Challenges and Opportunities for Designing Tactile Codecs from Audio Codecs

Xun Liu, Mischa Dohler, Toktam Mahmoodi, Hongbin Liu
Department of Informatics, King's College London
London WC2R 2LS, United Kingdom
{xun.2.liu, mischa.dohler, toktam.mahmoodi, hongbin.liu}@kcl.ac.uk

Abstract—**Haptic communications allow physical interaction over long distances and greatly complement conventional means of communications, such as audio and video. However, whilst standardized codecs for video and audio are well established, there is a lack of standardized codecs for haptics. This causes vendor lock-in and thereby greatly limits scalability, increases cost and prevents advanced usage scenarios with multi-sensors/actuators and multi-users. The aim of this paper is to introduce a new approach for understanding and encoding tactile signals, i.e. the sense of touch, among haptic interactions. Inspired by various audio codecs, we develop a similar methodology for tactile codecs. Notably, we demonstrate that tactile and audio signals are similar in both time and frequency domains, thereby allowing audio coding techniques to be adapted to tactile codecs with appropriate adjustments. We also present the differences between audio and tactile signals that should be considered in future designs. Moreover, in order to evaluate the performance of a tactile codec, we propose a potential direction of designing an objective quality metric which complements haptic mean opinion scores (h-MOS). This, we hope, will open the door for designing and assessing tactile codecs.**

Keywords—Haptic; texture; audio codec; tactile codec; objective quality metrics; haptic mean opinion score

I. INTRODUCTION

Audio and video communications allow users to see and talk to each other over long distances. With the development of multimedia technology, high quality audio-visual communication makes users feel present remotely to some extent. However, physical interaction/operation and a strong sense of immersion remains deficient to date. In fact, humans heavily count on haptic interaction with the environment in our daily life [1]. With the development of 5G in mobile and wireless networking, and the promises of higher data rate and close-to-zero latency (ultra-low latency) in communications, there are stronger motivations than in the past to complement the audio and video communication with other human senses and deliver a fully immersive experience remotely. Enabling such fully immersive remote experience is the first step in achieving the Internet of Skills [2], as desired by different industry sectors such as Healthcare and manufacturing.

It is currently known that involving haptic perception can significantly increase the degree of immersion for distant communications [3]. Haptics perception relies on two different human receptors that are kinesthetic and tactile. The former refers to the movement/activation of muscles and joints while the latter includes pressure, temperature, texture, among others. Design and development of codec for kinesthetic data has been well studied using different compression approaches such as sampling and quantization technologies, perceptual deadband (PD), and predictive coding [4]. The relevant family of codecs are to be developed and standardized in IEEE P1918.1. Therefore, in this paper, we focus on the tactile, and design of the tactile codec. Among different information that shapes the sense of touch, we devise our attention to texture information since it is more complex to be modelled comparing with pressure and temperature.

The main motivation for designing codecs, clearly, is to enable higher performance when transmitting data over communication path. Today's audio and video codecs can compensate for various shortcomings of the communication path such as recovering the lost subset of data, but also they are used to reduce the load on the communication by reducing the rate of transmitted data. Audio and video codecs are specifically used in wireless communications given the higher probability of data loss and lower availability of bandwidth. Both audio and video codecs work based on understanding the limitation in human perception in order to manipulate the corresponding data at the transmitter, before sending over communication path. Despite improvements in lowering the data loss and increasing the data rate have been significant in the evolution of communication network, and there will be yet another step forward with 5G, the need for encoding data remains the same.

In order to record and display tactile signal, researchers made efforts to model it in recent years [5] - [9]. Based on developed models for tactile signal, some researchers are seeking the connections between audio and tactile signals for the purpose of designing tactile codec. Ref. [10]

978-1-5386-3873-6/17/$31.00 ©2017 IEEE

conducted subjective tests showing that masking phenomenon of audio signals applies to tactile signals, but it did not clearly describe the relation between audio and tactile signals from a fundamental perspective. In this paper, we demonstrate that tactile signals and audio signals are similar in time and frequency domains such that we can potentially adapt almost all of the well-developed audio compression approaches to tactile codecs. The differences that should be noticed during the transformation from audio codec to tactile codec are listed as well. Once the relation between audio and tactile signals are clear, we point a potential direction of designing and evaluating tactile codecs.

The remainder of this paper is organized as follows. In Sections II and III, we explicitly present the similarities and differences between tactile and audio signals. Afterwards, we compare the main audio coding standards and techniques in Section IV. Besides, this section investigates the possibility of designing objective quality metrics for tactile codecs. Finally, in Section V, the article is concluded and future works are discussed.

II. Audio-Tactile Similarities

A. Mechanisms

How do humans hear? Although the structure of the human ear is quite sophisticated made up of outer, middle, and inner ear, the hearing process is straightforward [11]. Originally, sounds are the vibrations of air with the frequency and intensity of the vibrations determining the pitch and volume. When oscillations occur in air, the generated sound waves are collected by pinna (a part of the outer ear) before they are transmitted to the middle ear through the ear canal. In the middle ear, there is a key part called tympanic membrane which is very sensitive to vibrations. When the sound waves hit the tympanic membrane, the vibrations are transferred to the inner ear. After that, different areas of the cochlea (a part of the inner ear) representing various frequencies get excited according to the tone of the sound waves and thereupon neural signals are generated which are transmitted to the brain. Eventually, humans hear the sounds and are capable of recognizing the location and category of the sound source.

TABLE I. Function, Applications And Respective Frequency Range Of Four Types Of Mechanoreceptors

	Merkel cell	Ruffini ending	Meissner corpuscle	Pacinian corpuscle
Best stimulus	Pressure, edges, corner, points	Stretch	Lateral motion	High-frequency vibration
Example	Reading Braille	Holding large objects	Sensing Slippage of objects	Sensing texture
Freq. range (Hz)	0-100	/	1-300	5-1000
Most Sensitive Freq. (Hz)	5	/	50	200

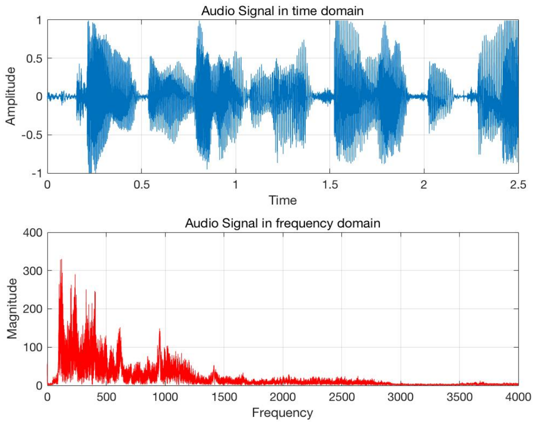

Fig. 1: Example of audio signal in time and frequency domain.

How do humans perceive touch? Whilst principles of hearing and vision are more or less established, touch is the least explored human perception. Reference [12] presents a thorough introduction of human somatosensory systems. Basically, human skin has eight types of mechanoreceptors controlling the sense of touch. Our research focuses on the glabrous skin of hands including four of the mechanoreceptors: Merkel cells, Ruffini endings, Meissner corpuscles and Pacinian corpuscles, because they play a vital role for humans to perceive the world. The main function of the mechanoreceptors and corresponding examples are shown in TABLE I. Humans can sense objects through an intermediate tool used to explore, based on which the authors of [13] have conducted experiments to improve contact realism. When the attendees used a probe to tap on objects, they feel high-frequency transient forces followed by stationary forces. Otherwise, if the attendees slid along the surface, they perceived continuous high-frequency vibrational forces. These results are also supported by [5] which shows that the high-frequency vibrational transients are the main cues for human to discriminate the texture of different materials.

Combining with the information in TABLE I, we can see that Pacinian corpuscles are responsible for sensing texture of various objects depending on the vibrations. In summary, humans have a sense of hearing as well as touch because a part of the human body perceives the high-frequency ambient physical vibrations.

B. Representations

Audio and tactile signals are the mathematic representations of sound and touch respectively; a comparison between both signals depends on the form of representations. In order to demonstrate the perceptual characteristics (pitch, loudness, timbre, duration) of sound, a joint expression in both time and frequency domain is chosen and typical waveforms are shown in Fig. 1.

As for tactile texture signals, initially, research was limited to virtual environments in which contact between users and virtual objects was simply modelled as a spring system. In [6], material properties were represented by stiffness and damping of the spring system; the authors of [7] generated a discrete height-field texture and mapped it to the virtual objects to simulate haptic texture using a well-known graphical technique called “bump mapping”. However, only few haptic texture models were built according to physical measurements of real objects until Okamura et al. built a decaying sinusoid model for acceleration transients recorded from taps on objects [8] and authors of [9] measured texture as height profiles using a subtle laser scanner. Moreover, [13] proved that purely kinesthetic systems lack the real feeling of the surface properties of objects and introduced high-frequency transients. In [5], the authors slid the surface of objects to measure the acceleration signals using a stylus mounted with an accelerometer. Furthermore, subjective experiments were conducted with the results demonstrating that involving high-frequency acceleration signals significantly increased the degree of immersion and rendering texture as vibrotactile signals improved the realism rate of tactile interaction. As a result, we believe that acceleration signals are the most appropriate and effective way to represent the sense of touch. Typical waveforms of an acceleration signal in time and frequency domain are shown in Fig. 2.

With a proper representation of sound and touch at hand, we can now compare the temporal and spectral properties. From time domain, despite the different range of amplitudes, we can see that both audio and tactile signal are vibrational waveforms with varying and relatively high frequencies. In another words, audio and tactile signals are temporally very similar.

In terms of spectral properties, audio signals can be periodic or aperiodic according to various sound sources while tactile signals are generally aperiodic due to the random nature of texture of object. Periodic audio signals are made up of a fundamental frequency and a series of multiples of the fundamental frequency. By contrast, aperiodic audio signals can be expressed as a sum of non-harmonically related sine waveforms with different frequencies and this also applies to aperiodic tactile signals. Actually, you can treat tactile signals as a kind of aperiodic audio signals; therefore, they are similar in the frequency domain as well. Consequently, it is possible to adapt well developed audio codecs to tactile codecs.

III. Audio-Tactile Differences

A. Bandwidth

The frequency range that humans normally can hear is from 20 Hz to 20 kHz as long as the sound intensity is above a certain level. According to the Nyquist-Shannon sampling theorem [14] - [15], the sampling frequency must be at least 40 kHz (Nyquist Rate) for perfect reconstruction. From Table I, we can see that the Pacinian corpuscles are able to sense vibrations of up to only 1 kHz, but we still call the tactile signal high-frequency vibrational signal because it is relatively high compared with the low-frequency (generally lower than 5 Hz) movement of human joints and muscles. Similarly, the sampling frequency of tactile signals needs to be over 2 kHz which is far below that of audio signals.

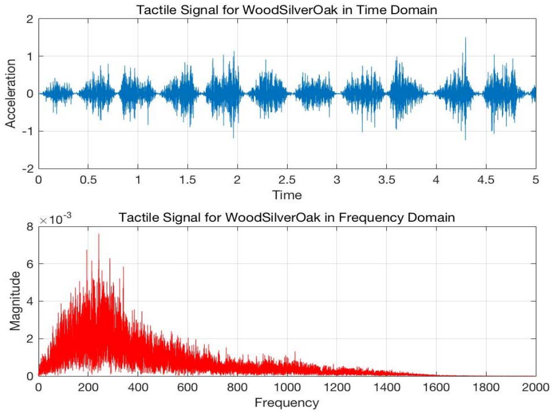

Fig. 2 Example of tactile texture signal in time and frequency domain

B. Dimensions

Audio signals are low in dimensionality, whilst Pacinian corpuscles are capable of sensing multi-dimensional vibrations. However, human are not quite sensitive to directions of the vibrations [16]. As a result, we need to transform three-axis vibrational tactile signal into one-axis signal which are convenient for modelling and rendering in haptic communication systems.

C. Requirements of Bitrate and Delay

The aim of lossy data compression is either to attain a performance as good as possible for a given bitrate or to reduce the bitrate as much as possible to obtain a given quality; hence, there is always a trade-off between fidelity and bitrate. Due to different sensitivity and characteristics of ear and skin, the requirements of fidelity and bitrate for audio and tactile codecs are not the same. Besides, to fulfil the requirements of various applications, the latency needs to be considered. According to the empirical results of [17], 70-80 ms time-delay is non-observable for ordinary hearing (excluding e.g. professional musicians) whereas the detection threshold of time-delay for tactile stimuli is just approximately 41 ms. The requirement for latency of tactile codecs is much more rigorous than that of audio codecs. Consequently, we need to improve the design of audio codecs for the sake of reducing time-delay as well as allowing for some buffer time due to the communications delay.

IV. Design Opportunities

A. Fundamental Audio Data Reduction Techniques

Lossy data reduction techniques are classified into waveform-following methods and analysis-by-synthesis approaches [18]. The most famous coders of the former type are pulse code modulation (PCM) and adaptive differential pulse code modulation (ADPCM) which aim to reconstruct the original signals as precisely as possible, hence the performance of them is called toll quality that sets the standard to which the performance of other codecs is compared. On the other hand, analysis-by-synthesis approaches only recreate the most important elements of the original signals depending on linear predictive models and perceptual weighting. The code-excited linear prediction (CELP) is the most widely used one but many variants of it, e.g. Conjugate-Structure algebraic-code-excited linear prediction (CS-ACELP), have also been used successfully. Note that all of the analysis-by-synthesis approaches are based on linear predictive coding (LPC) which is discussed in more details below.

B. Linear Predictive Coding

Since initiated by Robert M. Grey in 1966 [19], LPC has been developed rapidly and become one of the most useful coding techniques that achieves good performance at low bitrate. The core idea of LPC is to predict the signal $x(n)$ by means of a linear combination of previous samples:

$$\hat{x}(n) = \sum_{i=1}^{N} a_i x(n-i), \qquad (1)$$

where N is the number of previous samples used for prediction, $\mathrm{a_i}$ are the linear prediction coefficients and $\hat{\mathrm{x}}[\mathrm{n}]$ is the estimate of $\mathrm{x}[\mathrm{n}]$. Hence we can get the prediction error as:

$$e(n) = x(n) - \hat{x}(n) = x(n) - \sum_{i=1}^{N} a_i x(n-i). \qquad (2)$$

The objective is to obtain the optimal a_i that minimizes the square of prediction error,

$$\{a_i\}_{opt} = argmin[e^2(n)] \quad . \qquad (3)$$

The problem above can be solved via an autocorrelation or covariance method [20]. As of today, LPC is widely used as a fundamental technique in many audio coding standards, such as G.728/G.729/MPEG-4.

C. Codec Family

As stated above, audio and tactile signals are similar temporally and spectrally on the condition that we represent tactile signals as acceleration signals. As a result, it is theoretically possible to transform the audio codecs to tactile codecs regardless of the differences mentioned above. Indeed, in [10] and [21], an advanced audio codec called G.729 is successfully adapted to tactile and achieves a high data compression ratio of 8:1 with good performance. However, we believe it is essential to build a tactile codec family just as what we have done for audio data compression in the past decades to confront potential challenges of the upcoming haptic era. Here we list the core technique, bitrate, delay and quality of main audio codecs in Table II, in which the capital letters following the names stands for the core techniques, framing size stands for delay and MOS (mean opinion score) is the quality metric for audio codecs. Obviously, there is a trade-off between bitrate, delay and fidelity; for instance, G.729 has half bitrate as G.728 but has much larger delay than G.728. One may also notice that G.729 has a higher MOS than G.728, but it does not mean G.729 achieves a better performance than G.728 in the field of tactile signals because the MOS score is only the quality metric designed for audio signals. Designing an exclusive performance assessment system for tactile codecs is thus another goal and is discussed in next sub-section. Although some codecs are more sophisticated than others, each of the audio codec has its own advantages and is suitable for particular applications. We are convinced that this situation is also applicable to tactile codecs because haptic communications can be widely used in many areas in the future.

TABLE II. Comparison Between Main Audio Coding Standards

Audio Codec	Bitrate (Kbps)	Framing Size (ms)	MOS Score
G.711 PCM	64	0.125	4.1
G.726 ADPCM	32	0.125	3.85
G.728 LD-CELP	16	0.625	3.61
G.729 CS-ACELP	8	10	3.92
G.729a CS-ACELP	8	10	3.7
G.723.1 MP-MLQ	6.3	30	3.9
G.723.1 ACELP	5.3	30	3.65

D. Performance Assessment

There are two approaches to assessing the performance of a codec: either conducting subjective experiments or applying objective quality metrics. For the former one, it is not unlikely to set various subjective experiments since [22] provides an integrated haptic system which is able to record the acceleration data and display it to users. In [10], a typical setup of subjective test is proposed. Nevertheless, conducting such subjective experiments is always time- and money-consuming. On the other hand, designing objective quality metrics for tactile signals has drawn little attention to date despite many objective indicators to assess the intelligibility and quality of reconstructed audio signals being available; examples are diagnostic rhyme test (DRT), enhanced modified bark spectral distance (EMBSD), diagnostic acceptability measure (DAM), perceptual evaluation of speech quality (PESQ) and E-Model [18].

Due to the similarities between audio and tactile signals, we advocate for some equivalent objective quality metrics according to some unique features of tactile signals and human perceptual threshold. More importantly, the establishment of objective quality metrics is beneficial for developing better tactile codec since we can quantitatively compare different codecs. As an incipient attempt, [23] proposes the Haptic Perceptually Weighted Peak Signal-To-Noise Ratio (HPWPSNR) that is modified from Peak Signal-to-Noise Ratio (PSNR), but this indictor is merely applied to kinesthetic data rather than tactile data.

V. Conclusion

In this article, we focused on the design and development of codecs for tactile signals in order to enable digitization and delivery of sense of touch over long distance communication. We demonstrated that tactile signals and audio signals are inherently similar in time domain and frequency domain which means we can apply audio codec principles and other techniques to the tactile domain. We also discussed the differences between audio and tactile signals that need to be considered for designing tactile codecs. In order to make tactile codecs work for all the potential applications in the upcoming haptic era, we proposed the idea of building a tactile codec family that is similar to the audio codec family. Furthermore, and since there is no objective quality metric to evaluate the performance of tactile codec today, this paper provides some thread of designing objective performance indicators. Despite the clear direction for developing tactile codecs and corresponding objective quality metrics, a large number of challenges remain to be solved during the transformation from audio to tactile, such as the higher dimensionality; this, however, is left for future work.

Acknowledgment

We greatly appreciate that Matti Strese et al. from Technical University of Munich share the database of texture signals they recorded.

Bibliography

[1] K.E. MacLean, *Haptic interaction design for everyday interfaces.* Reviews of Human Factors and Ergonomics, 2008. **4**(1): pp. 149-194.

[2] M. Dohler. *Internet of Skills – where robotics meets AI, 5G and the Tactile Internet*. IEEE COMSOC TECHNOLOGY NEWS 2017 [cited 2017 19/02]; Available from: http://www.comsoc.org/ctn/global-reach-will-tactile-internet-globalize-your-skill-set.

[3] E. Steinbach, et al., *Haptic Communications.* Proceedings of the IEEE, 2012. **100**(4): pp. 937-956.

[4] E. Steinbach, S. Hirche, J. Kammerl, I. Vittorias, and R. Chaudhari, *Haptic Data Compression and Communication.* IEEE Signal Processing Magazine, 2011. **28**(1): pp. 87-96.

[5] W. McMahan, J.M. Romano, A.M.A. Rahuman, and K.J. Kuchenbecker. *High frequency acceleration feedback significantly increases the realism of haptically rendered textured surfaces*. in *Haptics Symposium, 2010 IEEE*. 2010. IEEE.

[6] C.B. Zilles and J.K. Salisbury. *A constraint-based god-object method for haptic display*. in *Intelligent Robots and Systems 95.'Human Robot Interaction and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on*. 1995. IEEE.

[7] C. Basdogan, C. Ho, and M.A. Srinivasan. *A raybased haptic rendering technique for displaying shape and texture of 3D objects in virtual environments*. in *ASME Winter Annual Meeting*. 1997.

[8] A.M. Okamura, M.R. Cutkosky, and J.T. Dennerlein, *Reality-based models for vibration feedback in virtual environments.* Mechatronics, IEEE/ASME Transactions on, 2001. **6**(3): pp. 245-252.

[9] S. Okamoto and Y. 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*. 2010. IEEE.

[10] R. Chaudhari, C. Schuwerk, M. Danaei, and E. Steinbach, *Perceptual and Bitrate-Scalable Coding of Haptic Surface Texture Signals.* IEEE Journal of Selected Topics in Signal Processing, 2015. **9**(3): pp. 462-473.

[11] P.W. Alberti, *The anatomy and physiology of the ear and hearing.* Occupational exposure to noise: Evaluation, prevention, and control, 2001: pp. 53-62.

[12] E.R. Kandel, J.H. Schwartz, T.M. Jessell, S.A. Siegelbaum, and A.J. Hudspeth, *Principles of Neural Science, Fifth Edition.* 2013: McGraw-Hill Education. pp. 499-510

[13] K.J. Kuchenbecker, J. Fiene, and G. Niemeyer, *Improving contact realism through event-based haptic feedback.* Visualization and Computer Graphics, IEEE Transactions on, 2006. **12**(2): pp. 219-230.

[14] C.E. Shannon, *A mathematical theory of communication.* The Bell System Technical Journal, 1948. **27**(4): pp. 623-656.

[15] C.E. Shannon, *A mathematical theory of communication.* The Bell System Technical Journal, 1948. **27**(3): pp. 379-423.

[16] N. Landin, J. Romano, W. McMahan, and K. Kuchenbecker, *Dimensional reduction of high-frequency accelerations for haptic rendering.* Haptics: Generating and perceiving tangible sensations, 2010: pp. 79-86.

[17] S. Okamoto, M. Konyo, S. Saga, and S. Tadokoro. *Identification of cutaneous detection thresholds against time-delay stimuli for tactile displays*. in *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*. 2008. IEEE.

[18] J.D. Gibson, *Speech coding methods, standards, and applications.* IEEE Circuits and Systems Magazine, 2005. **5**(4): pp. 30-49.

[19] D.O. Shaughnessy, *Linear predictive coding.* IEEE Potentials, 1988. **7**(1): pp. 29-32.

[20] L.R. Rabiner and R.W. Schafer, *Introduction to digital speech processing.* Foundations and trends in signal processing, 2007. **1**(1): pp. 1-194.

[21] R. Chaudhari, B. Çizmeci, K.J. Kuchenbecker, S. Choi, and E. Steinbach. *Low bitrate source-filter model based compression of vibrotactile texture signals in haptic teleoperation*. in *Proceedings of the 20th ACM international conference on Multimedia*. 2012. ACM.

[22] H. Culbertson, J.J. Lopez Delgado, and K.J. Kuchenbecker. *One hundred data-driven haptic texture models and open-source methods for rendering on 3d objects*. in *Haptics Symposium (HAPTICS), 2014 IEEE*. 2014. IEEE.

[23] N. Sakr, N. Georganas, and J. Zhao. *A perceptual quality metric for haptic signals*. in *Haptic, Audio and Visual Environments and Games, 2007. HAVE 2007. IEEE International Workshop on*. 2007. IEEE.