DEGREE PROJECT IN INFORMATION AND COMMUNICATION TECHNOLOGY,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, *SWEDEN 2016*

Real-time Coding for Kinesthetic and Tactile Signals

LIYANG ZHANG

KTH ROYAL INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL ENGINEERING

Real-time Coding for Kinesthetic and Tactile Signals

LIYANG ZHANG

Master in Wireless Systems
Date: October 2016
Supervisor: Volodya Grancharov (Ericsson),Saikat Chatterjee (KTH)
Examiner: Mikael Skoglund
School of Electrical Engineering

Abstract

The Tactile Internet is at the core of the 5G era, when the world will experience paradigm shift from content-delivery networks to service/labour-delivery ones. Systems that enable wireless communications of haptic data feature bi-directionality, high packet rate and resolution, large degrees of freedom, and above all, strict latency requirements in many applications, aggravating the shortage of wireless resources. Thus, more efficient haptic data reduction techniques are continuously summoned for. Previous studies on haptic compression mostly resort to DPCM/ADPCM plus entropy coding and perception-based down-sampling for real-time scenarios, and model-based techniques such as DCT and LP for the rest. However, with few exceptions they always segregate tactile signals from kinaesthetic signals, employing only kinaesthetic feedbacks in real-time compression experiments. In addition, these techniques are not optimized for efficient performance at scale.

This thesis project proposes a novel multi-channel real-time haptic compression system aimed at teleoperation applications with both kinaesthetic and tactile feedbacks. It consists of a lossy compression layer featuring predictive coding and a lossless layer featuring channel reshuffle and group transmission. By using different quantizer designs in the lossy layer, it abates the need for entropy coding, and leave room for future perception-based data compression modules. The lossless layer exploits inter-channel sparsity for further data reduction. The system is evaluated on a tactile texture database published by University of Pennsylvania in MATLAB. The performance measurements are in both time and frequency domain, mostly objective, but include subjective considerations as well.

Keywords: Tactile Internet, real-time data reduction, predictive coding, sparse coding, kinesthetic and tactile signal processing, multi-channel.

Sammanfattning

Haptisk kommunikation är central för de nya tjänster och användningsområden, t.ex. fjärrstyrning, som de kommande 5G-nätverken möjliggör. Haptiska system kräver av sin natur tvåvägskommunikation, ofta med hög bandbredd och upplösning, många frihetsgrader, och framför allt är latenskraven kritiska i många tillämpningar. Därför är intresset för kompressionsmetoder av haptisk data av stort intresse. Tidigare studier av haptisk kompression är mestadels baserade på DPCM/ADPCM med entropikodning samt perceptionsbaserad nedsampling för realtidsscenarier, medan modellbaserade tekniker såsom DCT och LP förekommer i övriga fall. Med några få undantag särskiljer dessa studier dock alltid taktila och kinestetiska signaler, och avhandlar endast kinestetisk återkoppling för kodning i realtid. Dessutom är dessa tekniker inte optimerade för effektiv prestanda i stor skala.

Detta examensarbete föreslår ett nytt haptiskt flerkanals-komprimeringssystem för realtid ämnat för applikationer med både kinestetiska och taktila återkopplingar. Det består av ett lager med destruktiv prediktiv kodning samt ett lager förlustfri kompression genom omstrukturering och gruppering av data från flera kanaler. Genom att använda olika kvantiseringsmetoder i det förstörande kodningslagret minskar behovet av entropikodning och lämnar utrymme för framtida perceptionsbaserade komprimeringsmoduler. Det icke-förstörande lagret utnyttjar också glesheten i datasekvenserna för de gemensamt kodade kanalerna för ytterligare datareduktion. Systemet utvärderas i MATLAB med taktil data som publicerats av University of Pennsylvania. Prestandan bedöms främst i objektiva mått i såväl tids- som frekvensdomän men även utifrån subjektiva aspekter.

Nyckelord: haptisk kommunikation, realtid, flerkanalig, kompression, prediktiv kodning, gles kodning, kinestetisk och taktil signalbehandling.

Acknowledgements

The bulk of this thesis project is accomplished at Ericsson Research, Stockholm in year 2016, between January and June. I would like to express my heartfelt gratitude to Volodya Grancharov, my supervisor at Ericsson, for his great effort and patience in guiding me through all the difficulties and confusions. Without his wise insights and criticism, the goal of the project would not have been achieved. I would also like to thank José Araújo, for his constant advice and detailed comments during the writing of the thesis report. His passion towards new technologies and solving problems has encouraged me a great deal.

The two years at KTH has witnessed the growth of me in both expertise and experience. I am truly thankful for Saikat Chatterjee, my thesis supervisor at the university, and the amiable and strict teacher for two of my courses. Last but not least, my sincere gratefulness to my parents and friends, for their selfless love and support.

Contents

Chapter 1

Introduction

1.1 Tactile Internet and the 5G Era

Like with the past four generations, the emerging 5G technology is again aimed at revolutionizing the world, by providing reliable transmission of an exponentially rocketing volume of data across the wireless network. A general vision of the 5G Era is to augment the data capacity of 4G by 1000 times and lower the end-to-end round-trip latency to the order of 1ms, about one twentieth of that of 4G [1].

On the other hand, the birth of various haptic devices and the telepresence and teleoperation (TPTO) systems since the end of the 20th century has prompted the concept of 'Tactile Internet', a network that can deliver the sense of touch, in addition to sound and vision, in an ultra-responsive, ultra-reliable and bilateral manner, such that human can indulge in physical interaction with remote or virtual environments just as if they were on site. Tactile Internet has enabled applications of significance in many fields, such as tele-surgery and surgery tutoring, remote driving, tele-assembly, tele-conference and games with higher level of human-computer interaction [2]. A typical TPTO system is considered having a master-slave structure (Figure 1.1), which consists of a human operator (or multiple, in a collaborative case) actively controlling the haptic device and a robot teleoperator (TO) in a remote place mimicking the action of human and send visual and haptic feedbacks from the environment. Ideally, the TO should behave as human commands and human should receive high-fidelity haptic feedbacks that harmonize with other modal displays (e.g. visual or audio). In a VR/AR setting, the teleoperator is replaced by a virtual avatar in the virtual environment and haptic feedbacks are computed instead of directly captured by sensors.

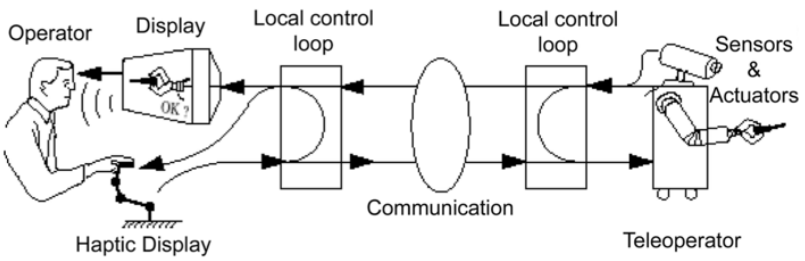


Figure 1.1: Structure of a TPTO system

However, Tactile Internet also imposes ever stringent requirements on data rates and latency in order to minimize immersion dropout. The requirements somewhat differ between the transmission of kinaesthetic and tactile feedbacks (for the definitions of the two terms, see Section 2.1.1). For instance, the former is usually used in closed-loop scenes and thus has stricter delay constraints, while the latter has been used mostly in probing, rendering and recognition tasks that have loose delay constraints, but overall, such systems feature round-trip latency of 1ms, as well as large Degrees of Freedom (DoF) and high sample resolutions [3], and we should be prepared to accommodate hundreds of thousands of such systems in the future.

As probably the most promising new application scenario in the dawning 5G era, Tactile Internet stands in the spotlight of many technology and innovation companies worldwide. For instance, Ericsson has announced 5G research in collaboration with King's College London, and the setup of a 5G Tactile Internet Laboratory. The infrastructure and protocol design of 5G should strive to be more suitable for haptic communications, i.e., use smarter bandwidth allocation and reuse, lighter packet header, low-latency modulation, etc. Meanwhile, viewing from the haptic systems' perspective, coding the data efficiently can surely alleviate the network load and in turn allow the network to serve more devices. And after all, it has been proved that for wireless telepresence applications, higher traffic means shorter lifetime of the mobile agents and higher energy consumption [4]. On these grounds, I feel both interested and incumbent to work on a novel compression algorithm design for haptic systems.

1.2 Haptic Coding: A Review

1.2.1 Development and Classification

Haptic devices and systems utilize haptic signals such as force, position, rotation and acceleration to enable physical interaction between some parts of human body (usually hand or arm) and a virtual or remote inaccessible environment. Hence, haptic interfaces are usually responsible for both measuring the haptic signals produced by human that are intended to exert on the environment and displaying the haptic feedback signals. For example, the PHANToM haptic interface first developed in 1994 at MIT that tracks the movement of a user's fingertip in terms of position and gives force feedback is one of the most renowned early commercialized product [5].

During past few decades, great effort in multimedia technology design has been devoted to creating immersive experience, i.e. computers displaying an inclusive, extensive, surrounding and vivid illusion of reality to the senses of a human participant [6]. Traditional work exploits mostly the possibilities lie in hearing and vision, and accomplishments including the Dolby stereo, 3D film technology, as well as devices like head-mounted-display (HMD) and Microsoft Kinect have been acknowledged globally. As a result, efficient coding and compression schemes for audio and visual signals have been extensively explored.

However, then people began to be no longer satiable with two sensory modalities. Cases are abundant in which mere audio and visual displays could trigger incoherence between real environments and synthesized ones in human perception that will cause immersion drop out or task performance deterioration. In fact, haptic interfaces that pro-

vide haptic drive and feedback adapted for different parts of the body and different uses are constantly being created since the 1970s, but it was in more recent time that people realize the significant role of haptic signals in addition to audio and visual signals in boosting the subjective feeling of immersion and ability to perform complex tasks [7].

Although the processing of haptic signals and that of audio and visual signals should have much in common, the former is bounded with higher restrictions and complexity because of the following factors:

1. *Haptic activities are bi-directional* – While data transmission in audio and visual applications are unidirectional, which means users interact with these systems on a passive or half-duplex basis, information flows in haptic systems are almost always bidirectional, since the touch feedbacks very much depend on how we reach out to the environment. Consequently, the amount of data to be transmitted over the network is doubled, and latency is more critical than in audio and visual systems, for the immersiveness and stability of a closed-loop system to be achieved.

2. *Human haptic perception is complicated and hard to be incorporated in haptic coding* – The trick for creating immersive experience while also transferring information economically is by studying the spatial and temporal limitation of human haptic perception and employing sensors and algorithm parameters compatible to those limits. While the temporal and spatial perception capability of human hearing and vision are much understood, e.g. 30 frames/sec is sufficient for human to see continuous movement and 4kHz sampling rate is basic for speech telephony [8], human haptic perception is rather poorly studied. While we can study audio or visual perception by solely looking at human auditory or visual systems, we cannot do the same thing for haptic perception, since it overwhelms every inch of our body. Large DoF is needed for haptic devices to mimick this dexterity and perceptual limit and resolution vary with different devices that involve participation of different parts of the body. Haptic perception comprises a variety of signals (e.g. pressure, friction, vibration, tapping, damping and so on for force signal alone). It is already much to study human perception of these genres individually, let alone the fact that they are often interweaved when forming complete touch stimuli.

The design of haptic codecs has been following two trains of thought, for online TPTO applications and non-real-time applications each: TPTO applications require 'transparency', that human operator does not perceive the presence mediating technology when experiencing the target environment [3], and thus only real-time schemes based on sample-wise processing are considered. These schemes can be subcategorized into lossy and lossless ones, based on whether the signal can be reconstructed perfectly. Lossy compression do not guarantee perfect reconstruction but can achieve satisfying performance in terms of signal-to-noise-ratio (SNR) and human perception using various downsampling and quantization methods. Lossless compression algorithms like Huffman Coding and Arithmetic Coding exploit the temporal redundancy in signals and allow for exact reconstruction. Interestingly, nearly all previous studies on haptic coding in TPTO systems assume kinesthetic feedback only. In [3] a passive extrapolative downsampling strategy is raised and evaluated on a TPTO system with velocity(feedforward)-force(feedback) architecture. In [9] Shahabi et.al analyzed distortion of several sampling methods and Adaptive Differential Pulse Code Modulation (ADPCM) individually on kinesthetic data and tested the effectiveness and efficiency when combining ADPCM and Adaptive sampling. In [10]

and [11] DPCM/ADPCM combined with quantization or lossless compression on force feedback data are studied.

On the other hand, block-based compression methods that have noticeable buffer delay like Linear Prediction (LP) and Discrete Cosine Transform (DCT) frequently employed in audio and visual compression are applicable to non-real-time tasks as in [12] and [13], or tasks with less strict delay requirements [14] [15] [16]. Here is where the tactile feedback use cases, or vibrotactile feedbacks, are employed more often, for tactile signals are believed to be similar to speech signals, and are less seen in teleoperation scenarios.

However, in this thesis study, we try to make our proposed compression algorithms as general-purpose and scalable as possible. This means we strive to address the TPTO scenario for both kinesthetic and tactile signals, and optimize the algorithm on large number of inputs. With 1ms latency constraint in mind, we only consider sample-based coding schemes.

While most of the abovementioned literature designed the compression algorithms with no constraint in the subjective human perception dimension at all or only use it as a performance measurement in the analysis phase, many other studies exhibited enormous interest in the huge data reduction potential lying in human perception limit. Hinterseer et.al were the first to apply a perceptual deadband (PD) sampling approach to compression of velocity and force signals, in which an incoming sample is only transmitted when it is considered perceptually different from the previously transmitted one [17]. They later motivated a first-order linear-prediction-based PD method and a multi-DoF PD method as extensions [18]. In successive studies the simplest form of PD is modified to boost performance for different signal types and take care of the passivity requirement in the closed-loop teleoperation systems [19] [20] [21] [22] [23]. In [24] Kammerl et.al got inspiration from psychophysical study on dynamic haptic perception limit and came up with a hand-velocity-dependent PD algorithm for task-oriented contexts with success. Sakr et.al use a least-squares prediction-based PD scheme in combination with uniform quantization and adaptive Golomb-Rice codes for haptic data reduction in tele-monitoring systems [25]. Studies [12] [26] conducted studies using vibrotactile texture signals instead of kinesthetic signals, and both of them proposed a perception-based compression scheme that doesn't go in the direction of time domain PD. In a recent article [27], cutaneous feedback is experimented on using PD with demonstrated feasibility.

1.2.2 Reflection on Previous Work

Admittedly, haptic coding is still a rather under-exploited area in many dimensions, despite the fruitful research over the years. In the course of literature reading for this thesis study, several of the dimensions have been clearly noticed:

1. *Tactile feedbacks are rarely considered in teleoperation scenario, while the forward path is completely forgotten.*

   Like mentioned in previous section, kinesthetic signals are much more frequently studied than tactile ones for real-time applications. [27] is the only work we find that evaluate PD compression on cutaneous signal. Another thing is that most research focus on experiments with feedback signals, not signals in the feed-forward paths, regardless of the fact that their compression will certainly affect the feedback signal quality. Moreover, some signal types like force is more frequently used for

perceptual compression study than others like acceleration and position signals. While the Just Noticeable Difference (JND) for force signal has been extensively researched and reached an consensus independent of body site and test condition [18] [28], we have faint ideas of what should be the JND for acceleration, position or velocity.

The haptic limit of human can be very complicated in the sense that it is dependent of body sites, passive/active exploration, visual/audio complementary information, mediating tools, subjects' concentration level [29], etc. However, past studies controlled these above, especially when it comes to the visual/audio information. Most studies cannot exempt visual/audio modalities completely from the haptic perceptual experiments, since the virtual environment must be displayed to the participants through the PC in many tasks. Since human beings are expert at integrating multiple modalities in an optimal way, it is a very demanding to guarantee that the compression solution we designed for one type of tasks will be suitable for others as well.

2. *Research on real-time non-perceptual compression has little diversity.*

   To the best of our knowledge, for lossy compression only ADPCM/DPCM with fixed scalar codebooks have been proposed and only Huffman Coding and Golomb-Rice Coding have been attempted for the lossless compression part so far. Hinterseer et.al have pointed out that methods that use differential coding followed by entropy coding in previous work can suffer from a bad packet header to payload ratio [18] and are not entirely suitable for TPTO systems. Therefore it is reasonable to encourage more diversity to see if any improvement can be made.

3. *Studies that take non-perceptual and perceptual paths are mostly isolated.*

   Studies that use agile quantization and sampling methods do not employ much study on perception when designing algorithms and those that design from a PD point of view seldom consider smarter quantization methods or lossless compression. Although this may be attributed to the fact that outcomes from these two paths are judged and criticized by different criteria (distortion rate and perceptual discriminability respectively), we believe there is room for merging the two branches for even better efficiency, since they have their own advantages and are not in alternative positions.

4. *Existing studies seldom explore relevance among input channels.*

   Just think of a haptic glove with distributed sensors on the palm and figures, not only the force signals in three dimensions at each sensing point can be coded and compressed more efficiently, but also the sensing locations are likely to be correlated to each other in signal variation. There might be another leap in haptic compression algorithm design if we start to take channel-wise relevance into consideration.

5. *The subjective measurement of perception-based compression is not standardized.*

   Researchers usually design their own psychophysical experiments to evaluate the transparency performance of their compression algorithms based on different subjective grading standards. For instance, in [24] and [27] participants are required to

tune the deadband parameters freely and report the point when a disturbance or difference in feeling is detected; in [13] researchers adopt the grading scale recommended by International Telecommunication Union (ITU) for subjectively assessing the impairment of audio/visual content, which allows subjects to choose from four subjective statements, i.e. 'same' 'possibly same' 'possibly different' and 'different' when comparing the uncompressed and compressed signals; [21] and [23] use a measure similar to the ITU style, but replace the four statements with a grading scale ranging from 0-100, with 0 representing 'strongly disturbing/different', and 100 representing 'no difference'; a three-interval forced choice (3IFC) paradigm is used in [17] [18] and [26], where participants go through several repetitions of experiments asking them to pick out a different signal among a group, correct answers lead to a decrease in the deadband coefficient/ masking threshold and vice versa; whereas in [20] and [22] objective criteria like SNR and reduction rate are used and no subjective experiments are conducted. Although the experiments in all of the studies above try their best to make the process more impartial through careful choice of participants, elimination of excessive audio/visual information, training phase, repetition of experiments and other scientific research methodology, I found discussion of neither the convincing power of each method nor the superiority of one method over another in haptic perception context.

1.3 Objectives

Ideally we wish to make an improvement on all five aspects mentioned above through our efforts. However, due to the time limit of this thesis project and immature infrastructure, we will leave the fifth aspect, i.e., the choice and standardization of subjective measurement methods for haptic codecs, for future research. Our main objectives are listed as follows:

- Design a compression system suitable for TPTO systems with both kinesthetic and tactile feedbacks, and assess its performance on both types of signal.
- Explore multiple real-time lossy compression techniques similar to those employed for audio and image processing, and make conclusions about their performance on haptic signals.
- Make the system scalable to and efficient for systems with large number of input channels.
- Design the system mainly using non-perceptual techniques, but in a way that is compatible with perceptual compression plug-ins. The performance evaluation are mainly based on objective indexes, but include subjective considerations as well.

The round-trip latency constraint throughout this study is 1ms, and signal sampling rate is kept at 1kHz.

1.4 Outline

The remainder of this thesis report is structured as follows. Chapter 2 is a comprehensive background study of haptic technology and the mathematics inside existing haptic data

real-time coding schemes. Especially, we incline our vision towards human perceptual capacity perspective, which we believe is the foremost and ultimate principle underlying a successful compression system.

Chapter 3 is the description of the proposed compression system. First the overview of the whole system is given, followed by details of subsections of the system.

Chapter 4 introduces the database we use for evaluate the compression system.

Chapter 5 presents all experiments and their corresponding results for testing the system.

And finally, chapter 6 summarizes over the entire design and analysis process, and gives retrospective and future thinking on haptic coding.

Chapter 2

Background

2.1 Haptic Perception and Haptic Technology

2.1.1 Characteristics of Human Haptic Perception

The touch modality of human is managed by somatosensory system and can be subcategorized into kinesthesis, tactile/cutaneous sense and proprioception depending on the location of the sensory receptors involved. When it comes to engineering we are content with the first two types of senses, where the stimuli come from outside of the body. Quoting Loomis and Lederman's clear-cut boundary line of the two types of senses in [29], tactile/cutaneous sense should solely rely on cutaneous stimuli processed by mechanoreceptors and thermoreceptors in the skin, such as recognition of a Braille character, or feeling the temperature or pinching of a needle; whereas kinesthetic sense refers to perception through mechanoreceptors in muscles, bones and joints, in which if not completely so the cutaneous stimuli should only serve as an indication of contact between human and the object, such as judging the size of a ball by grabbing it or detecting a wall by reaching out an arm when blind-folded and with gloves on. Haptic perception is then defined as 'perception in which both cutaneous sense and kinesthesis convey significant information about distal objects and events'.

Human haptic modality, like hearing and vision, has spatial and temporal sensitivity, as well as the discriminating power of signal intensity. The spatial acuity of skin is found to be generally better than the ear's and poorer than the eye's, and dependent of body sites (see Figure 2.1). That accounts for the fact that finger pads tend to be used to explore fine object information like surface textures much more than other body parts. The temporal acuity of skin is considered to be better than the eye's and poorer than the ears, and human can perceive stimuli frequency up to the order of 1kHz with a peak sensitivity at around 300Hz [8]. That is why 1kHz packet rate, which is higher than the frame rate of most videos and lower than the sampling rate of narrowband telephony, is considered satisfactory in most haptic applications.

The Just-Noticeable-Difference (JND), also named as difference threshold, is a frequently used term to describe human discriminating power of the variation of sensory stimuli strength. Namely it means the minimum amount of difference in strength perceptually detectable in at least 50% of trials. Weber's Law states the fact that the size of the JND appeared to be proportional to the initial stimulus magnitude for many sensory modalities, which is expressed as:

$$\Delta I/I = k \tag{2.1}$$

where ΔI is the magnitude of JND and I is the magnitude of stimulus. JND is sometimes represented by ratio since in most cases it is approximately a constant within a specific task modality, although it can fluctuate among individuals.

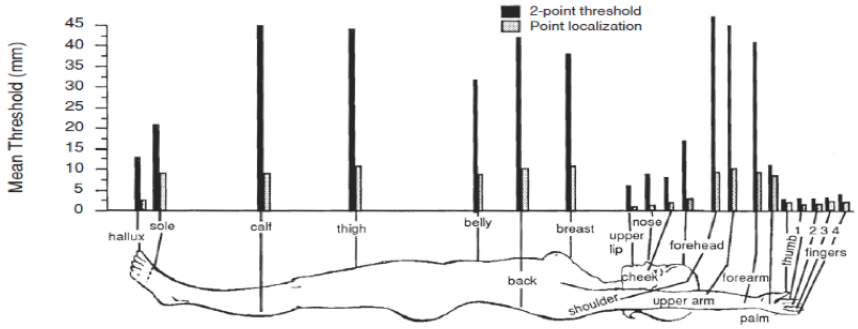


Figure 2.1: Spatial haptic acuity across human body [30]

For haptic modality, there have been cognitive and psychophysical studies proving the existence of JND in both intensity and frequency dimension. The former inspires the main branch in perception-based haptic compression schemes, the perceptual deadband approach, which will be covered more in section 2.2. [8] gives experimental values for JNDs of pressure and joint angles (see Table 2.1), and [31] proposes JND ratios of multiple haptic related properties (see Table 2.2). However, these JND ratios are mostly obtained in passive and pseudo-static conditions, and thus may not be applicable to circumstances with active haptic participation and multi-modal feedbacks. Also, studies on JNDs for cutaneous feedbacks are rare. [27] is one of the few who apply deadband compression to cutaneous force feedback, but didn't propose a JND range.

Body Site	**Contact Area (cm^2)**			**Average Joint Angle JNDs**
	1.27	**5.06**	**20.27**	
Elbow (Volar)	16.7%	6.2%	4.0%	2.0°
Elbow (Dorsal)	11.3%	5.2%	3.3%	
Wrist (Dorsal)	18.8%	4.4%	--	2.0°
Shoulder (side)	--			0.8°
Shoulder (front)				0.8°
Overall Average JND	15.6%	5.3%	3.7%	--

Table 2.1: Joint angle JNDs for some body sites

Meanwhile, there are interest findings of JNDs for vibrotactile signals. Both JNDs in amplitude (10% - 20%) [32] and frequency dimension (around 18% regardless of amplitude) [33] have been found, and frequency is believed to be the dominant source of information in the perception of vibrotacile signals. That gives incentive to the design of

Physical property	JND [%]	Experimental Conditions
Force	7±1 ca. 15 ca. 10	arm/ forearm arm, static force arm/ forearm
Movement	8±4	arm/ forearm
Position	8±2	arm/ forearm
Stiffness	23±3 8	arm/ forearm pinch-fingers
Viscosity	34±5 13.6±3	arm/ forearm, >20Ns/m pinch-fingers, 120Ns/m
Inertia	21±3.5	pinch-fingers, 12kg

Table 2.2: JNDs for some other haptic properties

perception-based compression schemes based on frequency masking and DCT for vibrotactile signals instead of amplitude JND as in article [13] and [26].

Moreover, there is an obvious asymmetry between human haptic control and haptic perception [8]. For example, while we are able to perceive high frequency vibrotactile stimuli, the frequency of haptic signals we are capable of generating is upper bounded by 20Hz-30Hz. However, our force control resolution is generally better than our perception JND ratio. This can imply that in the design of deadband-based compression algorithm for a bilateral haptic application, signal inputs from the two directions should be treated somewhat differently.

2.1.2 Haptic Devices

In order to create life-like experience or allow more natural human-computer interaction (HCI), multi-modal display is preferred, along with high quality of each modality to make participants feel on site. We can find the development of film technology a specific example of this. It started with only silent, monochrome films, followed by films with sound and color, to modern day 3D films which provide stereo sound and visual effects, and even 4D films which combine 3D films with various physical effects like vibration enabled by the seats or smoke, rain or smell in the hall. The rain effect in 4D films is one 'cheating' example of haptic display and neither extensible nor adjustable. The vibrating chair, on the other hand, counts as haptic display, for it exerts controllable force onto human body through electromechanical actuators and motors. Another daily example would be the prevalent use of touchpad on laptops and touchscreens and vibration in mobile phones. They either use touch events of human fingers (click, double click, swipe, long press, etc.) as commands for PCs or mobile devices or use different vibration feedbacks as alerts to users when muted mode is on.

However, the above two haptic applications belong to those rare cases where pure haptic input or output is sufficient. As is mentioned earlier, haptic perception is usually associated with active exploration of the environment and receiving haptic feedbacks accordingly. The feedbacks can serve as cues for familiarizing surface or geometric properties of the object in contact, or reassuring confirmation for your probing [34]. On the other hand, combination of haptic and visual feedbacks can reduce task completion time, error incidence and excessive force application in many situations in which human inter-

act with virtual or remote environments [7] [35]. Therefore, haptic devices we discuss in this thesis are always bi-directional.

Existing haptic devices can be divided into kinesthetic and cutaneous devices, in the same manner as we classify haptic perception. Kinesthetic devices use tools to mediate force and positions that are further passed onto human operators. Popular general-purpose kinesthetic devices include manipulandum type devices like the PHANToM series (Phantom Desktop, Phantom Omni, Phantom Premium, now named 3D Systems Geomagic series)[1], Force Dimension Omega and Sigma series[2], Falcon[3] and Virtuose[4]; Grip/Grasp type devices like CyberGrasp[5] and exoskeleton type like KINARM Exoskeleton Lab[6]. [36] proposes a reconfigurable module-based wooden haptic device called WoodenHaptics which is a manipulandum type kinesthetic device much cheaper than most of its commercial counterparts and whose haptic fidelity is comparable to Phantom Desktop. It also allows users to add sensors and vibrotactile actuators to improve perception of textures. The device group in Ericsson employ quite a few of these WoodenHaptics for research use.

The kinesthetic devices mentioned allow different ranges of motion from finger joint movement to full arm movement pivoting at shoulder. Other important specifications that we may be concerned with include work space size, resolution, peak force allowance, DoF and update rate and so on. Table 2.3 gives a summary of those specifications of some devices.

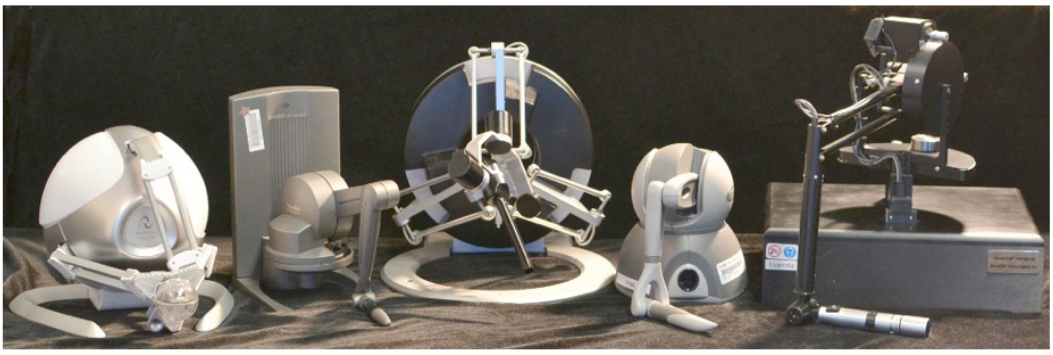


Figure 2.2: Common spatial haptic devices. From the left: Novint Falcon, Phantom Desktop (now 3D Systems Geomagic Touch X), Force Dimension Omega, Phantom Omni (now 3D Systems Geomagic Touch), and Phantom Premium 6-DOF (now 3D Systems Geomagic Phantom Premium)

Cutaneous interfaces are different from kinesthetic ones in the sense that they apply often distributed forces or displacements directly to the skin, and enable the feeling of static surface pattern, roughness, temperature and so on which cannot be perceived

[1]Detail information available at: `http://www.geomagic.com/files/1714/4842/0629/Haptic-Device_EN_Web.pdf`

[2]Detail information available at: `http://www.forcedimension.com/products`

[3]Detail information available at: `http://www.novint.com/index.php/novintfalcon`

[4]Detail information available at: `http://www.haption.com/site/index.php/en/products-menu-en/hardware-menu-en/virtuose-6d-menu-en`

[5]Detail information available at: `http://www.cyberglovesystems.com/cybergrasp/`

[6]Detail information available at: `http://www.bkintechnologies.com/bkin-products/what-is-kinarm-lab/`

Figure 2.3: $CyberGrasp$[5] (left), KINARM $Exoskeleton7$[6] (middle) and WoodenHaptics [36] (right)

in absence of receptors in the skin. Thus they can come in a variety of forms such as indentation-type shape displays, wearable devices with force or vibration actuators in contact with skin [37] and lateral skin stretching devices. For instance, CyberTouch glove is one cutaneous device equipped with 6 vibrotacile actuators on fingers and the palm that can produce vibration up to 125Hz[7]. In [38] Hayward et al. design a membrane-like cutaneous display device that produces lateral skin stretch on fingerpads realized by piezoelectric actuator arrays, while in [39] a stylus-like skin stretch device with 4-bar crank-slide mechanism is developed, and it has been proved by many that shear skin deformation can serve as a satisfactory substitution for kinesthetic feedbacks in a many cases including interpretation of surface texture, pressure and stiffness. Also notably in [40], researchers proposed a tactile display called T-pad phone[8] that leverage ultrasonic frequency vibration to modulate friction between human finger and the glass, therefore creating the sensation of different textures. It is currently being used at Ericsson for texture rendering related research.

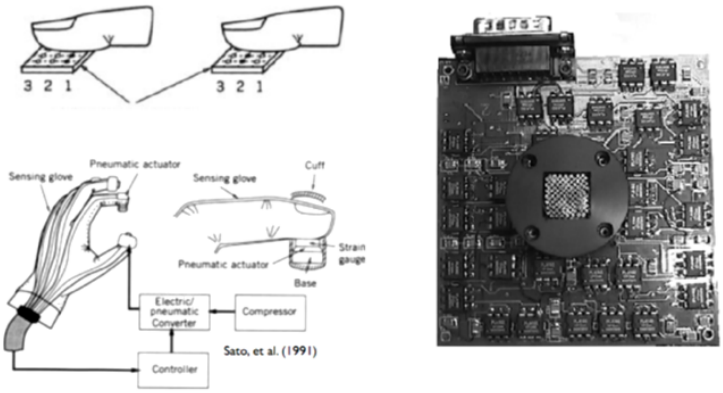


Figure 2.4: Shape display (upper left), device with pneumatic actuators at fingerpads (lower left) and skin stretching device by Hayward et al.[38]

[7]Detail information available at: www.cyberglovesystems.com/cybertouch/
[8]Detail information available at: http://www.thetpadphone.com/

	Workspace	Position Resolution	Max Force	DoF	Refresh Rate (Hz)
Phantom Omni	$160 * 120 * 70mm$	$0.055mm$	$3.3N$	6 (pos) 3(force)	$1k$
Falcon	$102 * 102 * 102mm$	$0.006mm$	$8.9N$	3(pos) 3(force)	$1k$
Omega.7	$\phi 160 * 110mm$ (translation) $240 * 140 * 180deg$ (rotation) $25mm$(grasp)	$< 0.01mm$ $0.09deg$	$12N$ $\pm 8N$(grasp)	7(pos) 4(force)	up to $4k$
CyberGrasp	1m radius from actuator module	$< 1deg$	$12N$/finger	22/hand (pos) ; 1/finger (force)	90
KINARM Exoskeleton	119.4cm diagonal plane	$0.0006deg$	$12Nm$ torque	2	$2k$(control) $1k$(data acquisition)
WoodenHaptics	Adjustable 3D workspace	–	$9.9N/19N$	3	$> 1k$

Table 2.3: Specifications of some kinesthetic devices

2.2 Real-time Compression for Haptic Signals

Real-time compression algorithms introduce no buffer delay and unnoticeable computation delay by performing sample-based compression and turning away from time consuming but highly adaptive machine learning process. Since we intend to build compression system for real-time haptic display systems, our goal is less than 1ms latency. Therefore, frame-based compression schemes such as DCT or Wavelet Transform that have been practiced before on haptic signals will not be considered in this study. Depending on whether the compression is perfectly irreversible, real-time compression are lossless or lossy. Real-time lossy compression can be achieved either through quantization or resampling, and lossless compression is mainly entropy coding that exploits redundancy.

2.2.1 Quantization

Scalar Quantization (SQ)

SQ is a basically a mapping $P : \mathbf{R} \rightarrow I$, where $I = \{0, 1, ..., N-1\}$ is a finite set of binary code words to be transmitted representing original symbol x that contain much more than N values. There are N representation levels $\{y_0, y_1, ..., y_{N-1}\}$ that correspond to the codewords, and $(N+1)$ decision points $\{x_0, x_1, ..., x_N\}$ that set the boundaries of the decision regions. If x falls in (x_k, x_{k+1}), it will be assigned level y_k (x_0 and x_n are usually chosen to be $-\infty$ and ∞ respectively). The decision points are placed midway between the representation points so that a symbol will be assigned to the quantization value closest to its value. For a given set of data, the average quantization error is obvi-

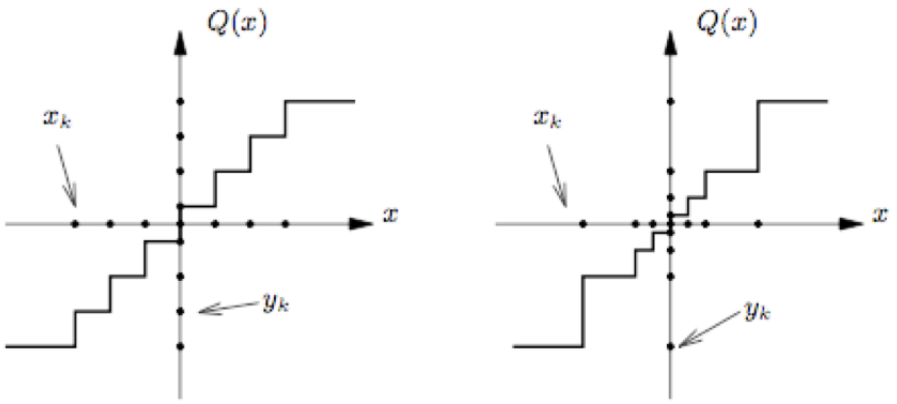


Figure 2.5: Uniform (left) and non-uniform (right) SQ

ously related to the number and position of the representation points. If the behavior of the data is prior knowledge, then intuitively we would place more levels at regions with dense data. A scalar quantizer with equal step size is called uniform scalar quantizer, but non-uniform scalar quantizers are also common (see Figure 2.5).

Vector Quantization (VQ)

VQ, on the other hand, is a mapping $P : \mathbf{R}^M \rightarrow I(M > 1)$, where N binary code words in I represent the N vector representations $\{y_0, y_1, ..., y_{N-1}\}$, $y_i \in \mathbf{R}^M$. Correspondingly, there are also N decision regions of equal or different sizes in the M-dimensional space and each incoming vector data $x \in \mathbf{R}^M$ will be assigned its nearest representation level y_j in the space by:

$$j = \arg_j \min d\left(x, y_i\right) \tag{2.2}$$

where $d(.)$ is the Euclidean distance. Figure 2.6 illustrates a 2-dimensional vector quantization with uniform and non-uniform resolution, respectively . For haptic signal compression, there is currently no study using VQ to the best of our knowledge, since they did not explore relevance among multiple channels. However, if there is indeed relevance among inputs, VQ would be a better choice.

Lloyd-Max Algorithm

The Lloyd-Max algorithm is used in quantizer design for approximating the optimal regions (R) and representatives (C) in the MSE sense for a particular dataset:

$$\left(R^{opt}, C^{opt}\right) = \arg_{R,C} min \int_{-\infty}^{\infty} f_X\left(x\right)\left(x - Q\left(x\right)\right)^2 dx \tag{2.3}$$

where $C = \{y_0, y_1, ..., y_{N-1}\}$. Generally, it works in an iterative manner that follows three steps [41]:

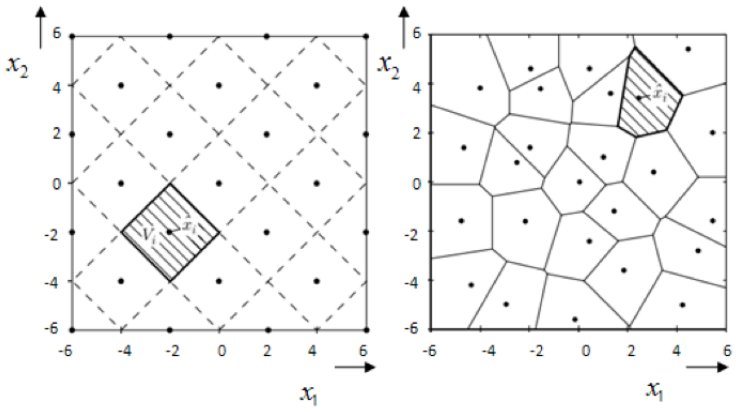


Figure 2.6: 2-dimensional vector quantization: uniform (left), non-uniform (right)

1. Given current codebook $C^{(t)} = \left\{y_0^{(t)}, y_1^{(t)}, ..., y_{N-1}^{(t)}\right\}$, obtain the optimal regions $R_i^{(t)} = \left\{x : \left(x - y_i^{(t)}\right)^2 \leq \left(x - y_j^{(t)}\right)^2, j \neq i\right\}, i = 0, 1, ..., N-1$;

2. Update the codebook according to: $y_i^{(t+1)} = \frac{\int_{R_i^{(t)}} f_X(x) x dx}{\int_{R_i^{(t)}} f_X(x) dx}, i = 0, 1, ..., N-1$;

3. Repeat step 1 and 2 until a terminate criterion is met: $\max_i \left(y_i^{(t)} - y_i^{(t+1)}\right)^2 \leq \epsilon$, where ϵ denotes the threshold.

Obviously, the probability density function (PDF) of the data f_x must be specified in advance. LM quantizers tend to give more representatives at regions where the data is denser, so naturally it will have an appeasing effect on the histogram of quantization indexes. When this happens, the entropy of the quantization index information will be larger than with uniform quantizers and a successive entropy coding will be less necessary (see Section 2.2.3).

2.2.2 Differential Pulse Code Modulation (DPCM)

DPCM is a non-perceptual, predictive coding method, and is with some exceptions a lossy scheme. It has been selected for the compression of kinesthetic feedback signals in [9] [10] and [11], and its diagram is shown as in Figure 2.7. Under the assumption that neighbouring samples are correlated, it takes the prediction error e, the difference between the current sample s and its predicted version $\hat{s}$ for quantization. $\hat{s}$ is calculated based on past reconstructed samples using identical predictor at the encoder and decoder, and reconstructed sample s' is output at the decoder. The predictor is in fact a recursive linear filter of order of the output s':

$$\hat{s}_n = s'_n - e' = \sum_{k=1}^{p} a_{n-k} s'_{n-k} \tag{2.4}$$

where n denotes the current instant and a_i are the optimal coefficients for the predictor. Therefore, prediction error e can be written as:

$$e = s_n - \hat{s}_n = s_n - \sum_{k=1}^{p} a_{n-k} s'_{n-k} \tag{2.5}$$

When $p = 1, a_{n-1} = 1$, the quantized prediction error e' is just the difference between current output sample and the one preceding it. As for the quantizer, only uniform scalar quantizer has been used for the DPCM compression on haptic signals, but usually cascaded by entropy coding as a further compression. Delta Modulation (DM) is a special case of DPCM where the predictor is of first order and the quantizer is 1-bit, and has been used in voice telephony applications.

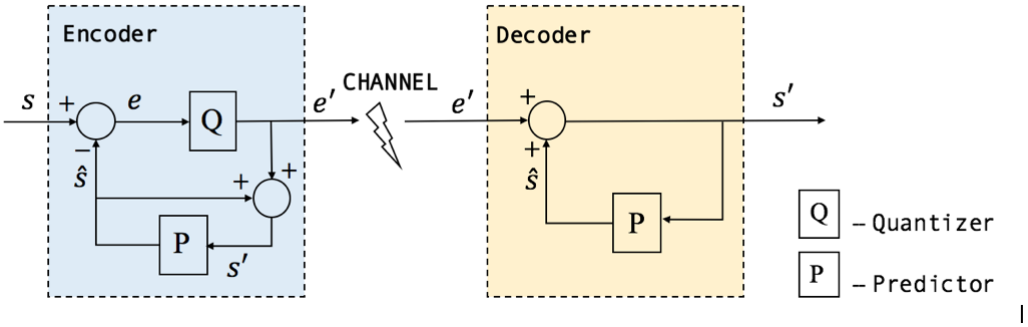


Figure 2.7: Schematic of DPCM

Previously widely employed for audio, image and video coding, DPCM is efficient for handling signals that are correlated at least at neighbouring positions. While that has been long known to us to hold truth for speech, music, image or video, it has not been much validated on haptic signals. The studies mentioned above all affirm the validity of first order DPCM on force signal. In [11], the second order DPCM is reported to outperform first order DPCM on fast moving force feedback signals. In [9], an adaptive DPCM (ADPCM) coding is used which can adapt the quantization step size automatically according to speed of motion. The optimal predictors and quantizers for different haptic signals remain to be explored by more studies.

2.2.3 Entropy Coding

In information theory, entropy describes the amount of information contained in a signal, which can be calculated as:

$$H(s) = -\sum_{i=1}^{n} p(s_i) \log(p(s_i)) \tag{2.6}$$

where n is the total number of symbols in signal s, and p stands for probability. It can be proved mathematically that the entropy of a signal will be at peak with a perfectly even-distributed histogram of all values. And when entropy is below its maximum, the signal can be losslessly compressed using entropy coding. For DPCM, when the quantizer is a uniform scalar quantizer, the subject quantized value e' will have a digitized distribution similar to e, which is roughly some zero-mean symmetrical distribution. Therefore, entropy coding is almost always concatenated with DPCM as a further data reduction.

Popular entropy coding schemes that have been applied to haptic signals include Huffman Coding and Golomb-Rice Coding [10] [11].

Huffman Coding assigns code with different lengths to each symbol, based on their statistical probability. Symbols with larger probabilities of occurrence are supposed to use shorter codewords than those that does not commonly appear. It adopts a bottom-up tree approach to choose the representation for each symbol such that a prefix code in which one codeword representing a symbol is never a prefix of the one representing any other symbol. Huffman coding is considered optimal among all entropy codes when a source consists of unrelated symbols with a known probability distribution, and thus often follows after DPCM, given that the predictor in DPCM will lead to prediction errors with little correlation.

Golomb-Rice Coding divides a symbol by a constant M to get the quotient q and remainder r. Later q is encoded using unary coding and r is encoded using truncated binary coding, and the two parts are concatenated to form the codeword for the symbol. The codeword length is only dependent on the magnitude of the symbol and constant M, and therefore is highly suitable for situations in which small values occur much more likely than large values.

In general, entropy coding is based on the assumption that the histogram of the input data is highly uneven. If it is not the case, the entropy of the data would be rendered high and there would be little motivation in using entropy coding for reduction. In this thesis, a different path which adopts histogram equalization plus sparse coding will be used instead of entropy coding, which will be described in Chapter 3.

2.2.4 Perceptual Deadband (PD)

Unlike the previously mentioned methods, PD is a class of adaptive sampling scheme that takes advantage of the perceptual limit of human. As its name suggests, it defines a deadband whose width is proportional to the magnitude of current value using Weber's law (Equation 2.1), the changes inside which will unlikely be felt. The principle is that current sample will only be transmitted if its value exceeds the deadband of the last transmitted sample, otherwise only zero-hold copies or predictions from the previous sample will be output at the decoder (Figure 2.8). PD can also be extended to vector samples of higher dimensions.

In haptic compression, many studies have been devoted to time domain PD. Besides the conventional zero-hold PD first proposed in [17] for velocity and force signals in bilateral TPTA system, numerous studies have improved the PD approach based on dif-

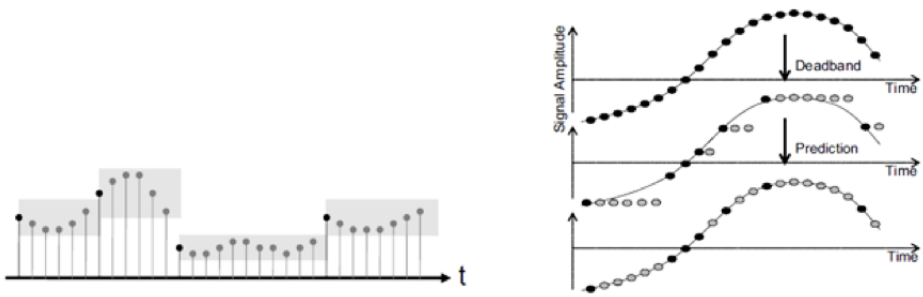


Figure 2.8: Illustration of PD (left) [18] and first-order predictive PD approach (right) [19]

ferent prediction algorithms. Among them, first-order prediction PD [18] is the simplest form of prediction which states that sample is only transmitted when the difference between the current value and the linearly predicted value of the current value is out of the deadband of the predicted value. The rest of sample values between the two updates are linearly extrapolated locally at the decoder using the slope information from two latest updates.

Apparently, the underlying implication of this prediction is that the change of the signal is unidirectional within a short duration, and therefore it will only outperform zero-hold PD method on smoothly-varying data.

There are other considerations on haptic PD. Human operator is not simply a passive receiver of information in a haptic TPTA system but also an active participator, and psychophysical experiments have validated that the perceptual ability of human is related to their activeness and level of attention. Therefore, Weber's law may not be applicable to all cases and the PD ratio can be purposely designed to be variable [24]. In this thesis, the PD method for compression will not be incorporated explicitly. Instead, we make it possible to be embedded into the system we propose that mainly adopts non-perceptual compression.

2.3 Performance Evaluation Methods

As haptic compression schemes can be either non-perceptual like the conventional DPCM method or perceptual like the PD method, the evaluation of performance should also divide between objective and subjective measurements. The former looks at only calculations when judging the system, while the latter relies on subjective reports from actual human operators and is hard to quantize.

2.3.1 Objective Measurements

Signal Distortion

Signal-to-Noise Ratio (SNR) is one of the most common methods to measure signal distortion, which basically is the ratio between the energy of the original signal and that of the difference between the original and the reconstruction expressed in decibels:

$$SNR_{dB} = 10\log_{10}(\frac{P_{signal}}{P_{noise}}) = 10\log_{10}(\frac{\sum_{i=1}^{N} s(i)^2}{\sum_{i=1}^{N}(s(i)-\hat{s}(i))^2}) \tag{2.7}$$

where s and $\hat{s}$ represent original and reconstructed signal in time domain, respectively, and N is the total number of samples. While equation 2.7 only represents the average signal distortion over a long time, we can always chunk up the signal and calculate the average SNR on segments (SegSNR):

$$SegSNR_{dB} = \frac{1}{N}\sum_{j=1}^{N} 10\log_{10}(\frac{\sum_{i=1}^{M} s(jM+i)^2}{\sum_{i=1}^{M}(s(jM+i)-\hat{s}(jM+i))^2}) \tag{2.8}$$

Where N is the number of frame and M is the frame length. In this way, the short-term variation in signal distortion will also influence the overall SNR.

Although we care much about time-domain signal distortion of haptic signals, the frequency domain matters a lot for vibrotactile signals in particular. In Section 2.1.1 it is mentioned that frequency component plays the dominant role when perceiving rapidly fluctuating vibrotactile signals, and deadband in the frequency dimension also exists. Since the system should also be suitable for vibrotactile signal transmission, spectral distortion is something we should be cautious when judging the system. One measurement of spectral distortion is known as the log-spectral distance (LSD):

$$D_{LS} = \sqrt{\frac{1}{2\pi}\int_{-\pi}^{\pi}[10\log_{10}\frac{P(\omega)}{\hat{P}(\omega)}]^2 d\omega} \tag{2.9}$$

where $P(\omega)$ and $\hat{P}(\omega)$ are the power spectra of the original and reconstructed signal, respectively. The larger the distance is, the more the distortion in the frequency domain. Again, LSD can also be calculated on a segment basis and average over all segments.

In this thesis, we use SNR results as the main criterion for signal distortion, since we are designing a real-time compression system that can exert a direct influence on SNR. But we will also calculate and compare spectral distortion and check on the shape of the spectra. Then we should be able to tell the SNR threshold for our system at which the spectral distortion is also acceptable.

Compression Efficiency

The efficiency of a compression algorithm lies in its power of data reduction. Usually compression ratio (CR) or data rate savings (DRS) is used to describe such power:

$$CR = \frac{\textit{uncompressed data in bps}}{\textit{compressed data in bps}} \tag{2.10}$$

$$DRS = 1 - \frac{1}{CR} \tag{2.11}$$

Apparently, there is a trade-off between compression efficiency and signal distortion. In networks with high traffic, we tend to weigh efficiency over distortion, while other times we focus more on reconstruction quality. Therefore, people take care of both specifications by drawing SNR-DRS coordinates and find the proper parameters for compression systems in regions that suit their specific circumstances.

2.3.2 Subjective Measurements

Despite the fact that objective measurements offer a clearly quantifiable evaluation of the reconstruction quality of compression algorithms, these numbers cannot have our entire faith. As the ultimate goal of designing a haptic codec or any audio or visual codec is to let human operators barely feel any degradation in the signal compared to an uncompressed version, it makes more sense to have the final say entitled to human. Therefore, subjective tests are widely recognized as the most reliable way for evaluating codecs, especially for the low bitrate ones.

There has been quite many standardized subjective tests for audio quality, such as MUSHRA (Multiple Stimuli with Hidden Reference and Anchor) defined by ITU-R recommendation BS.1534-3[9] and MOS (Mean Opinion Score) specified by ITU-T P.800[10]. But

[9]Detail information available at: `http://www.itu.int/rec/R-REC-BS.1534-3-201510-I/en`
[10]Detail information available at: `http://www.itu.int/rec/T-REC-P.800-199608-I/en`

haptic coding is a relatively new field with no such standards available. In fact, a subjective test standard for haptic systems would be more complex due to the bilateral nature of and drastic differences among haptic activities. So far, researchers have used diversified methods for subjective tests inspired either by ITU standards for audio codecs or by other psychophysical experiments, as covered in Section 1.2. In this thesis, subjective studies are not conducted. Only objective measurements are used complemented by direct observation of waveform.

Chapter 3

The Compression System

3.1 System Overview

The compression system proposed is a sample-based source coding scheme that exploits both intra-channel and inter-channel redundancy. It is comprised of two major parts: an outer layer of differential pulse coding or pulse coding that reduces resolution of each input and an inner layer of sparse coding that rearranges the channels and transmits the bit streams in an efficient way. The two parts are symmetrical about the transmission channel. The DPCM/PCM coding part is lossy while the sparse coding part is lossless. A schematic diagram of the whole system is presented in Figure 3.1 below.

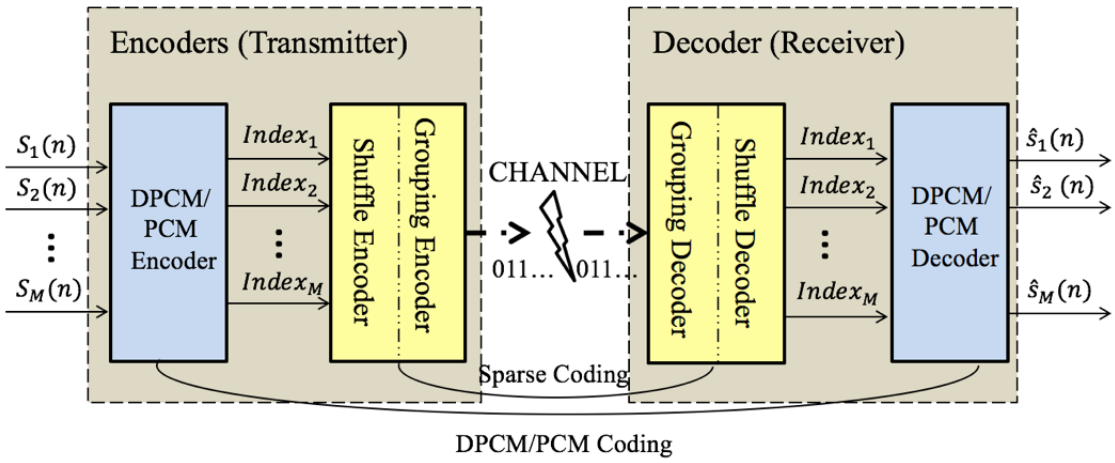


Figure 3.1: Schematic diagram of the proposed compression system

The blue modules are the DPCM/PCM coding part while the yellow modules are the sparse coding part. The S_1 through S_M on the left side are M signal inputs that can be interpreted as all data inputs on a haptic device, for example the x, y and z direction force, position, velocity or acceleration signal. In our experiments we treat each direction of each type of signal as one channel because we are using scalar quantizers, but one can always combine some channels as one if vector quantization is found to be effective in the outer layer coding in the future. The M samples come out from the DPCM/PCM encoder as M binary representations of the quantization indexes and they are again fed into the sparse encoder which takes two steps, shuffling and grouping, to take advantage

of inter-channel redundancy and transform those binary notations into a more concise single bit stream. Since we do not consider channel coding in this study, the transmission channel between the transmitter and the receiver is assumed to be ideally error-free and transparent, which means the data immediately on two sides of the transmission channel are identical and transmission latency is neglected. Therefore, the bit stream for one sample interval arrives at the receiver's sparse decoder exactly the same as how it leaves the transmitter's sparse encoder. The sparse decoder then recovers the quantization indexes as well as their original channel locations and finally the DPCM/PCM decoder outputs the M reconstructed signal samples. The following sections 3.2 and 3.3 will dissect the system into two layers and describe them accordingly.

3.2 DPCM/PCM Coding Layer

This is the outer layer of the compression system that takes in all input signal samples at the transmitter and outputs all reconstructed samples at the receiver, as the two parts colored in blue in Figure 3.1. The structure of basic DPCM has been illustrated in Figure 2.7, and our current scheme is no more than parallel versions of those basic structures. Any of the structures can be replaced by PCM effortlessly when necessary. Since the inner sparse coding layer and the transmission channel are assumed to be lossless and transparent, a simplified schematic which skip those parts will be used in the rest of this section. The DPCM/PCM coding layer is the only lossy part in the system, so we experiment with different designs at both the quantizer and the predictor to see the SNR performance at several bitrates.

3.2.1 Quantizer Design

Previous studies that use DPCM for haptic data reduction all resorted to uniform scalar quantizer, with either fixed or adaptive codebooks. This is often for the sake of simplicity or attaching entropy coding with variable length code words. If the chosen predictor performs well on the data, the consequent prediction error will be substantially reduced in range and correlation compared to the original sample and thus will have a distribution centralized around zero value. We can then assign the shortest code word to the index that occurs significantly more than others.

However, we are designing a compression system for multiple channels, and we further reduce the data rate through a means other than entropy coding, so there is no direct motivation in sticking to uniform quantizer step and fixed codebook. Instead we can have flexible quantization and try to maximize SNR at certain bitrates at this stage.

Non-uniform Scalar Quantizer with Lloyd-Max Codebook

The first plan for the quantizer in DPCM (or PCM) we propose is to train a Lloyd-Max codebook for each input channel, which has been described in Section 2.2.1. This can result in non-uniform quantization steps and ideally optimal representation points for the data from a particular channel in terms of MSE, but not a sparse histogram of quantization index. The training process can be either offline or online, leading to fixed or varying codebooks. In this thesis, the codebooks are trained offline since we do not have access to fresh data from haptic devices at the moment, and we only train one codebook

for one type of channel, i.e. three for force signal and three for acceleration signal (correspond to x, y, z direction). Figure 3.2 is a schematic view of this design.

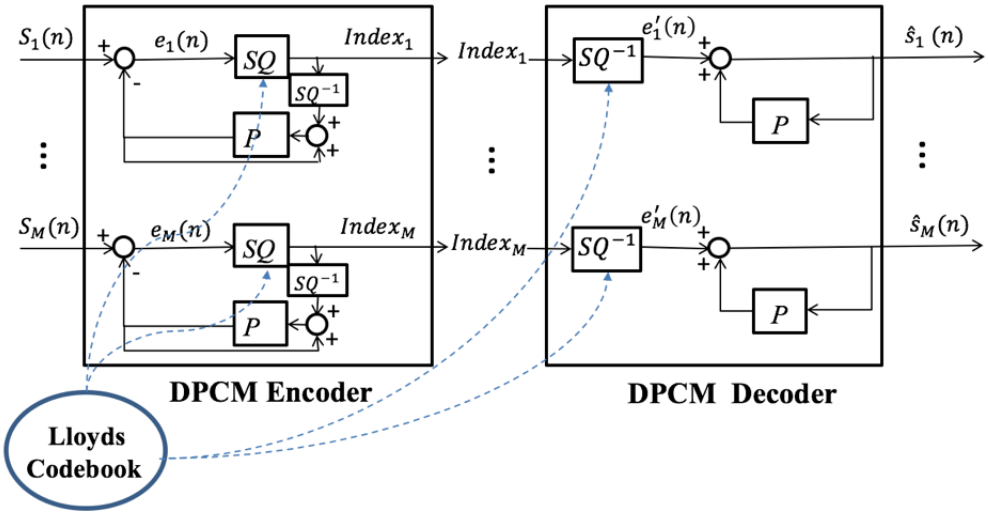


Figure 3.2: Schematic of DPCM part with Lloyd-Max codebook

Two design highlights of this plan should be mentioned. Firstly, only the magnitude of the input is used to train the Lloyd-Max codebook. The sign is represented by an extra bit regardless of the resolution, with '1' for positive values and '0' for the rest (Figure 3.3). The symmetric shape of DPCM sample distribution makes such mechanism equally efficient as the codebook that covers both signs. Moreover, as first-order DPCM samples are actually the variation between consecutive signal samples, it would be natural to plug in the concept of perceptual deadband later on. If we use the codebook without sign, we can set the first decision region to be the deadband width and then all variations that do not exceed the deadband will have index 0 no matter what number of quantization level we use. Secondly, the first representation point for the quantizer in PCM and DPCM is manually forced to zero. The advantage is that when there is an interval of zero inputs from a particular haptic channel (e.g. a person stops moving his hand for a second with a haptic glove on), there will be no drifting of the reconstructed signals on the receiver side. However, the time consuming characteristic of training good codebooks makes this design of quantizer less flexible in terms of resolution. Also, the performance of the codebooks will very much degrade if training data and real data do not resemble each other perfectly.

Uniform Scalar Quantizer with Compander

With a goal of avoiding the major drawbacks of having well-trained Lloyd-Max codebooks, i.e. resolution flexibility problem and suboptimal/overloaded quantizer problem, a second plan for quantization in the DPCM/PCM part featuring histogram compander and uniform quantization is proposed. This method has commonly been seen in the PCM of speech to avoid the pain of finding the optimal non-uniform quantizer for different inputs. The combined effect of compander and uniform quantizer is just that of a non-uniform quantizer, but the complex iterative codebook training phase is removed,

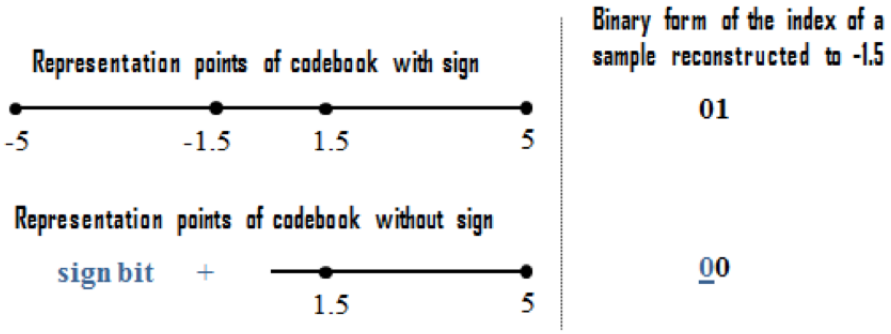


Figure 3.3: Comparison of signed and unsigned codebook (2-bit resolution)

while the price is the computational cost for the compander[1]. Instead of a predefined codebook for each channel, it first uses a compressor, a mathematic transformation that maps all DPCM/PCM samples to a known interval (e.g. [0, 1]). Such transformation is not difficult to conceive. The cumulative distribution function (CDF) of the data samples will exactly serve the purpose, for it maps each sample value to its corresponding percentile in CDF, which is between 0 and 1. For example, if the data is subjected to some Gaussian distribution with zero mean $\mathbf{N}(0, \sigma)$:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{x^2}{2\sigma^2}) \tag{3.1}$$

then the compressor function can be:

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du = \frac{1}{2} + \frac{1}{2} erf(\frac{x}{\sqrt{2}\sigma}) \tag{3.2}$$

where $erf(.)$ is a notation of the Gauss error function:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{3.3}$$

After transforming the DPCM/PCM samples with a compressor such as in equation 3.2, the histogram of the data will be greatly equalized and the range is strictly constrained between 0 and 1. The processed samples then go to the uniform quantizer set at a resolution of N bits that assign each sample its quantization index:

$$I = \left\lfloor x \cdot 2^N \right\rfloor \tag{3.4}$$

where x represents the sample after the compressor. On the receiver side, the index is mapped to the reconstruction point value through:

$$\hat{x} = \frac{2I + 1}{2^{N+1}} \tag{3.5}$$

And finally the original sample is recovered through an expander, the inverse function of the compressor. Its schematic is shown in Figure 3.5. C and E represent the compressor and expander, respectively. If the compressor function and the CDF of the data

[1] http://ocw.mit.edu/courses/aeronautics-and-astronautics/16-36-communication-systems-engineering-sp
lecture-notes/MIT16_36s09_lec04.pdf

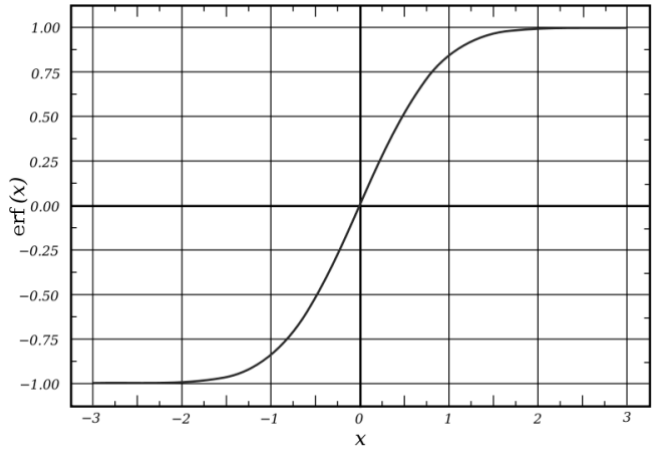


Figure 3.4: A Gauss error function

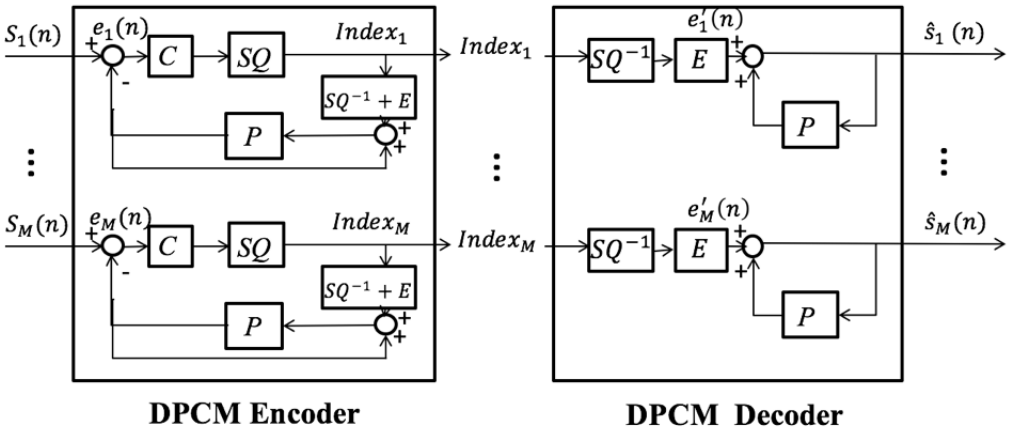


Figure 3.5: Schematic of DPCM part with compander

are perfectly identical, the distribution of the subjective samples will be perfectly uniform in [0,1], and a uniform quantizer can be hence applied. But it has to be made clear that although the exact CDF of the DPCM/PCM data is not always the optimal compressor function in terms of the overall SNR when followed by a uniform quantizer and an expander.

I will briefly explain the reason in an example of a dataset that has a normal distribution with zero mean and unit variance. Suppose an extreme circumstance where the uniform quantizer uses only 1 bit (2 reconstruction points). The reconstruction points fall at 0.25 and 0.75 before the expander will shift them, and certainly we wish the two shifted values to be the centroids of the two symmetrical halves of the $\mathbf{N}(0,1)$, which are about $pm0.8$. However, if we use the CDF of standard Gaussian as the compressor and the inverse of it as the expander, the reconstruction points will end up at ± 0.67, and the consequent SNR will be sub-optimal. If we adjust the compressor function to:

$$F_2(x) = F_X(\frac{x}{a}) \tag{3.6}$$

where F_X is the CDF according to Equation 3.2 and is some constant other than 1, the

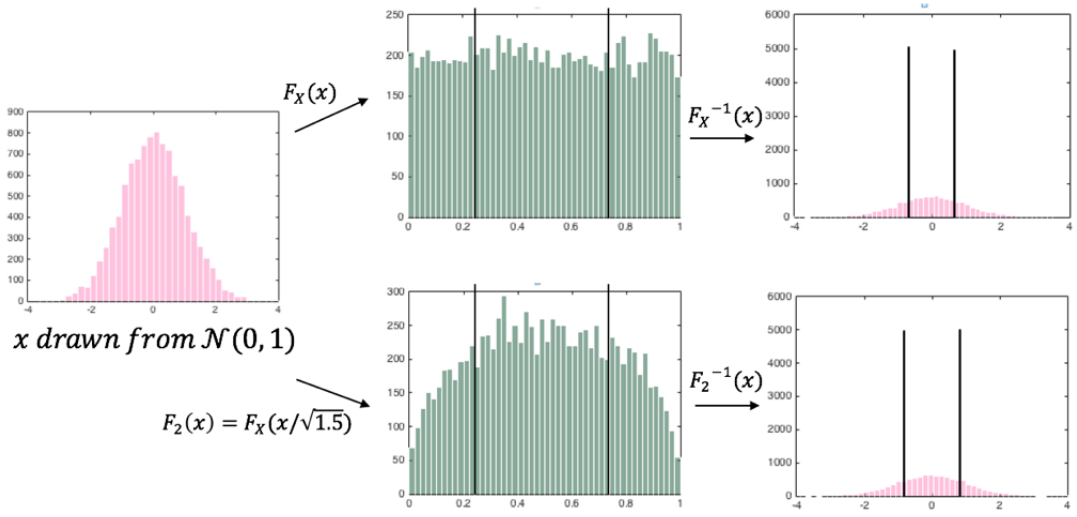


Figure 3.6: Histogram comparison with different compander with 1-bit uniform quantizer

reconstruction points will be likely to get closer to the optimal points. For example, with 1-bit uniform quantizer, if $a = \sqrt{1.5}$, then the reconstruction values after the expander will be ± 0.82, which are closer to ± 0.8. An illustration is shown as in Figure 3.6. With different Gaussian distributions and different resolutions of the quantizer, the optimal adjusting factor a for the compressor will also be different. And above all, the distribution of DPCM/PCM samples for haptic data may be modelled better by other distributions. In speech PCM modulation, the μ-law algorithm is frequently used as the compander function as recommended by ITU-T G.711[2]. In this thesis, Gaussian-CDF-like companders and Laplacian-CDF-like companders are experimented for the haptic signals. For a Laplacian distribution with zero mean and parameter b, the CDF is:

$$G(x) = \frac{1}{2} + \frac{1}{2} sgn(x)(1 - e^{\frac{-|x|}{b}}) \tag{3.7}$$

and the compressor function looks like:

$$G_2(x) = G(\frac{x}{a}) \tag{3.8}$$

where a is the adjusting factor.

3.2.2 Predictor Design

The predictor part in DPCM as shown both in Figure 3.2 and Figure 3.5 is crucial for producing highly de-correlated prediction error signals that are much more centralized around zero than original samples. Past studies mainly applied first order DPCM on slowly varying haptic signals. In this thesis, both first and second order DPCM as well as PCM are applied to multiple types of signals from the database combined with either the Lloyd-Max codebook or the compander approach. The distortion-rate results of the DPCM/PCM part alone will be evaluated in Section 5.2.

[2] http://www.itu.int/rec/T-REC-G.711-198811-I/en

3.3 Sparse Coding Part

This is the inner part of the entire compression system, labelled yellow in Figure 3.1. It is lossless because it further reduces data transmission only by mitigating the chance of transmitting zero quantization indexes (which also have zero reconstruction values and thus meaningless to transmit with full bits) from all input channels. In order to indicate whether to transmit a sample, we need flag bits to label the channels. However, chances are the proportion of input channels with zero indexes is low and one flag bit per channel is deemed superfluous. Since we also assume a bunch of inputs in our haptic application, we decide to divide the channels into small groups and each group is given one flag bit. The flag bit of a group is 0 only in the condition that all inputs in that group have zero indexes at one instant.

3.3.1 Shuffle Encoder and Decoder

We certainly wish the channels of zero and non-zero indexes flock separately in the first place so that the number of 0 flags can be maximized, and that give inspiration to the channel shuffling step before grouping. Basically, it rearranges the input channels to separate non-zero and zero channels. And on the decoder side, the channels are resumed to their original order. The shuffling rule is based on the most recent sample history of the channels for it is the only reasonable way to synchronize the transmitter and receiver without extra data transmission. In this study, we believe that neighboring samples from the same channel tend to have the same type of quantization indexes (zero/non-zero), considering the high sampling frequency of the haptic capturing devices. Thus, the simplest shuffling rule can be described as: rearrange the input channels at instant t the same way as how the channels can be perfectly arranged at instant $(t-1)$ (Figure 3.7).

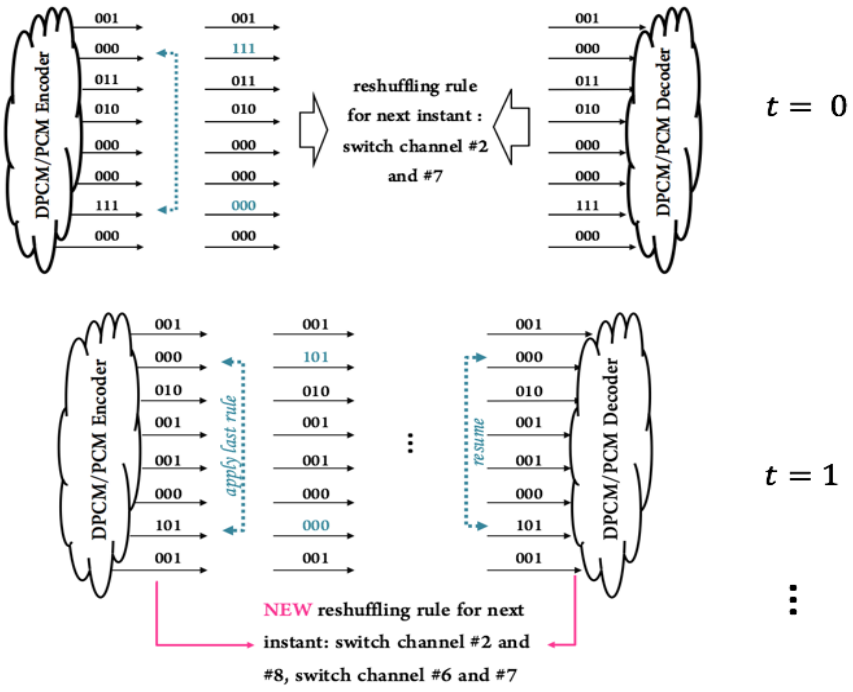


Figure 3.7: Channel Shuffle and Resume Process (8 channels)

3.3.2 Grouping Encoder and Decoder

After channel shuffling step the input samples are ready to be grouped. The group size is yet another important parameter to be decided. We do a calculation first. Suppose there are N input channels, the resolution of the signal for transmission is k bits/sample (including the sign bit in the Lloyds codebook quantizer design), and the group size is M channels/group. This would lead to a sum of $\left\lceil \frac{N}{M} \right\rceil$ groups and thus $\left\lceil \frac{N}{M} \right\rceil$ flag bits per transmission. According to the design logic, if a group has at least one non-zero channel, the flag bit would be 1 and all channels in the group will be transmitted in full resolution k. And all channels in a group will be omitted for transmission if otherwise. Therefore, if l out of the $\left\lceil \frac{N}{M} \right\rceil$ groups ($0 < l < \left\lceil \frac{N}{M} \right\rceil$) have flag bit 1, the total number of bits needed for one transmission is ($k \cdot lM + \left\lceil \frac{N}{M} \right\rceil$). The product of l and M reflects the sparsity of data and can be deemed as a constant when N and k remain unchanged. Therefore, the number of bits consumed is approximately a function of group size M. An illustration of the grouping step in the sparse coding part is shown in Figure 3.8.

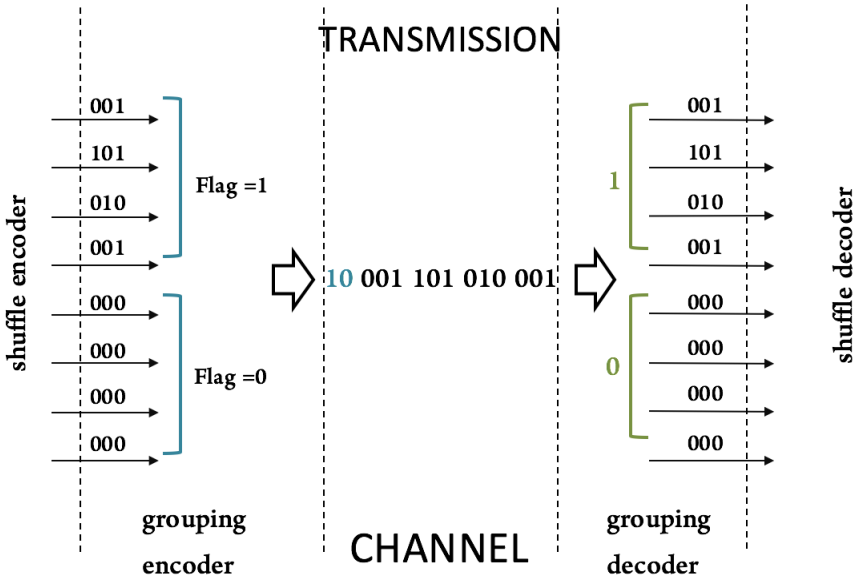


Figure 3.8: Channel Grouping Process (8 channels, $M = 4$)

To conclude, the sparse coding part explores the sparsity in the quantized samples in an inter-channel way and is a kind of lossless data reduction. The advantages of this design of sparse coding part over the traditional entropy coding are twofold. On the one hand, it is more efficient for a haptic system with a large number of inputs, for it leverages the sparsity across all inputs. On the other hand, it is more suitable for our quantizer design with either Lloyds codebook or compander, since the appearance of all indexes will be approximately equi-probable after these two quantizers. Things that matter the functionality of this part are the tuning of the shuffling rule and the group size. In the testing and evaluation chapter, experiment results with different parameters will be presented and compared.

Chapter 4

The Testing Database

Haptic data from previous research is used for testing the compression system instead of freshly collected data due to several reasons. First of all, currently the haptic infrastructure that has been set up for research use in our laboratory is a WoodHaptics device in connection with virtual haptic environments simulated through CHAI3D, therefore the number and type of inputs are quite limited (only 3D position data in the feedforward channel and 3D force data in the feedback channel because of lack of accelometers). Secondly, CHAI3D is a C++ framework, which means that all coding for the compression algorithm has to be translated into C++. But importing existing database allows for direct analysis with MATLAB and is much more convenient. Finally, since we aim at designing a universal compression system for both kinesthetic and cutaneous signals, of which the latter is smaller ranged, faster changing and more algorithm demanding (e.g. there is perceptual deadband in both amplitude and frequency dimension for vibration signals), we would certainly want to evaluate the system on those signals. We soon find the Haptic Texture Toolkit (HaTT) published by University of Pennsylvania [42] a great fit to our need.

This open source toolkit[1] is originally aimed at research for texture rendering in virtual environments. It includes a collection of 10-second raw recordings of the 3D force, orientation, position and acceleration signals experienced by an experimenter strolling on 100 real textures unconstrainedly with a SensAble Phantom Omni. The data comes in XML format with the categories for each material shown in Table 4.1. A summary of the details of collection process of raw data is presented in Table 4.2. From thes summary we clearly notice that the recording of 3D force and 3D acceleration data satisfies our requirements for refresh rate (>1kHz) for haptic applications. Therefore, we only use force and acceleration signals as inputs in experiments. The two types are treated separately, since their features are very different. Also note that although the force and acceleration data are used for modelling virtual textures by researchers from UPenn, they can be just regarded as feedback tactile signals from teleoperators or virtual environments on random 10-second trials, and thus the concept of perceptual deadband makes sense on this database.

[1]The toolkit is available for download at: `http://repository.upenn.edu/meam_papers/299/`

Field	Description
Accel_*MATERIAL* Files	
<AccelUnits>	Units of acceleration
<SampleRate>	Sample rate in Hz
<Accel>	Combined acceleration signal using DFT321 algorithm
<Accel_x>	Array of acceleration values in x-direction
<Accel_y>	Array of acceleration values in y-direction
<Accel_z>	Array of acceleration values in z-direction
Position_*MATERIAL* Files	
<SpeedUnits>	Units of speed
<PositionUnits>	Units of position
<SampleRate>	Sample rate in Hz
<Speed>	Array of absolute speed values, low-pass filtered at 100 Hz
<Position_x>	Array of position values in x-direction
<Position_y>	Array of position values in y-direction
<Position_z>	Array of position values in z-direction
Force_*MATERIAL* Files	
<ForceUnits>	Units of force
<SampleRate>	Sample rate in Hz
<ForceNormal>	Array of force normal to surface
<ForceTangential>	Array of force values tangential to direction of motion
<Force_x>	Array of force values in x-direction
<Force_y>	Array of force values in y-direction
<Force_z>	Array of force values in z-direction

Table 4.1: Data categories in HaTT

	Sensor	Sampling Rate	Resolution	Other
Position (Orientation)	8mm Ascension 3D Guidance TrakSTAR	240Hz	0.5mm 0.1°	Upsample to 10kHz using spline interpolation
Force	ATI Nano17 SI-25-0.25	10kHz	16 bits	
Acceleration	ADXL321 ±18 g	10kHz	0.1256 m/s^2, 16 bits	Low pass filtered with Bessel Filter at 1kHz

Table 4.2: Detail of data gathering in HaTT

Chapter 5

Experiments and Results

5.1 Data Pre-processing and Analysis

As a first step, 3D force and 3D acceleration data from UPenn 100-material texture database are extracted from corresponding labels in XML files mentioned in Table 4.1, downsampled to 1kHz (so 10000 instead of 100000 samples for each recording), and stored in .MAT structures suitable for MATLAB processing. After that, the two types of data will each go through the same compression system. The 3D force signals from 100-materials are regarded as parallel input channels in this study, simulating inputs from multiple sensors in a haptic system. Hence we are able to choose from 1-300 input channels. The 3D acceleration data is treated the same way. Although inputs from different textures may not resemble the behavior of multi-input on a single haptic device, it does no harm to this study, since exploring correlations between inputs is out of scope of this thesis.

Plotting the force data range of 100 materials gives Figure 5.1. Obviously, the x and y direction force data are almost symmetrical around zero value for each material, whereas the z direction data only span negative value. It implies that for the codebook training method, it is unnecessary to use codebooks with an extra sign bit for z direction force signal. The acceleration data range plots are shown in Figure .For acceleration signals, all three directions tend to have symmetric distribution around zero value, but the signal range for different materials vary drastically. This means we should better train the codebook on individual channels in order to achieve good performance.

Other properties including variance and autocorrelation patterns up to 2000 samples (2 seconds) of the signals as well as of their residuals of different orders are presented in Table 5.1. The plots of autocorrelation functions give us general impression about how much neighboring samples are correlated in each type of input. Apparently, a huge decorrelation takes place when we move from direct signal amplitude to first order difference for the force signal, and the variance also drops immensely. As for acceleration signals, however, the signal amplitude is barely correlated itself and the variance increases as we increase the order of difference. This indicates that we should better operate on first order difference signals of force and amplitude signal of acceleration, in order to utilize sample-wise signal correlation in the channel shuffling part, while minimizing the range of their values.

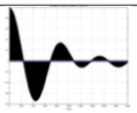
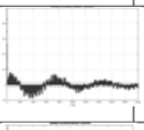
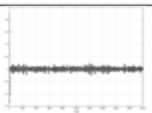
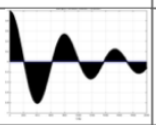
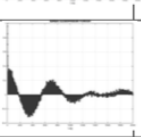
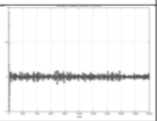
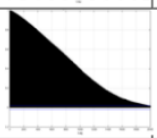
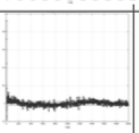
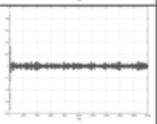
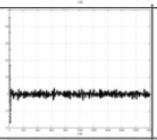
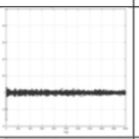
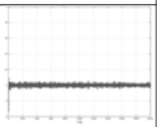
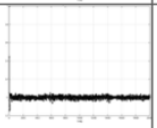
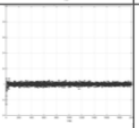
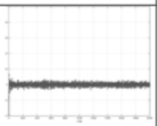
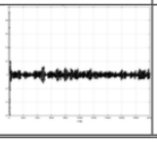
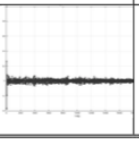
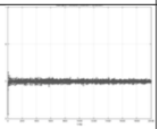
residual order / data type	0^{th} (x_n)		1^{st} $(x_n - x_{n-1})$		2^{nd} $(x_n - 2x_{n-1} + x_{n-2})$	
	AC	*Var*	*AC*	*Var*	*AC*	*Var*
Force.X		0.53		6.8×10^{-4}		6.1×10^{-5}
Force.Y		0.49		7.1×10^{-4}		6.2×10^{-5}
Force.Z		2.4		7.0×10^{-4}		8.0×10^{-5}
Acceleration.X		5.8		10.2		29.4
Acceleration.Y		6.5		9.5		25.7
Acceleration.Z		10.5		19.6		57.1

Table 5.1: Autocorrelations and variances of data

5.2 Performance of DPCM/PCM Coding Part

Following the explanation in Section 3.2, I analyse the performance of different designs of quantizer, i.e., non-uniform Lloyds SQ and uniform SQ with companders, combined with several orders of predictor, on both type of signals. Lloyds quantization require pre-trained codebooks. Considering the difference in behaviour between force and acceleration signals, I train one codebook per dimension for the force signal, while one codebook per material for the acceleration signal. On the other hand, for quantizer with companders one must use compressor functions that match the distribution of input signal. I run both Gaussian and Laplacian compressors and tune the parameters until it reaches peak performance.

5.2.1 Time-domain SNR

As the key performance indicator, the time-domain SNR of the abovementioned system designs at different quantization resolutions are obtained.

Force Signals

For force signals, the result for uniform quantization with Gaussian companders is shown in Figure 5.3. It is clear from the figure that the SNR results from predictive coding (DPCM)

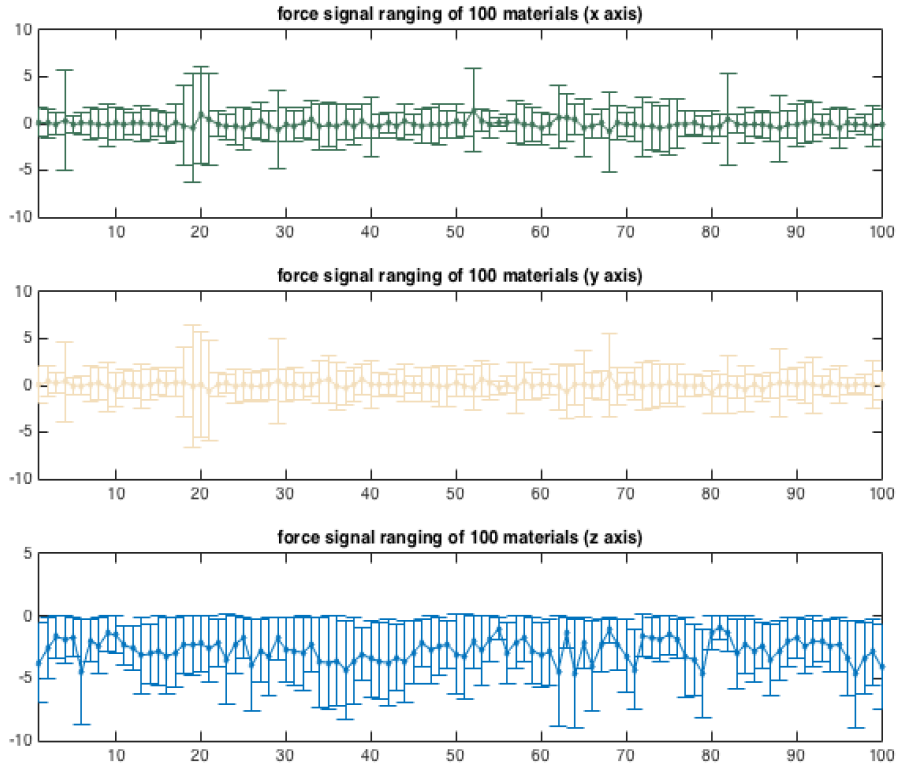


Figure 5.1: Range of 3D force signal (100 materials)

is way above those from non-predictive coding (PCM) at lower resolutions, although the SNR from PCM rise faster with the increase in resolution. Moreover, second order DPCM does only small improvement to first order DPCM at very low resolutions, and therefore first order DPCM is the best on force signals, which coincides with our inference in Section 5.1. Next, the red curve in is compared against uniform quantization with Laplacian compressor function, also first order DPCM. As is shown in Figure 5.4, with a Laplacian compressor function instead of a Gaussian one, there is at least a 2 dB SNR improvement at a certain resolution for force signal.

Non-uniform data quantization with Lloyds codebook is then also performed on the first order residuals of force. I train one codebook for one dimension and every resolution, and two sets of training data of different sizes are used on separate trials. The length of one is 20000 samples (2% of data) and another is 50000 samples (5% of data). The consequent SNR-rate curves plotted in the same graph with the best curve from uniform quantization with companders (the blue line in Figure 5.4) is shown in Figure 5.5.

It is clear from the figure that there is a huge SNR gap between non-uniform SQ with more and less training data. When Lloyds codebook is trained with 5% of the data, the SNR performance is sufficiently better than uniform quantization with Laplacian compressor on bit rates larger than 2 bit/sample. However, with the proportion of training data lowered to 2%, the SNR curve deteriorates to way below that of USQ. This means that the USQ method is much more robust to data from different textures and dimensions, since we only use one compressor function throughout all force residuals.

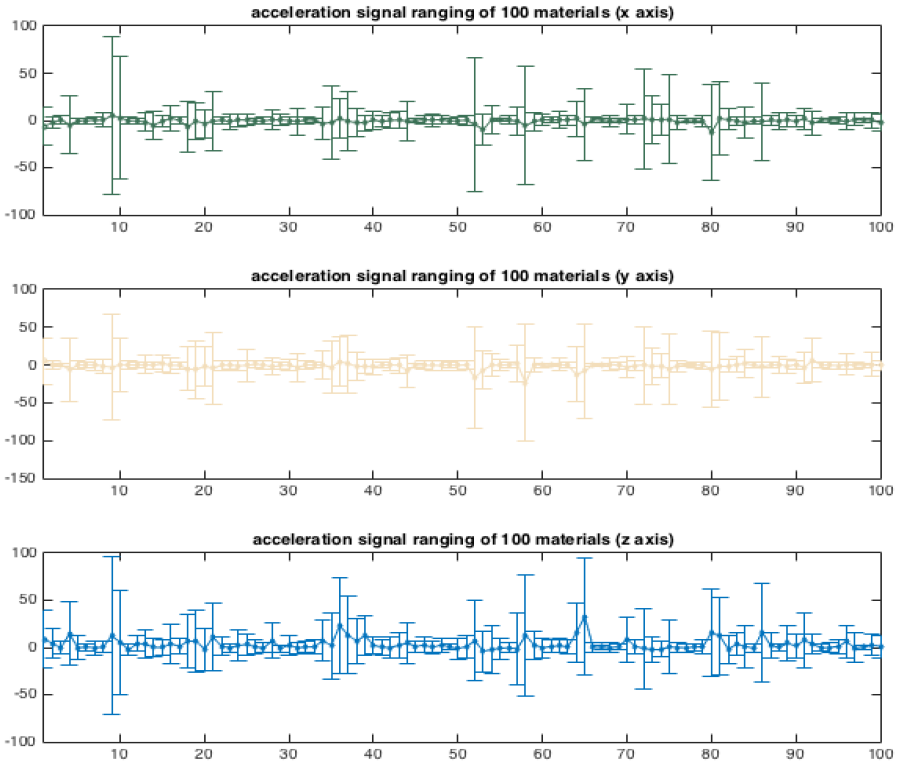


Figure 5.2: Range of 3D acceleration signal (100 materials)

Acceleration Signals

Due to the characteristics of acceleration signals, the standard deviation parameter is tuned individually for each material, whereas the constant factor (see equation 3.6 and 3.8) is still kept constant for all inputs of the same order. USQ with Gaussian compressors on amplitude signals and its first and second order residuals have the following SNR curves shown in Figure 5.6. Obviously, quantization on PCM samples always outperforms predictive coding, which again corroborates the inference in Section. Again, the best curve in this experiment is compared against the result of USQ on PCM samples with Laplacian compressor, see Figure 5.7. We can tell from the figure that for resolution above 4 bits/sample, Laplacian companders perform better than Gaussian companders. Since it is only for bit rates above 4 bits/sample that the average SNR values are satisfying (>20dB), a Laplacian compressor function is still preferred over a Gaussian function.

And lastly, non-uniform SQ with trained Lloyds codebooks are performed on acceleration PCM samples. I train one codebook for each material (3 dimensions), and 3%, 15% and 30% of total data are used as training data in separate trials. The results are compared with uniform quantization with Laplacian compressor function, as shown in Figure 5.8. Here non-uniform SQ performs worse than uniform SQ, even with a training data proportion as high as 30%. The advantage of uniform quantization is therefore self-evident.

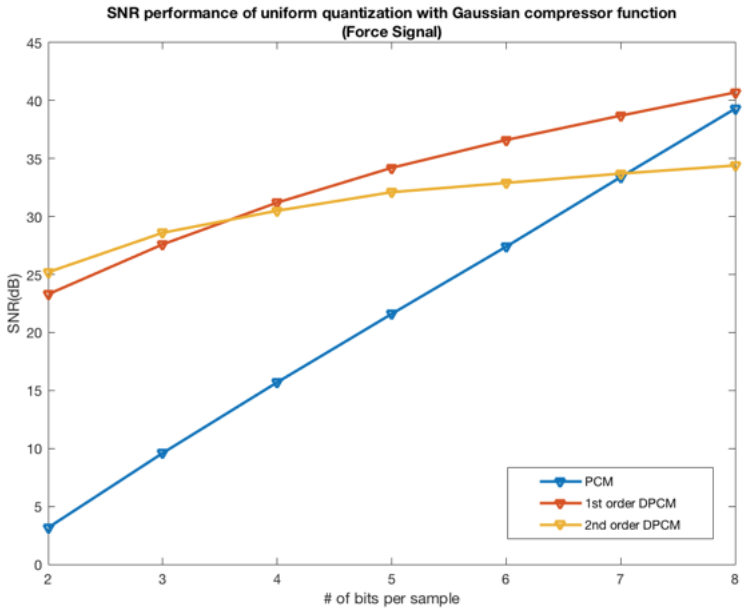


Figure 5.3: SNRs of force signal compression with uniform SQ and Gaussian compander

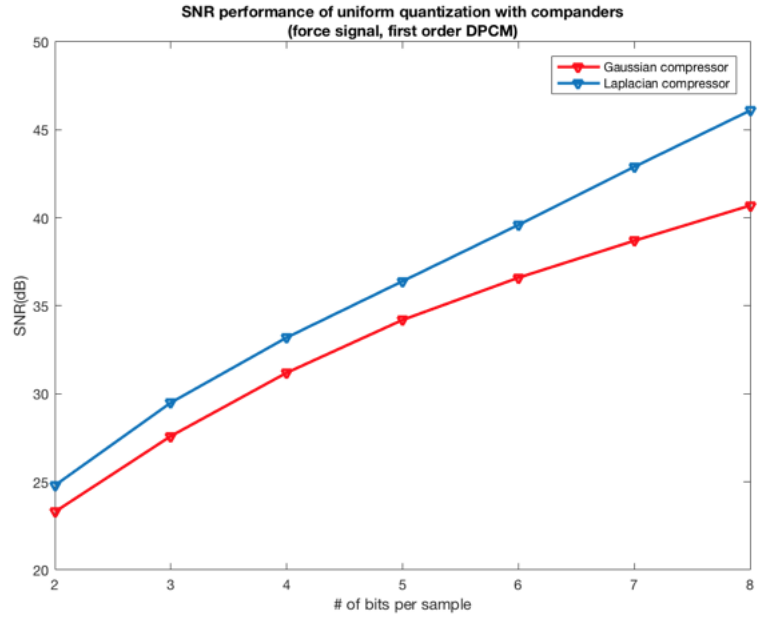


Figure 5.4: SNR results of uniform SQ with different compressors(force signal, first order DPCM)

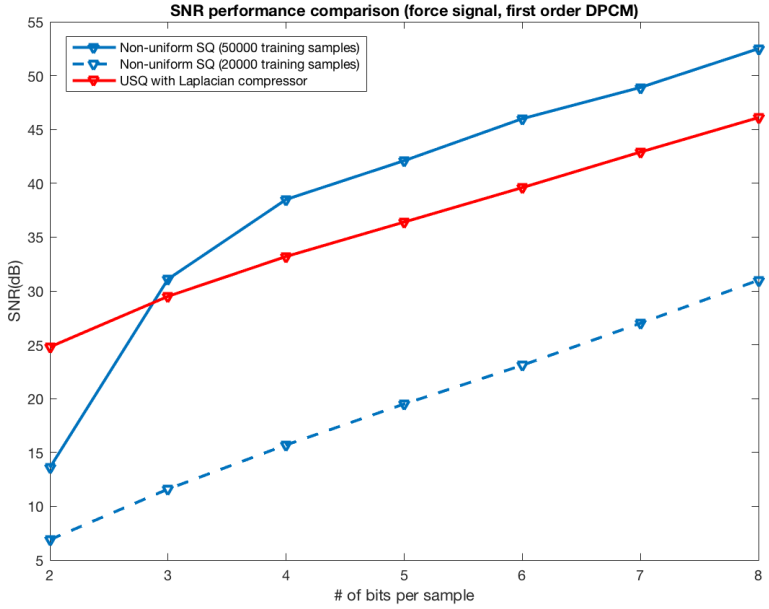


Figure 5.5: Comparison of SNR results between non-uniform Lloyds quantization and uniform SQ with Laplacian compressor (force signal, first order DPCM)

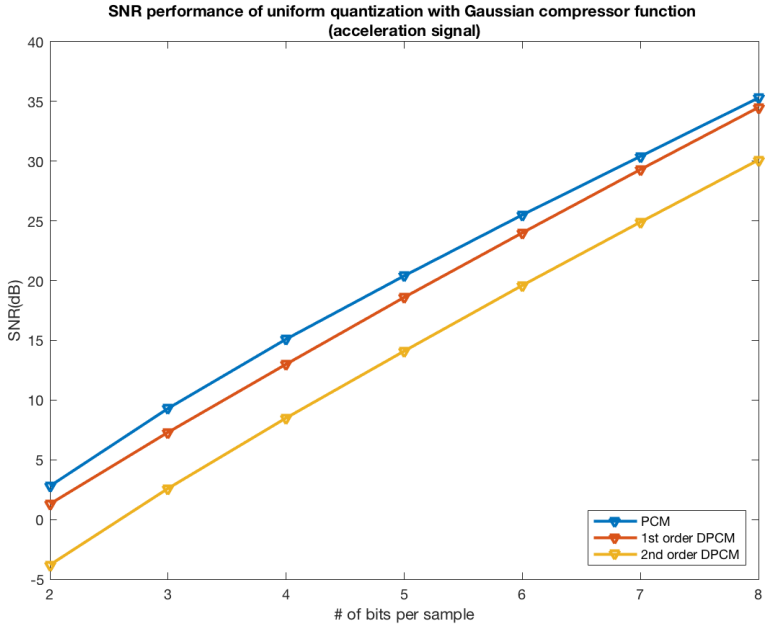


Figure 5.6: SNR results of acceleration signal compression with uniform quantization and Gaussian companders

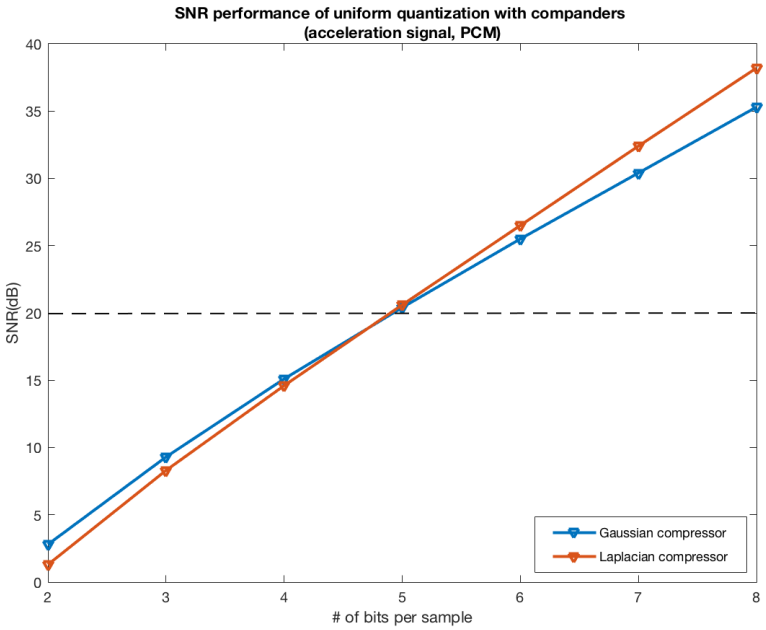


Figure 5.7: SNR results of uniform SQ with different compressors(acceleration signal, PCM)

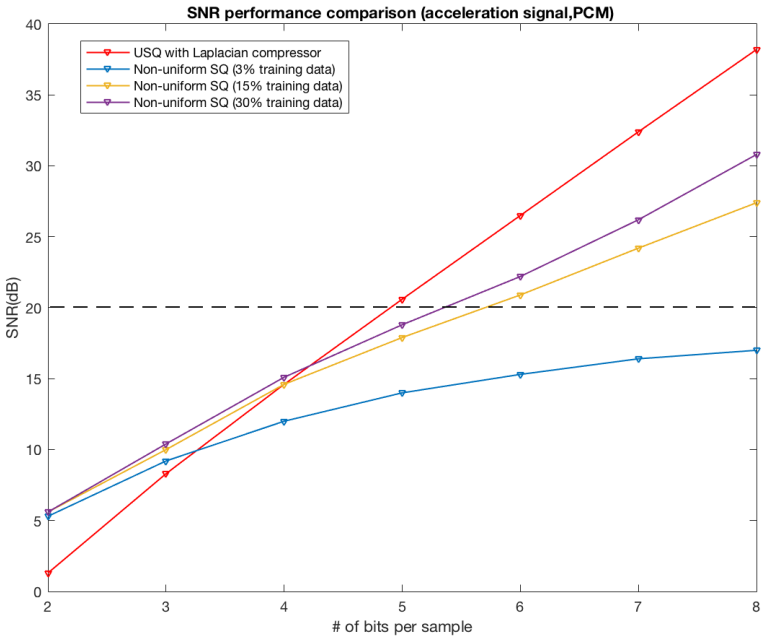


Figure 5.8: SNR results of non-uniform Lloyds quantization and uniform SQ (acceleration signal, PCM)

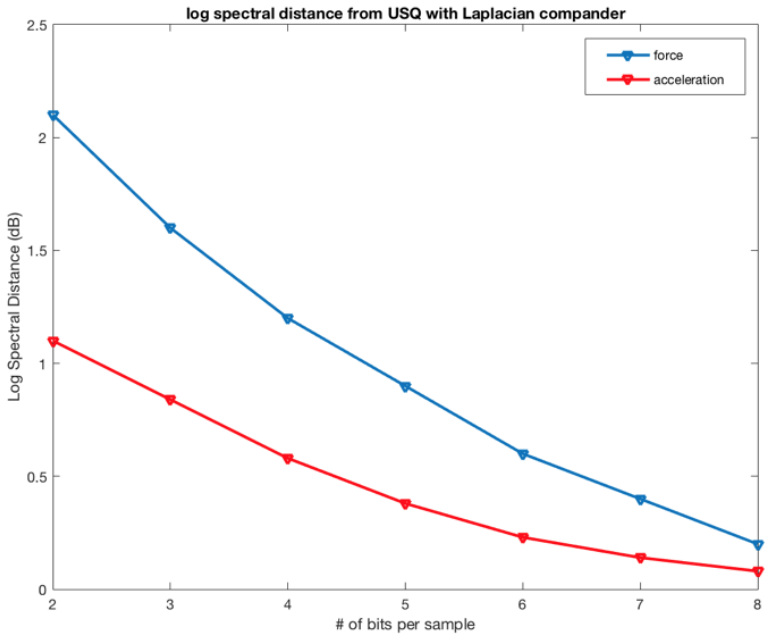


Figure 5.9: Log spectral distortion of uniform SQ with Laplacian companders

5.2.2 Spectral Distortion

As explained earlier, we need to watch out to the spectral distortion characteristics of the compression system on the data as well, for perceptual deadband in the frequency domain is crucial for human to vibrotactile signals . According to the paper, the deadband width is around 18% in the frequency dimension for vibrotactile signals, independent of the reference frequency. Since the system is intended for both kinesthetic and tactile devices, we need to guarantee that the spectral distortion does not go beyond the deadband. The quantitative analysis of the spectral distortion is based on equation 2.9, and the results from the optimal uniform SQ with Laplacian function on force and acceleration data are shown in 5.9. The log spectral distortion (LSD) decreases with the increase in resolution for both force and acceleration, and contrary to time domain SNR, the distortion on acceleration is outperforms distortion on force.

However, quantitative analysis is meaningless if not accompanied by the inspection on the change of major frequency components. As a further step I create another criterion called spectral variation ratios (SVR), which is an array that measures the absolute ratios between the differences of Discrete Fourier Transform (DFT) terms of the original and compressed signals and the DFT terms of original signal:

$$SVR = \left| \frac{DFT_N(x_{hat}) - DFT_N(x)}{DFT_N(x)} \right| \tag{5.1}$$

where N stands for N-point DFT. Consequently, the array will contain N terms in total, with a symmetry by halves. Firstly, I check on the general shape of spectrum for force and acceleration signals by plotting out the 200-point DFT magnitude terms of 6 randomly chosen channels for each type, as shown in Figure 5.10 and Figure 5.11. Force signals from the particular database we have, as can be seen, are not mathematically 'vi-

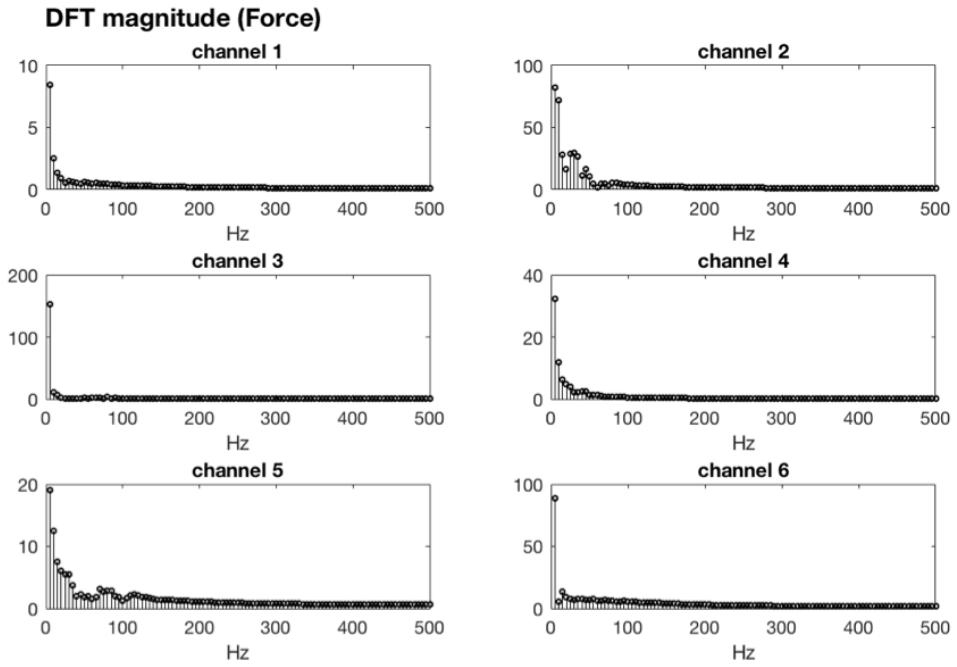


Figure 5.10: DFT magnitudes of force signals

bro' signals. Major frequency components almost always reside within 50Hz with few exceptions. Acceleration signals, on the other hand, are quite active across the 0-500Hz spectrum, with different channels featuring different major components. Therefore, the goal should be that the SVR be small enough for low frequency part of force signal and all frequencies of acceleration signal. Especially, since acceleration signals are vibrotactile signals, the SVRs should get as small as at least 0.2 to be deemed harmless for perception.

Then, average SVR values for compression ratio from 3 bits/sample to 8 bits/sample on force and acceleration signals are calculated under uniform SQ with their respective best Laplacian companders, whose results are shown as curves of different colors in Figure 5.12 and Figure 5.13. We can clearly tell from the contrast between the figures that the spectral distortion is generally smaller towards lower frequency (< 100Hz) force signals and higher frequency (> 20Hz) acceleration signals. For force signals, average SVRs easily drop below 0.2 for frequencies under 100Hz when resolution equals or above 4 bits/sample. However, for acceleration signals the compression scheme only achieves the minimum spectral distortion requirements across all frequencies at resolution as high as 7 bits/sample.

5.2.3 Summary

In Section 5.2.1 and 5.2.2, the lossy compression layer of the compression system is judged through its performance on HaTT database's force and acceleration data, from two perspectives: time domain SNR and spectral distortion. The former is prioritized over the latter as the main criterion at this stage, because there are few spectral distortion measures for haptic signal and not much research of its impact on the perceiving discrepancy.

For each type of signal, I first decide the optimal order of signal residual to imple-

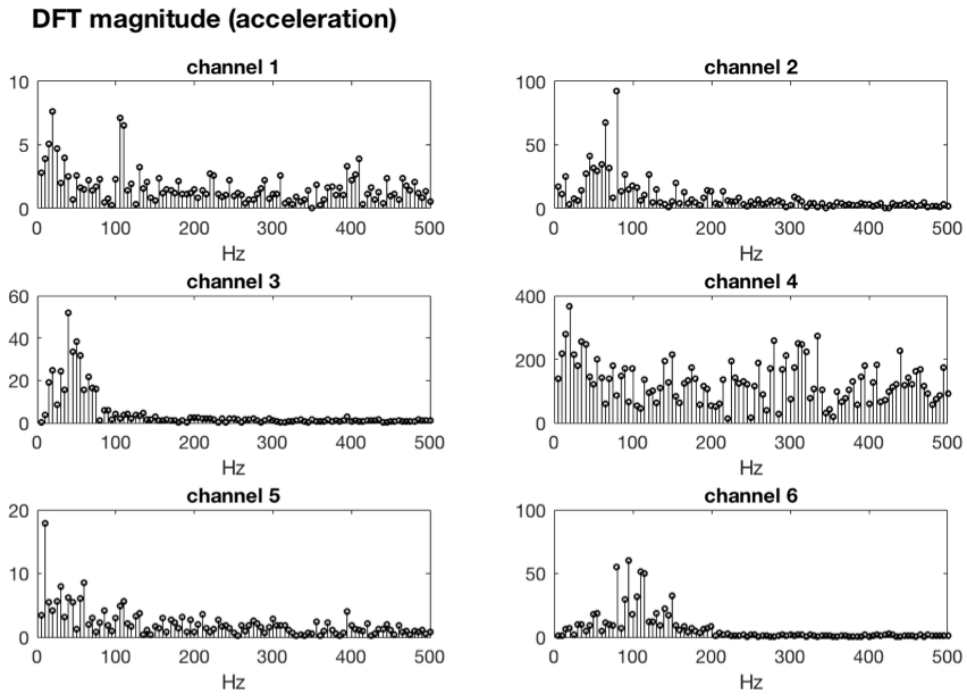


Figure 5.11: DFT magnitudes of acceleration signals

ment compression on by comparing the SNR results of PCM, first and second order DPCM processed with uniform SQ with Gaussian companders. The conclusion is that first order DPCM is the most suitable for coding force data, and PCM is the most suitable for coding acceleration.

Next, I compared Gaussian- and Laplacian- shaped compressor functions for USQ. The results show more or less advantages of Laplacian compressors on both types. And finally, USQ with Laplacian compressors is compared to non-uniform SQ with trained Lloyds codebooks. The results suggest that although non-uniform SQ yields better SNRs from time to time, the SNR is vulnerable to the representation power of the training data. Considering that USQ with Laplacian companders is also computationally simpler and more flexible, it is decided to be the most desirable mechanism amongst all. Its average SNR reaches 25dB from 2 bits/sample for force, and is above 20dB from 5 bits/sample for acceleration.

The spectral analysis following the time domain analysis uses both LSD and SVR as performance indexes. Although LSD calculation is more optimistic about the performance of the USQ on acceleration than force, SVR calculation says otherwise. While for force the spectral distortion is generally satisfying towards the densest frequency region, for acceleration it only falls within the perceptual deadband (20%) at high resolutions. It thus remains a mystery whether perception of the processed vibrotactile signals will be affected by bit rates lower than 7 bits/sample, without further psychophysical studies of the compression system on human participants.

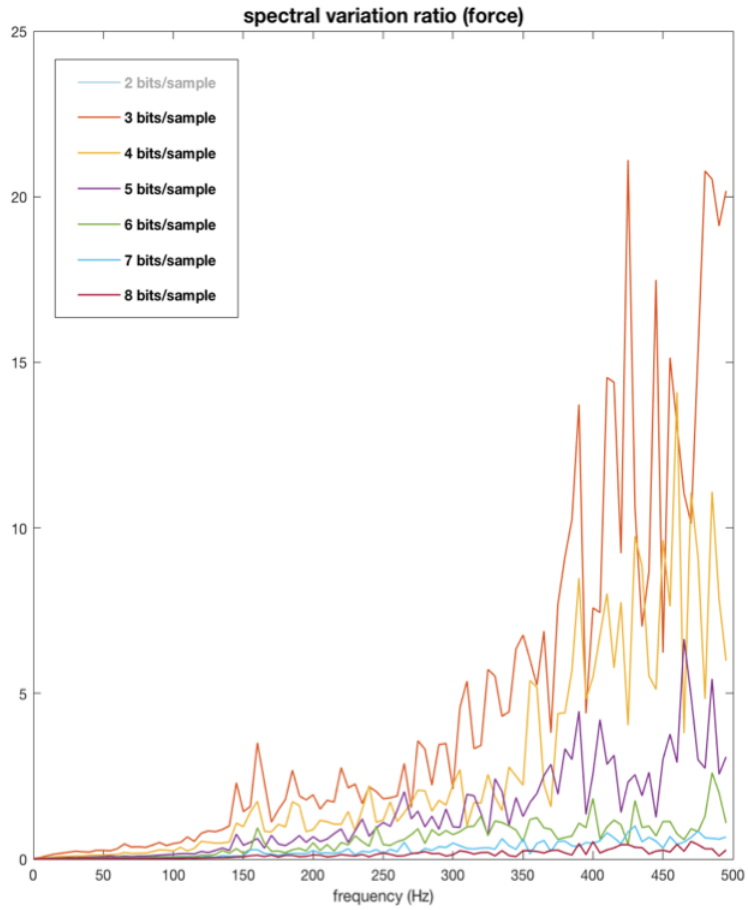


Figure 5.12: SVR values of force signal (USQ with Laplacian companders)

5.3 Performance of Sparse Coding Part

Sparse coding part does lossless compression, so the performance evaluation is based on its effectiveness in enhancing DRS (Equation 2.11). As discussed in Section 3.3, this layer would work better for sparser data, signal inputs with a large zero-index percentage, while the shuffling rule and grouping size depend on the correlation level and sparsity of inputs. If we look at Table 5.1, we may conclude that force DPCM signals as well as acceleration signals in general are largely de-correlated. Therefore, we use a shuffling rule based on 1-sample memory for all data, and start with measuring the efficacy of the sparse coding part alone, that is, without any improvement on the SQ.

The resulting DRSs are calculated for a series of grouping sizes (out of a sum of 300 inputs) run at 2-8 bits/sample, for acceleration PCM signals. The DRS at each resolution is compared to that with pure SQ, which is presented in Figure 5.14. Obviously, the shuffling rule seems to work fine with our data, since DRSs are greatly improved for all resolutions, ranging from 4% to 15%.

I repeat the same calculations on 1st order DPCM force samples (Figure 5.15), and see similar results, with DRS enhancement 6%- 13%. Thus, the efficacy of the sparse coding part alone on our operating data is affirmative. The DRS records of the abovementioned tests can be found in Appendix A1.

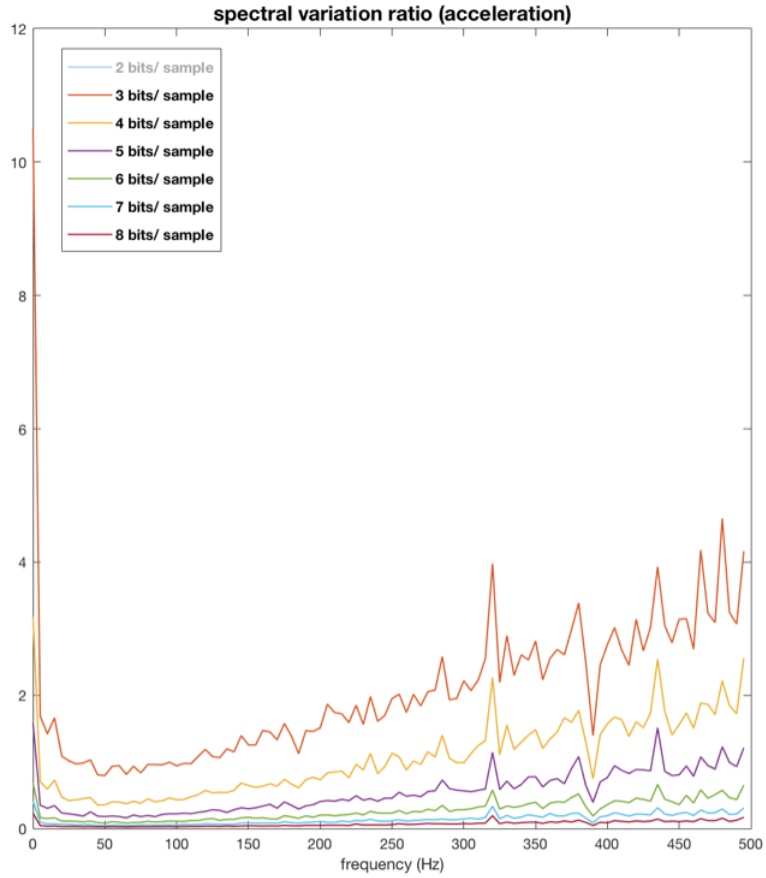


Figure 5.13: SVR values of acceleration signal (USQ with Laplacian companders)

As we know, the outer layer of the proposed compression system equalizes quantization indexes. The SNRs will be improved by such equalization, but the sparsity of signals will decrease. Therefore, additional test on the DRS performance of sparse coding part combined with our proposed DPCM/PCM layer (USQ with Laplacian compander) is necessary. Indeed, the results show compromised DRSs for both force and acceleration, compared to the previous experiment. For force DPCM samples, the best DRS performance is still above that of pure SQ, only that the best improvement is lowered to 2%-9% (See Appendix A2 for details). For acceleration PCM samples, however, there is only degradation in DRS compared to conventional SQ. This means that while first order residuals of force can be better coded with the entire proposed system, acceleration signals are only suitable for either an improved PCM compression alone, or a normal PCM plus sparse coding.

5.4 Overall Performance

To judge the aggregated performance of the two layers combined, the distortion-rate curve is used, i.e., the time domain SNR against the effective resolution. One curve is plotted for force signals processed by the entire compression system working at optimal points. Two curves are plotted for acceleration signals processed by improved PCM and normal PCM with sparse coding respectively. The curves are presented in Figure 5.16.

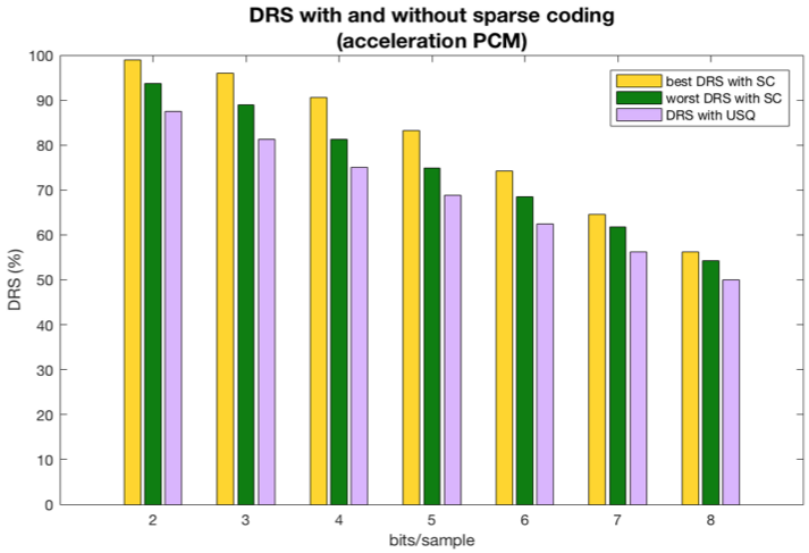


Figure 5.14: DRS comparison results (acceleration PCM)

Apparently, sparse coding coupled with conventional PCM on acceleration signal enhanced the SNR performance of PCM with Laplacian compander a great deal at low bit rates, raising the SNR to 20 dB starting from about 4 bits/sample. Therefore, PCM without companders in the outer layer fits the acceleration data better. However, a gap of more than 15 dB still remains between the optimal curve of force and acceleration at least from 1.5 bits/sample to 7 bits/sample, meaning that the proposed system is clearly more preferable with force data than acceleration data. From a latitudinal point of view, force data saves more than 3-bit resolution on average at a certain SNR value, compared to acceleration data.

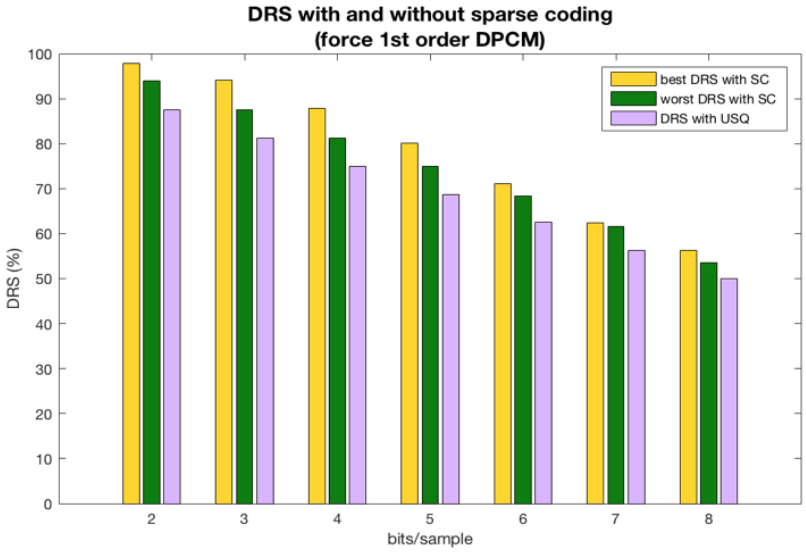


Figure 5.15: DRS comparison results (force 1^{st} DPCM)

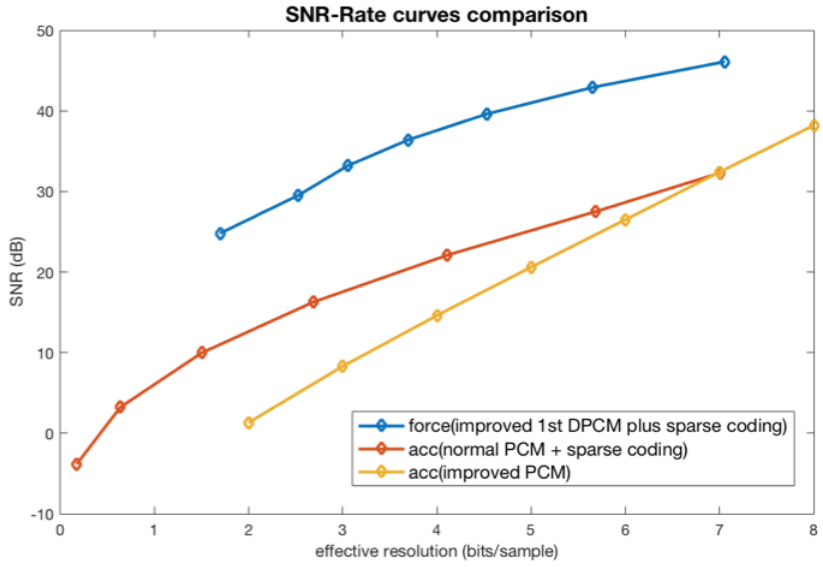


Figure 5.16: Distortion-rate curves of several coding schemes

Chapter 6

Conclusions and Future Work

6.1 Conclusions

A sample-based coding solution for mass input of kinaesthetic and tactile signals that will satisfy the strictest latency requirement of haptic applications is proposed and elaborated in this thesis. The system is implemented in MATLAB and evaluated both in separate modules and as a whole, on a texture tactile database, of which force and acceleration data are our focus. The update rate is 1kHz and sample resolution is 16 bits for the uncompressed data. All input channels are transmitted in one packet and we accommodate up to 300 channels in the experiments. Therefore, the data rate of a single channel is 16kbps, the maximum packet size is 600 Bytes and the packet rate of the system is 1000 packets/s. Individual samples are compressed to 2b – 8b in all experiments followed. Packet overheads, channel coding, network delay and noise are not included in our discussion. The performance measurements used are SNR, LSD, SVR, DRS and distortion-rate curve as mentioned in Chapter 2 and 4.

Two schemes for modifying the conventional DPCM/PCM, i.e. non-uniform Lloyds SQ and uniform SQ with companders, are discussed and compared. Results show that although non-uniform SQ is competitive in terms of SNR, its performance is extremely sensitive to the choice of training set. Uniform SQ with optimal Laplacian compander, on the other hand, yields good average SNR even with the same parameter shared among all channels, and is flexible for switching among resolutions. Therefore, uniform SQ with Laplacian compander is the recommended scheme. Its spectral performance is then analysed, mainly for perceptual considerations of vibrotactile signals. The results imply that perceptual transparency may be harder to achieve on acceleration data with the current scheme. However, we may only make conclusions after future perceptual studies are conducted.

The lossless layer is judged with DRSs first calculated for lossless layer combined with conventional PCM/DPCM, and then for the same lossless layer combined with the recommended lossy layer design, in which the resulting data is less sparse. This results suggests that force signal is suitable for the entire proposed compression scheme, while acceleration signal will be better coded with the lossy layer or the lossless layer alone.

Finally, the distortion-rate curves of several schemes are plotted. It suggests that the acceleration PCM data should best be coded with the proposed lossless layer alone, and force DPCM data should be compressed with both the lossy and lossless layer.

6.2 Future Work

Haptic coding research is still in its childhood, as has been discussed in details in Section 1.2. To the best of my knowledge, this thesis is a pioneer work in proposing and analysing a real-time haptic coding design aimed to cover both kinaesthetic and tactile devices, and optimized for multiple channels. There are still many underexplored aspects and deficiencies in experimental setup. I hereby list some insights for future work on this topic.

First of all, more databases and online data should be employed in the study of the proposed system. In this thesis study, the whole design of the system is implemented only in MATLAB, which certainly contains idealism in modelling the reality, neglecting some key factors that would influence compression performance. We use only one database for testing, but as far as we know, [43] offers similar surface tactile database containing tool-mediated acceleration and friction recordings of 43 textures. If more of such databases can be included, the conclusions made about this system in this thesis can be more convincingly validated. Of course, the optimal solution is implementing the system with executable codes in real haptic systems, and gather fresh online data. It will allow us to learn more about the arriving patterns, distributions and features of data in real haptic applications, and the study of parameter tuning such as the constant in compressor function and group size in sparse coding will make sense.

Secondly, perceptual analysis of the compression system based on online experiments with human is not performed in this thesis, and PD approach is not explicitly incorporated into the design, although we leave space for plug-ins of JNDs. We take advantage of data sparsity in the proposed scheme, and by sparsity we mean the percentage of zero index in one transmission. Why should groups with zero indexes be omitted during transmission? It is because they barely contain any information, since we know they are all zeros once we see the flags. It is the same as the PD. We know the samples within the PD contains superfluous information for human perception and can be reconstructed without knowing their values. Therefore, in future studies we can set the indexes of samples within deadbands to zero, which is particularly convenient for the DPCM case. Then we can evaluate its efficacy on human. But all these can only happen when we have access to online data.

Thirdly, parameter tuning is not the focus of this study and thus all relevant parameters are tuned manually for the optimal points throughout the experiments. In the future, when more haptic data are readily available to us, we can conclude more on the relation between data and parameters and search for ways for intelligent parameter tuning.

All in all, I consider real-time haptic coding for mass input and various signal types a promising topic in the coming 5G era. I heartily wish more fruits to be harvested in this field in the coming years.

Bibliography

[1] Jeffrey G Andrews, Stefano Buzzi, Wan Choi, Stephen V Hanly, Angel Lozano, Anthony CK Soong, and Jianzhong Charlie Zhang. What will 5g be? *IEEE Journal on Selected Areas in Communications*, 32(6):1065–1082, 2014.

[2] Adnan Aijaz, Mischa Dohler, A Hamid Aghvami, Vasilis Friderikos, and Magnus Frodigh. Realizing the tactile internet: Haptic communications over next generation 5g cellular networks. *arXiv preprint arXiv:1510.02826*, 2015.

[3] Martin Kuschel, Philipp Kremer, and Martin Buss. Passive haptic data-compression methods with perceptual coding for bilateral presence systems. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 39(6):1142–1151, 2009.

[4] Sandra Hirche, Peter Hinterseer, Eckehard Steinbach, and Martin Buss. Transparent data reduction in networked telepresence and teleaction systems. part i: Communication without time delay. *Presence: Teleoperators and Virtual Environments*, 16(5): 523–531, 2007.

[5] Thomas H Massie and J Kenneth Salisbury. The phantom haptic interface: A device for probing virtual objects. In *Proceedings of the ASME winter annual meeting, symposium on haptic interfaces for virtual environment and teleoperator systems*, volume 55, pages 295–300. Chicago, IL, 1994.

[6] Mel Slater and Sylvia Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and virtual environments*, 6(6):603–616, 1997.

[7] Thomas B Sheridan. Musings on telepresence and virtual presence. *Presence: Teleoperators & Virtual Environments*, 1(1):120–126, 1992.

[8] Hong Z Tan, Mandayam A Srinivasan, Brian Eberman, and Belinda Cheng. Human factors for the design of force-reflecting haptic interfaces. *Dynamic Systems and Control*, 55(1):353–359, 1994.

[9] Cyrus Shahabi, Antonio Ortega, and Mohammad R Kolahdouzan. A comparison of different haptic compression techniques. In *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*, volume 1, pages 657–660. IEEE, 2002.

[10] A Ortega and Y Liu. Lossy compression of haptic data. *Touch in virtual environments: Haptics and the design of interactive systems*, pages 119–136, 2002.

[11] Christoph W Borst. Predictive coding for efficient host-device communication in a pneumatic force-feedback display. In *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics Conference*, pages 596–599. IEEE, 2005.

[12] Shogo Okamoto and Yoji Yamada. Perceptual properties of vibrotactile material texture: Effects of amplitude changes and stimuli beneath detection thresholds. In *System Integration (SII), 2010 IEEE/SICE International Symposium on*, pages 384–389. IEEE, 2010.

[13] Shogo Okamoto and Yoji Yamada. Lossy data compression of vibrotactile material-like textures. *IEEE transactions on haptics*, 6(1):69–80, 2013.

[14] Rahul Chaudhari, Burak Çizmeci, Katherine J Kuchenbecker, Seungmoon Choi, and Eckehard 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*, pages 409–418. ACM, 2012.

[15] Hiroyuki Tanaka and Kouhei Ohnishi. Lossy compression of haptic data by using dct. *IEEJ Transactions on Industry Applications*, 130(8), 2010.

[16] Tomohiro Nakano, Seiji Uozumi, Rolf Johansson, and Kouhei Ohnishi. A quantization method for haptic data lossy compression. In *Mechatronics (ICM), 2015 IEEE International Conference on*, pages 126–131. IEEE, 2015.

[17] Peter Hinterseer, E Steinbach, Sandra Hirche, and Martin Buss. A novel, psychophysically motivated transmission approach for haptic data streams in telepresence and teleaction systems. In *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 2, pages ii–1097. IEEE, 2005.

[18] Peter Hinterseer, Sandra Hirche, Subhasis Chaudhuri, Eckehard Steinbach, and Martin Buss. Perception-based data reduction and transmission of haptic data in telepresence and teleaction systems. *IEEE Transactions on Signal Processing*, 56(2):588–597, 2008.

[19] Michael F Zaeh, Stella Clarke, Peter Hinterseer, and Eckehard Steinbach. Telepresence across networks: A combined deadband and prediction approach. In *Tenth International Conference on Information Visualisation (IV'06)*, pages 597–604. IEEE, 2006.

[20] Iason Vittorias, Julius Kammerl, Sandra Hirche, and Eckehard Steinbach. Perceptual coding of haptic data in time-delayed teleoperation. In *EuroHaptics conference, 2009 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2009. Third Joint*, pages 208–213. IEEE, 2009.

[21] Loi Huynh and Mehrdad H Zadeh. Effects of haptic data compression on user performance in collaborative haptic virtual environment. In *Haptic, Audio and Visual Environments and Games (HAVE), 2014 IEEE International Symposium on*, pages 106–111. IEEE, 2014.

[22] Jae-Young Lee and Shahram Payandeh. Performance evaluation of haptic data compression methods in teleoperation systems. In *World Haptics Conference (WHC), 2011 IEEE*, pages 137–142. IEEE, 2011.

[23] Julius Kammerl, Rahul Chaudhari, and Eckehard Steinbach. Combining contact models with perceptual data reduction for efficient haptic data communication in networked ves. *IEEE Transactions on Instrumentation and Measurement*, 60(1):57–68, 2011.

[24] Julius Kammerl, Iason Vittorias, Verena Nitsch, Berthold Faerber, Eckehard Steinbach, and Sandra Hirche. Perception-based data reduction for haptic force-feedback signals using velocity-adaptive deadbands. *Presence: Teleoperators and Virtual Environments*, 19(5):450–462, 2010.

[25] N Sakr, J Zhou, ND Georganas, J Zhao, and X Shen. Prediction-based haptic data reduction and compression in tele-mentoring systems. In *Instrumentation and Measurement Technology Conference Proceedings, 2008. IMTC 2008. IEEE*, pages 1828–1832. IEEE, 2008.

[26] Rahul Chaudhari, Clemens Schuwerk, Mojtaba Danaei, and Eckehard Steinbach. Perceptual and bitrate-scalable coding of haptic surface texture signals. *IEEE Journal of Selected Topics in Signal Processing*, 9(3):462–473, 2015.

[27] Asad Tirmizi, Claudio Pacchierotti, Irfan Hussain, Gianluca Alberico, and Domenico Prattichizzo. A perceptually-motivated deadband compression approach for cutaneous haptic feedback. In *2016 IEEE Haptics Symposium (HAPTICS)*, pages 223–228. IEEE, 2016.

[28] Xiao-Dong Pang, Hong Z Tan, and Nathaniel I Durlach. Manual discrimination of force using active finger motion. *Perception & psychophysics*, 49(6):531–540, 1991.

[29] Jack M Loomis and Susan J Lederman. Tactual perception. *Handbook of perception and human performances*, 2:2, 1986.

[30] Susan J Lederman and Roberta L Klatzky. Haptic perception: A tutorial. *Attention, Perception, & Psychophysics*, 71(7):1439–1459, 2009.

[31] Sandra Hirche and Martin Buss. Human-oriented control for haptic teleoperation. *Proceedings of the IEEE*, 100(3):623–647, 2012.

[32] James C Craig. Difference threshold for intensity of tactile stimuli. *Perception & Psychophysics*, 11(2):150–152, 1972.

[33] Helena Pongrac. Vibrotactile perception: examining the coding of vibrations and the just noticeable difference under various conditions. *Multimedia systems*, 13(4):297–307, 2008.

[34] Yuan D. Enhancing your device design through tactile feedback.

[35] Jack Tigh Dennerlein, David B Martin, and Christopher Hasser. Force-feedback improves performance for steering and combined steering-targeting tasks. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 423–429. ACM, 2000.

[36] Jonas Forsslund, Michael Yip, and Eva-Lotta Sallnäs. Woodenhaptics: A starting kit for crafting force-reflecting spatial haptic devices. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 133–140. ACM, 2015.

[37] Domenico Prattichizzo, Francesco Chinello, Claudio Pacchierotti, and Monica Malvezzi. Towards wearability in fingertip haptics: a 3-dof wearable device for cutaneous force feedback. *IEEE Transactions on Haptics*, 6(4):506–516, 2013.

[38] Vincent Hayward and M Cruz-Hernandez. Tactile display device using distributed lateral skin stretch. In *Proceedings of the haptic interfaces for virtual environment and teleoperator systems symposium*, volume 69, pages 1309–1314, 2000.

[39] Samuel B Schorr, Zhan Fan Quek, Robert Y Romano, Ilana Nisky, William R Provancher, and Allison M Okamura. Sensory substitution via cutaneous skin stretch feedback. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 2341–2346. IEEE, 2013.

[40] Laura Winfield, John Glassmire, J Edward Colgate, and Michael Peshkin. T-pad: Tactile pattern display through variable friction reduction. In *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC'07)*, pages 421–426. IEEE, 2007.

[41] Peter Vary and Rainer Martin. *Digital speech transmission: Enhancement, coding and error concealment*. John Wiley & Sons, 2006.

[42] Heather Culbertson, Juan Jose Lopez Delgado, and Katherine J Kuchenbecker. The penn haptic texture toolkit for modeling, rendering, and evaluating haptic virtual textures. 2014.

[43] Matti Strese, Jun-Yong Lee, Clemens Schuwerk, Qingfu Han, Hyoung-Gook Kim, and Eckehard Steinbach. A haptic texture database for tool-mediated texture recognition and classification. In *Haptic, Audio and Visual Environments and Games (HAVE), 2014 IEEE International Symposium on*, pages 118–123. IEEE, 2014.

Appendix A

DRS Record Sheet

A.1 DRS of sparse coding layer

Res. \ Grp. Size		1	2	3	6	10	20	30	50	75	100
2b	(0.875)	0.937	0.968	0.978	0.987	0.989	0.989	0.986	0.981	0.976	0.972
3b	(0.8125)	0.929	0.954	0.960	0.959	0.951	0.933	0.921	0.905	0.895	0.889
4b	(0.75)	0.897	0.906	0.903	0.884	0.865	0.841	0.830	0.819	0.815	0.813
5b	(0.6875)	0.832	0.820	0.809	0.783	0.767	0.754	0.750	0.749	0.749	0.750
6b	(0.625)	0.743	0.718	0.707	0.692	0.686	0.685	0.686	0.687	0.687	0.687
7b	(0.5625)	0.645	0.624	0.619	0.618	0.620	0.622	0.623	0.624	0.625	0.625
8b	(0.5)	0.553	0.543	0.546	0.553	0.557	0.560	0.561	0.562	0.562	0.562

Figure A.1: Acceleration PCM, 300 parallel inputs

Res. \ Grp. Size		1	2	3	6	10	20	30	50	75	100
2b	(0.875)	0.935	0.964	0.973	0.979	0.978	0.972	0.966	0.958	0.952	0.948
3b	(0.8125)	0.918	0.938	0.941	0.934	0.922	0.902	0.891	0.882	0.877	0.876
4b	(0.75)	0.874	0.879	0.874	0.856	0.839	0.820	0.814	0.812	0.812	0.812
5b	(0.6875)	0.801	0.794	0.786	0.768	0.756	0.749	0.748	0.749	0.749	0.749
6b	(0.625)	0.711	0.701	0.697	0.688	0.684	0.685	0.685	0.686	0.687	0.687
7b	(0.5625)	0.620	0.616	0.616	0.617	0.619	0.622	0.623	0.624	0.624	0.624
8b	(0.5)	0.536	0.540	0.545	0.553	0.556	0.559	0.560	0.561	0.562	0.562

Figure A.2: Force 1st order DPCM, 300 parallel inputs

A.2 DRS records of the proposed compression system

Res. \ Grp. Size		1	2	3	6	10	20	30	50	75	100
2b	(0.875)	0.875	0.891	0.894	0.890	0.883	0.876				
3b	(0.8125)	0.841	0.842	0.838	0.825	0.815					
4b	(0.75)	0.809	0.799	0.789	0.769	0.754					
5b	(0.6875)	0.769	0.734	0.715	0.690						
6b	(0.625)	0.717	0.663	0.639							
7b	(0.5625)	0.647	0.632	0.588							
8b	(0.5)	0.559	0.509								

Figure A.3: Force 1st order DPCM, 300 parallel inputs

- Red and green regions represent the highest and lowest DRS values of that resolution respectively. Both are larger than the DRS with pure SQ.
- Grey regions means the DRS values lower than the value with pure SQ of that resolution.

TRITA EE 2016:175