconstruct the texture signal $\hat{x}[n]$.

3.5 Codec implementation

While the previous sections conceptually explained the linear prediction based texture codec, this section gives key implementation details about the codec.

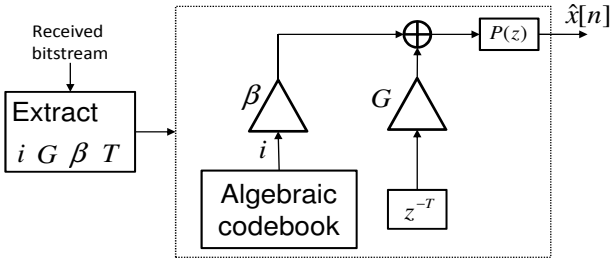


Figure 3.14: The decoder side. Figure reproduced from [CcK+12].

Parameter	Bits per frame
Quantized LPC coefficients	17
Pitch delay T	14
Algebraic-codebook index i	26
Quantized codebook gains β and G	14
Total	71

Table 3.1: Bit budget allocation

The human vibrotactile perception mediated by the Pacinian neural channel is sensitive to vibrations upto a bandwidth of 1 kHz [BH03]. Hence, the texture signal was sampled at the Nyquist rate of 2 kHz. In order to remove DC-bias, the sampled signal was then digitally high-pass filtered with a 4^{th} order Butterworth filter with a cutoff frequency of 40 Hz.

Texture signals are short-term stationary for natural texture scan motions (slow changes in scan velocity and operator applied force), and when the surface features do not change drastically. For natural scans, it was found that an LPC analysis buffer duration of 0.06 s is a good choice for time/frequency-domain tracking performance. The sampled texture signal was therefore subdivided into frames of length $L = 120$ samples (60 past + 40 current + 20 look-ahead samples) that were processed in succession.

The LPC analysis took place per frame and had an order of $N = 10$, with which 5 distinct spectral peaks could be captured. The excitation search took place on a per-subframe basis, with each subframe being 20 samples-long. The open-loop T determination used 143 past samples and 40 samples from the current frame. The algebraic codebook contains a set of generic 20 sample-long vectors. Each of these vectors contains four non-zero pulses, with the rest of the samples set to zero. The overall bit budget of 71 bits/frame was distributed among the various codec components as shown in Table 3.1.

The texture codec described here was implemented as a Simulink model. Such a model-based design allowed for efficient optimization due to the signal recording and plotting facilities of Matlab/Simulink. Also, a Linux Realtime Application Interface (RTAI)-interface for Simulink facilitated the generation of a realtime executable for the subjective experiments.

3.6 Codec performance

This section describes the evaluation of the codec for various objective and subjective performance criteria.

3.6.1 Bitrate

The uncompressed bitrate for texture signals:

$$\begin{aligned} \text{bitrate}_{\text{uc}} &= \text{sampling rate} \times \text{number of bits/sample} \\ &= 2000\,\text{samples/sec.} \times 16\,\text{bits/sample} \\ &= 32\,\text{kbps} \end{aligned} \tag{3.7}$$

The codec algorithm described above encodes every frame of 40 samples with 71 bits. Accordingly, the compressed bitrate can be calculated as follows:

$$\text{frames/sec.} = \frac{2000\,\text{samples/sec.}}{40\,\text{samples/frame}} = 50 \tag{3.8}$$

$$\begin{aligned} \text{bitrate}_{\text{c}} &= 50\,\text{frames/sec.} \times 71\,\text{bits/frame} \\ &= 3.55\,\text{kbps} \end{aligned} \tag{3.9}$$

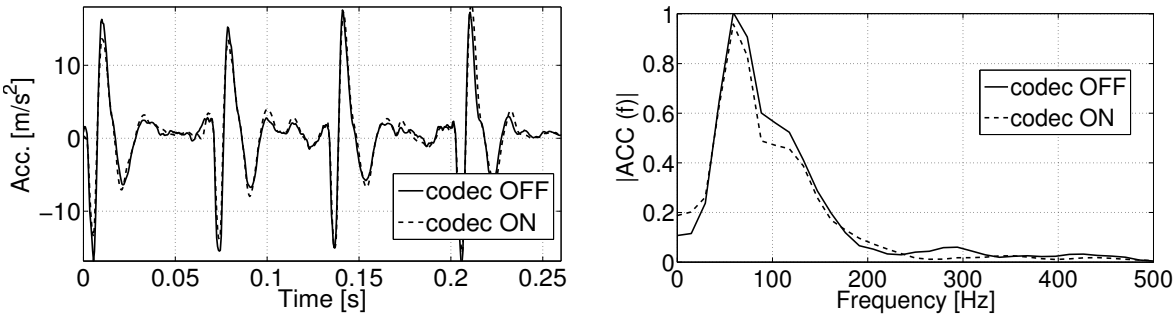
This amounts to a compression ratio of approximately 8:1. To match this performance, for every frame of 40 samples, the Okamoto codec described in Section 2.4.2 [OY10, OY13] should also generate 71 bits. That is, the corresponding 40 DCT coefficients should be quantized with a 2-bit (4-level) linear quantizer. However, [OY13] reported a compression ratio of approximately 4:1 with a 12-level quantizer (fewer than 12 levels trigger significant perceptual differences between the original and quantized textures). This comparison shows that in terms of bitrate reduction, the CELP-based codec described above performs better by at least a factor of 2 than the texture codec from [OY13].

3.6.2 Delay

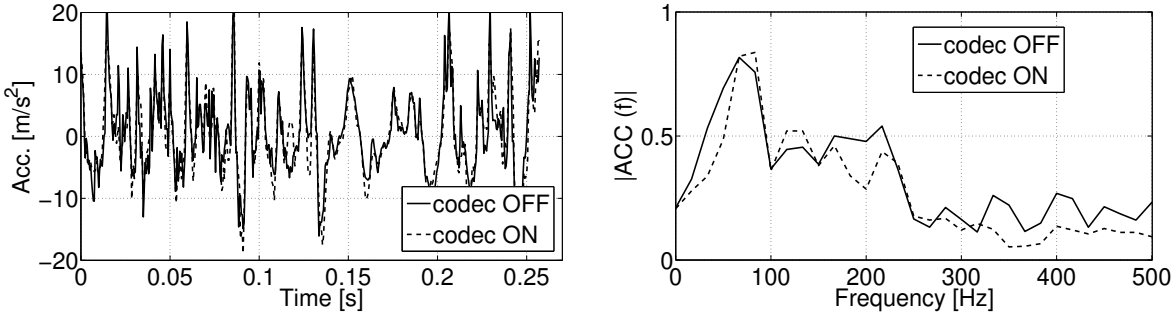
The codec introduces a delay of 20 ms into the signal stream due to frame buffering (40 samples per frame at a sampling rate of 2 kHz). Additionally, a look-ahead delay of 10 ms occurs due to buffering of 20 future samples. On a common PC configuration (Intel(R) Xeon(R) CPU EE5506, quad-core @2.13 GHz), the average encoding time was measured to be around 230 µs, and the average decoding time was around 18 µs. Thus, the encoder and decoder introduce a negligible delay in comparison with the buffering delay of 30 ms. A summary of the main codec performance parameters is given in Table 3.2.

During haptic manipulation, the presence of delay in the haptic signal stream is detected by humans at approximately 40 ms [OKST08]. Furthermore, subjective degradation in the perception of textures creeps in at about 60 ms of delay (also see [OKST08]). Thus, a further channel transmission delay of around 10 ms is tolerable before the human delay-detection threshold is reached. A further channel transmission delay of around 30 ms is tolerable before subjective degradation in texture perception results.

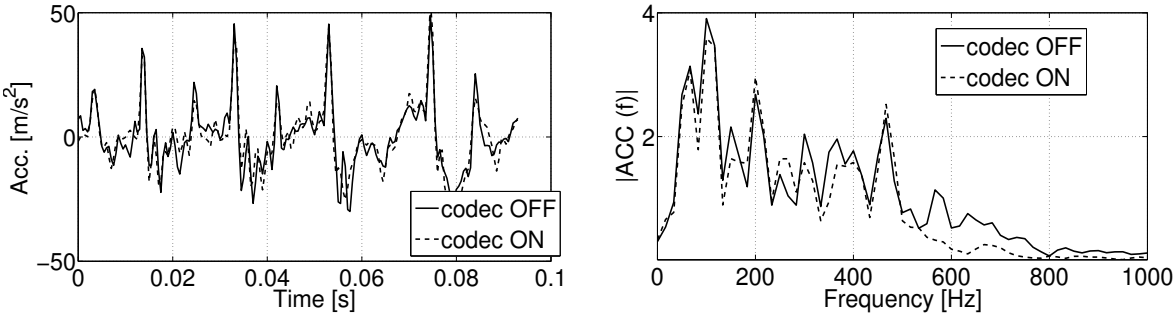
Figure 3.15 shows the time-domain and frequency-domain performance of the CELP-based codec.



(a) Coarse (here, regularly patterned) texture



(b) Fine random texture



(c) Mixed texture

Figure 3.15: [Left panels] Time-domain view of the codec operation. Codec input (codec OFF) and output (codec ON) are shown. It can be seen that the codec has a very good temporal tracking performance. [Right panels] Frequency-domain view of the codec operation for one texture frame (120 samples, i.e., 0.06 seconds long). It can be seen that the spectral signature is faithfully captured. Figure reproduced from [CcK$^+$12].

Parameter	Value
Output bitrate	3.55 kbps
Buffering delay	20 ms
Look-ahead delay	10 ms
Average encoding time	230 µs
Average decoding time	18 µs

Table 3.2: Summary of the codec performance parameters

3.6.3 Perceptual quality

Let us firstly outline the requirements of subjective evaluation for assessing the perceptual performance of the CELP-based codec. The objective was to prove that the codec is perceptually *transparent*. Perceptual transparency means that, *ceteris paribus*, subjects are unable to discriminate between the tactile stimuli perceived:

1. when the codec was switched OFF. Let us denote this stimulus set as $C1_i$ (where i is an index into this set): tactile signals sensed by the accelerometer presented directly to the subject without any intermediate processing — in other words the codec input, and,

2. when the codec was switched ON. Let us denote this stimulus set as $C2_j$ (where j is an index into this set): compressed-decompressed tactile signals — in other words, the codec output.

The subjects needed to make a binary decision if pairs of stimuli, with one stimulus from each set above, felt the *same* or *different*. In practice, such choices are made by subjects in the presence of uncertainty. For instance, non-idealities like internal noise in the human sensory process might lead to different percepts for the same physical stimulus in two different trials separated by time and/or space. Additionally at a cognitive level, a bias in the subject's decision process may push her/his response in a certain direction.

To achieve robustness toward uncertainties and to account for response bias, a Signal Detection Theory (SDT) approach was taken for perceptual performance evaluation [MC05]. In particular, the two-alternative forced-choice (2AFC) experiment paradigm was employed. Herein, subjects were presented a pair of tactile stimuli and had to make a choice between two alternatives — do the two stimuli feel the "same" or "different". This was followed by an SDT analysis of the results.

Another consideration unique to haptics should be mentioned here. Contrary to audio and video where information flows unidirectionally from a technical system to the human, we can actively affect the feedback we perceive in case of the haptic modality (see Figure 3.16

(a) (b)

Figure 3.16: (a) Teleoperation setup used for this work. The operator (MASTER) controls the activities of the remote SLAVE robot during interactions with the remote environment. The accelerometer on the SLAVE robot picks up wideband acceleration signals arising in the interactions, which capture the surface texture of the remote objects. These signals are then transmitted to the MASTER and displayed to the human operator hand through a vibrotactile actuator. Figure reproduced from [CcK+12]. (b) Voice-coil actuator embedded inside the haptic device stylus on the MASTER side.

(a)). This makes it difficult to compare two stimuli perceptually in realtime teleoperation. This is because it cannot be made certain that the human elicits two *temporally* comparable stimuli in two separate trials. To work around this problem, the subjective evaluation was broken down into two experiments:

- Offline experiment: Subjects judged *recorded* pairs of stimuli haptically, with no visual feedback. Thus in this experiment, manipulability of the haptic interface was sacrificed in the interest of a fair comparison between two stimuli.

- Online (realtime) teleoperation experiment: Subjects performed realtime teleoperation with visual-haptic feedback. For a given texture, they compared the general feel of the surface for two modes of operation — when the codec was switched OFF versus when it was switched ON.

3.6.3.1 Experiment setup

The perceptual experiments were based on a teleoperation setup similar to the one presented by McMahan *et al.* in [MRRK10]. As shown in Figure 3.16, the master and slave robots were both PHANToM Omni haptic devices (SensAble Technologies, Inc., USA). To limit the motions used by the subjects and to simplify the setup, the elbow and gimbal joints of both Omnis were immobilized using physical constraints. The shoulder joint's two rotational degrees of freedom allowed the handle to move on a spherical surface. As a result, the slave end-effector traversed the remote object surface along an arc.

The master Omni carried a custom handle 3D-printed in ABS plastic. Inside the handle, a NCM02-05-005-4JB linear voice-coil actuator from H2W Technologies, Inc. was attached (see Figure 3.16 (b)). The actuator coil was centered in its 7.6 mm workspace by a pair of preloaded compression springs. It could move up and down on linear bearings. Passing current through the coil created equal and opposite electromagnetic interaction forces between the handle and the coil, which induced high-frequency accelerations on the user's hand.

The slave Omni end-effector was instrumented with a three-axis LIS344ALH accelerometer from ST Microelectronics. This MEMS-based analog accelerometer has a range of ± 6 g (± 58 m/s^2), and its output is filtered with an on-board analog low-pass filter (cutoff frequency 956 Hz).

Since the actuator had only one degree-of-freedom, the three-axis accelerometer output needed to be reduced to a single dimension. One solution to this issue could be to sum the three acceleration components together sample-by-sample. However, this approach leads to a distortion of vital texture information due to constructive or destructive interference whenever the acceleration components are positively or negatively correlated, respectively. To avoid this, the DFT321 algorithm proposed by Landin *et al.* in [LRMK10] was employed for dimensionality reduction. The DFT321 algorithm combines the three-axis data in the frequency-domain in a perceptually optimal manner, while preserving both important temporal and spectral properties of the texture signals.

A quad-core 2.13 GHz PC with an NI PCI 6221 data acquisition card was used to sample the accelerometer output with a 16-bit ADC at 2 kHz. The PC also drove the voice-coil actuator via a 16-bit DAC output at a 2 kHz update rate. The Sensable OpenHaptics 3.0 library was used to control the teleoperation system. It operated a servo-loop at approximately 1 kHz, sensing the 3D position of each Omni and specifying its 3D force output at each time step.

A cardboard barrier shielded the haptic device and the subject's hand from her/his vision during the actual experiment to prevent visual observation of the moving coil. Moreover, auditory cues from the voice coil were masked out by playing pink noise through headphones, since they tend to interfere with haptic perception [SGS08]. Thus, isolation of haptic perception was ensured in the experiment.

Subjects The same nine subjects participated in both experiments, 1 female and 8 males, ranging in age from 23–30 years. None of them reported having any ailments that would hamper their sensorimotor performance. Two of them were researchers in haptics, using commercial kinesthetic devices like the PHANToM Omni on a regular basis. However, none of them had prior experience with tactile feedback systems. In a preliminary session, all the subjects were made familiar with the experiment setting and the operation of the hardware.

3.6.3.2 Offline experiment

In this experiment, only the master side of the teleoperation system shown in Figure 3.16 (a) was used to present pre-recorded texture stimuli to the subjects. As mentioned before at the beginning of Section 3.6.3, the subjects were asked to give binary "same/different" judgments for each pair of texture stimuli presented. This choice of the experiment paradigm was driven by the fact that binary decisions are the easiest for subjects to make and tend to yield results that are cleaner than those from other methods like magnitude estimation. Magnitude estimation suffers from response bias, whereas the Signal Detection Theory-based "same/different" approach does not. The scope of this experiment was to prove the perceptual indiscriminability of uncompressed and compressed textures. Subjects could switch between the two stimuli in a pair as many times as they wanted before reaching a decision and moving on to the next pair.

There were a total of 80 pairs of stimuli to be judged. These were divided as follows (32 + 32 + 16):

1. Ground truth: "different": $\{C1_i \Leftrightarrow C2_j | i = j \in \{1, 2, ...32\}\}$ (corresponding codec input and output signals),

2. Ground truth: "same": $\{C1_i \Leftrightarrow C1_i | i \in \{33, 34, ...64\}\}$ (same stimulus repeated twice for each of the 32 pairs),

3. Ground truth: "different": 16 randomly selected pairs from $\{C1_i \Leftrightarrow C1_k | i \neq k \in \{33, 34, ...64\}\}$

The 80 pairs were randomly ordered, and the order changed randomly from subject to subject. Each stimulus was an approximately 1 s long texture signal segment recorded while scanning four different textures (ABS plastic, shoe sole, rock, and vinyl, see Figure 3.17). The recording and segmentation was done by the experimenter in advance before the experiment. The segments were carefully selected to cover representatives of a wide variety of stimuli from every category of texture stimuli — coarse, fine random and mixed (see Section 3.1).

Since manipulability of the haptic device was removed in this experiment, fairness was maintained in the comparisons, and every subject judged the same stimulus pairs with exactly the same time-frequency profiles. Moreover, subjects were not required to remember long durations of stimuli when making comparisons.

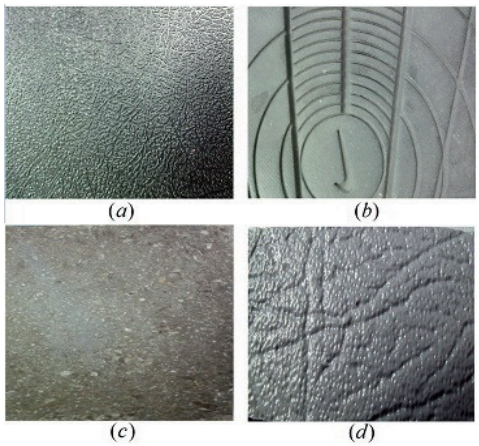


Figure 3.17: Textures used: (a) rough ABS plastic (fine random), (b) shoe sole (coarse regularly patterned), (c) rock (mixed), and (d) vinyl (mixed). Figure reproduced from [CcK+12].

Results

For every subject, the number of "hits" (h) and "false alarms" (f) were counted across all stimuli as follows:

$$\begin{aligned} h &= \sum(\text{decision "same"} \mid \text{ground truth "same"}) \\ f &= \sum(\text{decision "same"} \mid \text{ground truth "different"}) \end{aligned}$$

Accordingly, the "hit rate" (H) and the "false-alarm rate" (F) were calculated as $H = h/32$ and $F = f/32$ (eight 1 s recordings per texture $\times$ 4 textures = 32 stimuli).

A stimulus sensitivity parameter from Signal Detection Theory called d' was used for the analyses [MC05]. The parameter d' represents the ability of a subject to discriminate between stimuli belonging to two different classes. Low values of d' indicate that the stimuli are not easily discriminable, whereas high values indicate higher discriminability. SDT commonly assumes a Gaussian noise-model of the human sensory process. Accordingly, the sensitivity parameter can be computed as $d' = z(H) - z(F)$, and the response bias as $c = \dfrac{-[z(H) + z(F)]}{2}$, where $z()$ denotes the inverse Gaussian cumulative distribution function.

Figures 3.18 show the results of the analysis of data from the offline experiment. It is generally accepted that if $|d'| < 1.0$, the difference between the stimuli is below human perceptual thresholds. From Figure 3.18(a), it can be seen that across all subjects, $|d'| < 0.5$ consistently. In informal accounts from the subjects, it was noted that all of them found the stimuli to be "very difficult to distinguish" in general.

3.6.3.3 Online (realtime) teleoperation experiment

In this experiment, subjects teleoperated the slave omni through the master (see Figure 3.16). They could directly observe the remote teleoperator side visually. Visual observation of the

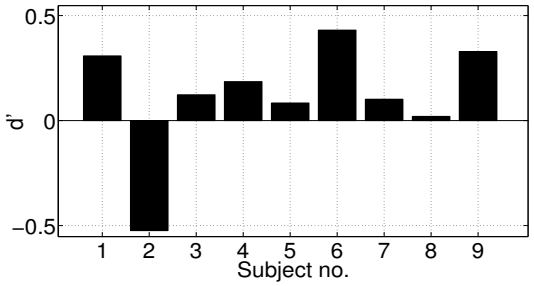


Figure 3.18: Offline experiment result: d'. Figure reproduced from [CcK+12].

local voice-coil movements by the subject was prevented by a cardboard shield. Headphones playing pink noise masked out any auditory cues from the voice coil actuator.

One by one, the subjects scanned each of the four textures, while turning the intermediate codec between the teleoperator and operator sides ON and OFF using a keyboard. They could explore the four textures as long as they liked. They were then asked to judge if the textures felt the "same" or "different" in the two modes of operation (codec ON, codec OFF).

At an average, the experiments required a total of 45 min to complete per subject, excluding training sessions at the beginning. No subjects except one found any perceivable differences in the feel of the textures between the two modes of operation (codec ON/OFF). This further corroborates the conclusions drawn from the results of the offline experiment.

3.7 Bitrate scalability

The haptic texture codec from [CcK+12] described in the previous sections has a constant output bitrate of 3.55 kbps. In this section, work is presented towards turning this codec bitrate-scalable, enabling us to operate in a range of different bitrates under varied network traffic conditions and quality demands.

Out of the total 71 bits per frame (3.55 kbps), the LPC coefficients in the codec account for 18 bits per frame (0.9 kbps). Furthermore, the algebraic codebook excitation codevector accounts for 26 bits per frame (1.3 kbps). Since these two components together contribute more than 50% of the total bitrate, they were the best targets for bitrate reduction.

3.7.1 Scalable Vector-Quantization for the Linear Prediction coefficients

The LPC coefficient vector determined for a frame at the encoder was vector quantized before transmission to the decoder side. This quantization took place in the Line Spectral Pair (LSP) domain, in order to ensure the stability of the LPC IIR filter after quantization [RS07]. Vector Quantization (VQ) codebooks corresponding to different VQ resolutions were generated from a pre-recorded texture dataset. Anywhere from 17 down to 0 bits could be allocated to the

LPC VQ (see Table 3.3) for achieving bitrate scalability. Mean Square Error (MSE) was used as the error criterion for generating the codebooks as well for searching them.

The VQ implementation was divided into two stages. Stage 1 coarsely quantized the entire input vector. Stage 2 separately and finely quantized the lower and upper halves of the quantization error vector from stage 1. Such splitting allows high-fidelity quantization, while reducing the memory requirements for the same total number of bits allocated to the VQ.

There are multiple ways in which bits can be allocated to the two VQ stages, based on the total available number of bits. Table 3.3 shows two such bit allocation schemes possible. Scheme 1 shares the available bits (almost) equally between the lower and higher splits, whereas Scheme 2 allocates more bits to the lower-split, thus favoring lower frequencies over higher ones.

LPC bitrate (bits/frame)	Scheme1 $(1^{st}, 2^{nd}_{low}, 2^{nd}_{high})$	Scheme2 $(1^{st}, 2^{nd}_{low}, 2^{nd}_{high})$
17	(7,5,5)	(7,5,5)
16	(7,5,4)	(7,5,4)
15	(7,4,4)	(7,5,3)
14	(7,4,3)	(7,5,2)
13	(7,3,3)	(7,5,1)
12	(7,3,2)	(7,5,0)
11	(7,2,2)	(7,4,0)
10	(7,2,1)	(7,3,0)
9	(7,1,1)	(7,2,0)
8	(7,1,0)	(7,1,0)
7	(7,0,0)	(7,0,0)
6	(6,0,0)	(6,0,0)
5	(5,0,0)	(5,0,0)
4	(4,0,0)	(4,0,0)
3	(3,0,0)	(3,0,0)
2	(2,0,0)	(2,0,0)
1	(1,0,0)	(1,0,0)
0	(0,0,0)	(0,0,0)

Table 3.3: LPC bit allocation

3.7.2 Scalable algebraic codebook

By changing the number of non-zero pulses in the algebraic codebook, the reconstructed texture-quality can be traded off against the number of bits allocated to this codebook.

3.7.2.1 4-pulse structure

Here, each algebraic codevector of length 20 samples (size of the subframe) contained four non-zero pulses. Each pulse p_i could have the amplitude of ± 1 and could assume the positions indicated in the table below.

Pulse	**Sign**	**Positions**
p_0	$s_0 : \pm 1$	$m_0 : 0, 5, 10, 15$
p_1	$s_1 : \pm 1$	$m_1 : 1, 6, 11, 16$
p_2	$s_2 : \pm 1$	$m_2 : 2, 7, 12, 17$
p_3	$s_3 : \pm 1$	$m_3 : 3, 8, 13, 18,$
		$4, 9, 14, 19$

The excitation codevector $v[n]$ wass constructed by summing the four pulses as follows:

$$v[n] = \sum_{i=0}^{3} p_i[n] = \sum_{i=0}^{3} s_i \delta[n - m_i]; \quad n = 0, ..., 19 \tag{3.10}$$

Each pulse required 1 bit for the sign. Two bits were required for encodeing the positions of pulses p_0, p_1 and p_2, and 3 bits for the position of pulse p_3. Thus, a total of 13 bits were required to index the entire codebook for each subframe in the texture codec in Section 3.3 (4 bits for signs and 9 bits for the positions). Since a frame consists of two subframes, 26 bits were necessary to encode each frame.

3.7.2.2 2-pulse structure

Here, the algebraic codevector was configured to have only two pulses per subframe, instead of four. The two signed pulses were searched in two overlapping tracks, as shown in the track table below. Like the 4-pulse structure, pulses could have amplitudes of ± 1. Thus, 2 bits were required for the signs, 3 bits for the position of pulse p_0, and 4 bits for the position of pulse p_1. Thus, a total of 9 bits were required to index the entire codebook for a subframe, and 18 bits were required for a frame.

Pulse	Sign	Positions
p_0	$s_0 : \pm 1$	$m_0 : 1, 3, 6, 8, 11, 13, 16, 18$
p_1	$s_1 : \pm 1$	$m_1 : 0, 1, 2, 4, 5, 6, 7, 9, 10$
		$11, 12, 14, 15, 16, 17, 19$

3.7.3 Performance

The performance of the bitrate scalable codec was again evaluated objectively and subjectively.

3.7.3.1 Objective evaluation

Since texture communication is a relatively new area of research, there is a lack of reliable and robust objective quality metrics for wideband vibrotactile signals in the literature. Hence, the audio-domain metrics Log-spectral Distortion and Segmental SNR were used to get an impression of the increase in distortion as the codec bitrate was scaled down.

The frequency-domain performance was evaluated with the standard form average Log-spectral Distortion (SD):

$$SD(i)[dB] = \sqrt{\frac{1}{f_h - f_l} \int_{f_l}^{f_h} \left[10 \cdot log_{10} \frac{S_i(f)}{\hat{S}_i(f)} \right]^2 df} \tag{3.11}$$

where $f_l = 40\,\text{Hz}$ and $f_h = 1000\,\text{Hz}$ are the lower and upper frequency limits, respectively, corresponding to the sensitivity bandwidth of the Pacinian mechanoreceptors in the human hand, $S_i(f)$ and $\hat{S}_i(f)$ denote LP power spectra for the i^{th} texture frame corresponding to the unquantized and quantized LPC coefficients, respectively.

Figure 3.19 shows the averaged SD results across all texture signals in the test dataset. Due to the absence of haptic perceptual thresholds for spectral distortion in the literature, it is currently not possible to comment on the significance of a given SD value. However, the figure serves as an illustration of how SD increased monotonically as the LPC VQ bitrate was decreased.

The time-domain performance was evaluated using the Segmental Signal-to-Noise ratio metric defined as:

$$SSNR[dB] = \frac{10}{N_f} \sum_{i=0}^{N_f - 1} log_{10} \left(\frac{\sum_{j=0}^{N_s - 1} s^2(k)}{\sum_{j=0}^{N_s - 1} [s(k) - \hat{s}(k)]^2} \right) \tag{3.12}$$

where $k = (i \cdot N_s + j)$, N_f is the number of frames in a texture signal, $N_s = 40$ is the number of samples per frame, $s(n)$ and $\hat{s}(n)$ are the codec input and output signals, respectively.

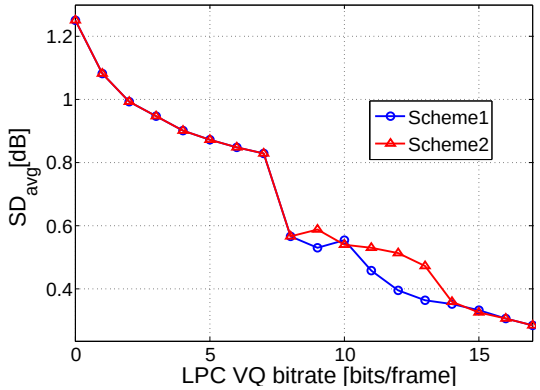
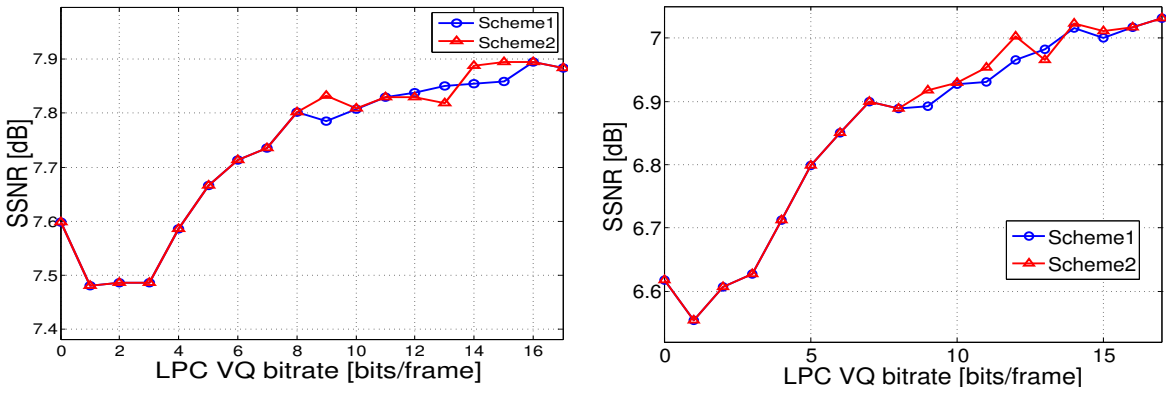


Figure 3.19: Average spectral distortion for the LPC VQ, evaluated for the two bit allocation schemes (Table 3.3) over the LPC VQ bitrates ranging from 17 to 0 bits per frame. Figure reproduced from [CSDS15], ©2014 IEEE.

Figure 3.20 shows the monotonic rise of the SSNR for the 4-pulse and 2-pulse codebooks as the LPC VQ bitrate increases. The SSNR profile for the 2-pulse codebook is similar to the one for the 4-pulse codebook, but with lower SSNR values.

3.7.3.2 Subjective evaluation

Subjects Seven right-handed male subjects in the age-range of 21 to 44 years participated in the experiment designed to evaluate the perceptual performance of the codec. None of them reported having any ailments that would affect their sensorimotor performance. Their experience with haptic devices ranged from "none" to "extensive". Two had participated in haptic psychophysical experiments before, while the rest were inexperienced.



(a) (b)

Figure 3.20: (a) SSNR when a 4-pulse FCB is used, plotted against LPC VQ bitrate, (b) SSNR when a 2-pulse FCB is used, plotted against LPC VQ bitrate. Both subfigures show SSNR for the two LPC VQ bit allocation schemes from Table 3.3. Figure reproduced from [CSDS15], ©2014 IEEE.

Textures Five different textures, shown in Figure 3.21 were used to perform the experiment. Textures (a), (b), (c) in the figure were used before in generating the training set to design the codebooks of the VQ encoder, while new textures (d) and (e) were included for the experiment. The experiment setup remained the same as described in Section 3.6.3.1.

Method Pilot tests were conducted with five subjects to get a first impression of the codec quality, before a formal subjective experiment was performed using a standard psychophysical method. While the subjects explored each of the five textures with the teleoperation setup, sometimes the codec was turned OFF, and at other times, it was turn ON and set to the maximum compression strength. After these pilot tests, the subjects reported that while none of them could "perceive any distortion" introduced by the codec, all of them could "clearly distinguish" between different textures.

With this outcome, a simple psychophysical experiment was designed based on the classical psychophysical method of 1 Up–1 Down (1U–1D) staircases [Lev70]. The need for a higher order adaptive staircase method (e.g. 1U–3D, as in Section 3.2.1.4) was not felt due to the clear impressions of the codec performance from the pilot tests.

The stimuli used for the experiment consisted of the codec highest-bitrate setting (3.55 kbps), the lowest-bitrate setting (2.30 kbps, achievable with the LPC VQ bit allocation from Table 3.3 and a 2-pulse algebraic codebook), and several settings corresponding to intermediate compression levels. A total of 26 stimuli corresponding to 26 different bitrate settings (50 bps steps) were chosen by making various combinations of the LPC and the algebraic codebook bit-allocation.

In each experiment run, the subject compared two bitrate settings of the codec bitrate while exploring a remote texture, judging them to feel the "same" or "different". One of the two settings was fixed to be the highest-bitrate or the "reference" (3.55 kbps), while the other "comparison" bitrate corresponded to one of the remaining 25 settings. To begin with, the "comparison" bitrate was set to the lowest value of 2.3 kbps. Depending upon whether the subject responded with a "same" or a "different", the "comparison" bitrate was decreased or increased, respectively, through a step-size of 50 bps (the 1U–1D method). If the subject responded with "same" when the "comparison" bitrate was 2.30 kbps or with "different" when it was 3.55 kbps, that "comparison" bitrate was only repeated, since it was not yet possible to go beyond these bitrates.

An experiment run was terminated when the staircase had undergone three reversals (up to down) of direction. It was also terminated if the "comparison" bitrate got saturated to the highest bitrate of 3.55 kbps or the lowest bitrate of 2.3 kbps over the last three steps of the staircase. In case of reversal-based termination, the threshold bitrate was calculated as the mean of all the staircase points across the reversals, while for saturation-based termination, the threshold bitrate was selected to be the saturation value.

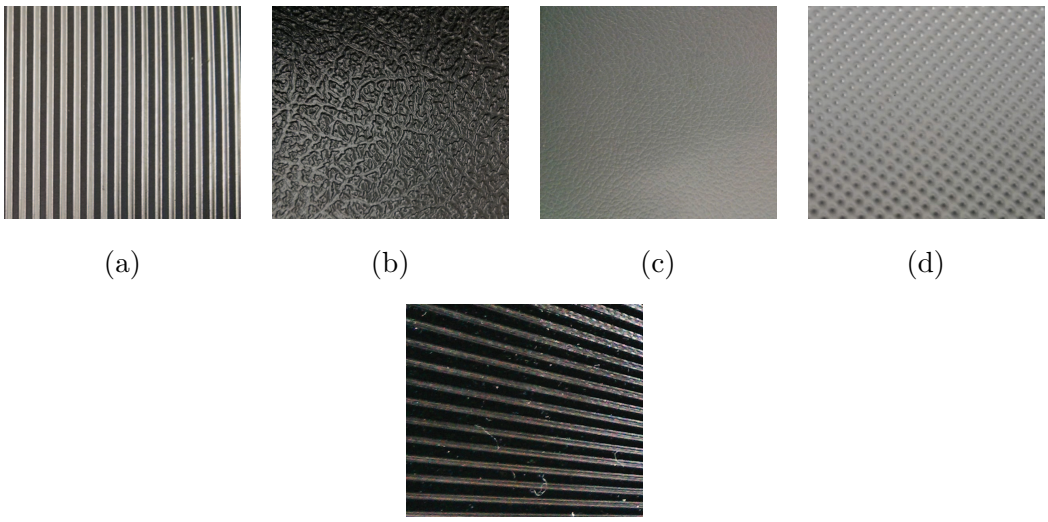


Figure 3.21: Textures used for the haptic interaction. (a) quasi-sinusoidal aluminium grating, (b) random-structured plastic, (c) random-structured fine vinyl, (d) peaky metal-case, (d) circular-sinusoidal plastic grating. Figure reproduced from [CSDS15], ©2014 IEEE.

Results Figure 3.22 shows staircase patterns for all 7 subjects for Texture (a) from Figure 3.21. It can be seen that the top six subjects could not reliably distinguish between the "reference" bitrate of 3.55 kbps and the lowest "comparison" bitrate of 2.3 kbps, while the

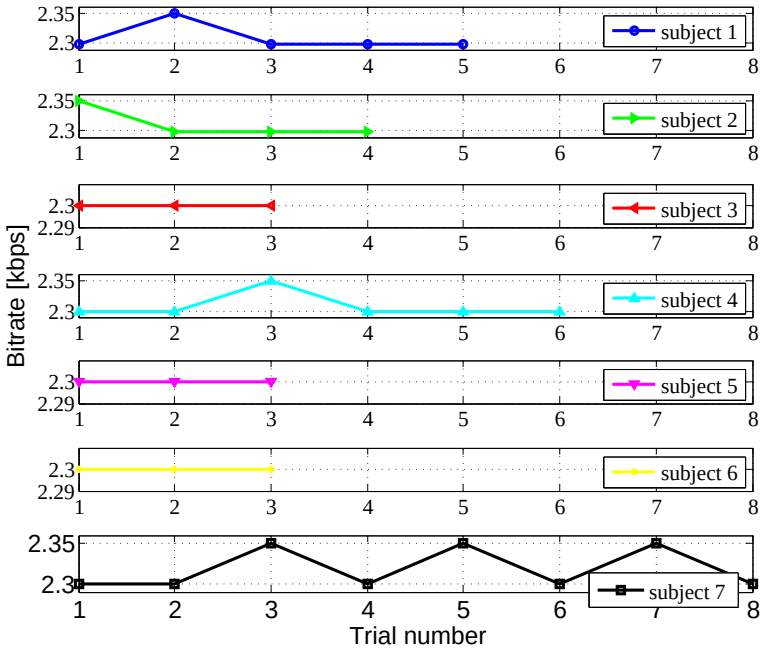


Figure 3.22: Staircase patterns for all 7 subjects for Texture (a) in Figure 3.21. Figure reproduced from [CSDS15], ©2014 IEEE.

last subject flipped between 2.3 and 2.35 kbps three times in a row.

In total 35 experiment runs (7 subjects and 5 textures) were performed. Thirty-one of these runs showed that the subjects could not distinguish between the reference and the lowest bitrates. All these 31 runs ended after receiving three "same" responses from the subjects. The remaining 4 runs stopped after six reversals of direction between 2.3 and 2.35 kbps, giving a threshold of 2.325 kbps. Overall, these experiments show that subjects could not reliably detect coding distortion even at the lowest bitrate possible at the time.

3.8 Conclusion

This chapter presented some of the most important contributions of this thesis. Mathematical models for the production and perception of haptic texture signals were described. They were employed in the development of a bitrate-scalable perceptually transparent haptic texture codec. The codec was evaluated both subjectively and objectively to test for operational as well as perceptual performance.

Here is a summary of the conclusions:

- The source–filter model based on Linear Predictive Coding fits well to the problem of mathematically describing haptic texture signal production. A frame of 40 samples can be described by a 14-parameter model (see Table 3.1), encoded with only 71 bits.

- The nature of the Simultaneous Masking phenomenon for vibrotactile perception was found to be very similar to that of its auditory counterpart. Masking thresholds peak when the maskee is collocated with the masker. They decrease as the function of the frequency difference between the masker and the maskee. In the texture codec, a weighting filter exploiting this phenomenon guides the search for the model parameters, ensuring that the coding noise stays below human perceptual thresholds.

- Bitrate scalability is achieved by scaling the number of pulses used by the algebraic-codebook, and by having a scalable vector quantizer for the LPC coefficients. With the bitrate scalable codec, it is possible to drive down the output bitrate to as low as 2.3 kbps without the subjects being able to reliable discriminate between the codec input and distorted output texture signals. The next chapter addresses further reduction of the output bitrate until users clearly perceive coding artifacts.

3.9 Outlook

The presence of multiple frequency-domain peaks in a haptic texture signal is a common occurrence in a real-world texture system. However, the simultaneous masking experiments

in this chapter dealt only with a single-peak masker. Future work in this area should address the characterization of the shape of the masking curves in the presence of more than one wideband masker.

As seen in Section 3.7.3.2, even with a codec bitrate of 2.3 kbps, subjects could not reliably detect perceptual distortion in the compressed haptic signals. Future work should address further reduction of bitrate in order to observe how the perceived quality degrades as a function of the bitrate.

The subjective tests in this chapter studied the deterioration of the perceptual performance for a given texture as the coding distortion rises. However, it is also necessary to answer the following question: can subjects still discriminate between two different textures as the coding distortion rises? Such a "task performance" evaluation of the codec will give a more comprehensive picture of the codec performance. Audio-video research also usually progresses along the same lines.

This chapter demonstrated how difficult and time-consuming it is to test perceptual quality aspects for haptic signal processing methods via subjective tests. This drives the need for development of model-based prediction of haptic perceptual quality, which is the subject of the next chapter.

Chapter 4

Model-based quality prediction for haptic communications

Efficient communication of haptic data over a network in a teleoperation system necessitates data compression [HHC+08a, CcK+12]. Lossy data compression may lead to the degradation of the perceptual quality of the haptic signals.

Traditionally, perceptual quality in media compression schemes is evaluated through subjective user tests. A human user introduces a large number of variables into the evaluation. These variables may be psychological (e.g. response bias), or cognitive (attention, intention, misinterpretation of experiment instructions, etc.). In addition to these, haptics involves physical variables (e.g. body posture, neuromuscular dynamics, grip-force, etc.). The experimenter has to very carefully control these variables, so as to standardize subjective tests and isolate the focus variables.

Thus, subjective tests (or user studies) need to be designed meticulously, and usually turn out to be difficult, time-consuming, and expensive. Such difficulties have led to the development of a number of perceptual model-based methodologies for quality evaluation in the audio and video fields [Jek05, WSB03], something that haptics research currently lacks. Computer modeling and simulation of entire haptic teleoperation sessions should be developed to address this.

To eliminate the need for difficult and time-consuming subjective tests, this chapter develops a Model-based Quality Prediction (MBQP) methodology for compressed haptic signals. Vibrotactile and kinesthetic signals are treated separately. Such a quality-prediction methodology aims to simulate a dedicated human who gives honest and consistent opinions about the system, thus removing unnecessary inter-subject variability appearing in subjective tests.

4.1 Vibrotactile textures

Certain haptic properties of a textured surface can be captured by scanning it with a tool and recording the wideband acceleration signals generated. It was shown in [MRRK10, KH96] that displaying these texture signals to the human through a vibrotactile actuator (see Section 3.6.3.1) improves the perceived haptic realism of object surfaces as well as the task performance.

Chapter 3 of this thesis presented a Code-Excited Linear Predictor (CELP)-based codec for haptic texture signals. In this codec, a linear predictive (LP) filter captures the spectral envelope, while a pair of codebooks capture the salient time-domain features of the texture signal. The LP and codebook parameters are vector quantized before transmission to the receiver side. For the work presented in this section, the CELP texture codec was used to generate compressed haptic texture signals at five decreasing bitrates — 3.15, 2.75, 2.45, 2.15, and 1.75 kbps — leading to five perceptual quality degradation levels for the subjective experiments performed in this section. Then, a methodology for the model-based (objective) prediction of the texture quality degradation was developed and validated.

In Section 3.6.3.3, the perceptual quality of the texture codec was evaluated in an online teleoperation system. Therein, subjects actively controlled the teleoperator robot while being displayed vibrotactile feedback, which they were asked to perceptually evaluate. Such an experimental setup presents a number of difficulties. The human-in-the-loop nature of the haptic interaction introduces variability in the texture scan velocity and the applied force due to the individual and time-varying active human behavior. The dynamics of the coupling between the human and the haptic interface may also change with time and from user to user. Furthermore, cognitive resources that should ideally be dedicated to the perceptual task alone are shared with the action task. The subject therefore cannot pay full attention to the perception task, potentially leading to perceptual thresholds higher than the actual ones.

The above issues introduce unnecessary and undesirable hurdles in an initial subjective evaluation of the codec, as well as in the development of an MBQP methodology. Therefore in this chapter, the experimental setup from Section 3.6.3 is simplified further in order to focus on just the perceptual performance of the codec, rather than on the combination of the human action and codec performance.

To this end, the human action-related variability in the texture signals is eliminated by recording texture signals under robotically controlled constant force and velocity conditions. The codec then operated on these recorded signals to produce distorted output signals. Due to this offline recording approach, exactly the same codec input and output signals could be displayed to all the subjects.

A conference paper based on parts of the work presented in this section has been submitted for peer review [CYS+15].

LPC vector quantizer (per frame)	Algebraic codebook (per subframe)	Gain vector quantizer (per subframe)	Bitrate (kbps)
12	13	7	3.15
12	13	3	2.75
2	13	5	2.45
0	9	7	2.15
0	9	3	1.75

Table 4.1: Codec bit allocation

4.1.1 Codec perceptual quality degradation

This section describes the psychophysical experiments conducted to generate ground-truth perceptual quality degradation data for the textures signals processed by the codec. Perceptual quality degradation is expressed as perceptual dissimilarity between the codec input and output signals, as measured through psychophysical tests based on the "same/different" experiment paradigm from Signal Detection Theory [MC05].

By an appropriate choice of bit allocation for a texture signal frame, five monotonically decreasing codec output bitrates — 3.15, 2.75, 2.45, 2.15, and 1.75 kbps — were obtained. Table 4.1 shows how bits were distributed across various codec parameters — the LPC vector quantizer, the algebraic codebook, and the vector quantizer for the codebook gains — to achieve those bitrates.

Subjects

Eight right-handed subjects participated in the experiments — 2 female, and 6 male. Their ages ranged from 22 to 27, with an average of 24 years. All of them were right handed. None of them reported having any sensory ailments that would affect their perceptual performance. Six of them had participated in haptic psychophysical experiments before, while the rest were inexperienced. Their self-reported experience with haptic devices ranged from "none" to "extensive".

Hardware/software setup

The hardware consisted of a Geomagic Touch haptic device (3D Systems, Inc., USA) with a three-axis ±6 g accelerometer (LIS344ALH from ST Microelectronics) mounted near the stylus tip (see Figure 4.1). The texture to be recorded was fixed onto a rotatory plate driven by a DC motor. The tip of the stylus rested on the texture material. The speed of the rotatory

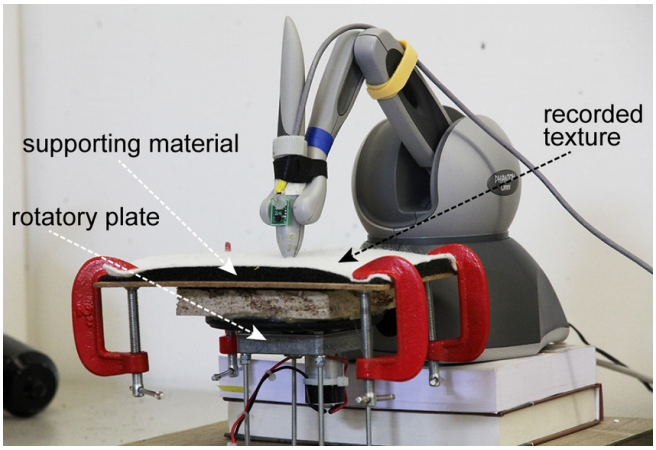


Figure 4.1: Hardware setup for recording texture signals. A MEMS acceleration sensor board can be seen near the stylus tip. Figure reproduced from [SLS+14], ©2014 IEEE.

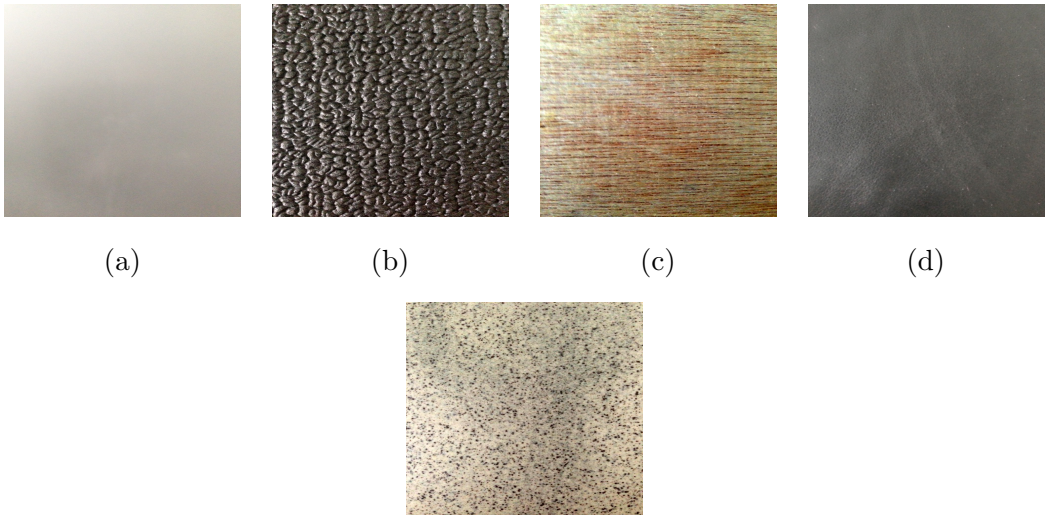
plate, and thus the texture scan speed was controlled to be constant. Moreover, the position of the stylus tip was controlled with a PID controller. This controller forced the tip to try to reach a reference position slightly below the texture material on the vertical axis. Thus the force with which the stylus pressed onto the texture could also be indirectly controlled to be constant.

The texture recorder software was implemented as a Simulink model. A realtime executable was generated for the Realtime Application Interface (RTAI)-Linux target operating system running on a standard PC. The realtime program sampled the accelerometer output through the Analog-to-Digital converter on a NI PCI-6221 Data Acquistion (DAQ) card, and saved the recorded data to a file on the PC.

Stimuli

Five textures varying widely in material and surface patterns were used - a steel texture, a rubber pad, hard wood, leather, and marble (see pictures in Figure 4.2). Each texture signal was then processed with the bitrate-scalable codec from [CSDS15], with parameter values from Table 4.1. This resulted in five distorted output signals per input reference signal. The input-output signals were divided into 3 segments of 0.8 s each. Figure 4.3 shows pairs of input-output signals for three of the five distortion levels for the first segment of the steel texture. Each of these pairs of input-output segments constituted a separate stimulus pair.

Psychophysical trials with the ground-truth answer “different” presented the reference and the distorted stimuli from a given pair to the subject one after the other, with a gap of 0.6 s. Additionally, trials with the ground-truth answer “same” were presented, in which the reference stimulus was repeated twice. The subject then indicated if she/he perceived the two stimuli to be the “same” or “different”.



(a) (b) (c) (d)

(e)

Figure 4.2: Textures used for the haptic interaction. (a) quasi-sinusoidal aluminium grating, (b) random-structured plastic, (c) random-structured fine vinyl, (d) peaky metal-case, (d) circular-sinusoidal plastic grating. Figure reproduced from [CSDS15], ©2014 IEEE.

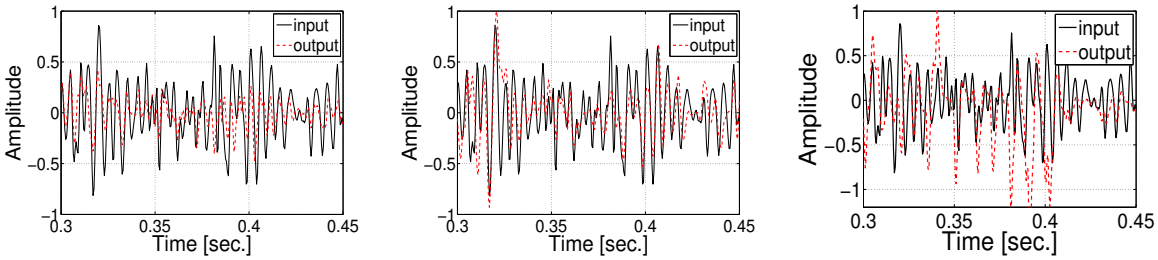


Figure 4.3: Three distortion levels (codec output bitrate, left to right: 3.15, 2.45, and 1.75 kbps) for a segment of the "steel" texture signal. The distorted output segment (- - -) is shown overlapped on the input reference segment (—).

The stimulus duration was chosen to be 0.8 s to avoid adaptation to the stimulus. The pause duration was chosen to be 0.6 s, to avoid haptic enhancement and summation effects [VG75] in the display of consecutive stimuli. Each trial thus lasted about 2.2 s, which falls within the constraint of several seconds imposed by the haptic working memory [ASG08].

Experiment setup

A custom-made stylus-like handle similar to that of the Geomagic Touch haptic device (3D Systems, Inc., USA) (Figure 4.1) haptic device was mounted on the K2004E01 minishaker (The Modal Shop, Inc., USA, see Figure 4.4). The subjects were instructed to hold the stylus like a pen in their dominant hand in a comfortable 3-finger grip. This grip was standardized across subjects. Their elbow and forearm rested on a wooden plank, which supported a neutral wrist position. Subjects wore acoustic noise-canceling headphones (QuietComfort 15,

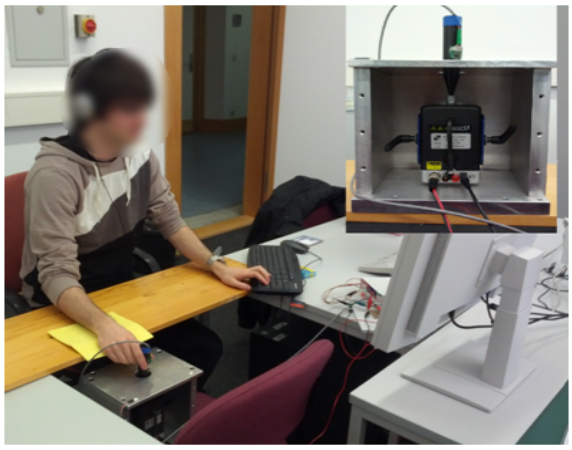


Figure 4.4: Experiment setup (minishaker device inset). Picture reproduced from [CSDS15], ©2014 IEEE.

Bose Corp., USA) that played pink noise to mask out auditory cues from the experiment apparatus. Moreover, a cardboard barrier (not shown in the figure) visually shielded the minishaker from the subject. It was thus ensured that the subject only had a haptic contact with the minishaker. Subjects interacted with the experiment's GUI through a keyboard.

Method

As mentioned before, subjects were asked to give binary "same/different" judgments for each pair of texture stimuli presented. Each stimulus pair was tested 6 times to have higher reliability of the results. For each test, the subjects could try each stimulus pair as many times as required before making a decision and moving on to the next pair.

Results

For a given stimulus-pair, all subject responses were aggregated, and the number of "hits" (h) and "false alarms" (f) were counted:

$$h = \sum(\text{decision "different"} \,|\, \text{ground truth "different"})$$
$$f = \sum(\text{decision "different"} \,|\, \text{ground truth "same"})$$

The "hit rate" (H) and the "false-alarm rate" (F) were then calculated as $H = h/n$ and $F = f/n$, where $n = 144$ was the number of ground-truth "same" or "different" pairs (8 subjects $\times$ 3 segments $\times$ 6 repetitions/segment/subject) for a given texture signal and bitrate setting.

For the analysis of the results, concepts from the Signal Detection Theory [MC05] were employed, similar to the subjective evaluation in Section 3.6.3.2. The d' parameter that represents perceptual distance between stimuli belonging to two different classes, was computed. The significance of d' (occasionally refered to as "dissimilarity" in this chapter, similar

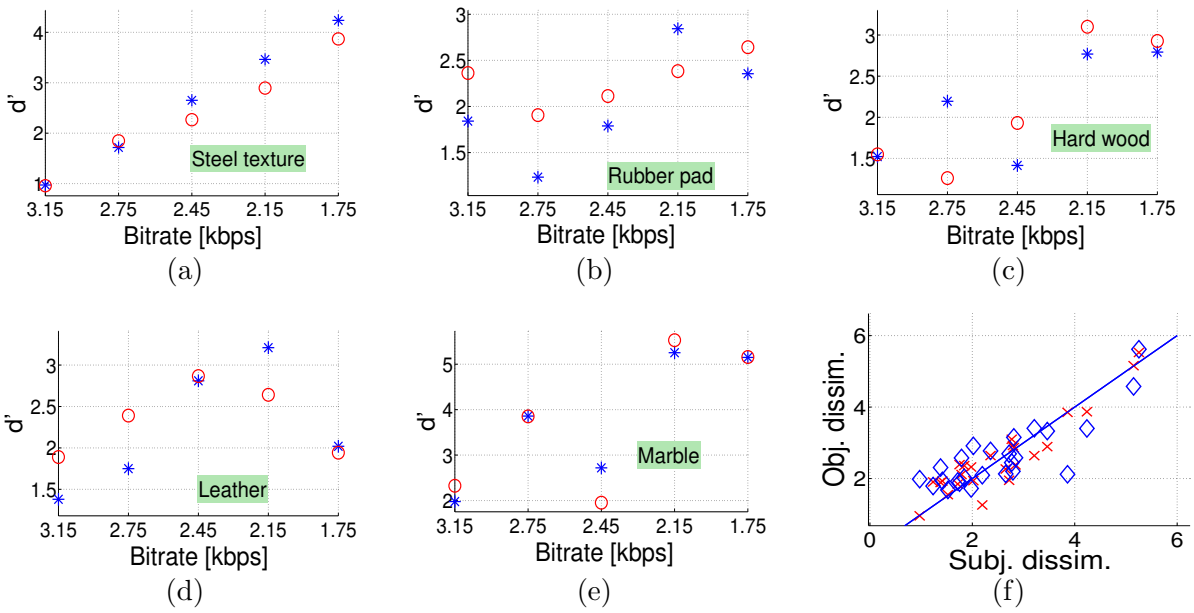


Figure 4.5: (a) – (f) Objective predictions '○' in comparison with the corresponding subjective results '★' for all five textures, (g) comparison of the prediction performance of all features combined together '×' versus the Okamoto model (see Section 2.5.2, [OY11]) alone '◇'.

to [BHY05]) was explained in detail in Section 3.6.3.2; it is repeated here briefly for convenience. Low values of d' indicate that the stimuli are not easily discriminable, whereas high values indicate that they are easily discriminable. The parameter d' is calculated as $d' = z(H) - z(F)$, where $z()$ denotes the inverse Gaussian cumulative distribution function (assuming a Gaussian noise model for the human sensory process).

Figure 4.5 shows the d' results for all five textures. It can be seen that, except for the steel texture, d' does not always monotonically increase as the bitrate reduces. A possible reason for this maybe the simultaneous manipulation of multiple unrelated codec parameters (see Table 4.1) for tuning the bitrate, and the closed-loop nature of the codec.

4.1.2 Model-based quality prediction

The following work aims to algorithmically predict the perceptual dissimilarities obtained from the subjective tests described above. For this purpose, perceptually meaningful features were extracted from the texture signals and combined within a linear model. More complex nonlinear models may be learned in the future.

A number of perceptual tactile features could potentially be exploited — e.g., 'slow-fast', 'sparse-dense', 'blunt-sharp', 'bumpy-smooth', 'hard-soft', 'dark-bright', 'dull-clear', 'thick-thin', 'vague-distinct', 'heavy-light', etc. as proposed in [HC10]. However, algorithmic quality prediction requires features to have well-defined mathematical models, which are not available in most of the above cases. This limits the spectrum of features that can currently be considered for quality prediction.

4.1.2.1 Perceptually-motivated texture features

A list of the following eight features that could be modeled mathematically was compiled, combining ones from the literature and the ones from the subjective feedback of the subjects — roughness [FL12], coarseness [BH05], brightness [HC10], regularity, stimulus vertical asymmetry, and time-envelope pattern [PC11]. The Bensmaïa power-spectral model, and the Okamoto amplitude-spectral model [OY11] (see Section 2.5.2) were also included as additional features. The eight feature definitions are given below:

1. Roughness Haptic roughness can be calculated as the logarithm of the average signal power, as defined in [FL12].

2. Coarseness The signal-analytic coarseness can be defined as the *waviness* of the fundamental signal oscillation, which also corresponds to the height profile of the surface texture [Str14]. The characteristic low-frequency profile of a texture was extracted using the DB20 wavelet transformation. The coarseness was then calculated as the standard deviation of the short-term slopes of the low-frequency profile. A coarse texture will have a wider distribution of these slopes around the mean than fine ones.

3. Spectral centroid ("Brightness") The spectral centroid is commonly defined as the "center of mass" of the magnitude spectrum of a signal (sum of frequency-weighted spectral magnitude samples). For haptic textures, higher spectral centroids convey "lively" or "bright" textures.

4. Regularity The regularity of a textured surface gets reflected in the periodicity of the acceleration signals, which can be captured with the autocorrelation function.

5. Vertical asymmetry Natural texture scans usually generate acceleration signals that are symmetric about the time-axis. A deviation from such a symmetry causes the hand to be forcefully "pushed" or "pulled" by the haptic device, rather than vibrate — a perceptually very strong and disturbing artifact. To capture such asymmetry, the standard statistical measure of *skewness* of the signal amplitude samples was used.

6. Bensmaïa spectral model and 7. Okamoto model These features were based on psychophysical detection thresholds, as described in Section 2.5.2.

8. Time-envelope Most of the subjects reported memorizing the undulations of a signal as audio patterns in their minds, and using them for discriminating between two stimuli. This

characteristic can be encoded in the form of time-domain signal envelopes [PC11]. The cross-correlation between the time-envelopes of the codec input and output signals was used as a measure of perceptual distance between them.

4.1.2.2 Regression modeling and validation

A 25×8 feature-distance matrix $\boldsymbol{X}$ was formed for the 25 stimulus pairs (5 textures × 5 distortions) and the 8 features defined above. Each element in a row of this matrix contained the Euclidean distance between the input and output stimuli for that particular feature. The corresponding dissimilarity results from the user studies were stacked in a 25×1 vector $\boldsymbol{y}$.

A feature-selection algorithm ("sequentialfs" Matlab function) was then used to select a minimal subset of features that best predict the data in $\boldsymbol{y}$. The algorithm started by including all eight features at the beginning. Then it dropped features one-by-one until there was no further improvement in the prediction performance. The sum of squared linear regression errors was used as the criterion function to select features. Another parameter specified a threshold on the error, thus determining when to stop. This linear-regression based feature-selection process (described below) filtered the list of 8 features down to the following 4 features — roughness, brightness, the Bensmaïa spectral model, and the time-envelope pattern.

With the reduced feature subset, the data was partitioned into 3 textures for training a linear regression model, and the remaining 2 textures for validation. The textures were exhaustively rotated through the training and validation sets. Regression coefficients for the rotation that led to the least regression error for all textures were selected as final feature weights.

4.1.3 Quality prediction results

The model-based quality predictions are shown in comparison with the perceptual ground-truth for all textures in Figure 4.5. Using the Okamoto amplitude-spectral model, an R^2 (Goodness-of-Fit) of 0.69 was obtained (R^2 adjusted for number of predictor variables was 0.64). When the features described in Section 4.1.2.1 were included on top, an improved R^2 of 0.90 (adjusted R^2 of 0.84) was obtained. The bottom-right panel in Figure 4.5 shows the improvement of the prediction performance when all the features combined together against when the Okamoto model alone is considered.

The weights selected for the four features were [roughness: 5.2, brightness: 3.8, the Bensmaïa spectral model: 6.3, and the time-envelope pattern:3.3]. It is evident that now, in combination with other features, the Bensmaïa model plays a vital role in explaining the variance in the perceptual data, followed by Roughness, then the brightness, and finally the time-domain signal envelope of the texture signal.

The MBQP methodology developed for haptic texture signals in Section 4.1 intentionally avoided including human action into the system. This was done to prevent the active behavior from interfering with texture perception tasks, which have a complex multidimensional nature [ONY13]. In comparison to that, kinesthetic signals and their perception are relatively simpler, and hence can afford the inclusion of human action models for MBQP. Work that addresses action-perception modeling for predicting subjective quality for kinesthetic haptic signals is presented in the next section.

4.2 Kinesthetic haptics

Quality evaluation in the context of kinesthetic haptic systems can be a vast topic, depending upon the system aspects considered — the control architecture (e.g. Position–Position vs. Position–Force), communication-induced artifacts (packet loss, delay, and jitter), the data reduction/compression scheme (e.g., perceptual deadband-based data reduction, see Section 2.4.1), the interplay between the audio-video-haptic modalities, the design of human-system interfaces (haptic devices), etc. [CB94, Law92, LPD+00]. This work however, specifically concentrates on the evaluation of perceptual quality considering the artifacts arising out of the perceptual deadband-based data reduction presented in [HHC+08a], and described in Section 2.4.1.

Haptics involves bidirectional interaction between the human and the machine. So, not only can humans perceive the haptic feedback offered - similar to other modalities - but also physically alter the environment and what they perceive. Unlike audio and video, this tight coupling of the human to the system necessitates the development of a holistic human model that accounts for human haptic perception as well as human action (models for the Central Nervous System and the Neuromuscular System of the arm).

4.2.1 A quality evaluation framework for kinesthetic data reduction

While analyzing the performance of different data reduction/compression schemes or even different settings of the same scheme, fairness demands having a common basis for comparison. This is automatically ensured in audio, video, and even haptic texture coding schemes since exactly the same media signals are subjected to compression to produce deterministic output, which can be evaluated subjectively as well as objectively.

However, the additional manipulative ability to be considered while dealing with kinesthetic haptics may result in the generation of different haptic signals due to the within- and across-user variability of human action. Thus there is a need for standardization of the task to be performed by the user, in order to enforce experiences similar as far as possible across users. A trajectory tracking task, for which detailed human user models have been derived

in the past [Hes80, PCAB98] fulfills this requirement. For such a task, a reference position signal to be followed haptically by the user in a real/virtual environment can standardize measurements across all users to a large extent.

For the experimenter, a trajectory tracking task buys a definite degree of control over physical haptic stimuli that different subjects experience. By controlling the reference trajectory to be followed, the user trajectory input to the system can be regulated over a reasonably wide dynamic range in amplitude and frequency. This range is determined by the limits of the haptic device and those of human manipulation. For regulated trajectory inputs, a simple haptic environment (e.g. a linear spring) should respond with comparable force feedback profiles across users.

Trajectory tracking tasks can take two forms:

- Pursuit tracking, wherein the instantaneous reference trajectory is displayed to the user visually. She/he then follows it with the haptic device.

- Compensatory tracking, wherein the instantaneous tracking error (= reference trajectory - actual trajectory) is displayed to the user visually. She/he then tries to keep this error at a minimum by appropriately steering the haptic device.

The tracking task is a kind of negative feedback control, where the user applies corrections through the physical motion of her/his muscles to achieve the control objective of keeping the tracking error at a minimum. In both cases above, the user is subjected to force feedback generated from the haptic environment in response to her/his actions. A compensatory tracking task was chosen for this work, so as to discourage learning or prediction of the reference trajectory.

In the following, position trajectories are replaced by angular trajectories, and forces by torques. A rotatory-domain implementation was simpler, since the haptic device used had rotatory actuators. Figure 4.6 shows an abstract block diagram of the proposed kinesthetic Quality Evalution framework. In this framework, the instantaneous error between the reference trajectory θ_{ref} to be followed and the actual trajectory θ_D of the operator is fed back to the operator visually. The operator responds to this error by producing a torque, τ_h while perceiving the torque reflected from a virtual environment by the haptic device. The operator torque τ_h and the resistive torque τ_{env} generated by the environment in response to user motion collectively drive the hand-device model to produce a position signal at its output.

The output position θ_D of the device on the human operator side is communicated to the virtual/remote environment over the forward communication channel. In the reverse direction, before the environment response torque τ_{env} is sent over the reverse channel to the operator side for display, the data compression/reduction scheme operates on it to minimize the data transmission rate.

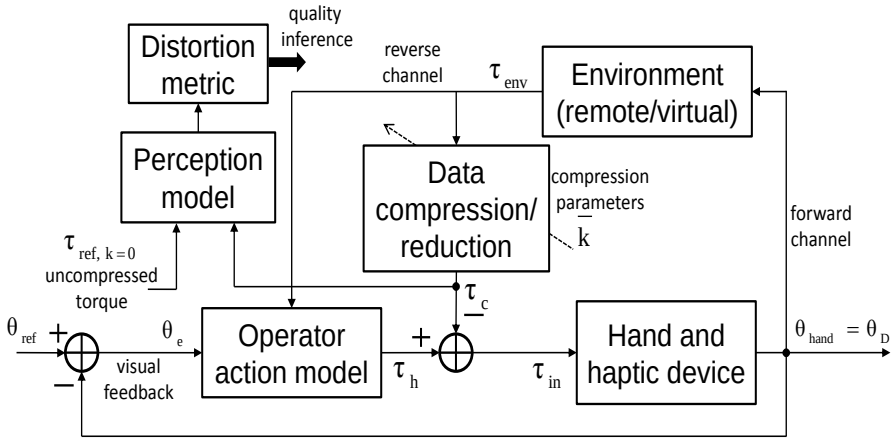


Figure 4.6: Block diagram of the model-based quality evaluation framework for haptic data compression/reduction. Figure reproduced from [CSH11], ©2011 IEEE.

Altering the compression scheme parameter vector $\vec{k}$ allows us to tune the strength of data compression/reduction as required. For lossy compression schemes, it usually holds that the stronger the compression, the larger the savings in the transmission rate, but larger the distortion introduced into the feedback signal, i.e. the lower the perceptual quality. For a given parameter setting of the data reduction parameters $\vec{k}$, based on pyschophysical laws, the compressed feedback signal τ_c can be mapped to the perceptual domain along with the reference uncompressed one τ_{env}. The distortion of the compressed signal τ_c with respect to the uncompressed one τ_{env} can then be measured in the perceptual domain, inverting it to produce an estimate of the feedback signal quality. Mathematical models of the dynamics of such a haptic system are developed in the following.

4.2.2 Modeling the human operator

The model-based quality evaluation framework can be divided into two main parts:

1. a model of the active control behavior of the human operator, and

2. a model of the human haptic perceptual mechanism.

It is important to note that depending upon the task performed by the user, the human control model will change. The human intention may also need to be modeled, in addition to action. But these advanced topics are out of scope for this initial investigation into kinesthetic model-based quality evaluation.

4.2.2.1 Action model

A well established isomorphic structural model (shown in the dashed box in Figure 4.7) of the human operator for compensatory manual control was presented by Hess in [Hes80]. It was

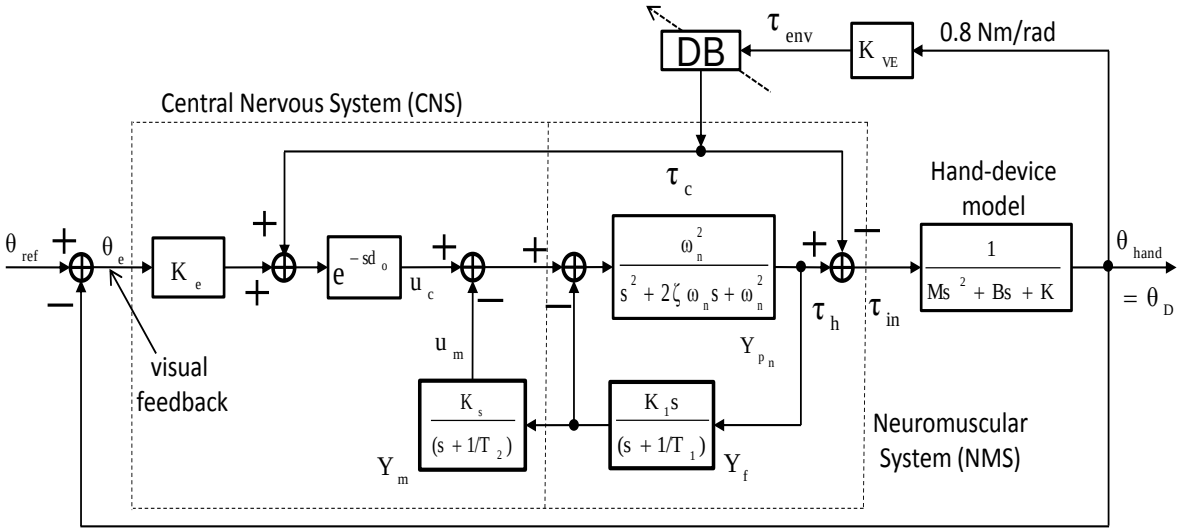


Figure 4.7: Structural operator model for the compensatory tracking task. The perceptual deadband scheme from Section 2.4.1 is used for minimizing the haptic data transmission rate on the reverse channel. Figure based on [Hes80, PCAB98], and reproduced from [CSH11], ©2011 IEEE.

applied by Penin *et al.* for predicting human motion signals in a teleoperation system with kinesthetic feedback in [PCAB98]. The model is divided into the *Central Nervous System (CNS)* and the *Neuromuscular System (NMS)*. This division clearly delineates the signal processing involved in the human body.

According to this structural model, control signals, proportional to the tracking error (the proportionality constant being K_e) are produced by the human. They are delayed by time d_o due to the inherent response delay (e.g. visual stimulus processing delay) and the transport time along the neuromotor pathways. The delayed control signal $u_c(t)$ is the input command to the closed-loop system consisting of the open-loop neuromuscular dynamics Y_{p_n} of the arm, and elements Y_f and Y_m. Y_f and Y_m represent in an approximate fashion, the combined processing of the muscle spindles, the Golgi tendon organs, and the joint angle receptors. As proposed by Hess in [Hes80], for the neuromuscular system, the following characteristic values were selected: $\omega_n = 10$ rad/s and $\zeta = 0.707$.

In a haptic task, the hand is always coupled to the haptic device. The passive hand-

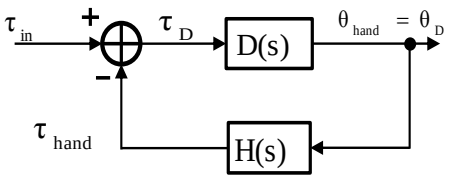


Figure 4.8: Hand-device model structure. Figure reproduced from [CSH11], ©2011 IEEE.

device characteristics can be represented by the block diagram shown in Figure 4.8. In this figure, τ_{in} denotes the superposition of the environment feedback torque τ_c and the torque τ_h commanded by the CNS. Output $\theta_{hand} = \theta_D$ denotes the collective position of the hand and the device. The hand impedance $H(s)$ was modeled as a linear time-invariant 2^{nd} order model with effective mass M_h, damping B_h, and stiffness K_h:

$$H(s) = \frac{\tau_{hand}(s)}{\theta_{hand}(s)} = M_h s^2 + B_h s + K_h \tag{4.1}$$

The transfer function of the haptic device dynamics $D(s)$ with mass M_d and damping B_d was modeled as:

$$D(s) = \frac{\theta_D(s)}{\tau_D(s)} = \frac{1}{M_d s^2 + B_d s} \tag{4.2}$$

From (4.1) and (4.2), the overall transfer function is:

$$\begin{aligned} HD(s) &= \frac{\theta_D(s)}{\tau_{in}(s)} = \frac{D(s)}{1 + D(s) \cdot H(s)} \\ &= \frac{1}{Ms^2 + Bs + K} \end{aligned} \tag{4.3}$$

where $M \triangleq (M_h + M_d)$, $B \triangleq (B_h + B_d)$ and $K \triangleq K_h$. The hand-device model fits into the operator model as shown in Figure 4.7.

Human operator modeling has been studied in great depth and much more detailed models for tracking tasks have been developed, for example by McRuer in [McR80]. More recently, researchers have also begun to focus on tracking tasks with haptic feedback [LMvP09]. In comparison to those models, this work neither insists on nor claims rigor. The primary intention here is to lay the foundations of a first easy-to-understand framework which should be subject to refinements in the future.

Parameter estimation

The operator model parameters were estimated in two successive experiments. In the first experiment, the hand-device model was identified for a specific grip-force typical for desktop haptics. With the parameter values determined from the first experiment, the operator model that drives the hand-device model was identified in the second experiment.

Experimental setup Experiments were performed using a PHANToM Omni haptic device (Sensable Technologies Corp., USA) equipped with a FlexiForce sensor (TekScan Corp., USA) to read the hand grip-force. A USB sensor-interface kit from Phidgets Inc. (Calgary, Canada) acquired data from the grip-force sensor. Only the first (base) joint of the haptic device was used. A real-time executable application was built using MATLAB/Simulink and QUARC Real-Time Control Software (Quanser Inc., Ontario, Canada). On a standard Windows PC (dual core, 2.13 GHz), the executable ran a servo loop that read the high-precision position

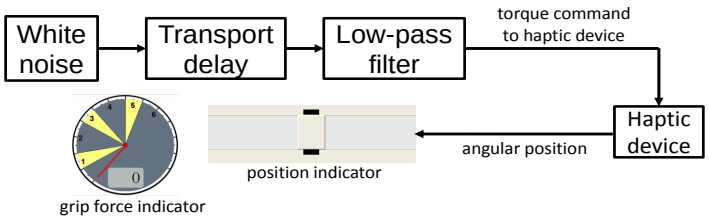


Figure 4.9: Identification of the hand-device model. A transport delay is added to allow the subject 1 second to position the slider exactly at the center before the experiment began. Figure reproduced from [CSH11], ©2011 IEEE.

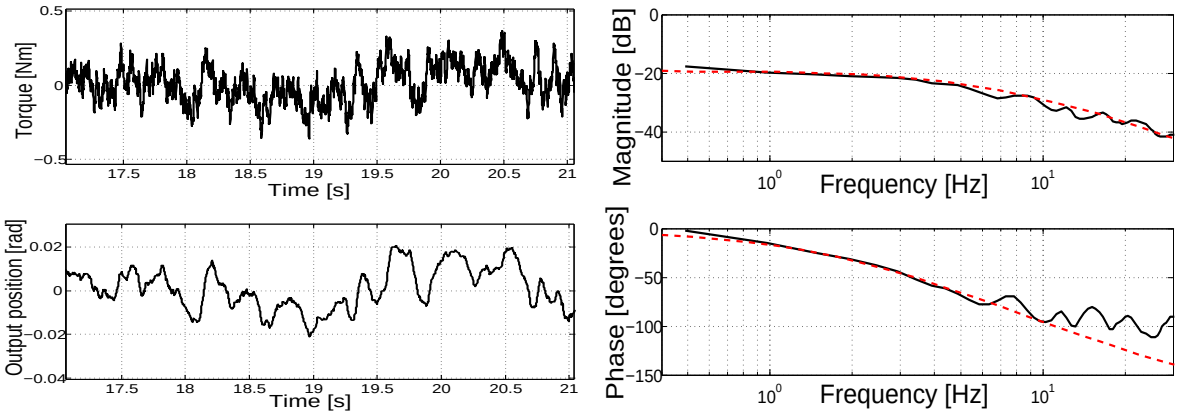
encoders of the haptic device, and commanded torque values to it at the standard 1 kHz haptic sampling rate.

Subjects Nine right-handed subjects (7 male, 2 female, ages 20–30 years, with an average of 26 years) participated in this study. All of them were right handed. None of them reported having any sensory ailments that would affect their perceptual performance. Seven of them had participated in haptic psychophysical experiments before, while the rest were inexperienced. Their self-reported experience with haptic devices ranged from "none" to "extensive". The same subjects participated in both the experiments.

Experimental procedure In the hand-device identification experiment, the subjects were displayed a slider on the screen which corresponded directly to the position of the hand-device coupling (see Figure 4.9). A Gaussian white noise force-input, lowpass filtered with a cutoff frequency of 30 Hz was commanded to the haptic device [FC10]. When the human hand is subjected to such a random force disturbance, it is less likely to trigger reflexes, which our model does not account for.

The subjects were instructed to attempt to hold the visual slider position at the center as far as possible, while maintaining a constant grip-force of roughly 3 N typical to desktop haptic applications. A grip-force measurement in the form of a dial was shown on the screen too, so that the subjects could tighten/loosen their grip-force as necessary. The input torque commanded and the corresponding output position of the haptic device were recorded for subsequent analysis. Figure 4.10 (a) shows an example of the time-domain input and output signals recorded during the experiment. The low pass character of the hand-device combination (also see Equation 4.2) is evident in this plot.

In the operator model identification experiment (see Figure 4.11), subjects were displayed the trajectory tracking error as the deviation of the slider from its center. They were then asked to keep the error at a minimum by adjusting the hand position so that the slider approached its center, which corresponded to zero tracking error. They were well trained for this tracking task before the actual experiment.



(a) (b)

Figure 4.10: Experiment for identifying the hand-device model: (a) time-domain plots, (b) Bode plots. The dashed curve represents the transfer function fitted to the solid experimental one. Figures reproduced from [CSH11], ©2011 IEEE.

Force feedback was computed according to a spring model (with stiffness $K_{VE} = ...$ $0.8\,\mathrm{N\,m\,rad^{-1}}$) for the environment. The force signal then underwent Deadband data reduction (see Section 2.4.1), before being commanded to the hand through the haptic device. The subjects were asked to maintain a consistent grip-force on the device. Similar to the first experiment, grip force was indicated visually. The reference input trajectory, the actual trajectory traced by the user and the corresponding feedback torque were recorded. Figure 4.12 (a) shows an example of the reference trajectory that was to be followed, and the corresponding trajectories traced by the users in response.

Each subject was initially trained to recognize the step-like deadband artifacts haptically. She/he then performed the tracking task (100 s long) multiple times for the range of deadband parameter k from 0% to 20%. To avoid learning effects, the k values were chosen randomly across experimental runs. After each run, a subjective rating corresponding to the perceived distortion in the haptic feedback because of the compression scheme was solicited. The rating scale shown in Table 2.5 was used for this purpose.

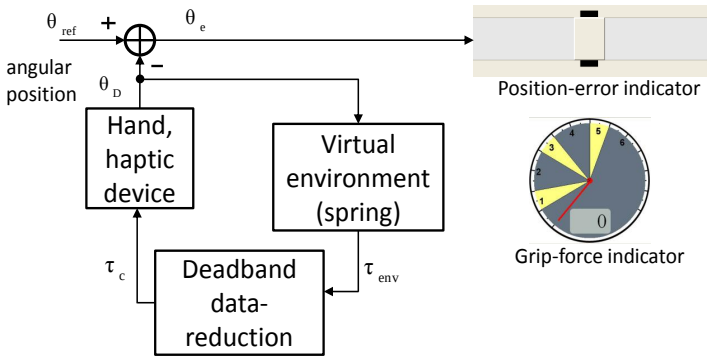


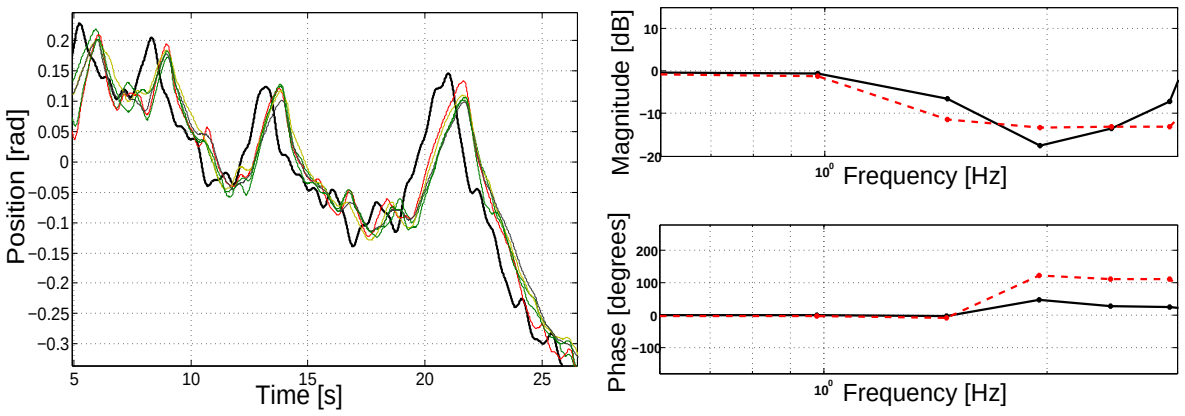
Figure 4.11: Human model identification. Figure reproduced from [CSH11], ©2011 IEEE.

For the range of k used here, no operational instabilities were observed, even though no special control schemes were implemented to ensure stability (see [HHSB05a] for details). This may be attributed to the hand and device damping, which dissipate energy injected by the Deadband data reduction into the the haptic loop.

System identification Figure 4.10 (b) shows the Bode plot of the transfer function estimated from the first half of input-output data measured in the hand-device identification experiment, averaged across all users. Averaging was possible since exactly the same input torque was used for all the users. A second-order transfer function (see Equation 4.1) was then fitted to the experimental data. The estimated parameters are listed in Table 4.2. The second half of the experimental data was used for model validation.

For the second (operator model identification) experiment, the input signal was designed to be the summation of thirteen sinusoids to ensure a random appearing input to the user. The input frequencies were distributed evenly over the frequency range for manual control (0 $\leq f_{in} \leq 3$ Hz) [Hes80]. Figure 4.12(a) shows the time-domain plots of the reference input and the user output position signals.

The presence of delay d_o in the operator model made parameter adjustment in the frequency-domain difficult. Therefore, the parameter values in Table 4.3 were estimated in the time-domain by least squares-based optimization. Nevertheless, the frequency-domain model-fit is shown in Figure 4.12(b). To make the parameter estimation tractable given the complexity of the action model in Figure 4.7, some a-priori information or simplifications were used. As mentioned before, the neuromuscular system parameters were set as $\omega_n = 10$ rad/s and



(a) (b)

Figure 4.12: Experiment for identifying the human action model: (a) time-domain plots. The thick leftmost curve shows the reference position signal to be followed, all other curves are the position signals traced by subjects for a range of deadband k values. Variances in model parameters are shown in Table 4.3, (b) Bode plots. The dashed curve shows the transfer function fitted to the solid experimental one. Figure reproduced from [CSH11], ©2011 IEEE.

$M(kgm^2)$	$B(kgm^2/rad)$	$K(Nm/rad)$
0.003 ± 0.002	0.45 ± 0.21	9.25 ± 3

Table 4.2: Hand-device model parameters

K_e	$d_o(s)$	K_1	K_2	$\frac{1}{T_1}, \frac{1}{T_2}(s^{-1})$
141 ± 34	0.28 ± 0.07	65 ± 15	38 ± 10	23 ± 7

Table 4.3: Operator structural model parameters

$\zeta = 0.707$. The mechanoreceptor time constants T_1 and T_2 were set to be equal [PCAB98]. Furthermore, an initial guess of $d_o = 0.2\,\mathrm{s}$ was used for the operator delay [Hes80].

4.2.2.2 Perception model

This section presents a haptic perception model based on the well-established Weber-Fechner law of psychophysics [Yil64]. Inspired by this law, it was assumed that the psychological sensation registered by the brain when experiencing haptic stimuli varies logarithmically with the magnitude of the physical haptic stimulus:

$$S = c \cdot ln(\frac{x}{x_o}) \tag{4.4}$$

where S denotes the perceived sensation, determined experimentally, x represents the stimulus magnitude, and x_o denotes the absolute detection threshold of the stimulus, and c is the proportionality constant.

For an N-long time-sampled signal, a distortion metric called *Perceptual Mean Square Error* or *PMSE* in the perceptual domain can be defined as:

$$\begin{aligned} PMSE &= \frac{1}{N}\sum_{i=0}^{N-1}\left[S(i) - \hat{S}(i)\right]^2 \\ &= \frac{c^2}{N}\cdot\sum_{i=0}^{N-1}\left[ln(\frac{x(i)}{\hat{x}(i)})\right]^2 \end{aligned} \tag{4.5}$$

where S is the undistorted reference, while $\hat{S}$ is the signal distorted by data reduction/compression, both in the perceptual domain, x and $\hat{x}$ are the corresponding time-domain signals.

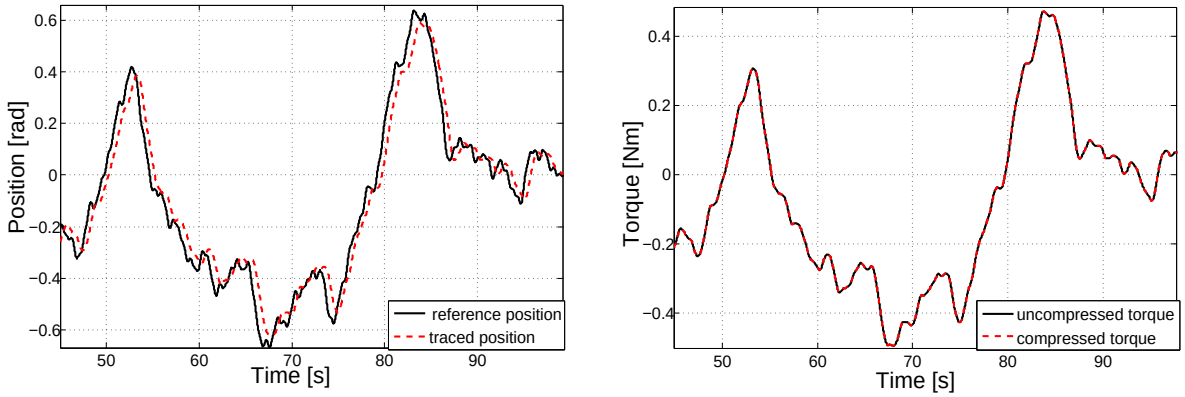
4.2.3 Quality prediction results

With the model parameters obtained before, the model of the human operator interacting haptically with a linear spring environment was simulated with the deadband data reduction scheme parameter k ranging from 0% to 20%. The quality of the resultant compressed force

feedback signals, distorted due to the lossy data reduction scheme was evaluated using the PMSE metric for perceptual distortion.

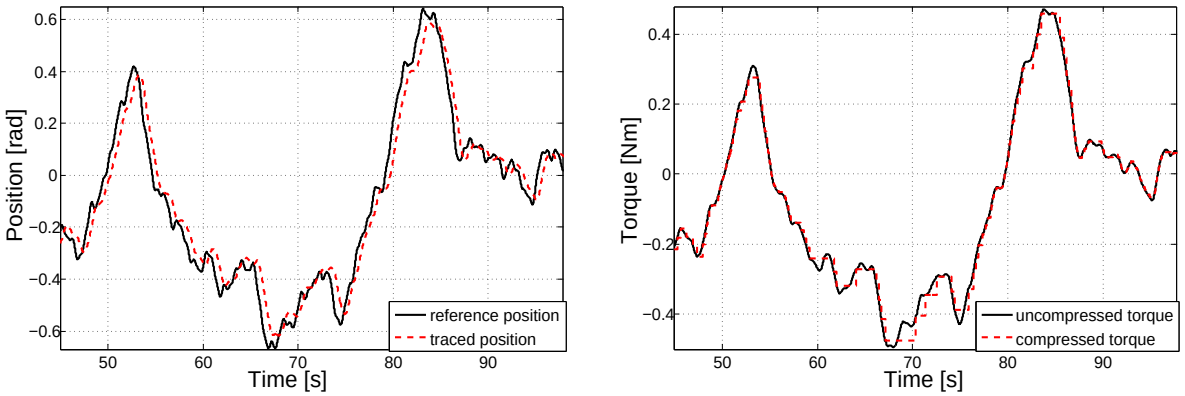
Figures 4.13 and 4.14 show the time-domain position and feedback torque waveforms for deadband parameter k values of 0% and 15%. Figure 4.15 (a) shows subjective quality ratings obtained from the experiments as the deadband parameter k increases. Figure 4.15 (a) also shows results of the simulations performed for increasing values of the deadband parameter k. The trend of the subjective quality curve is coherent with that of the predictions from the human model simulation. Moreover, there is high correlation between the corresponding subjective quality ratings and the objective predictions (correlation coefficient of 0.97). It can be seen from Figure 4.15 (b) that with increasing k, the predicted perceived quality as well as transmission packet rates decreased monotonically.

(Figures 4.13– 4.15 appear on the next page.)



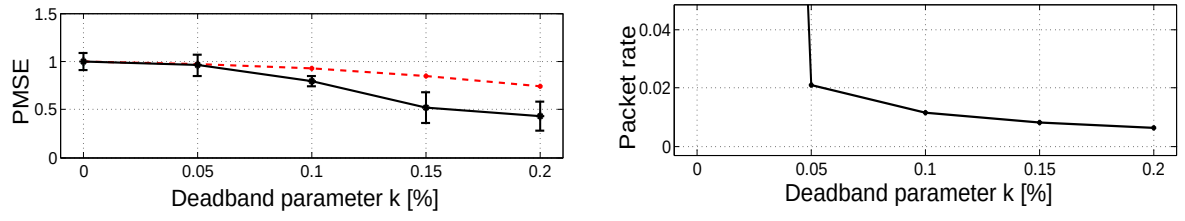
(a) (b)

Figure 4.13: Simulation waveforms for deadband parameter $k = 0\%$: (a) positions, (b) torques (the uncompressed and compressed waveforms coincide for $k = 0$). Figure reproduced from [CSH11], ©2011 IEEE.



(a) (b)

Figure 4.14: Simulation waveforms for deadband parameter $k = 15\%$: (a) positions, (b) torques. Figure reproduced from [CSH11], ©2011 IEEE.



(a) (b)

Figure 4.15: (a) Normalized subjective ratings (—) from the perceptual experiments, and the corresponding normalized quality predictions using PMSE (- - -), and (b) packet rate simulation results. The normalized packet rate at $k = 0$ is 1. Figure reproduced from [CSH11], ©2011 IEEE.

4.3 Conclusion

Perceptual user studies in haptics are difficult and time-consuming, slowing down the progress of haptic signal processing research. To address this challenge, this chapter presented novel methods for model-based prediction of perceptual quality degradation in compressed haptic signals.

Vibrotactile haptics

For haptic texture signals, perceptually-motivated features were extracted that were vital clues for texture signal discrimination. It was seen that abstract features like roughness, brightness, and time-envelope pattern play an important role in haptic signal perception, in addition to low-level psychophysical models. A combination of all the above features led to about 30% improvement in the performance of the quality prediction algorithm over the state-of-the-art.

For future work, a significant challenge will be to construct mathematical models for haptic perceptual features that have not been accounted for in the present work. These features could, for example, try to capture subjective characteristics like 'slowness-fastness', 'sparseness-denseness', 'bluntness-sharpness', etc. of the haptic texture signals.

Furthermore, the performance of the proposed approach should be investigated for a wider variety of textures. The major challenge here will be to compile a comprehensive database of subjective dissimilarity data for a variety of textures.

Kinesthetic haptics

A framework and a methodology for the model-based Quality Evaluation (MBQP) of a haptic data reduction scheme was proposed and validated by experiments and simulations. Here is a summary of the most important points:

- In addition to a perception model, a human action model forms an integral part of the MBQP framework for kinesthetic haptic interaction.

- The action model presented in Section 4.2.2.1 is based on a compensatory trajectory tracking task, which to a large extent makes haptic signals uniform across users. The action model presented contains sub-models for the Central Nervous System and the Neuromuscular system of the human arm.

- A perception model based on the well-established Weber-Fechner law of psychophysics was proposed to aid in the perception-domain comparison of clean original kinesthetic signals and the corresponding compressed ones.

- Simulation of a haptic communication system with the above models leads to perceptual quality predictions that match in trend with subjective ratings obtained from psychophysical tests, as well as show a high correlation with them. The optimal compression parameter k can be chosen from the simulation results depending upon the operation quality regarded as acceptable.

When haptic tasks less constrained than trajectory tracking are studied in the future, human intention will also need to be modeled based on task description. The detection of more complex features (e.g. artifacts arising out of the presence of network packet loss, delay and jitter) in the haptic signals may necessitate more complex psychophysical models. Remnant signals that account for the non-linearities may also improve the prediction of perceptual quality for more complex non-linear artifacts.

Chapter 5

Conclusion and outlook

Perceptual coding schemes for kinesthetic data streams have been intensively researched upon over the past decade. However, such investigations for vibrotactile texture signals have not yet received much attention, although texture signals contribute significantly to the realism of haptic interaction mediated by technical systems. In Chapter 3, this thesis has addressed the development of a texture codec based on models for the mechanisms of production and perception of vibrotactile texture signals. Furthermore in Chapter 4, perceptual quality aspects for haptic communication systems have been studied. A summary of the main contributions is given below, along with ideas for future research.

5.1 Production and perception models for haptic texture signals

Section 3.1 qualitatively analyzed the mechanism behind the production of vibrotactile signals when a mechanical tool is used to scan a textured surface. It was shown that the source-filter signal model, similar to that used for speech production, can be used to effectively describe this mechanism in the texture codec. Section 3.2 presented novel masking data, which motivate the usage of a perceptual weighting filter in the texture codec based on the Simultaneous Masking phenomenon.

Limitations and Future work

The simultaneous masking investigations presented in Section 3.2 dealt with single-peak noise maskers. However, vibrotactile signals obtained by scanning a texture surface may be wide-band in nature, with multiple peaks in the frequency domain. Therefore, an interesting direction for future research is the characterization of the shape of the masking curves in the presence of more than one narrowband maskers.

Furthermore, an interesting topic to be addressed in the future is the sensing, coding, and display of multi-point haptic texture signals, for example to the entire palm of the hand. For the development of a perceptually motivated coding scheme for multiple spatial channels in such a system, the spatial aspects of haptic perception will have to be considered. In particular, the changes in the perceptual threshold of a given channel due to the presence of a strong signal on another channel will have to be studied.

5.2 Haptic texture compression

To address the cost-efficient transmission and storage of haptic texture signals, Section 3.3 presented a novel texture compression scheme that exploits the production and perception models mentioned above.

An encoder based on a Linear Predictive source-filter model of the texture signal was presented. A perceptual weighting filter based on simultaneous masking was used to guide the automatic codec parameter selection, so that the coding noise remained imperceptible to the user. The performance of this codec was analyzed both objectively as well as subjectively. It was shown that the compression algorithm achieves, for a comparable perceptual quality, a compression ratio two times better than the state-of-the-art.

Furthermore, a bitrate-scalable codec to enable operation in a range of different bitrates under varied network traffic conditions and quality demands was presented in Section 3.7. Subjective evaluation of the performance of the bitrate-scalable codec showed that it is possible to drive down the codec output bitrate to as low as 2.3 kbps (down from 32 kbps) without the subjects being able to reliable discriminate between the codec input and distorted output texture signals.

Limitations and Future work

The subjective evaluations of the codec studied the deterioration of perceptual performance for a given texture as the coding distortion rises. However, in the future it is also necessary to study the "task performance" of the codec. Such an evaluation will study if the subjects can still reliably discriminate between two different textures as the coding distortion rises.

The above texture codec was developed for a single-channel haptic texture signal. The surface was sensed with a single stylus tip, and the resulting vibrations were displayed to the entire hand. Future research could address the development of a codec for multi-point haptic texture signals that can be displayed to the hand surface with a high spatial resolution. Under this topic, in addition to the exploiting temporal redundancies on a given haptic channel, spatial redundancies across neighboring haptic channels could be exploited to achieve efficient data compression.

5.3 Model-based perceptual haptic quality prediction

To address the challenge of difficult, time-consuming, and expensive subjective user studies, Chapter 4 presented model-based methods to predict haptic perceptual quality. Prediction methods were developed for both vibrotactile texture and kinesthetic signals.

Section 4.1 presented a model that predicts perceptual quality based on high-level features for vibrotactile texture signals. A combination of high-level features like roughness, brightness, and time-envelope pattern, in addition to low-level psychophysical models, led to 30% improvement in the performance of the quality prediction algorithm over the state-of-the-art.

Section 4.2 treated the quality prediction for kinesthetic signals. A human action-perception model accounting for the bidirectional closed-loop nature of kinesthetic haptics was presented. A high correlation between the corresponding subjective quality ratings and objective predictions (correlation coefficient of 0.97) was obtained.

Limitations and Future work

For vibrotactile signals, a number of high-level perceptual features like 'slowness-fastness', 'sparseness-denseness', 'bluntness-sharpness', etc. have been identified. However, there is a lack of mathematical descriptions of such features in the literature. This issue should be addressed for further improvement of the perceptual quality prediction methods. Another issue that needs to be addressed is creating a comprehensive database of subjective data for the validation of the proposed prediction methods across many classes of textures.

In case of kinesthetic haptics, the action model presented in Section 4.2 is constrained by the definition of the haptic task, namely trajectory tracking. Generalization of the human action model will demand the study of other common tasks like pick-and-place, free natural exploration, etc. This in turn will require models of the human intention while performing these tasks.

Finally, an end-to-end system model would allow the simulation of the entire kinesthetic-vibrotactile teleoperation system on a computer. Any signal-processing algorithms operating between the remote and local side, could then be designed/evaluated in simulation, without the need for carrying out human subject experiments every time. Such an end-to-end model will require the development of a action model together with a unified kinesthetic and vibrotactile perception model. Under this topic, physical models for the subsystems constituting the haptic teleoperation system shown in Figure 1.1 should be developed. This would involve models for the entire chain of human-robot coupling, robot-robot coupling, and the robot-to-remote-surface interaction.

Bibliography

Publications by the author

Journal publications

[CSDS15] R. Chaudhari, C. Schuwerk, M. Danaei, and E. Steinbach. Perceptual and bitrate-scalable coding of haptic surface texture signals. *Selected Topics in Signal Processing, IEEE Journal of*, 9(3):462–473, Apr. 2015. [cited on page(s) 7, 42, 43, 44, 45, 46, 47, 48, 49, 54, 55, 68, 69, 70, 71, 72, 78, 79, 80, 117, 120, 121, 122]

[KCS11] J. Kammerl, R. Chaudhari, and E. Steinbach. Combining contact models with perceptual data reduction for efficient haptic data communication in networked VEs. *IEEE Trans. on Instrumentation and Measurement*, 60(1):57–68, 2011. [cited on page(s) 7]

[SHE+12] E. Steinbach, S. Hirche, M. Ernst, F. Brandi, R. Chaudhari, J. Kammerl, and I. Vittorias. Haptic communications. *Proceedings of the IEEE*, 100(4):937–956, 2012. [cited on page(s) 1, 3, 7, 15, 23]

[SHK+11] E. Steinbach, S. Hirche, J. Kammerl, I. Vittorias, and R. Chaudhari. Haptic data compression and communication. *IEEE Signal Processing Magazine*, 28(1):87–96, Jan. 2011. [cited on page(s) 7, 15, 22, 23, 24, 25, 120]

[SPCS14] C. Schuwerk, G. Paggetti, R. Chaudhari, and E. Steinbach. Perception based traffic control for shared haptic virtual environments. *Presence: Teleoperators and Virtual Environments*, 23(3):320–338, 2014. [cited on page(s) 7]

Book chapter

[CAS14] R. Chaudhari, E. Altinsoy, and E. Steinbach. *in Quality of Experience: Advanced Concepts, Applications and Methods*, chapter Haptics, pages 261–276. Springer, Heidelberg, 2014. [cited on page(s) 7, 12, 15, 24, 119, 120]

Conference publications

[BCS11] F. Brandi, R. Chaudhari, and E. Steinbach. Texture compensation for haptic feedback signal compression. In *2011 IEEE Int. Conf. on Virtual Environments, Human-Computer*

Interfaces, and Measurement Systems (VECIMS), pages 1–4, Ottawa, Canada, Sep. 2011. [cited on page(s) 7]

[CcK+12] R. Chaudhari, B. Çizmeci, K. J. Kuchenbecker, S. Choi, and E. Steinbach. Low bitrate source-filter model based compression of vibrotactile texture signals in haptic teleoperation. In *Proceedings of the 20th ACM International Conference on Multimedia*, MM '12, pages 409–418, New York, NY, USA, 2012. ACM. [cited on page(s) 2, 3, 7, 20, 23, 28, 38, 39, 40, 41, 54, 55, 57, 58, 60, 61, 62, 63, 64, 65, 75, 102, 119, 120, 121]

[CCX+14] B. Cizmeci, R. Chaudhari, X. Xu, N. Alt, and E. Steinbach. A visual-haptic multiplexing scheme for teleoperation over constant-bitrate communication links. In Malika Auvray and Christian Duriez, editors, *Haptics: Neuroscience, Devices, Modeling, and Applications*, Lecture Notes in Computer Science, pages 131–138. Springer Berlin Heidelberg, 2014. [cited on page(s) 7]

[CKS09] R. Chaudhari, J. Kammerl, and E. Steinbach. On the compression and rendering of event-triggered force transients in networked virtual environments. In *IEEE Inte. Workshop on Haptic Audio visual Environments and Games 2009. HAVE 2009.*, pages 147–152, Lecco, Italy, Nov. 2009. [cited on page(s) 7]

[CSH11] R. Chaudhari, E. Steinbach, and S. Hirche. Towards an objective quality evaluation framework for haptic data reduction. In *2011 IEEE World Haptics Conf. (WHC)*, pages 539–544, Jun. 2011. [cited on page(s) 7, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 122, 123]

[CSN+11] R. Chaudhari, C. Schuwerk, V. Nitsch, E. Steinbach, and B. Farber. Opening the haptic loop: Network degradation limits for haptic task performance. In *2011 IEEE Int. Workshop on Haptic Audio Visual Environments and Games (HAVE)*, pages 56–61, 2011. [cited on page(s) 7]

[CYS+15] R. Chaudhari, Y. Yoo, C. Schuwerk, S. Choi, and E. Steinbach. Objective quality prediction for haptic texture signal compression. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2015, submitted for review*, Apr. 2015. [cited on page(s) 7, 77]

[KCS10] J. Kammerl, R. Chaudhari, and E. Steinbach. Exploiting directional dependencies of force perception for lossy haptic data reduction. In *2010 IEEE Int. Symp. on Haptic Audio-Visual Environments and Games (HAVE)*, pages 1–6, Phoenix, Arizona, USA, Oct. 2010. [cited on page(s) 7]

[SCS12] C. Schuwerk, R. Chaudhari, and E. Steinbach. Perceptually robust traffic control in distributed haptic virtual environments. In Poika Isokoski and Jukka Springare, editors, *Haptics: Perception, Devices, Mobility, and Communication*, volume 7282 of *Lecture Notes in Computer Science*, pages 469–480. Springer Berlin Heidelberg, 2012. [cited on page(s) 7]

[SCS14] C. Schuwerk, E. Chaudhari, and E. Steinbach. Delay compensation in shared haptic virtual environments. In *2014 IEEE Haptics Symposium (HAPTICS)*, pages 371–377, Feb. 2014. [cited on page(s) 7]

[XKCS11] X. Xu, J. Kammerl, R. Chaudhari, and E. Steinbach. Hybrid signal-based and geometry-based prediction for haptic data reduction. In *2011 IEEE Int. Workshop on Haptic Audio Visual Environments and Games (HAVE)*, pages 68–73, Qinhuangdao, Hebei, China, Oct. 2011. [cited on page(s) 7]

E-letter

[BCH+11] F. Brandi, R. Chaudhari, S. Hirche, J. Kammerl, E. Steinbach, and I. Vittorias. Perceptual haptic data reduction in telepresence and teleaction systems. *Multimedia Communication Technical Committee (MMTC) E-Letter*, 6(1):7–10, 2011. [cited on page(s) 7]

General publications

[ABC+07] R. Aracil, M. Buss, S. Cobos, M. Ferre, S. Hirche, M. Kuschel, and A. Peer. The human role in telerobotics. In M. Ferre, M. Buss, R. Aracil, C. Melchiorri, and C. Balaguer, editors, *Advances in Telerobotics*, volume 31 of *Springer Tracts in Advanced Robotics*, pages 11–24. Springer Berlin Heidelberg, 2007. [cited on page(s) 2]

[Alt12] M. E. Altinsoy. The quality of auditory-tactile virtual environments. *Journal of the Audio Engineering Society*, 60(1-2):38–46, Jan.–Feb. 2012. [cited on page(s) 13]

[AM12] M. E. Altinsoy and S. Merchel. Electrotactile feedback for handheld devices with touch screen and simulation of roughness. *IEEE Trans. on Haptics*, 5(1):6–13, Jan.–Mar. 2012. [cited on page(s) 13]

[ASG08] L. K. Amanda, M. C. Stoeckel, and R. Goebel. The neural bases of haptic working memory. In Martin Grunwald, editor, *Human Haptic Perception: Basics and Applications*, pages 113–129. Birkhäuser Basel, 2008. [cited on page(s) 46, 80]

[AYS98] N. Asamura, N. Yokoyama, and H. Shinoda. Selectively stimulating skin receptors for tactile display. *IEEE Computer Graphics and Applications*, 18(6):32–37, 1998. [cited on page(s) 10]

[BCSS00] C. Basdogan, C., M. A. Srinivasan, and M. Slater. An experimental study on the role of touch in shared virtual environments. *ACM Trans. on Computer-Human Interaction*, 7(4):443–460, 2000. [cited on page(s) 31]

[BH03] S. Bensmaïa and M. Hollins. The vibrations of texture. *Somatosensory Motor Research*, 20(1):33–43, 2003. [cited on page(s) 6, 23, 55]

[BH05] S. Bensmaïa and M. Hollins. Pacinian representations of fine surface texture. *Perception & Psychophysics*, 67(5):842–854, Jul. 2005. [cited on page(s) 33, 82]

[BHY05] S. Bensmaïa, M. Hollins, and J. Yau. Vibrotactile intensity and frequency information in the pacinian system: a psychophysical model. *Perception & Psychophysics*, 67(5):828–841, Jul. 2005. [cited on page(s) 31, 34, 81]

[BO07] B. G. Breitmeyer and H. Ogmen. Visual masking. *Scholarpedia*, 2(7):3330, 2007. [cited on page(s) 17]

[Bor05] C. W. Borst. Predictive coding for efficient host-device communication in a pneumatic force-feedback display. In *Proc. First Joint Eurohaptics Conf. and Int. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems (World Haptics)*, pages 596–599, Pisa, Italy, March 2005. [cited on page(s) 3, 25]

[BS94] K. Brandenburg and G. Stoll. Iso/mpeg-1 audio: A generic standard for coding of high-quality digital audio. *J. Audio Eng. Soc*, 42(10):780–792, 1994. [cited on page(s) 41]

[BS10] F. Brandi and E. Steinbach. Minimizing the number of support points in perceptual offline haptic compression. In *Proc. IEEE Int. Workshop on Haptic Audio Visual Environments and Games (HAVE)*, pages 7–12, Phoenix, AZ, USA, Oct. 2010. [cited on page(s) 25]

[BSH+06] F. Barbagli, K. Salisbury, C. Ho, C. Spence, and H. Tan. Haptic discrimination of force direction and the influence of visual information. *ACM Trans. on Applied Perception*, 3(2):125–135, 2006. [cited on page(s) 15, 16, 125]

[CB94] J. E. Colgate and J. M. Brown. Factors affecting the z-width of a haptic display. *1994 IEEE Int. Conf. on Robotics and Automation, Proceedings*, 4:3205–3210, May 1994. [cited on page(s) 85]

[CCK+12] S. Chauhan, R. F. Coelho, S. Kalan, R. M. Satava, and V. R. Patel. Evolution of robotic surgery: Past, present, and future. In V. R. Patel, editor, *Robotic Urologic Surgery*, pages 3–10. Springer London, 2012. [cited on page(s) 2]

[Cis14] Cisco Systems, Inc. TelePresence, 2014. http://www.cisco.com/c/en/us/products/collaboration-endpoints/telepresence-system-3000-series. [cited on page(s) 1]

[CK13] S. Choi and K. J. Kuchenbecker. Vibrotactile display: Perception, technology, and applications. *Proceedings of the IEEE*, 101(9):2093–2104, Sep. 2013. [cited on page(s) 2]

[Cra72] J. C. Craig. Difference threshold for intensity of tactile stimuli. *Perception & Psychophysics*, 11(2):150–152, 1972. [cited on page(s) 17, 18, 25, 120]

[CT08] J. Chen and J. Thyssen. Analysis-by-synthesis speech coding. In Y. Huang J. Benesty, M. M. Sondhi, editor, *Springer Handbook of Speech Processing*, pages 351–392. Springer Berlin Heidelberg, 2008. [cited on page(s) 41, 50, 52, 53, 121]

[EH50] J. P. Egan and H. W. Hake. On the masking pattern of a simple auditory stimulus. *J. Acoust. Soc. Am.*, 22(5):622–630, 1950. [cited on page(s) 19]

[FB01] A. Fischer and J. Barhak. Tele-design for manufacturing. *CIRP Annals - Manufacturing Technology*, 50(1):77–80, 2001. [cited on page(s) 2]

[FC10] M. J. Fu and M. C. Cavusoglu. Three-dimensional human arm and hand dynamics and variability model for a stylus-based haptic interface. *2010 IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 1339–1346, 2010. [cited on page(s) 91]

[FCR+70] J. L. Flanagan, C. Coker, L. Rabiner, R. W. Schafer, and N. Umeda. Synthetic voices for computers. *IEEE Spectrum*, 7(10):22–45, Oct. 1970. [cited on page(s) 38, 39, 120]

[FL12] J. A. Fishel and G. E. Loeb. Bayesian exploration for intelligent identification of textures. *Frontiers in Neurorobotics*, 6(4), 2012. [cited on page(s) 33, 82]

[FS67a] W. Ferrell and T. B. Sheridan. Supervisory control of remote manipulation. *IEEE Spectrum*, 4(10):81–88, Oct. 1967. [cited on page(s) 12, 119]

[FUS92] Y. Fujii, H. Usui, and Y. Shinohara. Development of multi-functional telerobotic systems for reactor dismantlement. *Journal of Nuclear Science and Technology*, 29(9):930–936, 1992. [cited on page(s) 2]

[FZ56] R. Feldtkeller and E. Zwicker. *Das Ohr als Nachrichtenempfänger*. S. Hirzel Verlag Stuttgart (Monographien der elektrischen Nachrichtentechnik), 1956. [cited on page(s) 17, 18, 19, 120]

[Gel10] S. A Gelfand. *Hearing: An Introduction to Psychological and Physiological Acoustics*, chapter Masking. Informa Healthcare, the UK, 2010. [cited on page(s) 5]

[Ges13] G. A. Gescheider. *Psychophysics: the fundamentals*. Psychology Press, 2013. [cited on page(s) 47]

[GLL11] V. L. Guruswamy, J. Lang, and W. Lee. IIR filter models of haptic vibration textures. *IEEE Trans. on Instrumentation and Measurement*, 60(1):93–103, Jan. 2011. [cited on page(s) 37]

[GSJBV89] G. A. Gescheider, Jr. S. J. Bolanowski, and R. T. Verrillo. Vibrotactile masking: Effects of stimulus onset asynchrony and stimulus frequency. *J. Acoust. Soc. Am.*, 85(5), May 1989. [cited on page(s) 20]

[GVD82] G. A. Gescheider, R. T. Verrillo, and C. L. Van Doren. Prediction of vibrotactile masking functions. *J. Acoust. Soc. Am.*, 72(5):1421–1426, Nov. 1982. [cited on page(s) 19, 21]

[HA96] V. Hayward and O. R. Astley. Performance measures for haptic interfaces. In *Robotics Research: the 7th International Symposium*, pages 195–207. Springer Verlag, 1996. [cited on page(s) 11]

[HAC+04] V. Hayward, O. R. Astley, M. CruzHernandez, D. Grant, and G. RoblesDeLaTorre. Haptic interfaces and devices. *Sensor Review*, 24(1):16–29, 2004. [cited on page(s) 2]

[Han89] B. Hannaford. Stability and performance tradeoffs in bi-lateral telemanipulation. In *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, volume 3, pages 1764–1767, May 1989. [cited on page(s) 30]

[Han06] M. Hansen. Lehre und Ausbildung in Psychoakustik mit psylab: freie Software für psychoakustische Experimente. *Fortschritte der Akustik – DAGA '06*, pages 591–592, 2006. [cited on page(s) 47]

[HB04] S. Hirche and M. Buss. Telepresence control in packet-switched communication networks. In *Proc. IEEE Int. Conf. on Control Applications CCA'04*, volume 1, pages 236–241, 2004. [cited on page(s) 3]

[HB06] S. Hirche and M. Buss. Transparenz haptischer telepräsenzsysteme mit konstanter zeitverzögerung (transparency of haptic telepresence systems with constant time delay). *Automatisierungstechnik*, 54(2):51–59, 2006. [cited on page(s) 30, 31, 32]

[HB07b] S. Hirche and M. Buss. Transparent data reduction in networked telepresence and teleaction systems part II: Time-delayed communication. *PRESENCE: Teleoperators and Virtual Environments*, 16(5):532–542, 2007. [cited on page(s) 3]

[HC10] I. Hwang and S. Choi. Perceptual space and adjective rating of sinusoidal vibrations perceived via mobile device. In *Proc. 2010 IEEE Haptics Symposium*, pages 1–8, Washington, DC, USA, 2010. IEEE Computer Society. [cited on page(s) 33, 81, 82]

[Hes80] R. A. Hess. Structural model of the adaptive human pilot. *Journal of Guidance and Control*, 3(5):416–423, 1980. [cited on page(s) 86, 88, 89, 93, 94, 122]

[HESG08a] A. Hamam, M. Eid, A. E. Saddik, and N. D. Georganas. A quality of experience model for haptic user interfaces. In *Proc. Workshop Haptic User Inter. (Ambient Media Syst.)*, pages 1–6, 2008. [cited on page(s) 30, 32]

[HESG08b] A. Hamam, M. Eid, A. El Saddik, and N. D. Georganas. A fuzzy logic system for evaluating quality of experience of haptic-based applications. In *Proc. 6th int. conf. on Haptics: Perception, Devices and Scenarios*, EuroHaptics '08, pages 129–138, Berlin, Heidelberg, 2008. Springer-Verlag. [cited on page(s) 30, 31, 32]

[Hew14] Hewlett-Packard Development Company, L.P. HP Video Conferencing and Halo Telepresence Solutions, 2014. http://www.polycom.com/solutions/solutions-by-business-uc-environment/solutions-for-hp/uc-services.html. [cited on page(s) 1]

[HFB+07] S. Hirche, M. Ferre, J. Barrio, C. Melchiorri, and M. Buss. *Advances in Telerobotics*, chapter Bilateral Control Architectures for Telerobotics, pages 163–176. STAR. Springer-Verlag, Berlin, 2007. [cited on page(s) 30]

[HHC+08a] 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 Trans. on Signal Processing*, 56(2):588–597, Feb. 2008. [cited on page(s) 3, 15, 22, 23, 25, 75, 85]

[HHSB05a] S. Hirche, P. Hinterseer, E. Steinbach, and M. Buss. Network traffic reduction in haptic telepresence systems by deadband control. In *Proc. 16th IFAC World*, 2005. [cited on page(s) 92]

[HHSB05b] S. Hirche, P. Hinterseer, E. Steinbach, and M. Buss. Towards deadband control in networked teleoperation systems. In *Proc. IFAC World Congress*, 2005. [cited on page(s) 3]

[HHSB07] S. Hirche, P. Hinterseer, E. Steinbach, and M. Buss. Transparent data reduction in networked telepresence and teleaction systems. Part I: Communication without time delay. *Presence: Teleoperators & Virtual Environments*, 16(5):523–531, Oct. 2007. [cited on page(s) 23]

[Hin09] P. Hinterseer. *Compression and Transmission of Haptic Data in Telepresence and Teleaction Systems*. PhD thesis, Technische Universität München, 2009. [cited on page(s) 3, 28]

[Hir05] S. Hirche. *Haptic Telepresence in Packet Switched Communication Networks*. PhD thesis, VDI-Verlag, Düsseldorf, Germany, 2005. [cited on page(s) 30]

[HMF+01] K. Hikichi, H. Morino, I. Fukuda, S. Matsumoto, Y. Yasuda, I. Arimoto, M. Iijima, and K. Sezaki. Architecture of haptics communication system for adaptation to network environments. In *Proc. IEEE Int. Conf. on Multimedia & Expo (ICME)*, pages 563–566, Los Alamitos, CA, USA, Aug. 2001. [cited on page(s) 3, 25]

[HS06] P. F. Hokayem and M. W. Spong. Bilateral teleoperation: An historical survey. *Automatica*, 42(12):2035–2057, 2006. [cited on page(s) 23]

[HS13] A. Hamam and A. El Saddik. Toward a mathematical model for quality of experience evaluation of haptic applications. *IEEE Trans. on Instrumentation and Measurement*, 62(12):3315–3322, 2013. [cited on page(s) 30, 31, 32]

[HSA14] A. Hamam, A. El Saddik, and J. Alja'am. A quality of experience model for haptic virtual environments. *ACM Trans. Multimedia Comput. Commun. Appl.*, 10(3):1–23, Apr. 2014. [cited on page(s) 30, 31, 32]

[HVZ83] R. D. Hamer, R. T. Verrillo, and J. J. Zwislocki. Vibrotactile masking of pacinian and non-pacinian channels. *J. Acoust. Soc. Am.*, 73(4):1293–1303, Apr. 1983. [cited on page(s) 20, 21]

[HZS01] K. Hashtrudi-Zaad and S. E. Salcudean. Analysis of control architectures for teleoperation systems with impedance/admittance master and slave manipulators. *International Journal of Robotics Research*, 20(6):419–445, 2001. [cited on page(s) 30]

[ICG+08] R. Iglesias, S. Casado, T. Gutirrez, A. Garca-Alonso, W. Yu, and A. Marshall. Simultaneous remote haptic collaboration for assembling tasks. *Multimedia Systems*, 13(4):263–274, 2008. [cited on page(s) 2]

[ICT07a] A. Israr, S. Choi, and H. Z. Tan. Mechanical impedance of the hand holding a spherical tool at threshold and suprathreshold stimulation levels. In *EuroHaptics Conf. 2007 and Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2007. Second Joint*, pages 56–60, Mar. 2007. [cited on page(s) 17, 18, 120]

[ICT07b] A. Israr, S. Choi, and H. Z. Tan. Mechanical impedance of the hand holding a spherical tool at threshold and suprathreshold stimulation levels. *In Proc. Symp. on Haptic*

Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 07, pages 56–60, 2007. [cited on page(s) 42, 45, 47, 48, 49, 121]

[Jek05] U. Jekosch. *Voice and Speech Quality Perception.* Springer Verlag, 2005. [cited on page(s) 75]

[Jon00] L. A. Jones. Kinesthetic sensing. *MIT Press, Cambridge*, 2000. [cited on page(s) 6, 9, 14]

[JT13] L. A. Jones and H. Z. Tan. Application of psychophysical techniques to haptic research. *IEEE Trans. on Haptics*, 6(3):268–284, Jul. 2013. [cited on page(s) 45]

[Kam12] J. Kammerl. *Perceptual haptic data communication for telepresence and teleaction.* PhD thesis, Technische Universität München, 2012. [cited on page(s) 3, 25, 28]

[KCS10b] J. Kammerl, R. Chaudhari, and E. Steinbach. Exploting directional dependencies of force perception for lossy haptic data reduction. In *Proc. IEEE Int. Symp. on Haptic Audio-Visual Environments and Games (HAVE)*, pages 1–6, Phoenix, AZ, USA, Oct. 2010. [cited on page(s) 15, 16, 24, 120, 125]

[KCS11] J. Kammerl, R. Chaudhari, and E. Steinbach. Combining contact models with perceptual data reduction for efficient haptic data communication in networked VEs. *IEEE Trans. on Instrumentation and Measurement*, 60(1):57–68, Jan. 2011. [cited on page(s) 25]

[KH96] D. A. Kontarinis and R. D. Howe. Tactile display of vibratory information in teleoperation and virtual environments. *Presence*, 4:387–402, 1996. [cited on page(s) 76]

[KKH06] M. Kuschel, P. Kremer, and S. Hirche. Lossy data reduction methods for haptic telepresence systems. In *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 2933–2938, 2006. [cited on page(s) 3]

[KRM11] K. J. Kuchenbecker, J. M. Romano, and W. McMahan. Haptography: Capturing and recreating the rich feel of real surfaces. In Cedric Pradalier, Roland Siegwart, and Gerhard Hirzinger, editors, *Springer Tracts in Advanced Robotics*, volume 70 of *Robotics Research: the 14th International Symposium (ISRR 2009)*, pages 245–260. Springer, 2011. [cited on page(s) 41, 50]

[KS08] J. Kammerl and E. Steinbach. Deadband-based offline-coding of haptic media. In *Proc. ACM Multimedia*, pages 549–558, Vancouver, Canada, Oct. 2008. [cited on page(s) 25]

[KSP+04] A. Kron, G. Schmidt, B. Petzold, M. I. Zäh, P. Hinterseer, and E. Steinbach. Disposal of explosive ordnances by use of a bimanual haptic telepresence system. In *Proc. IEEE Int. Conf. on Robotics and Automation*, volume 2, pages 1968–1973, 2004. [cited on page(s) 2]

[Law92] D. A. Lawrence. Stability and transparency in bilateral teleoperation. *Proc. 31st IEEE Conf. on Decision and Control*, pages 2649–2655, 1992. [cited on page(s) 85]

[Law93] D. A. Lawrence. Stability and transparency in bilateral teleoperation. *IEEE Trans. on Robotics and Automation*, 9(5):624–637, Oct. 1993. [cited on page(s) 29, 30]

[Lev70] H. Levitt. Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.*, 33:467–476, 1970. [cited on page(s) 16, 47, 70]

[LKK12] S. Lim, K. Kyung, and D. Kwon. Effect of frequency difference on sensitivity of beats perception. *Experimental Brain Research*, 216:11–19, 2012. [cited on page(s) 42]

[LL86] J. M. Loomis and S. J. Lederman. *Tactual perception*, volume II of *Handbook of Perception and Human Performance*. Wiley-Interscience, 1986. [cited on page(s) 9]

[LMvP09] T. M. Lam, M. Mulder, and M. M. van Paassen. Haptic interface in uav tele-operation using force-stiffness feedback. *IEEE Int. Conf. on Systems, Man and Cybernetics, 2009. SMC 2009.*, pages 835–840, 2009. [cited on page(s) 89]

[LPD+00] D. A. Lawrence, L. Y. Pao, A. M. Dougherty, M. A. Salada, and Y. Pavlou. Rate-hardness: a new performance metric for haptic interfaces. *IEEE Trans. on Robotics and Automation*, 16(4):357–371, Aug 2000. [cited on page(s) 85]

[LRMK10] N. Landin, J. M. Romano, W. McMahan, and K. J. Kuchenbecker. Dimensional reduction of high-frequency accelerations for haptic rendering. In A. M. L. Kappers, J. B. F. van Erp, W. M. Bergmann Tiest, and F. C. T. van der Helm, editors, *Haptics: Generating and Perceiving Tangible Sensations*, volume 6192 of *Lecture Notes in Computer Science*, pages 79–86. Springer Berlin Heidelberg, 2010. [cited on page(s) 61]

[LS01] G. F. Lang and D. Snyder. Understanding the physics of electrodynamic shaker performance. *Sound and Vibration*, 1(1):2–10, October 2001. [cited on page(s) 13, 119]

[MC05] N. A. Macmillan and C. D. Creelman. *Detection theory: A user's guide (2nd Ed.)*. Mahwah, NJ: Erlbaum, 2005. [cited on page(s) 29, 59, 64, 77, 81]

[McR80] D. T. McRuer. *Human Dynamics in Man-Machine Systems*, volume 16. Automatica, 1980. [cited on page(s) 89]

[MFCJV95b] J. C. Makous, R. M. Friedman, and Jr. C. J. Vierck. A critical band filter in touch. *The Journal of Neuroscience*, 15(4):2808–2818, 1995. [cited on page(s) 20]

[Min80] M. Minsky. Telepresence. *Omni Magazine*, Jun. 1980. [cited on page(s) 29]

[MRRK10] W. McMahan, J. M. Romano, A. M. Abdul Rahuman, and K. J. Kuchenbecker. High frequency acceleration feedback significantly increases the realism of haptically rendered textured surfaces. In *Proc. IEEE Haptics Symp.*, pages 141–148, Mar. 2010. [cited on page(s) 12, 60, 76]

[OIFS12] J. Osada, Y. Ishibashi, N. Fukushima, and S. Sugawara. Qoe assessment in haptic tele-operation systems:position-position versus position-force. *David Publishing: Computer Technology and Application*, 3:756–764, 2012. [cited on page(s) 32]

[OKST08] S. Okamoto, M. Konyo, S. Saga, and S. Tadokoro. Identification of cutaneous detection thresholds against time-delay stimuli for tactile displays. In *IEEE Int. Conf. on Robotics and Automation*, pages 220–225, May 2008. [cited on page(s) 23, 57]

[ONY13] S. Okamoto, H. Nagano, and Y. Yamada. Psychophysical dimensions of tactile perception of textures. *IEEE Trans. on Haptics*, 6(1):81–93, Jan. 2013. [cited on page(s) 85]

[OSSP12] M. Orozco, J. Silva, A. El Saddik, and E. Petriu. *Haptics rendering and applications: The role of haptics in games*, chapter 11, pages 217–234. InTech, 2012. [cited on page(s) 2]

[OY10] S. Okamoto and Y. Yamada. Perceptual properties of vibrotactile material texture: Effects of amplitude changes and stimuli beneath detection thresholds. In *IEEE/SICE Int. Symp. on System Integration*, pages 384–389, Dec. 2010. [cited on page(s) 57]

[OY11] S. Okamoto and Y. Yamada. An objective index that substitutes for subjective quality of vibrotactile material-like textures. In *2011 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 3060–3067, Sep. 2011. [cited on page(s) 31, 32, 33, 34, 81, 82, 84, 120, 122]

[OY13] S. Okamoto and Y. Yamada. Lossy data compression of vibrotactile material-like textures. *IEEE Trans. on Haptics*, 6(1):69–80, Jan. 2013. [cited on page(s) 3, 25, 26, 27, 28, 57, 120, 125]

[PAF+01] D. Purves, G. J. Augustine, D. Fitzpatrick, L. C Katz, A. LaMantia, J. O. McNamara, and S. M. Williams, editors. *Neuroscience*. Sunderland (MA): Sinauer Associates, 2001. [cited on page(s) 10]

[PC11] G. Park and S. Choi. Perceptual space of amplitude-modulated vibrotactile stimuli. In *IEEE World Haptics Conference*, pages 59–64, Jun. 2011. [cited on page(s) 33, 82, 83]

[PCAB98] L. F. Penin, A. Caballero, R. Aracil, and A. Barrientos. Human behavior modeling in master-slave teleoperation with kinesthetic feedback. *Proc. 1998 IEEE Int. Conf. on Robotics and Automation, 1998.*, 3:2244–2249, 1998. [cited on page(s) 86, 88, 89, 94, 122]

[PFH+06] H. Pongrac, B. Färber, P. Hinterseer, J. Kammerl, and E. Steinbach. Limitations of human 3D force discrimination. In *Human-Centered Robotics Systems*, Munich, Germany, Oct. 2006. [cited on page(s) 15]

[PUB06] A. Peer, U. Unterhinninghofen, and M. Buss. Tele-assembly in wide remote environments. In *2nd Int. Workshop on Human-Centered Robotic Systems*, pages 2–8, 2006. [cited on page(s) 2]

[RCHP07] P. Ridao, M. Carreras, E. Hernandez, and N. Palomeras. Underwater telerobotics for collaborative research. In M. Ferre, M. Buss, R. Aracil, C. Melchiorri, and C. Balaguer, editors, *Advances in Telerobotics*, volume 31 of *Springer Tracts in Advanced Robotics*, pages 347–359. Springer Berlin Heidelberg, 2007. [cited on page(s) 2]

[RLH07] D. Reintsema, K. Landzettel, and G. Hirzinger. Dlrs advanced telerobotic concepts and experiments for on-orbit servicing. In M. Ferre, M. Buss, R. Aracil, C. Melchiorri, and C. Balaguer, editors, *Advances in Telerobotics*, volume 31 of *Springer Tracts in Advanced Robotics*, pages 323–345. Springer Berlin Heidelberg, 2007. [cited on page(s) 2]

[RS07] L. R. Rabiner and R. W. Schafer. *Introduction to Digital Speech Processing*. Foundations and Trends in Signal Processing. Now Publishers, 2007. [cited on page(s) 38, 50, 54, 65]

[RYK10] J. M. Romano, T. Yoshioka, and K. J. Kuchenbecker. Automatic filter design for synthesis of haptic textures from recorded acceleration data. In *2010 IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 1815–1821, May 2010. [cited on page(s) 50, 51, 121]

[Sad07] A. El Saddik. The potential of haptics technologies. *IEEE Instrumentation and Measurement Magazine*, 10(1):10–17, Feb. 2007. [cited on page(s) 1]

[SAH79] M. R. Schroeder, B. S. Atal, and J. L. Hall. Optimizing digital speech coders by exploiting masking properties of the human ear. *J. Acoust. Soc. Am.*, 66(6):1647–1652, Dec. 1979. [cited on page(s) 19, 41]

[SGS08] Y. Suzuki, J. Gyoba, and S. Sakamoto. Selective effects of auditory stimuli on tactile roughness perception. *Brain Research*, 1242:87–94, 2008. [cited on page(s) 62]

[She92a] T. B. Sheridan. Musings on telepresence and virtual presence. *Presence: Teleoperators & Virtual Environments*, 1(1):120–126, 1992. [cited on page(s) 29]

[She92b] T. B. Sheridan. *Telerobotics, automation, and human supervisory control*. MIT Press, Cambridge, MA, USA, 1992. [cited on page(s) 1]

[SHPH09] M. Sargardia, T. Hulin, C. Preusche, and G. Hirzinger. A benchmark of force quality in haptic rendering. *Proc. 13th Int. Conf. on Human-Computer Interaction (HCI)*, 2009. [cited on page(s) 30]

[SKB+01] C. Shahabi, M. R. Kolahdouzan, G. Barish, R. Zimmermann, D. Yao, K. Fu, and L. Zhang. Alternative techniques for the efficient acquisition of haptic data. *ACM SIGMETRICS Performance Evaluation Review*, 29(1):334–335, Jun. 2001. [cited on page(s) 25]

[SLS+14] M. Strese, J. Lee, C. Schuwerk, Q. Han, H. Kim, and E. Steinbach. A haptic texture database for tool-mediated texture recognition and classification. In *2014 IEEE Int. Symp. on Haptic, Audio and Visual Environments and Games (HAVE)*, pages 118–123, Oct. 2014. [cited on page(s) 78, 80, 122]

[SOK02] C. Shahabi, A. Ortega, and M. R. Kolahdouzan. A comparison of different haptic compression techniques. In *Proc. IEEE Int. Conf. on Multimedia & Expo (ICME)*, volume 1, pages 657–660, Lausanne, Switzerland, Aug. 2002. [cited on page(s) 3, 25]

[Sri10] Mandayam A. Srinivasan. What is haptics. tutorial/whitepaper, 2010. [cited on page(s) 1]

[ST92] P. D. Spudis and G. J. Taylor. The roles of humans and robots as field geologists on the moon. *In Lunar Bases and Space Activities of the 21st Century*, 1:307–313, 1992. [cited on page(s) 2]

[Ste75] S. S. Stevens. *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects*. New York: John Wiley, 1975. [cited on page(s) 28]

[Str14] M. Strese. Content based texture retrieval. Master's thesis, Technische Universität München, 2014. [cited on page(s) 82]

[TBS+06] H. Tan, F. Barbagli, K. Salisbury, C. Ho, and C. Spence. Force-direction discrimination is not influenced by reference force direction. *Haptics-e*, 4(1):1–6, 2006. [cited on page(s) 15, 16, 125]

[TSEC94] H. Tan, M. Srinivasan, B. Eberman, and B. Cheng. Human factors for the design of force-reflecting haptic interfaces. *Dynamic Systems and Control*, 55(1):353–359, 1994. [cited on page(s) 14]

[VG75] R. T. Verrillo and G. A. Gescheider. Enhancement and summation in the perception of two successive vibrotactile stimuli. *Perception & Psychophysics*, 8(2):128–136, Mar. 1975. [cited on page(s) 46, 80]

[vH54] A. von Holst. Relations between the central nervous system and the peripheral organs. *British Journal of Animal Behavior*, 2:89–94, 1954. [cited on page(s) 9]

[VKHS09] I. Vittorias, J. Kammerl, S. Hirche, and E. Steinbach. Perceptual coding of haptic data in time-delayed teleoperation. In *Third Joint Eurohaptics Conf. and Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems (Worldhaptics 2009)*, pages 208–213, 2009. [cited on page(s) 3]

[VRH10] I. Vittorias, H. B. Rached, and S. Hirche. Haptic data reduction in multi-dof teleoperation systems. In *Proc. Int. Symp. on Haptics Audio-Visual Environments and Games (HAVE)*, 2010. Phoenix, AZ, USA. [cited on page(s) 3]

[Web51b] E. H. Weber. *Die Lehre vom Tastsinn und Gemeingefühl, auf Versuche gegründet.* Vieweg: Braunschweig, Germany, 1851. [cited on page(s) 14]

[WOZ02] Y. Wang, J. Ostermann, and Y. Zhang. *Video Processing and Communications*. Prentice Hall, 2002. [cited on page(s) 2]

[WS98a] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence:Teleoperators & Virtual Environments*, 7:225–240, 1998. [cited on page(s) 27]

[WSB03] Z. Wang, H. R. Sheikh, and A. C. Bovik. Objective video quality assessment. In B. Furht and O. Marqure, editors, *The Handbook of Video Databases: Design and Applications*, chapter 41, pages 1041–1078. CRC Press, September 2003. [cited on page(s) 75]

[Yil64] H. Yilmaz. On the laws of psychophysics. *Bulletin of Mathematical Biology*, 26:235–237, 1964. 10.1007/BF02479044. [cited on page(s) 94]

[YY94] Y. Yokokohji and T. Yoshikawa. Bilateral control of master-slave manipulators for ideal kinesthetic coupling-formulation and experiment. *IEEE Trans. on Robotics and Automation*, 10(5):605–620, Oct. 1994. [cited on page(s) 29]

[ZÖ8] U. Zölzer. *Digital Audio Signal Processing*. Wiley, 2008. [cited on page(s) 2]

[ZG80] J. Zwislocki and D. Goodman. Absolute scaling of sensory magnitudes: A validation. *Perception & Psychophysics*, 28:28–38, 1980. [cited on page(s) 28]

List of Abbreviations

Abbreviation	Description	
DoF	Degree-of-Freedom	page 24
PDDR	perceptual deadband data reduction	page 22
HSI	Human System Interface	page 24
DCT	Discrete Cosine Transform	page 25
UX	User Experience	page 30
QoS	Quality-of-Service	page 30
QoE	Quality-of-Experience	page 30
CELP	Code-Excited Linear Prediction	page 41
LPC	Linear Predictive Coding	page 41
NBN	Narrowband Noise	page 42
BPF	Bandpass Filter	page 42
FRF	Frequency Response Functions	page 44
RTAI	Realtime Application Interface	page 44
MSE	Mean Square Error	page 51
2AFC	2-alternative Forced Choice	page 59
SDT	Signal Detection Theory	page 59
LSP	Line Spectral Pair	page 65
VQ	Vector Quantization	page 65
MBQP	Model-based Quality Prediction	page 75

List of Figures

List of Tables