

Setting the Yardstick: A Quantitative Metric for Effectively Measuring Tactile Internet

J.P. Verburg^{*§}, H.J.C. Kroep^{*§}, V. Gokhale^{*}, R. Venkatesha Prasad^{*}, and V. Rao[†]

^{*}Delft University of Technology, The Netherlands, [†]Cognizant Technology Solutions, The Netherlands

Abstract—The next frontier in communications is *teleoperation* – manipulation and control of remote environments. Compared to conventional networked applications, teleoperation poses widely different requirements, ultra-low latency (ULL) being the primary one. Teleoperation, along with a host of other applications requiring ULL communication, is termed as *Tactile Internet* (TI). A significant redesign of conventional networking techniques is necessary to realize TI applications. Further, these advancements can be evaluated only when meaningful performance metrics are available. However, existing TI performance metrics fall severely short of comprehensively characterizing TI performance. In this paper, we take the first step towards bridging this gap. To this end, we propose a method that captures the fine-grained performance of TI in terms of delay and precision. We take Dynamic Time Warping (DTW) as the basis of our work and identify whether it is sufficient in characterizing TI systems. We refine DTW by developing a framework called *Effective Time- and Value-Offset (ETVO)* that extracts fine-grained time and value offsets between input and output signals of TI. Using ETVO, we present two quantitative metrics for TI – *Effective Delay-Derivative (EDD)* and *Effective Root Mean Square Error*. Through rigorous experiments conducted on a realistic TI setup, we demonstrate the potential of the proposed metrics to precisely characterize TI interactions.

I. INTRODUCTION

Tactile Internet (TI) [1] is irrefutably at the forefront amid several emerging technological innovations that are foreseen to revolutionize future industries and the lifestyle of humans. The crux of TI is its potential to enable *teleoperation* – transportation of physical skills of humans for manipulation and control of remote environments. Sensory feedback, including vibrotactile, kinesthetic, audio, and video modalities, transmitted via TI, should provide a feeling of collocation between geographically distant locations. These can facilitate humans to remotely perform activities in uninhabitable or resource-constrained locations as if they are physically present there. TI is not just limited to teleoperation applications such as telesurgery, remote disaster management and telerepair of machinery, but also the other sectors like education, health, and manufacturing [2].

TI applications are significantly different from the existing real-time applications served by the Internet. The transmission of tactile sensor data and the sensory feedback have latency requirements in orders of magnitude lower than the current real-time applications on the Internet. Furthermore, some TI applications would also need ultra-high reliability, which is not yet supported on current networks. Thanks to the vision

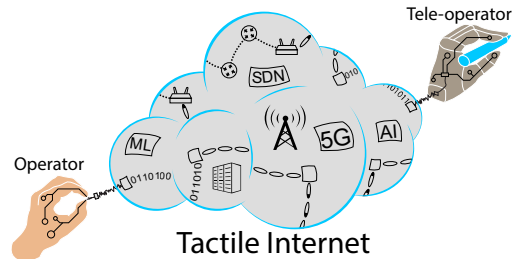


Fig. 1. A schematic representation of an E2E TI application.

and recent advancements in the field of 5G, particularly in the context of ultra-reliable, low-latency communication (URLLC), TI will benefit enormously [3]–[5].

As with real-time applications on the Internet, objectively measuring the quality of a session is important for several reasons: for adapting session quality based on network dynamics, estimating the quality of a session a priori (which is important before executing mission-critical applications), developing and benchmarking of new solutions at various layers of the protocol stack including application layer. Unlike the real-time applications on the Internet today, TI applications are complex due to the active involvement of humans in the entire control loop. However, it is imperative to establish objective metrics for measuring and comparing TI session performances.

Measuring TI performance. For fine-grained performance analysis, a TI system should be evaluated based on how quickly and accurately it can reproduce actions at the remote end. Hence, it is crucial to determine the degradation it introduces in terms of *time* and *value* collectively. While Quality of Service (QoS) metrics, such as latency, reliability, and throughput, are known to characterize real-time network applications well, they fall short for TI applications. For instance, a TI solution with an average latency of 15 ms is not necessarily inferior compared to another with 10 ms average latency. This is because the former could be intelligently delaying the signals to avoid over-provisioning of resources when the human perception is less sensitive to latency; for example, in case of medium dynamic environments [6]. Similar arguments can be constructed for other QoS metrics. On the other hand, literature provides several works that propose root-mean-square error (RMSE) based Quality of Experience (QoE) metrics for qualitative performance evaluation of TI [7], [8]. Essentially, RMSE is not designed to extract time

[§]J.P. Verburg and H.J.C. Kroep assert equal contributions.

and value offsets separately. Time offset is the delay variations due to network and value offset may be due to noise or packet loss (inducing loss of data). We demonstrate this limitation in Figure 2. While it is easy to observe that among y_1 and y_2 , the profile of y_1 is closer to x due to the presence of the dominant peak, the RMSE of y_1 is 19% higher than that of y_2 . Hence, the QoE metrics do not always indicate close to real performance of TI.

The lack of a framework for performance characterization of TI at a fine-grained level severely impedes the overall progress of TI. This forms the primary motivation for this work.

Our contributions. Our approach is to devise a method that is capable of extracting time and value offset between the input and output signals of TI. The strength of this approach lies in the fact that it can be applied to any end-to-end TI system (starting from sensors on one end all the way to the actuators on the other end) in a manner that is agnostic to the TI modules. We take Dynamic Time Warping (DTW), a common method for determining sample-wise similarity between two signals, as the starting point for our work. Our contributions in this work are as follows.

- We present a detailed discussion on DTW in characterizing TI sessions and show where we can further use it for TI applications.
- We present a concrete mathematical framework that we call *Effective Time- and Value-Offset (ETVO)* that extract time and value offset between input and output signals of a TI system, which hitherto was impossible.
- Based on ETVO, we propose two quantitative metrics – *Effective Delay-Derivative (EDD)* and *Effective RMSE (ERMSE)* that can jointly characterize the TI system in a fine-grained manner.
- Through experiments conducted on a realistic TI setup, we demonstrate the efficacy of the proposed method in supporting the characterization of a TI session close to real.

The remainder of the paper is organized as follows. We review the literature for potential performance metrics for TI in Section II. We then provide a brief overview of DTW and its insufficiency specifically for characterization of TI interactions in Section III. We present the complete design of the proposed performance metrics in Section IV. In Section V, we describe the experimental setup, and present our findings in Section VI. Finally, we state our conclusions and future directions in Section VII.

II. RELATED WORK

A. TI performance metrics

Standard QoS metrics have been extensively employed, and several QoE metrics have been devised for evaluating TI systems. In this section, we present an overview of these.

1) **QoS:** Traditional QoS metrics for network performance include latency, jitter, reliability, and throughput. Several modular designs of TI systems use QoS for characterizing TI performance. While *Admux*, an adaptive multiplexer for TI

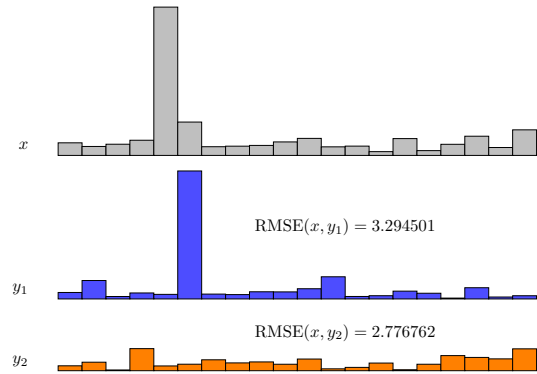


Fig. 2. Demonstration of wrong conclusions drawn by RMSE although the shapes of x and y_1 are more similar than the shapes of x and y_2 . The abscissa has units in time and the ordinate in amplitude of the signal.

proposed by Eid et al. [9], uses all of the above metrics, the multiplexing scheme by Cizmeci et al. focuses on throughput and latency [10]. Hinterseer et al. [11] proposed a haptic codec that focuses on reducing the application throughput. The codec is based on perceptual Deadband, where packets are not sent if the changes are below the level of human perception. The congestion control scheme by Gokhale et al. [12] aims to contain latency and jitter within their permissible QoS limits. Further, a string of works has emerged recently that attempt to address the ULL requirement of TI by leveraging the advancements in the field of 5G networks. The works in [13]–[15] provide a detailed discussion on the vision and progress in this direction. While QoS metrics play a vital role in characterizing the network performance, they are insufficient for comprehensive characterization of TI, as described in Section I.

2) **QoE:** Subjective QoE metrics aim to capture the quality of teleoperation by having the human users subjectively rate their experience. A couple of works that adopt this approach include [8], [16]. Since this method is cumbersome and resource-intensive, objective QoE metrics that estimate the quality of teleoperation as experienced by the human controller have also been designed. Hinterseer et al. [11] exploited the idea of human perception having a logarithmic relationship with the haptic stimulus. A framework was developed to validate this by using the traditional Peak Signal-to-Noise Ratio (PSNR) of the reconstructed haptic signal as the QoE metric. Later, Sakr et al. [17] introduced Haptic Perceptually Weighted Peak Signal to Noise Ratio (HPW-PSNR) that is aware of human insensitivity to Deadband of the haptic signal, the idea initially presented in [11]. Chaudhuri et al. followed up on this to propose Perceptual Mean Square Error (PMSE) that maps MSE to the human perceptual domain [7]. Recently, Hassen et al. proposed the Haptic Structure SIMilarity (HSSIM) index to improve the objective estimation of human perception. HSSIM extracts the similarity between original and reconstructed haptic signals. All of the above metrics are based primarily on RMSE, which works by combining degradation in both time and value domains.

B. Signal similarity metrics

Determining the similarity between two signals is an age-old problem in the signal processing domain and has been extensively researched due to its numerous applications, such as speech and gesture recognition [18]. In this section, we discuss some of the techniques devised for this purpose and examine their applicability in extracting time and value errors. Cross-correlation computes the time-offset (delay) between the two signals that maximizes their dot product [19]. It is well known that shared networks usually manifest highly non-deterministic and time-varying characteristics. Hence, a constant delay is an incorrect choice for representing the entire TI characteristics. Another popular method known as Dynamic Time Warping (DTW) exists for signals encountering a time-varying delay. DTW conducts an exhaustive search to achieve sample-wise matching between the two signals in a manner that minimizes the cumulative Euclidean distance [20]. It provides an extremely useful construct in determining *how similar two signals are*. Since in the case of teleoperation, input and output signals are expected to be similar, our problem is to find out *how two similar signals are different*. DTW falls short of achieving this objective of TI. Several follow-up works on DTW exist, with each of them attempting to outperform DTW in one or more aspects. The most widely recognized ones include Edit Distance on Real sequences (EDR) [21], Edit distance with Real Penalty (ERP) [22], and Longest Common Sub-Sequence (LCSS) [23]. However, they manifest the inherent characteristics of DTW and hence are unsuitable for TI as will be detailed in the following sections. We discuss DTW in detail in the rest of the paper as it forms the basis of the ETVO design.

III. DTW: BACKGROUND AND ANALYSIS

In this section, we present the necessary mathematical background for DTW and analyze its shortcomings in addressing the stated objectives of this work. DTW constructs a *warp path* that indicates a sample-wise mapping between two time-series that minimizes their cumulative Euclidean distance.

A. Mathematical Representation

Let $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{g}}$ denote the N -length discrete time-series corresponding to input and output signals, respectively. Hence, $\tilde{\mathbf{f}}, \tilde{\mathbf{g}} \in \mathbb{R}^N$. Let $\tilde{\mathbf{W}}$ denote the set of all possible warp paths between $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{g}}$, where a warp path represents the sample-wise alignment. Let the $(k+1)$ -th point of a warp path $\tilde{\mathbf{w}} \in \tilde{\mathbf{W}}$ be denoted as $\tilde{\mathbf{w}}(k) = (\tilde{\mathbf{w}}_0(k), \tilde{\mathbf{w}}_1(k))$, where $\tilde{\mathbf{w}}_0, \tilde{\mathbf{w}}_1 \in \mathbb{N}^K$ and $K \in [N, 2N-1]$. For example, the warp path in Figure 3 is given as $[(0,0), (1,0), (2,0), (3,0), (4,1), (5,2), \dots]$. Essentially, $\tilde{\mathbf{w}}_0$ and $\tilde{\mathbf{w}}_1$ return the indices of $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{g}}$, respectively.

The entries in $\tilde{\mathbf{w}} \in \tilde{\mathbf{W}}$ must meet the following conditions:

- 1) Monotonicity and continuity:

$$\begin{aligned} \tilde{\mathbf{w}}_0(k) &\leq \tilde{\mathbf{w}}_0(k+1) \leq \tilde{\mathbf{w}}_0(k) + 1, \\ \tilde{\mathbf{w}}_1(k) &\leq \tilde{\mathbf{w}}_1(k+1) \leq \tilde{\mathbf{w}}_1(k) + 1. \end{aligned}$$

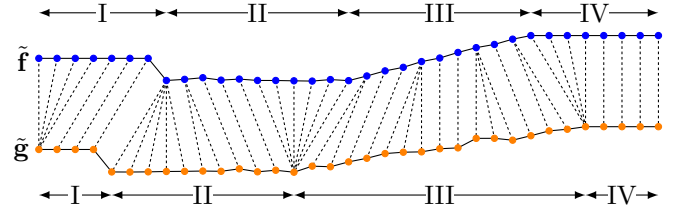


Fig. 3. Insufficient of sample-wise alignment of DTW. The solid lines indicate the input and output signals, and the dashed lines indicate the alignment between the samples.

- 2) Boundary:

$$\tilde{\mathbf{w}}(0) = (0, 0), \tilde{\mathbf{w}}(K-1) = (N-1, N-1). \quad (1)$$

DTW chooses the warp path that gives the minimum cumulative distance between $\tilde{\mathbf{f}}$ and $\tilde{\mathbf{g}}$. Taking the squared Euclidean Distance as the distance metric, as recommended in [24], we get the minimum cumulative distance computed by DTW as

$$\text{DTW}(\tilde{\mathbf{f}}, \tilde{\mathbf{g}}) = \min_{\tilde{\mathbf{w}} \in \tilde{\mathbf{W}}} \sum_{k=0}^{K-1} \tilde{\delta}(\tilde{\mathbf{w}}(k)), \quad (2)$$

where $\tilde{\delta}(\tilde{\mathbf{w}}_0(k), \tilde{\mathbf{w}}_1(k)) = (\tilde{\mathbf{f}}(\tilde{\mathbf{w}}_0(k)) - \tilde{\mathbf{g}}(\tilde{\mathbf{w}}_1(k)))^2$.

The computation of $\text{DTW}(\tilde{\mathbf{f}}, \tilde{\mathbf{g}})$ is carried out as follows:

- 1) Populate a cumulative distance matrix $\tilde{\mathbf{C}} \in \mathbb{R}^{N \times N}$. Every point in this matrix gives a value indicating the cheapest path to that point from the start. Every element is given by,

$$\tilde{\mathbf{C}}[i, j] = (\tilde{\delta}(i, j)) + \min \begin{cases} \tilde{\mathbf{C}}[i, j-1], \\ \tilde{\mathbf{C}}[i-1, j-1], \\ \tilde{\mathbf{C}}[i-1, j], \end{cases} \quad \forall i, j \in [0, N-1]. \quad (3)$$

- 2) Backtrack from $\tilde{\mathbf{C}}(N-1, N-1)$ to $\tilde{\mathbf{C}}(0, 0)$ to construct the warp path $\tilde{\mathbf{w}}$, where,

$$\tilde{\mathbf{w}}(\tilde{K}-1) = (N-1, N-1), \text{ and}$$

$$\begin{aligned} \tilde{\mathbf{C}}[\tilde{\mathbf{w}}(k-1)] &= \min\{\tilde{\mathbf{C}}[\tilde{\mathbf{w}}(k) - \mathbf{d}]\}, \\ \forall \mathbf{d} \in \{(0, 1), (1, 0), (1, 1)\}. \end{aligned} \quad (4)$$

The time complexity of DTW is $O(N^2)$, although several algorithms for speeding up the computations exist [18], [25], [26].

B. Insufficiency for TI applications

We now move to investigate the characteristics of DTW that prohibit it from being readily used for TI performance evaluation in terms of time and value errors. These should not be interpreted as limitations of DTW since DTW has not been designed for the question at hand.

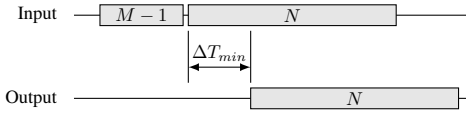


Fig. 4. Illustration of extending the input sequence by $M - 1$ samples to avoid the start and end artefacts in ETVO.

1) *Boundary conditions cause unrealistic artifacts:* DTW, which measures the similarity between two sequences, is extremely useful for sequences classification problems like correlation power analysis, DNA classification, and notably, speech recognition. In the latter, DTW can be used to correctly recognize a word even when spoken at different speeds or pitches. For achieving this, DTW assumes that every sample in one sequence is related to at least one sample in another sequence. While the boundary conditions in Equation (1) ensure that the extreme ends of the sequences are invariably aligned with each other, the monotonicity and continuity conditions ensure that none of the other samples in either sequence is skipped. Figure 3 illustrates the two conditions in action. Segments ‘I’ and ‘IV’ show how the boundary conditions lead to start and end artifacts. On the contrary, for TI applications, this assumption does not hold good as we expect some samples in the ideal signal to be missing or distorted in the observed signal. Instead of calculating how two sequences are similar, we want to calculate how the output is different from the input, where differences due to time or value offsets are distinguished.

2) *Unconstrained delay adjustments:* Since the objective of DTW is to find the best match between two sequences by minimizing the cumulative the Euclidean distances, as shown in Equation (2), delay adjustments are no concern for DTW. In practice, this means that in pursuit of picking the minimum Euclidean distance, DTW compromises on the delay adjustments. Indeed, for applications like speech recognition, this is a desirable property, as differences in delay do not matter when identifying words spoken. However, this is unacceptable for TI, as delay and delay adjustments cause more significant degradation in performance than noise in most cases. Examples are presented in Figure 3. In segment ‘II’, the time derivative of both sequences is low, so that small differences cause large fluctuations in delay. In segment ‘III’, one can see a large number of small changes that lead to minor improvements.

In the pursuit to identify similar sequences, DTW aims for the highest similarity it can find. It has been proven that DTW achieves its stated goal optimally [20]. As a consequence, the delay changes are not necessarily recognized at the time instants the real delay changes. In most instances, the changes in delay happen earlier. Multiple examples can be found in Figure 3. Segments ‘I’ and ‘IV’ start with an adjustment of delay, even though the changes are made toward the end of the corresponding segments. At the start of Segment ‘II’ there is a large delay change a few samples before a small peak that causes the change.

In order to resolve the above issues and design suitable performance measures for TI, we perform substantial refinements to DTW as we describe in the next section.

IV. DESIGN OF QUANTITATIVE METRIC

In this section, we begin by presenting the mathematical foundation of *Effective Time- and Value-Offset (ETVO)* framework that sets the stage for proposing the two quantitative metrics for TI. We then move to introducing the metrics – *Effective Delay-derivative (EDD)* and *Effective RMSE (ERMSE)*.

A. Design of ETVO framework

We now discuss our refinements for resolving the previously discussed issues of DTW for TI applications.

1) *Relaxation of boundary conditions:* Let \mathbf{f} and \mathbf{g} denote slices of the input and the output signals, respectively. For ease of explanation, we use the same notations as in DTW, however, without the accent ($\tilde{\cdot}$) to denote the ETVO counterparts. We consider a range of possible values for sample delays. The minimum delay is $\Delta T_{min} \in \mathbb{R}$, and the maximum delay is $\Delta T_{max} \equiv \Delta T_{min} + MT$, where $M \in \mathbb{N}^+$, and T the sample period. Let N denote the length of \mathbf{g} . Hence, \mathbf{f} should be of length $N + M - 1$. If the first sample of \mathbf{g} is at $t = 0$, then the first sample of $\mathbf{f}[k]$ should be at $t = -\Delta T_{min} - (M - 1)T$. This is illustrated in Figure 4.

Let $\mathbf{W} \subset \mathbb{N}^N$ denote the powerset of possible warp paths to align the output \mathbf{g} onto the input \mathbf{f} . The optimal warp path is denoted as $\mathbf{w} \in \mathbf{W}$, where $\mathbf{w}[k]$ indicates that $g[k]$ corresponds to $f[k - w[k]]$. We define ETO as the sample-wise time-offset corresponding to the alignment between \mathbf{g} and \mathbf{f} . ETO can be derived directly from the warp path and is expressed as

$$\text{ETO}[k] = \Delta T_{min} + w[k]. \quad (5)$$

We define the Cumulative Distribution Matrix as $\mathbf{C} \in \mathbb{R}^{N \times M}$, where the x -axis indicates the sample index of $\mathbf{g}[k]$, and the y -axis is the corresponding ETO. This is illustrated in Figure 5, wherein the value at each entry of \mathbf{C} indicates the cumulative cost of getting to that point. This is similar to the DTW counterpart $\tilde{\mathbf{C}}$ defined in Equation (3) except that the y -axis here denotes the ETO. The propagation through \mathbf{C} is

$$\mathbf{C}[i, j] = \delta(i, j) + \min \begin{cases} \mathbf{C}[i - 1, j], \\ \mathbf{C}[i - 1, j - 1], \\ \mathbf{C}[i, j + 1], \end{cases}$$

where

$$\delta(i, j) \equiv (\mathbf{g}[i] - \mathbf{f}[i - j + M - 1])^2, \forall i \in [0, N - 1], j \in [0, M - 1].$$

The three directions for calculating \mathbf{C} correspond directly to the three directions in DTW as defined in Equation (3). In one step of $\mathbf{g}[k]$, the delay can go down multiple steps, but can only go up one step at a time. For this translated system, the monotonicity and continuity condition is given as

$$0 \leq \mathbf{w}(k + 1) \leq \mathbf{w}(k) + 1.$$

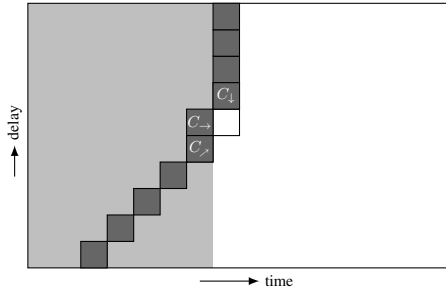


Fig. 5. The squares indicate all the values that need to be considered to calculate the optimum for the value in the white square. Gray indicates that the optimal value for that spot is already calculated. Dark gray indicates the cheapest source to get to the white square for three different methods. These are the only squares that need to be checked

This translation step accomplishes two things. Firstly, this structure is intuitively appealing for reporting time and value offsets for each output sample. Secondly, unlike in DTW, the search space does not scale quadratically with the signal length. Limiting the search space reduces the memory requirements and accelerates the algorithm. With the new foundation in place, we can prevent the start and end samples to have zero delay. The first column of \mathbf{C} is initialized as $\mathbf{C}(0, *) = [0]^M$. Every starting delay is equally expensive. To remove the ending artifact, we let the last sample of ETO be chosen as the cheapest option, so that

$$\mathbf{C}(N-1, \mathbf{w}[N-1]) \leq \mathbf{C}(N-1, j), \quad \forall j \in [0, M-1].$$

2) *Constraining delay adjustments*: In order to mitigate the issue of unconstrained delay adjustments in DTW, we come up with substantial refinements to its design. First, let us define what a delay adjustment is in the context of ETVO. It is the change in estimated delay per unit time. If multiple changes of the same type happen consecutively, we treat them as a single change of the cumulative magnitude. Recall that the diagonal arrow and down arrow represent an increase and decrease in delay, respectively. The dark grey squares in Figure 5 indicate this.

In the following, we present the mathematical foundation behind cumulative distribution matrix \mathbf{C} and describe the rationale behind the penalties.

$$C_{\rightarrow}[i, j] = [i-1, j] \quad (6)$$

$$C_{\downarrow}[i, j] = \min_{k \in \mathbb{N}^+} \left\{ \mathbf{C}[i, j+k] + \sum_{l=1}^{k-1} \delta(i, j+l) + kP_{\text{prop}} + P_{\text{fixed}} \right\} \quad (7)$$

$$C_{\nearrow}[i, j] = \min_{k \in \mathbb{N}^+} \left\{ \mathbf{C}[i-k, j-k] + \sum_{l=1}^{k-1} \delta(i-l, j-l) + kP_{\text{prop}} + P_{\text{fixed}} \right\}. \quad (8)$$

For every delay adjustment, we introduce two variables – P_{fixed} and P_{prop} , as shown in Equations (7) and (8). These correspond to fixed penalty for every delay adjustment and penalty proportional to size of the delay adjustment, respectively. P_{fixed}

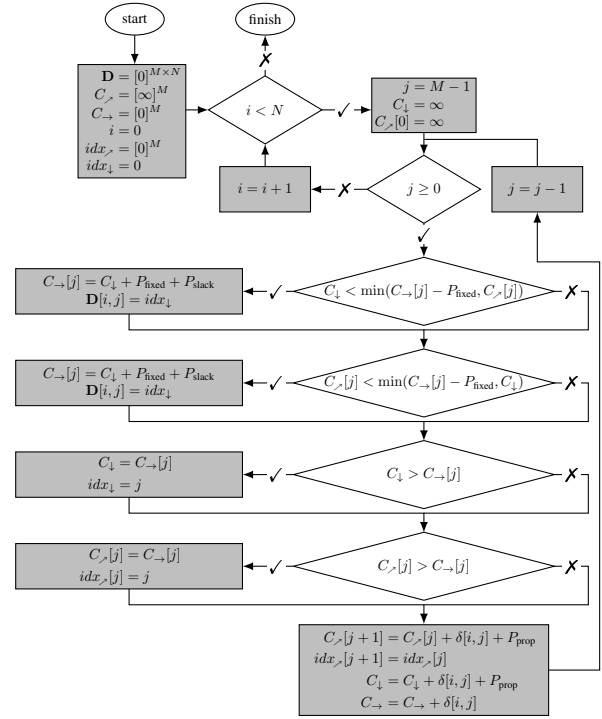


Fig. 6. Flowchart of the algorithm that finds the optimal way of traversing the delay, given the constraints specified for ETO.

affects the number of delay adjustments, and P_{prop} affects the magnitude of each adjustment. Together, these penalties suppress the delay adjustments estimated by the algorithm. The variable P_{prop} balances between time and value errors. A high P_{prop} reduces the time error and increases the value error, and vice-versa. Hence, ETVO performance approaches DTW as P_{prop} tends to zero. As a consequence, P_{prop} results in frequent delay adjustments even for minor changes in delay. From the standpoint of TI, this behavior is undesirable, since the human perception is insensitive to smaller delay changes. The addition of P_{fixed} controls the frequency of delay adjustments. The best candidate for each of the three directions is calculated as shown in Equation (6)-(8) and is illustrated in Figure 5.

3) *Improving the timing of delay adjustments*: In case of DTW, the delay adjustments do not align with the actual events that trigger the delay changes. For TI, such behavior not only makes analysis hard, but also makes the session quality estimates inaccurate. ETO should not be influenced by an event that occurs in the future. Note that P_{fixed} and P_{prop} do not address this issue of timing the delay adjustments. Therefore, we propose to introduce slack in delay adjustments where their timing is postponed until the slack penalty P_{slack} is breached. P_{slack} acts on top of P_{fixed} and P_{prop} for every delay adjustment, but is only added after an adjustment is made. The addition of P_{slack} increases the likelihood that the delay adjustments match the events that cause them. With this, the overall cumulative

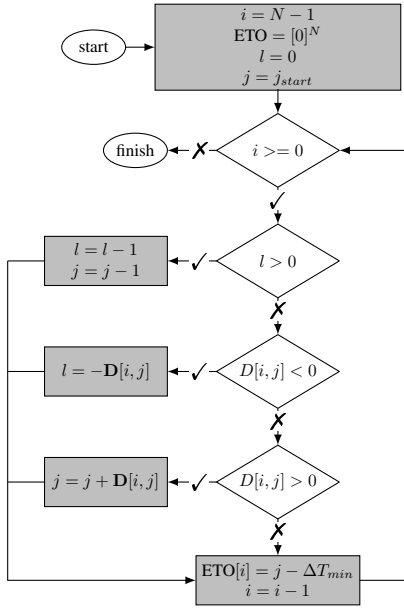


Fig. 7. Flowchart of the backtracking algorithm used to extract the ETO from direction matrix \mathbf{D}

distribution matrix \mathbf{C} is given as follows.

$$\mathbf{C}[i, j] = \delta(i, j) + \begin{cases} C_{\rightarrow}[i, j] & \text{if } C_{\rightarrow}[i, j] < \min\{C_{\downarrow}[i, j], C_{\nearrow}[i, j]\}, \\ C_{\downarrow}[i, j] + P_{\text{slack}} & \text{if } C_{\downarrow}[i, j] < \min\{C_{\rightarrow}[i, j], C_{\nearrow}[i, j]\}, \\ C_{\nearrow}[i, j] + P_{\text{slack}} & \text{otherwise} \end{cases}$$

4) *Defining EVO*: Unlike DTW where the residual distance for every sample in the warp path is aggregated into a single number, we represent the value-offset as a time-series that we call EVO. Every sample of EVO indicates the cumulative amount of distance from all samples of \mathbf{g} compared to the corresponding sample in \mathbf{f} , excluding the penalties. When ETO increases or stays the same, only one sample of \mathbf{g} is compared to \mathbf{f} , however, when ETO reduces, EVO is the cumulative distance between the output sample and several input samples. This enables obtaining fine-grained information on how samples contribute to the value-offset. The mathematical description for every element of EVO is given by

$$\text{EVO}[i] = \begin{cases} \sum_{l=\text{ETO}[k+1]}^{\text{ETO}[k]} \delta(i, l) & \text{if } \text{ETO}[i] > \text{ETO}[i+1], \\ \delta(i, \text{ETO}[i]) & \text{otherwise.} \end{cases}$$

Due to this, there are spikes in EVO every time the ETO reduces by a large amount.

5) *Computational complexity*: Besides presenting the ETVO framework, we also provide an efficient way of calculating ETO and EVO. The addition of P_{fixed} results in a larger set of values to consider when finding the optimal path. Instead of the three adjacent locations, one has to consider a total of M entries. Besides considering multiple entries, when backtracking to retrieve the delay, one has to consider how many steps were taken. To store that information, we propose

a direction matrix $\mathbf{D} \in \mathbb{Z}^{M \times N}$. The number stored in $\mathbf{D}(k, i)$ indicates that the next point is at $i + \mathbf{D}(k, i)$. Because the direction is stored, there is no need to store \mathbf{C} entirely. There are three directions to consider: up, down, and forward. Each direction has one optimal source from which to start. If we remember only the optimal continuations in each stage of the algorithm, we only need to store a value and the corresponding index $C_{\downarrow} \in \mathbb{R}$ and $idx_{\downarrow} \in \mathbb{N}$ for downward propagation, arrays $C_{\nearrow}[i, j] \in \mathbb{R}^M$ and $idx_{\nearrow}[i, j] \in \mathbb{N}^M$ for upward propagation, and array $C_{\rightarrow} \in \mathbb{R}^M$ for forward propagation. The resulting algorithm for populating \mathbf{D} is illustrated with a flow chart in Figure 6.

The backtracking algorithm needs to account for the multiple steps it can take. The calculation method matches how \mathbf{C} was populated. Figure 7 shows a flow chart of the backtracking algorithm. The complexity of the algorithm for populating \mathbf{D} is determined by two *for loops*. One of the *for loops* is looping through a fixed range M , which does not scale with signal length. Therefore the complexity is $\mathcal{O}(N)$. The complexity of the backtracking algorithm is bound by a single *for loop*, so the upper bound on the combined set is also $\mathcal{O}(N)$.

B. Quantitative Metrics for TI

Based on ETVO, we propose a set of two metrics that can characterize the TI session:

- *Effective Delay-Derivative (EDD)* expressed as the average rate of change of ETO, and given as $\frac{1}{N-1} \sum_{k=2}^{N-1} |\text{ETO}(k-1) - \text{ETO}(k)|$, and
- *Effective Root Mean Square Error (ERMSE)* defined as the deviation of value error, and given as $\sqrt{\frac{1}{N} \sum_{k=0}^{N-1} \text{EVO}[k]}$.

The parameters P_{prop} , P_{fixed} , and P_{slack} influence how EDD weighs against ERMSE. Out of these, P_{prop} is the most important of the three. The choice of the parameter values depends on the type of TI application.

The desirable feature of EDD and ERMSE is the following. For an application with high dynamic environment, where the movements are reasonably fast, a small amount of noise is insignificant, and hence should be attributed to ERMSE. On the other hand, in low dynamic environment, any amount of noise becomes significant and it makes sense to attribute it to EDD. Therefore, in order to accurately tune P_{prop} , we have to state how we rate time noise against value noise. One perspective is to consider time noise to be a source of value noise too. The amount of value noise is linearly proportional to the velocity. We suggest using the average velocity of the TI session so that $P_{\text{prop}} = T \dot{x}_{\text{average}}$, where T is the sampling period and \dot{x}_{average} is the average velocity of the signal. P_{fixed} suppresses unrealistic optimization and micro-adjustments. In practice, we found that a value of $P_{\text{fixed}} = 2P_{\text{prop}}$ yields good results. P_{slack} effectively cleans up the signal and makes the timing more accurate, but it only works if P_{slack} is by far the least significant. We found that $P_{\text{slack}} \leq P_{\text{prop}}$ yields good results. We believe a more sophisticated method or mathematical backing to tune the penalties is possible in the future.

V. EXPERIMENTAL SETUP

To validate our proposed metric, we develop a realistic TI testbed, where a human user can interact with a remotely rendered virtual environment (VE) over a network. We consider a recently proposed testbed for simulating TI interaction (Section V-A). This testbed lacks a network and is therefore not suited for mimicking TI interactions. Hence we carry out significant refinements to realize a networked TI testbed that is fast enough to provide a comfortable TI experience (see Section V-B).

A. Standard TI testbed

The standard TI testbed proposed recently by Bhardwaj et al., simulates a TI session by having the human user interact with a Virtual Environment (VE) in slave domain via haptic and visual feedback [27]. The master domain transmits position information of the haptic device sampling rate of 1 kHz to the slave domain. The resulting calculated force is sent back to the master domain along with the visual rendering of VE. A schematic overview of this setup is indicated as the gray blocks in Figure 8. In this figure, both master and slave domains are collocated for ease of experimentation. The haptic device used in this setup is a Novint Falcon [28]. Force calculation and rendering in the VE is implemented using Chai3D [29].

B. Networked TI testbed

We extend the testbed by decoupling the master and slave domains (each consisting of a workstation) across a network. Therefore, (i) the VE needs to be rendered on a separate workstation, and (ii) the two workstations are connected via a physical network. The local simulation is split into a renderer at the master domain and a physics engine at the slave domain. The physics engine translates force into movement and sends that as feedback. The real-network conditions between the master and slave are emulated using Netem on a separate workstation running Ubuntu 18.04 [30]. Netem can emulate network parameters such as the latency, jitter, and packet losses. A schematic overview of the entire system is shown in Figure 8. To emulate real network conditions, we use the Gilbert-Elliott model to handle packet loss, which is known to mimic real network behavior closely [31]. For delay jitter, we use a high correlation factor of 90% between latency of the packets. We specify the configured parameters of the loss model, average latency, jitter, and the algorithm parameters P_{prop} , P_{fixed} , and P_{slack} in Section VI as and when required.

VI. PERFORMANCE ANALYSIS

We conduct rigorous experimentation to comprehensively evaluate our metrics using the networked TI testbed. For the sake of simplicity, we consider only one axis of the position signal corresponding to the haptic device. Henceforth, we will refer to that axis signal as input. We compare the performances of DTW and ETVO. We formulate several scenarios which enable easy understanding of the working of ETVO. Finally, we demonstrate the strength of the proposed performance

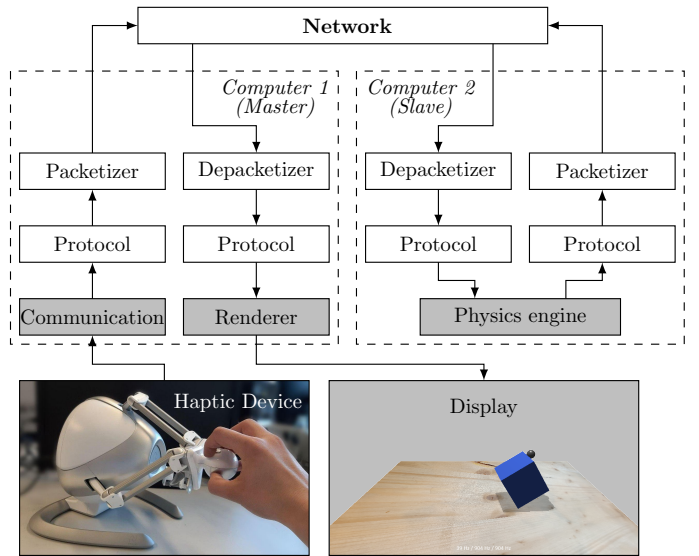


Fig. 8. Schematic overview of our measurement setup. The components present in the reference setup of Bhardwaj et al. highlighted in gray [27].

metrics along with describing how they can be arrived at based on the ETVO.

A. Comparison between DTW and ETVO

For this work we made alterations to the base DTW algorithm. In this section, we take a look at the effects of these alterations. To illustrate the differences in behavior between DTW and ETVO, we picked four fragments of time, each corresponding to a different signal. The fragments are picked to illustrate different aspects of decision making between DTW and ETVO. We start by gauging the sensitivity of each of the schemes to the signal variations. For this experiment, we set the following configuration: [average lat., jitter, P_{prop} , P_{fixed}] = [15 ms, 10 ms, 0.01, 0.005]. Figure 9(a) corresponds to this experimental setup. It can be observed that when the velocity is high (extremes of the plot), DTW shows vigorous fluctuations in time-offset estimation. On the other hand, ETVO sensibly changes its estimation of ETO exhibiting resilience to the noise in the signal. When the velocity is low in the middle, ETVO being context-aware stops adjusting the ETO since there is a negligible improvement in EVO. DTW however, does not care about how minor changes and keeps constantly adjusting the time-offset, irrespective of the context. The scenario for $P_{slack} = 0$ is also demonstrated (labelled as 'NoSlack'). In this signal, a change in delay is randomly made in the middle of the quiet period (as indicated with 1), while the ETO with P_{slack} postpones that decision to a more sensible moment. Note that in the value domain, ETVO and DTW perform similar, despite the significantly higher number of delay adjustments performed by DTW. Moreover, DTW lets the delay fluctuate regardless of what is going on, while ETVO modulates the delay only at high velocities. ETVO reasonably concludes that a higher update rate is desired when more is happening. This example shows how ETVO makes evaluations that are context-aware.

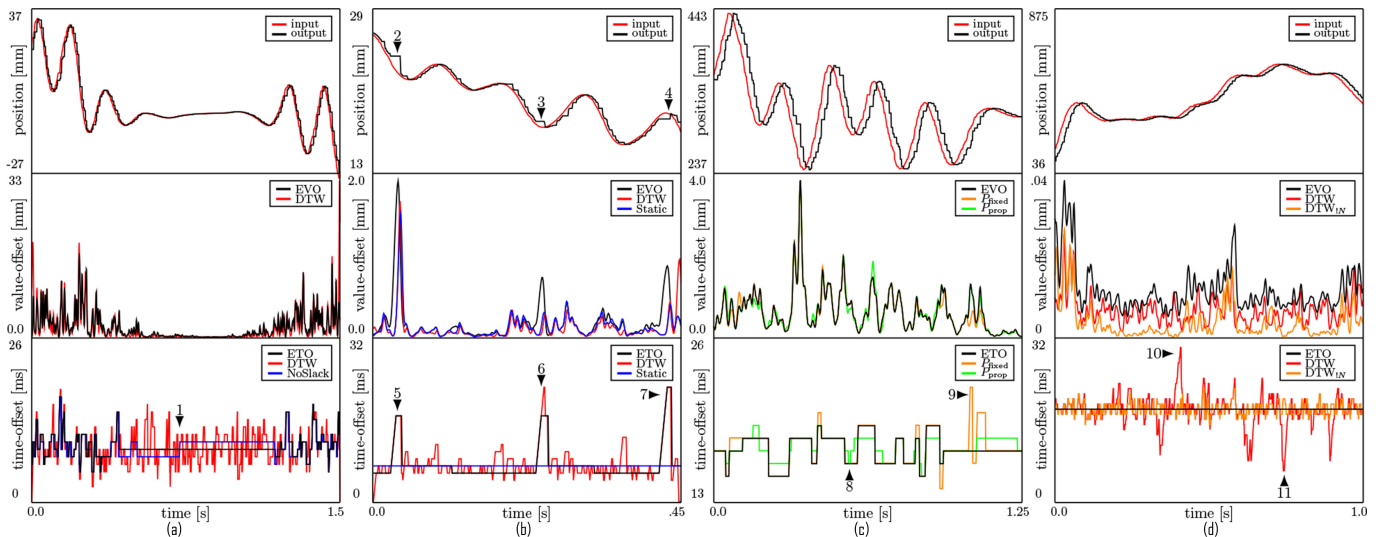


Fig. 9. Comparison of the performances of DTW and ETO frameworks in a variety of experimental setups under the conditions (a) Heavily varying position signal, (b) Packet losses and perceptual deadband, (c) Effects of P_{prop} and P_{fixed} , and (d) Effect of addition of noise to input signal.

We now investigate the ETVO performance in the presence of fewer haptic updates. This could be a consequence of either packet losses or employment of compression schemes. To capture the effects of these phenomena, we emulate this scenario through two processes: perceptual deadband [32] and packet losses. This enables us to evaluate the performance when the value-offset is considerable. We choose a deadband of 5% and bursty loss following the Gilbert-Elliott model with parameters $p = 5\%$ and $r = 50\%$, as prescribed in [33]. For this experiment, we set the following configuration: **[average lat., jitter, P_{prop} , P_{fixed}] = [0 ms, 0 ms, 0.005, 0.005]**. Figure 9(b) corresponds to this scenario. It can be clearly seen that the output signal is subject to a noticeable jitter due to the combined effect of packet loss and deadband. However, there are three specific instances (indicated by markers 2-4) where the combined effect of deadband and bursty losses prominently results in no change in the output signal. In this case, DTW relentlessly adjusts the time-offset as the deadband and losses are slightly degrading the signal. ETVO, however, remains robust to the jitter for the most part and sits at the average delay. But in cases where the effect is prominent, ETO is adjusted as indicated with markers 5-7. The value-offset is smoothed with a Gaussian distribution for visual clarity. As one expected, in the value domain ETVO is slightly higher than DTW. Although each of the schemes perform their corresponding tasks accurately, the behavior that is most favorable for TI applications is exhibited by ETVO. Especially considering that time variations are considered to be extremely detrimental to the session quality.

In this part we show the distinct effects of P_{prop} or P_{fixed} , and the need of considering both. We show the results in Figure 9(c). We consider three different settings for algorithm parameters:

- (i) $[P_{prop}, P_{fixed}] = [0.025, 0.05]$ (corresponds to black curve),
- (ii) $[P_{prop}, P_{fixed}] = [0.05, 0]$ (corresponds to amber curve), and

(iii) $[P_{prop}, P_{fixed}] = [0, 0.1]$ (corresponds to green curve). Marker 7 indicates an event where the ETO with $P_{fixed} = 0$ adjusts in multiple small steps. In this case, there is no difference in extra cost associated with using multiple steps, causing the amount of adjustments to increase. Marker 8 indicates an event where the ETO with $P_{prop} = 0$ causes a large step changes but limited in the number of steps. In this case, there is no extra cost associated with how much the ETO changes, giving rise to giving rise to unreasonably large ETO adjustments. The signal that has both components has a similar performance in the value domain, but a significantly less cluttered ETO.

Figure 9(d) shows the effect that high-frequency noise has on DTW and ETVO. For this purpose, we introduce AWGN with an SNR of 70 dB. The experimental configuration is as follows: **[average lat., jitter, P_{prop} , P_{fixed} , P_{slack}] = [15 ms, 1 ms, 0.005, 0.01, 0.005]**. Both DTW and EVO are plotted with the noise added, while DTW_{IN} is a version of DTW without the added AWGN. High-frequency noise is a good example of a common way of signal distortion that DTW cannot deal with properly. Note that ETO outperforms the best case DTW i.e. DTW_{IN} , demonstrating its noise resilience. Further, one can also notice the vulnerability of DTW to even a marginal amount of noise causing time-offset to fluctuate vigorously.

B. Performance of EDD and ERMSE

In previous subsections, we demonstrated how the performance of ETVO is desirable for TI in comparison to DTW. In this section, we demonstrate how the proposed measures EDD and ERMSE perform. We conduct four runs of the experiment with one value for following network jitter in each run – 0 ms, 5 ms, 10 ms, and 20 ms. In all experiments, the input and output signal durations are larger than 10 s. In each run, we calculate the EDD and ERMSE for a wide range of P_{prop}

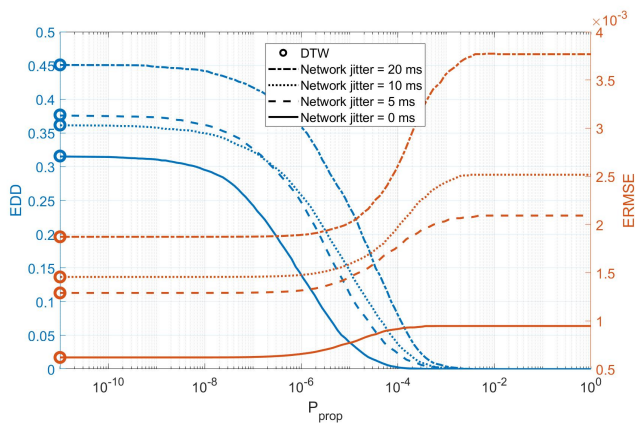


Fig. 10. Based on the data of our experiments, the EDD and ERMSE that ETVO calculates are plotted as a function of P_{prop} .

to demonstrate how the penalties balance out the EDD and ERMSE. The results are shown in Figure 10.

DTW assigns no value to EDD, which means $P_{\text{prop}} = 0$. The endpoints on the left match what DTW would report, but without accounting for the starting and the ending artifacts. When $N \rightarrow \infty$ the contribution of those artifacts become negligible, but for small N , the contributions can be significant in comparison with ERMSE. For high P_{prop} , the EDD becomes zero, as any change is too expensive. However, we can still have any constant delay, so the result is the ERMSE value for the optimal constant delay. This can be seen as the best scenario for ERMSE when the optimal constant delay is known. Existing measures do not compute this, so in practise the ERMSE will be orders of magnitude higher for the methods proposed in literature. For every experiment there is a pair of lines that indicates the balance between EDD and ERMSE for $P_{\text{prop}} \in [\approx 0, 1]$. For low values of P_{prop} , the signals reach a point where decreasing P_{prop} no longer leads to a reduction in ERMSE and therefore an increase in EDD. One can also see the *law of diminishing returns* in play, where a large increase in amount EDD is required for the same amount of ERMSE reduction. For high values of P_{prop} , signals reach a point where increasing P_{prop} no longer leads to a decrease in EDD. That is because at that point EDD is already zero. Again, we see the law of diminishing returns, where the ERMSE goes up quite a bit for only a marginal reduction in EDD. Both DTW and taking RMSE of the output directly (constant delay) attempt to account all errors leading a single metric such as mean error. As we see in Figure 10, DTW takes one extreme and constant delay takes another extreme. However, we believe that the optimal should be in between these extremes. ETVO provides a handle in the form of tunable penalties (P_{prop}) using which one can select the tolerable delay and corresponding least error.

VII. CONCLUSION AND FUTURE WORK

As the field of Tactile Internet (TI) is advancing fast, there is a strong need for quantifying performance sessions

objectively. In this paper, we first presented first lack of TI performance metrics in the state of the art and their limitations for using them to characterize TI sessions and systems. Specifically both time offsets and value offsets in the input signal and the received (or output) signal cannot be extracted separately in those methods. To overcome these limitations, we find Dynamic Time Warping (DTW) algorithm used in speech recognition as the most suitable. However, we also encountered certain issues in applying DTW directly to characterize performance of TI applications. We demonstrated the issues and provided a mathematical framework to address them. During the course, we proposed two novel concepts: Effective Time-Offset (ETO) and Effective Value-Offset (EVO) that enable the representation of TI performance at a fine scale. Based on ETO and EVO, we defined two quantitative performance metrics - Effective Delay-Derivative (EDD) and Effective RMSE (ERMSE) - for quantifying the impact of network effects on the quality of a TI session.

Through rigorous experiments using a haptic tool and experimental setup, we demonstrated their salient features. We demonstrate how ETO and EVO are context-aware and noise-resilient estimation of similarity between the input and output signals. We believe that objective metrics for measuring subjective quality will enable significant improvements in design of TI applications. While the current work looks at an offline session, we intend to solve the non-trivial problems of measuring objectively in real-time in the near future.

VIII. ACKNOWLEDGEMENTS

This project is supported by SCOTT <http://www.scottproject.eu> that has received funding from the Electronic Component Systems for European Leadership Joint Undertaking under grant agreement No 737422. This joint undertaking receives support from the European Unions Horizon 2020 research and innovation program and Austria, Spain, Finland, Ireland, Sweden, Germany, Poland, Portugal, Netherlands, Belgium, Norway. The authors would like to thank Prof. Eamonn Keogh, UC Riverside, and ir. Luc Hogervorst for providing valuable feedback on the manuscript.

REFERENCES

- [1] G. P. Fettweis, "The tactile internet: Applications and challenges," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 64–70, 2014.
- [2] A. Aijaz and M. Sooriyabandara, "The tactile internet for industries: A review," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 414–435, 2018.
- [3] V. Millnert, J. Eker, and E. Bini, "Achieving predictable and low end-to-end latency for a network of smart services," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–7.
- [4] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, "Federated learning for ultra-reliable low-latency v2v communications," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–7.
- [5] F. Gringoli, R. Klose, M. Hollick, and N. Ali, "Making wi-fi fit for the tactile internet: Low-latency wi-fi flooding using concurrent transmissions," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2018, pp. 1–6.
- [6] O. Holland, E. Steinbach, R. V. Prasad, Q. Liu, Z. Dawy, A. Aijaz, N. Pappas, K. Chandra, V. S. Rao, S. Oteafy *et al.*, "The ieee 1918.1 tactile internet standards working group and its standards," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 256–279, 2019.

- [7] R. Chaudhari, E. Steinbach, and S. Hirche, "Towards an objective quality evaluation framework for haptic data reduction," in *2011 IEEE World Haptics Conference*. IEEE, 2011, pp. 539–544.
- [8] Z. Yuan, S. Chen, G. Ghinea, and G.-M. Muntean, "User quality of experience of mulsemmedia applications," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 11, no. 1s, p. 15, 2014.
- [9] M. Eid, J. Cha, and A. El Saddik, "Admux: An adaptive multiplexer for haptic–audio–visual data communication," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 1, pp. 21–31, 2010.
- [10] B. Cizmeci, X. Xu, R. Chaudhari, C. Bachhuber, N. Alt, and E. Steinbach, "A multiplexing scheme for multimodal teleoperation," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 13, no. 2, pp. 21:1–21:28, Apr. 2017. [Online]. Available: <http://doi.acm.org/10.1145/3063594>
- [11] P. Hinterseer, E. Steinbach, and S. Chaudhuri, "Perception-based compression of haptic data streams using kalman filters," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 5, May 2006, pp. V–V.
- [12] V. Gokhale, J. Nair, and S. Chaudhuri, "Congestion control for network-aware telehaptic communication," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 13, no. 2, pp. 17:1–17:26, Mar. 2017. [Online]. Available: <http://doi.acm.org/10.1145/3052821>
- [13] C. Li, C.-P. Li, K. Hosseini, S. B. Lee, J. Jiang, W. Chen, G. Horn, T. Ji, J. E. Smee, and J. Li, "5G-Based Systems Design For Tactile Internet," *Proceedings of the IEEE*, pp. 1–18, 2018. [Online]. Available: <https://ieeexplore.ieee.org/document/8452975/>
- [14] J. Sachs, L. A. Andersson, J. Araújo, C. Curescu, J. Lundsjö, G. Rune, E. Steinbach, and G. Wikström, "Adaptive 5g low-latency communication for tactile internet services," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 325–349, 2018.
- [15] K. S. Kim, D. K. Kim, C.-B. Chae, S. Choi, Y.-C. Ko, J. Kim, Y.-G. Lim, M. Yang, S. Kim, B. Lim *et al.*, "Ultrareliable and low-latency communication techniques for tactile internet services," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 376–393, 2018.
- [16] C. Basdogan, C.-H. Ho, M. A. Srinivasan, and M. Slater, "An experimental study on the role of touch in shared virtual environments," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 7, no. 4, pp. 443–460, 2000.
- [17] N. Sakr, N. Georganas, and J. Zhao, "A perceptual quality metric for haptic signals," in *2007 IEEE International Workshop on Haptic, Audio and Visual Environments and Games*. IEEE, 2007, pp. 27–32.
- [18] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 561–580, 2007.
- [19] L. R. Rabiner and B. Gold, "Theory and application of digital signal processing," *Englewood Cliffs, NJ, Prentice-Hall, Inc., 1975. 777 p.*, 1975.
- [20] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.
- [21] L. Chen, M. T. Özsu, and V. Oria, "Robust and fast similarity search for moving object trajectories," in *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*. ACM, 2005, pp. 491–502.
- [22] L. Chen and R. Ng, "On the marriage of lp-norms and edit distance," in *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*. VLDB Endowment, 2004, pp. 792–803.
- [23] M. Vlachos, D. Gunopoulos, and G. Kollios, "Discovering similar multidimensional trajectories," in *icde*. IEEE, 2002, p. 0673.
- [24] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *KDD workshop*, vol. 10, no. 16. Seattle, WA, 1994, pp. 359–370.
- [25] D. F. Silva and G. E. Batista, "Speeding up all-pairwise dynamic time warping matrix calculation," in *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, 2016, pp. 837–845.
- [26] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowledge and information systems*, vol. 7, no. 3, pp. 358–386, 2005.
- [27] A. Bhardwaj, B. Cizmeci, E. Steinbach, Q. Liu, M. Eid, A. E. Saddik, R. Kundu, X. Liu, O. Holland, M. A. Luden, S. Oteafy, and V. Prasad, "A Candidate Hardware and Software Reference Setup for Kinesthetic Codec Standardization," in *2017 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE)*, 2017, pp. 53–58.
- [28] (2015) Novint falcon - product page. [Online]. Available: <https://web.archive.org/web/20150215032400/http://www.novint.com/index.php/products/novintfalcon>
- [29] (2019) Chai3d - features. [Online]. Available: <https://www.chai3d.org/concept/features>
- [30] (2019) networking:netem [wiki]. [Online]. Available: <https://wiki.linuxfoundation.org/networking/netem>
- [31] G. Haßlinger and O. Hohlfeld, "The gilbert-elliott model for packet loss in real time services on the internet," in *14th GI/ITG Conference-Measurement, Modelling and Evaluation of Computer and Communication Systems*. VDE, 2008, pp. 1–15.
- [32] I. Vittorias, J. Kammerl, S. Hirche, and E. Steinbach, "Perceptual coding of haptic data in time-delayed teleoperation," in *World Haptics 2009-Third Joint EuroHaptics conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*. IEEE, 2009, pp. 208–213.
- [33] M. Ellis, D. P. Pazaros, T. Kypraios, and C. Perkins, "A two-level markov model for packet loss in udp/ip-based real-time video applications targeting residential users," *Computer Networks*, vol. 70, pp. 384–399, 2014.