IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

NetLfD: Network-Aware Learning from Demonstration for In-Contact Skills via Teleoperation

Başak Güleçyüz[1,2], Vincent von Büren[1], Xiao Xu[1], Eckehard Steinbach[1,2]

***Abstract*—When providing task demonstrations to a remote robot over the network via bilateral teleoperation, communication impairments are unavoidable, hindering the human operator from delivering high-quality demonstrations. Poor-quality demonstrations can negatively impact the robot's ability to learn and generalize. In this work, we propose to enhance learning performance by introducing a network-aware confidence weighting strategy for remote learning from demonstration. Our approach extends the Hidden Semi-Markov Model (HSMM) and its task-parameterized version (TP-HSMM) to their confidence-weighted versions, WHSMM and WTP-HSMM. We evaluated various weight metrics that serve as teleoperation transparency measures and demonstration quality indicators under varying communication delays. We validated the proposed approach in two different in-contact tasks using data collected from 18 participants. The results show that weighting improves task performance in reproduction by up to 42% in the force precision and 63% in the success rate, demonstrating the potential of the proposed approach to enhance the effectiveness of robot learning from remote demonstrations.**

***Index Terms*—Learning from Demonstration; Telerobotics and Teleoperation**

I. Introduction

THE use of teleoperation with the emergence of the Tactile Internet (TI) opens up various possibilities for remote physical interactions, allowing global access to skills [1], [2]. Immersive teleoperation over the TI will pave the way to remote skill transfer between humans and robots, particularly in environments where the coexistence of the human and robot is not feasible such as in dangerous environments, underwater, in space, or when the skilled person is far away. In this regard, learning from demonstration (LfD) is a simple yet effective approach to learning new tasks with few demonstrations and generalizing to unseen scenarios [3].

The quality of demonstrations is prominent in LfD to achieve reliable performance and generalization ability [3], [4]. When using bilateral teleoperation with haptic feedback to provide demonstrations remotely, the communication delay between the human and the robot is unavoidable which plays a crucial role in teleoperation transparency and, as a result, in the quality of demonstrations. It is especially significant for manipulation tasks as the position and physical interaction with the environment should be safely controlled [5]. To ensure stability and transparency in bilateral teleoperation with haptic feedback, high haptic packet rates (typically 1 kHz or higher) are required [6], which causes a high load on the network. To mitigate this high load, the perceptual deadband (DB) based rate reduction approach has been proposed [7]. The DB approach or the communication delay may lead to instabilities and reduced realism in teleoperation. Although passivity-based approaches maintain stability, they often degrade the transparency [8], resulting in suboptimal demonstrations and poor learning outcomes. This degradation is expected to be even more pronounced for in-contact skills, as the perception of correct forces is crucial to perform the task accurately.

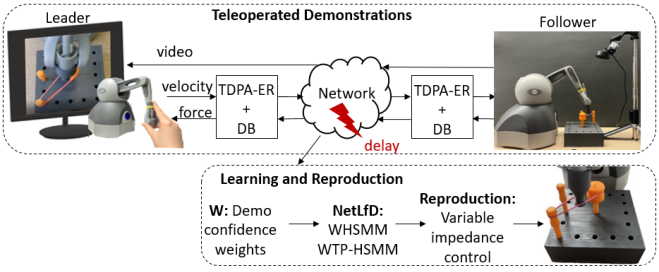


Fig. 1: Overview of NetLfD. Top: Bilateral teleoperation with haptic feedback for demonstrations. Bottom: NetLfD and reproduction.

We propose Network-Aware Learning from Demonstration (NetLfD) to overcome these limitations and advance the learning quality. NetLfD identifies the noisy or poor-quality parts in the demonstrations that are potentially harmful to learning while being aware of the network and tunes their contribution to learning. Fig. 1 visualizes our proposed approach. A human operator provides demonstrations to a remote follower robot using bilateral teleoperation with haptic feedback over the network in the presence of communication delay. We use the Geomagic Touch haptic device as the human-system interface to sense human motion, display force feedback, and as the follower robot in the remote environment. In the teleoperation setup, we consider the time domain passivity approach with energy reflection (TDPA-ER) with DB approach [9] as the joint solution for haptic rate reduction and passivity control for stability. By evaluating the teleoperation transparency using various metrics, we adjust the contribution of each demonstration or its sub-parts into the learned model. During

Manuscript received: May, 17, 2023; Revised July, 27, 2023; Accepted August, 26, 2023. This paper was recommended for publication by Editor Jee-Hwan Ryu upon evaluation of the Associate Editor and Reviewers' comments.

B. Gülecyüz, V. von Büren, X. Xu, and E. Steinbach are with the School of Computation, Information, and Technology, Department of Computer Engineering, Munich Institute of Robotics and Machine Intelligence (MIRMI), Chair of Media Technology, Technical University of Munich, e-mail: {basak.guelecyuez,xiao.xu,vincentvon.bueren,eckehard.steinbach}@tum.de

B. Gülecyüz and E. Steinbach are also with the Centre for Tactile Internet with Human-in-the-Loop (CeTI), TU Dresden, Germany.

Digital Object Identifier (DOI): see top of this page.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

reproduction, direct and indirect force controllers are used for different in-contact tasks.

Our contributions can be summarized as follows:

- We formulate the reliability-weighted log-likelihood function for HSMM and TP-HSMM to achieve weighted contribution of demonstrations in learning from teleoperated demonstrations, which we refer to as WHSMM and WTP-HSMM.
- We evaluate different reliability weight metrics as indicators of teleoperation transparency and demonstration quality under varying delays in the network.
- We validate the proposed approach using data collected from 18 participants and demonstrate that it improves the accuracy and success rate for two different in-contact tasks: letter writing and rubber band placement.

II. Related Work

In LfD, many approaches have been proposed to learn and generalize the demonstrated movements: Dynamical Systems (DS) [10], Dynamic Movement Primitives (DMP) [11], and generative models such as TP-GMM and TP-HSMM [12], [13]. In this work, we adopt and extend the HSMM and TP-HSMM, as they are well equipped to extract spatio-temporal features from the demonstrations and provide generalizability to novel conditions. Additionally, these models can incorporate multiple demonstrations, which is necessary in the case of teleoperated demonstrations, as they tend to include higher noise, errors, and spatio-temporal differences due to the inherent difficulty of using the teleoperation system and also the communication unreliabilities [14], [15].

It is essential to model both the demonstrated trajectories and the interaction with the environment for learning in-contact skills. In [16], the authors employed DMPs to learn position and force separately for a writing task and used direct force control during reproduction. In [17], [18], the authors proposed to learn both trajectories and interaction forces using HSMM and applied variable impedance control during reproduction. Rozo *et al.* [19], and Le *et al.* [20] combined position and force into an attractor model with stiffness estimation and variable impedance control. In [21], the authors unified position and force control in joint space for bilateral control-based imitation learning for a writing task. Our approach NetLfD is not limited to any controller and can be used with different controllers. To evaluate the applicability and effectiveness of NetLfD in different settings, we perform experiments with two different tasks using different control perspectives. In the first task, we focus on learning position and force separately and using direct force control for reproduction. In the second, we use attractor learning and indirect force control with variable impedance similar to [19], [20].

To account for the unreliabilities in the teleoperated demonstrations under varying network conditions, we consider a reliability-weighted contribution of the demonstrations to learning. This idea has been effective for clustering in the presence of outliers in the data. Newton *et al.* formulated the weighted likelihood bootstrap as the samples are produced by a reweighted maximum likelihood cost function [22]. Markatou down-weighted the observations with large residuals for a mixture of distributions in the presence of outliers [23]. Gebru *et al.* simplified the Weighted Likelihood Estimation (WLE) for a GMM by reflecting the weights to the covariance of the emission probabilities as a division factor for audio-visual scene analysis under observation unreliability [24]. In this work, we extend the prior work for learning HSMM and TP-HSMM by integrating the reliability weights directly to the log-likelihood function without any probabilistic simplification. The weights are attributed to various confidence metrics linked to teleoperation transparency.

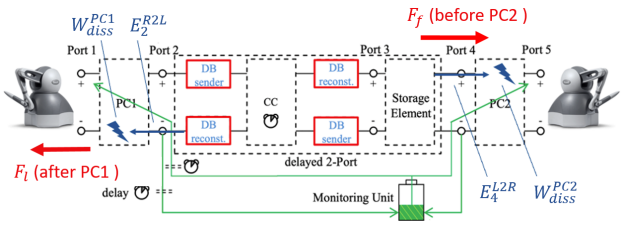


Fig. 2: TDPA-ER with perceptual DB approach, adapted from [9].

III. Preliminaries

A. Time-Delayed Bilateral Teleoperation

In this work, we employ the recently proposed passivity control, TDPA-ER, together with the perceptual DB approach (depicted in Fig. 2) to stabilize the teleoperator under delay and reduce the transmission rate [9]. This approach has been selected for standardization in IEEE P1918.1.1 [25].

The perceptual DB approach reduces the high packet rate by transmitting haptic packets only when there is a significant perceptual change for the human operator, compared to the previously transmitted data [7]. The amount of rate reduction is tuned by the deadband parameter (DBP). The TDPA-ER approach considers the follower controller as an energy storage element and allocates the stored energy to both the leader and follower as the upper bounds of the desired output energy at each time instant. The passivity controllers, PC1 and PC2, dissipate the excessive energy if the output energy is beyond the desired upper bound [8]. Although TDPA-ER is less conservative than the original formulation of TDPA [26], the reduced force magnitude due to damping and force jumps that occur during communication interruptions compromise the transparency of the teleoperation [9]. Hence, the amount of dissipated energy plays a crucial role in transparency and can impact the quality of remote demonstrations.

B. HSMM and TP-HSMM for LfD

To teach the robot the desired tasks, we adopt the approach of modeling demonstrations as a joint probability density using HSMM. HSMM encodes observations, $\{\boldsymbol{\xi}_t\}_{t=1}^{T}$, in relation with the hidden state sequence, $\{z_t\}_{t=1}^{T}$, and explicit state duration [13], [27]. The parameters of an HSMM are:

$$\theta = \left\{\Pi_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i, \{a_{ij}\}_{j=1,j\neq i}^{K}, \boldsymbol{\mu}_i^D, \boldsymbol{\Sigma}_i^D\right\}_{i=1}^{K}, \quad (1)$$

where K represents the number of states, a_{ij} the state transition probability from state i to j, and Π_i, the prior state

probability of starting in state i. The Gaussian distribution for state i is characterized by its mean vector and its covariance matrix $\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i$, respectively. The state duration is modeled as a Gaussian distribution with $\boldsymbol{\mu}_i^D$ and $\boldsymbol{\Sigma}_i^D$ for each state i.

The original formulation of HSMM provides an elegant and simple approach to learning individual movements; however, it suffers from limited generalizability. The task-parameterized formulation of HSMM (TP-HSMM) allows the skills to be interpolated or extrapolated beyond the demonstrations by describing task-specific parameters using coordinate frames [13]. The task parameters for F frames are represented with $\{\boldsymbol{A}_j, \boldsymbol{b}_j\}_{j=1}^{F}$ where $\boldsymbol{A}_j$ is the rotation matrix and $\boldsymbol{b}_j$ is the position vector. Each sample point $\boldsymbol{\xi}_t$ is observed from the perspective of the F frames with $\boldsymbol{\xi}_t^{(j)} = \boldsymbol{A}_j^{-1}(\boldsymbol{\xi}_t - \boldsymbol{b}_j)$. The parameters for TP-HSMM are:

$$\theta = \left\{ \Pi_i, \left\{ \boldsymbol{\mu}_i^{(j)}, \boldsymbol{\Sigma}_i^{(j)} \right\}_{j=1}^{F}, \{a_{il}\}_{i=1, l\neq i}^{K}, \boldsymbol{\mu}_i^D, \boldsymbol{\Sigma}_i^D \right\}_{i=1}^{K}, \quad (2)$$

where $\boldsymbol{\mu}_i^{(j)}$ and $\boldsymbol{\Sigma}_i^{(j)}$ represent the mean and covariance of the i-th component in frame j. The emission probability for the i-th state is the product of probabilities of the observation in F frames. For a novel condition with the new reference frames of $\left\{\tilde{\boldsymbol{A}}_j, \tilde{\boldsymbol{b}}_j\right\}_{j=1}^{F}$, we linearly transform the frame-specific Gaussians back to global coordinates and compute the product of Gaussians (see [12], [13] for more details).

The parameters of HSMM can be obtained using the Baum-Welch algorithm, a special case of Expectation Maximization (EM) for the Hidden Markov Model (HMM) [28]. The state durations are approximated with the empirical means and variances using the most likely state sequence [13]. For the TP-HSMM, the model parameters are estimated with the constraint that each frame belongs to the same source (see [12] for details). The state priors and transition probabilities are shared among different frames.

IV. Proposed Approach

A. NetLfD: WHSMM and WTP-HSMM

To enhance the quality of robot learning and autonomy by leveraging demonstration quality, we propose using reliability weight metrics that correlate with the network and teleoperation transparency. The higher the weight assigned to a particular demonstration, the more significant its contribution to the overall learning should be. To this end, the WLE [22], [23] with predetermined weights can be utilized to maximize the reliability-weighted likelihood of demonstrations. We will refer to the reliability-weighted versions of HSMM and TP-HSMM as WHSMM and WTP-HSMM, respectively.

We formulate the reliability-weighted expected complete log-likelihood, Q_w (auxiliary function), for HMM as:

$$\begin{aligned} Q_w\left(\theta, \theta^{old}\right) = & \sum_{i=1}^{K}\sum_{m=1}^{M} w_{m,1}\gamma_{m,1,i}\log \pi_i \\ & + \sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{m=1}^{M}\sum_{t=1}^{T_{m-1}} w_{m,t}w_{m,t+1}\zeta_{m,t,i,j}\log a_{ij} \\ & + \sum_{m=1}^{M}\sum_{t=1}^{T_m}\sum_{i=1}^{K} w_{m,t}\gamma_{m,t,i}\log \mathcal{N}\left(\boldsymbol{\xi_{m,t}} \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma_i}\right), \end{aligned} \quad (3)$$

where $w_{m,t}$ is the reliability weight assigned to the m-th demonstration at time t, which is either constant or time-varying depending on the delay variations in the network. K is the number of states, M is the number of total demonstrations, and T_m is the duration of the m-th demonstration. $\gamma_{m,t,i}$ and $\zeta_{m,t,i,j}$ are the intermediate variables in the HMM corresponding to states i and j. When updating the model parameters to maximize Q_w through EM, the expectation step is unchanged. However, in the maximization step, the weights are incorporated using the update equations:

$$\pi_i \leftarrow \frac{\sum_{m=1}^{M} w_{m,1}\gamma_{m,1,i}}{\sum_{m=1}^{M} w_{m,1}}, \quad (4)$$

$$a_{ij} \leftarrow \frac{\sum_{m=1}^{M}\sum_{t=1}^{T_{m-1}} w_{m,t}w_{m,t+1}\zeta_{m,t,i,j}}{\sum_{j=1}^{K}\sum_{m=1}^{M}\sum_{t=1}^{T_{m-1}} w_{m,t}w_{m,t+1}\gamma_{m,t,i}}, \quad (5)$$

$$\boldsymbol{\mu}_i \leftarrow \frac{\sum_{m=1}^{M}\sum_{t=1}^{T_m} w_{m,t}\gamma_{m,t,i}\boldsymbol{\xi}_{\mathrm{m},t}}{\sum_{m=1}^{M}\sum_{t=1}^{T_m} w_{m,t}\gamma_{m,t,i}}, \quad (6)$$

$$\boldsymbol{\Sigma}_i \leftarrow \frac{\sum_{m=1}^{M}\sum_{t=1}^{T_m} w_{m,t}\gamma_{m,t,i}\left(\boldsymbol{\xi}_{\mathrm{m},t} - \boldsymbol{\mu}_i\right)\left(\boldsymbol{\xi}_{\mathrm{m},t} - \boldsymbol{\mu}_i\right)^T}{\sum_{m=1}^{M}\sum_{t=1}^{T_m} w_{m,t}\gamma_{m,t,i}}. \quad (7)$$

For the task-parameterized version, WTP-HSMM, Eq. (4) and Eq. (5) remain the same. Similar to Eq. (6) and Eq. (7) the update steps for $\boldsymbol{\mu}_i^{(j)}$ and $\boldsymbol{\Sigma}_i^{(j)}$ are performed for each frame j using the observations from j-th frame, $\boldsymbol{\xi}_{m,t}^{(j)}$.

B. Weight Metrics

Here, we describe the weight metrics evaluated in our approach. Each metric is kept constant per demonstration, if the delay is constant during the demonstration. When the delay is time-varying, we compute the weights per predefined block length using the block mean for each metric.

1) Delay$_W$: The amount of delay in the network can be used as a straightforward weighting strategy that reflects the teleoperation transparency to some extent. The demonstration with the lowest delay is assigned the highest weight value of 1, while the other demonstrations receive lower weights inversely proportional to their delay:

$$\mathrm{Delay_W}(m) = \frac{\min\{d_m\}_{m=1}^{M}}{d_m}, \quad (8)$$

where d_m is the delay during the m-th demonstration.

2) ForceRMSE$_{LF}$: We examine the root mean squared error (RMSE) between the computed controller force at the follower, F_f, and the displayed force feedback at the leader, F_l, after temporally aligning them according to the delay:

$$\mathrm{ForceRMSE_{LF}} = \sqrt{\frac{1}{T_m}\sum_{t=1}^{T_m}(F_l(t) - F_f(t))^2}. \quad (9)$$

The computed control force F_f is retrieved before the PC2 at the follower, whereas the force feedback F_l after the PC1 at the leader (see Fig. 2). ForceRMSE$_{\mathrm{LF}}$ can take an arbitrary range of values. Thus, similar to Delay$_{\mathrm{W}}$, we perform an inversely proportional normalization to the range $(0-1]$.

3) HSSIM: This metric represents the haptic interpretation of the structural similarity index measure (SSIM) in the visual domain. The Haptic SSIM (HSSIM) differs from the sample error-based measure of $\text{ForceRMSE}_{\text{LF}}$ in the way that it considers neighboring sample dependencies. HSSIM considers the human haptic perception by using Steven's power law and is correlated to human subjective assessments [29]. We compute the HSSIM between F_f and F_l after temporal alignment with respect to the delay. HSSIM lies in the range of $(0-1]$. The higher the HSSIM, the more transparent the teleoperation is.

4) 1-RAT$_E$: RAT_{E} represents the energy dissipation ratio of the system and mirrors the system's conservatism [8]:

$$\text{RAT}_{\text{E}} = \frac{W_{diss}^{PC1} + W_{diss}^{PC2}}{E_2^{R2L} + E_4^{L2R}}. \tag{10}$$

The term E_2^{R2L} and E_4^{L2R} represent the output energy at port 2 and port 4, respectively (see Fig. 2). The sum of both terms is the total output energy from the system. To maintain the passivity and thus the stability of the teleoperator, the excess energies W_{diss}^{PC1} and W_{diss}^{PC2} are dissipated in the passivity controllers PC1 and PC2, respectively. RAT_{E} lies in the range of $(0-1]$. The lower the ratio, the less conservative the system is. Thus, we propose to use 1-RAT_{E} for weighting.

V. Experiments

To evaluate the proposed approach and different weight metrics, we present two in-contact tasks with different characteristics and notions of task performance. The first task is a letter-writing task in a virtual environment (VE), where the precision in the trajectory and the applied force indicate the performance. The second task is manipulating a rubber band for proper pick and placement onto target pillars, allowing us to evaluate the success rate and generalizability to novel conditions. These two tasks also differ regarding the force control strategy. We use direct force control in the first task where the movement and force directions are separate, and variable impedance control in the second task where force and motion is controlled in a unified manner.

We investigate the performance of NetLfD for constant and time-varying delays using the data collected from 18 participants. The Ethics Commission of the Faculty of Medicine of TUM approved our experiment procedure with the application ID "2023-39-S-KK". The participants' age range is 18-35. 22.2% of the participants have prior experience with haptic devices and robotics. During experiments, the tested delays were randomized. The DBP was kept constant at 15%, achieving substantial packet rate reduction and not exceeding the perceptual limits. The participants first performed a warm-up trial to familiarize themselves with each task. The data collection took 40 minutes on average per participant. We also performed one-way analysis of variance (ANOVA) to verify if there is any significant difference among the results. The teleoperation software was developed in C++ and deployed on two computers with Intel i7-6700HQ. The LfD algorithm was implemented in MATLAB 2020b.

A. Letter-Writing Task

Task Description: The letter-writing task is performed via teleoperation setup depicted in Fig. 3 where the task is implemented in a VE using the CHAI3D library version 3.1.1 [30] at the follower side. The goal is to trace the letter "C" using a virtual pencil and apply the correct amount of force in the Z-direction to achieve the desired gray color intensity which is linearly proportional to the applied force. The VE consists of two surfaces, the right one showing the reference letter and the desired color intensity, and the left only its outline. Fig. 5 illustrates the reference letter. During teleoperation, the acting force on the follower robot is the sum of the controller and computed virtual contact forces.

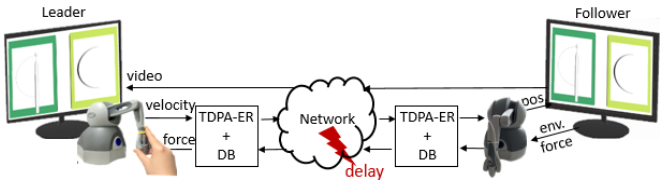


Fig. 3: Bilateral teleoperation setup for the letter-writing task demonstrations. The task is implemented in the VE at the follower.

Learning: We model the two distributions of $\mathcal{P}(\dot{\boldsymbol{x}}; \boldsymbol{x})$ and $\mathcal{P}(F^z; \boldsymbol{x})$ to learn both trajectory and force from the provided demonstrations, inspired by the approaches in [17], [18]. We utilize the proposed WHSMM with the number of states determined using the Bayesian information criterion (BIC).

Reproduction: We use Gaussian Mixture Regression (GMR) to compute the conditional distributions $\mathcal{P}(\dot{\boldsymbol{x}}|\boldsymbol{x})$ and $\mathcal{P}(F^z|\boldsymbol{x})$ and to obtain the position and force to be subsequently given to the follower robot for reproduction. During reproduction, we employ direct force control with a closed feedback loop for the motion-constrained Z-direction and position control for the X and Y-directions. The Cartesian control command is computed with the following:

$$\begin{aligned}\boldsymbol{F}_{cmd} &= \boldsymbol{K}^P\left(\boldsymbol{x}_{d,t} - \boldsymbol{x_t}\right) + \boldsymbol{K}^V\left(\dot{\boldsymbol{x}}_{d,t} - \dot{\boldsymbol{x}}_{\boldsymbol{t}}\right) + \boldsymbol{F}_{d,t} \\ &+ \boldsymbol{K}^{PF} * \left(\boldsymbol{F}_{d,t} - \boldsymbol{F}_{e,t}\right) + \boldsymbol{K}^{IF} * \int_0^t \left(\boldsymbol{F}_{d,t} - \boldsymbol{F}_{e,t}\right), \end{aligned} \tag{11}$$

where $\boldsymbol{x}_{d,t}$ and $\dot{\boldsymbol{x}}_{d,t}$ are the desired and $\boldsymbol{x_t}$ and $\dot{\boldsymbol{x}}_{\boldsymbol{t}}$ are the actual robot position and velocity at time t. $\boldsymbol{F}_{d,t} = [0, 0, F_{d,t}^z]^T$ is the learned force used as the direct feedforward term in the controller [31], and $\boldsymbol{F}_e = [0, 0, F_{e,t}^z]^T$ is the environment force. $\boldsymbol{K}^P$, $\boldsymbol{K}^V$, $\boldsymbol{K}^{PF}$ and $\boldsymbol{K}^{IF}$ are controller constants whose values we have determined empirically.

We use position and force RMSE between the reproduction and the reference to evaluate the accuracy of the reproduction. The reference trajectory is sampled with a fixed number of equidistant points. The distance from these sampled points to the closest points on the reproduced trajectory is then used to compute the position RMSE. The force RMSE is evaluated using the closest points' force error.

Constant Delay Experiments: First, we evaluated the accuracy of demonstrations under constant network delay and examined their correlations with weight metrics. We gathered one demonstration from each participant at each constant delay of $0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$ seconds. Fig. 4 illustrates the scatter plots for the position and force RMSE of demonstrations vs. each metric across all participants. We observed a decline in demonstration accuracy with increasing delay, starting at a delay of 0.2 seconds. Although there are low

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

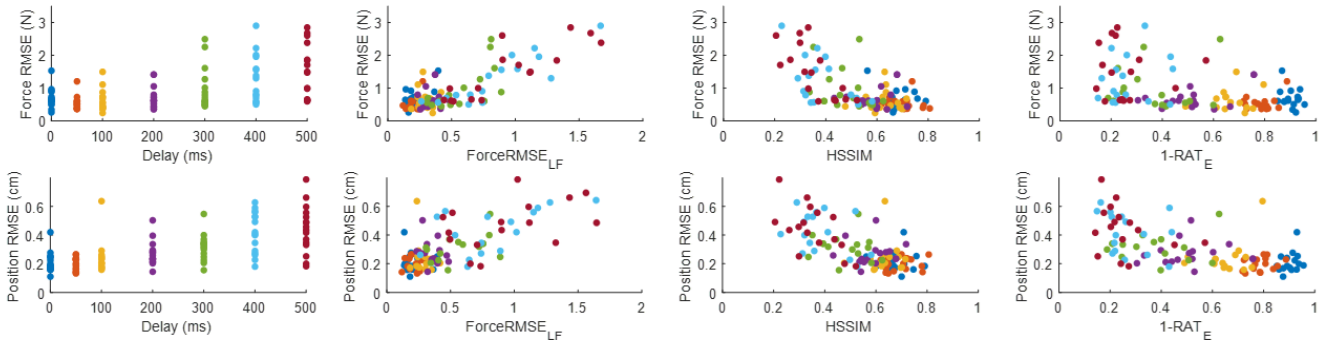


Fig. 4: Letter-writing task scatter plots: Force (top) and position (bottom) errors vs. different metrics for constant delay.

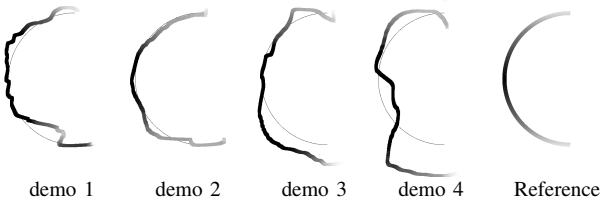


Fig. 5: Constant delay example demonstrations and the reference. From demo 1 to 4 the delays are 0.2s, 0.3s, 0.4s, 0.5s, respectively.

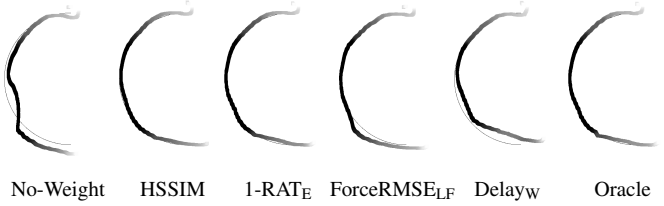


Fig. 6: Constant delay example reproduction results.

TABLE I: Constant Delay Example: Weights

Demo	Delay (s)	HSSIM	Force RMSE_{LF}	1-RAT_{E}	Delay_{W}	Oracle force\pos.
1	0.2	0.69	0.47	0.75	**1.0**	0.38\0.48
2	0.3	**1.0**	**1.0**	**1.0**	0.67	**1.0****1.0**
3	0.4	0.54	0.45	0.45	0.50	0.28\0.24
4	0.5	0.38	0.33	0.28	0.40	0.28\0.11

TABLE II: Constant Delay Example: Repro. Errors

	No Weight	Lowest delay demo	HSSIM	Force RMSE_{LF}	1-RAT_{E}	Delay_{W}	Oracle
Pos. RMSE (cm)	0.33	0.30	0.11	0.15	0.13	0.26	**0.10**
Force RMSE (N)	1.97	1.92	1.43	1.40	1.45	1.51	**1.35**

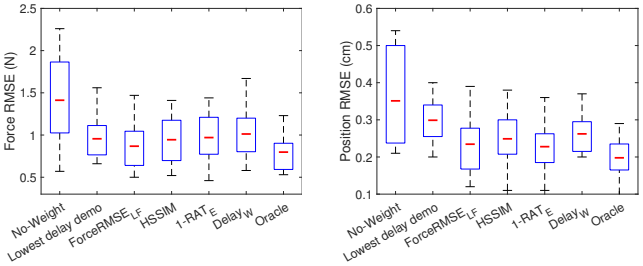


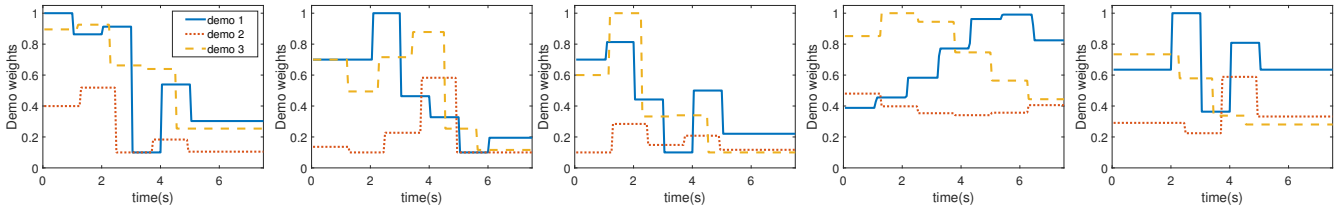
Fig. 7: Constant delay Force and Position RMSE box plots.

error demonstrations at high constant delays, the range of errors expanded. All weight metrics display correlations with position and force RMSE. HSSIM and $\text{ForceRMSE}_{\text{LF}}$ show stronger correlation in force with correlation coefficients of -0.67 and 0.80 than 1-RAT_{E} and delay with -0.52 and 0.47, respectively. In position, the correlation coefficient magnitudes are in the interval of $[0.5, 0.55]$ for all metrics.

Next, we applied the NetLfD approach with WHSMM for each participant using four demonstrations with different delays from 0.2s to 0.5s, as the demonstration accuracy started to decrease with a minimum delay of 0.2s. We define "Oracle" weighting for benchmarking purposes, first to validate the concept of weight-adjusted learning and later to compare it with other proposed metrics. As we have a reference force and position profile, we can compute the actual error in a demonstration and use as the Oracle weight metric. Accordingly, we use both position and force RMSE between the demonstration and the reference separately for weighted learning of $\mathcal{P}(\dot{\boldsymbol{x}}; \boldsymbol{x})$ and $\mathcal{P}(F^z; \boldsymbol{x})$, respectively.

Fig. 5 depicts four demonstrations from one of the participants. The corresponding weight metrics for each demonstration are given in Table I. For Oracle, we provide the weights for force and position learning individually. In this example, all weight metrics, except Delay_{W}, exhibit a similar pattern, assigning the highest weight to the second demonstration. Fig. 6 illustrates the reproduced letters. The corresponding errors and the lowest delay demo error are provided in Table II. The lowest errors are attained by Oracle. There is a noticeable performance enhancement for all weight metrics compared to learning without weighting which we denote as No-Weight. No-Weight corresponds to the approach of using classic HSMM for position and force learning as in [18]. $\text{ForceRMSE}_{\text{LF}}$ resulted in the lowest force error, and HSSIM yields the lowest position error in this particular example.

The results in Fig. 7 indicate that our findings are not only limited to a single participant but mostly consistent across all. Notably, the weighted methods reduce the average position tracking error by up to 34% and the force tracking error by up to 42% compared to the No-Weight baseline. Weighting also exhibits better performance than the demonstration with the lowest delay. Our benchmark comparison, Oracle, produced the lowest mean errors, while the Delay_{W} resulted in the highest mean errors for both position and force tracking. $\text{ForceRMSE}_{\text{LF}}$ and HSSIM weighting approached the results of Oracle the closest. Statistical analysis with ANOVA revealed a significant difference between the results for both force and position RMSE with $F(6, 119) = 6, p < 0.001$, and $F(6, 119) = 6.78, p < 0.001$, respectively. Pairwise comparisons with Tukey's Honestly Significant Difference (HSD) indicated a significant difference between No-Weight



HSSIM 1-RAT$_E$ ForceRMSE$_{LF}$ Delay$_W$ Oracle

Fig. 8: Time-varying delay weights for example demonstrations shown in Fig. 9.

demo 1 demo 2 demo 3

Fig. 9: Time-varying delay example demonstrations.

No-Weight HSSIM 1-RAT$_E$ ForceRMSE$_{LF}$ Delay$_W$ Oracle

Fig. 10: Time-varying delay example reproductions

TABLE III: Time-Varying Delay Example: Repro. Errors

	No Weight	HSSIM	Force RMSE$_{LF}$	1-RAT$_E$	Delay	Oracle
Pos. RMSE (cm)	0.42	0.28	0.29	0.28	0.30	**0.21**
Force RMSE (N)	0.97	0.55	0.48	0.61	0.85	**0.44**

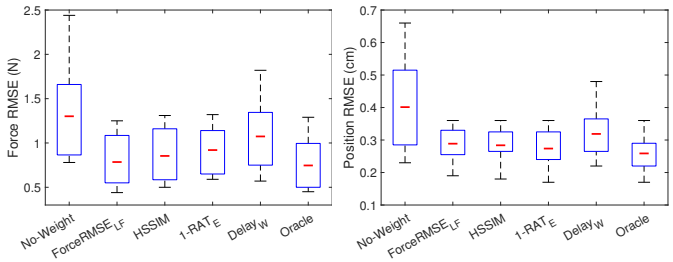


Fig. 11: Time-varying delay Force and Position RMSE box plots.

baseline and all metrics. However, there is no significant difference detected among the weight metrics.

Time-Varying Delay Experiments: In the first scenario, we simplified the occurrence of delay-induced network problems to constant delay. To explore the adaptability of NetLfD to delay fluctuations, we define a sinusoidal time-varying delay that oscillates between 0.15s and 0.45s with a period of 15s. We recorded for each participant three demonstrations where the delay starting point was shifted by a phase angle of 0.7π relative to the start time to generate variability on the amount of delay at different parts of each demonstration. Since the transmission quality varies during the demonstration, we cannot calculate the weights for the entire demonstration but rather for a time window of 1s, which we found to exhibit the highest correlations between weight metrics and errors.

Fig. 9 and Fig. 8 display the three demonstrations and the corresponding time-varying weights for different metrics for one participant, respectively. The reproduced letters are shown in Fig. 10, and the corresponding errors are given in Table III. No-Weight resulted in the highest position and force errors. The three weighting schemes, HSSIM, ForceRMSE$_{LF}$, and 1-RAT$_E$, showed comparable performances. The ForceRMSE$_{LF}$ slightly outperformed the others regarding force error and approached Oracle the most.

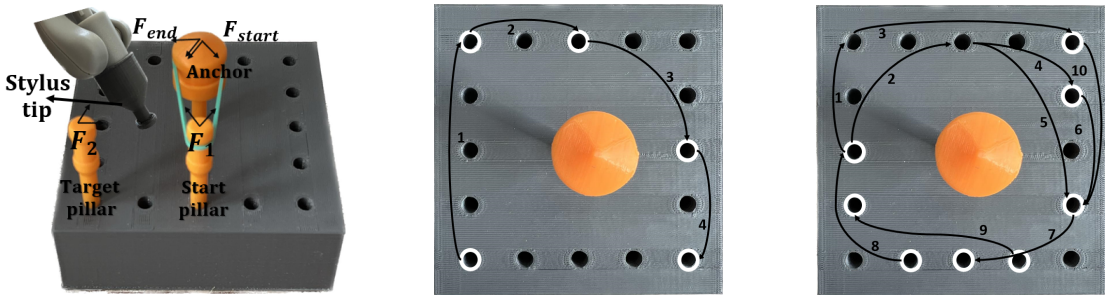
Finally, we aggregated and averaged the results of all participants. Fig. 11 shows that the reproductions of the No-Weight baseline obtained the worst results across all participants. For time-varying delay, weighting strategies reduced the average position tracking error by up to 30% and force tracking error by up to 41%. The approaches based on HSSIM and ForceRMSE$_{LF}$ have the lowest mean force errors and perform similarly to Oracle, whereas 1-RAT$_E$ achieved slightly better performance in position tracking. Similar to the constant delay scenario, we see relatively higher mean errors with Delay$_W$. ANOVA test indicated a significant difference between the results for both force and position RMSE with $F(5, 102) = 5.54, p < 0.001$, and $F(5, 102) = 8.05, p < 0.001$, respectively. Pairwise comparisons with Tukey's HSD showed a significant difference between No-Weight and all metrics except for Delay$_W$ in force RMSE. There is no significant difference detected among the metrics.

B. Rubber Band Placement Task

Task Description: The second task aims to evaluate the effectiveness of NetLfD with varying impedance requirements and test the generalizability to novel conditions. Inspired by the belt drive replacement, we designed a testbed with adjustable pillars and one anchor pillar with a rubber band stretched in between. The objective is to remove the rubber band from the start pillar and install it on the target pillar while maintaining the required tension. We designed an end-effector attachment for the haptic stylus to ease the handling of the rubber band. Fig. 12a shows our 3D-printed testbed, pillars, and stylus tip. As for the first task, a human operator demonstrates the task via teleoperation and receives the video stream captured with a camera at the follower robot side.

To learn and generalize to novel points, we collect four demonstrations with start and end points shown in Fig. 12b. We have empirically determined these points such that the novel test conditions in Fig. 12c can be accomplished in reproduction when clean demonstrations are provided for learning. The test conditions start or end at a point that is not

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.



(a) Task frames (b) Demonstrations (c) Novel tests

Fig. 12: Task frames, demonstrations and novel tests for the rubber band task. Four task frames are used in TP-HSMM. Each novel test either starts or ends at a non-demonstrated point displayed in white in (c).

visited during demonstrations (shown in white in Fig. 12c) and cover different length trajectories and orientations.

Learning: Given the demonstrations with different start and end pillar configurations, we train the proposed WTP-HSMM. For the task parameters, we defined four frames: one for each start and target pillar directed towards the anchor pillar and two for the anchor pillar looking at the start and target pillars (see Fig. 12a). We use a unified attractor model as in [20] for learning and later estimate the stiffness for each state in HSMM by minimizing the residual between the demonstration attractors and the state means. The virtual attractor $\boldsymbol{y_t}$ is formulated using a spring-damper system with:

$$\boldsymbol{y}_t = \boldsymbol{x}_t + \boldsymbol{K}_t^{-P}\left(\boldsymbol{K}_t^{V}\dot{\boldsymbol{x}}_t + \ddot{\boldsymbol{x}}_t - \boldsymbol{f}_t^{e}\right), \tag{12}$$

where $\boldsymbol{x}_t$, $\dot{\boldsymbol{x}}_t$, and $\ddot{\boldsymbol{x}}_t$ represent the end-effector position, velocity, and acceleration, respectively. $\boldsymbol{f}_t^e$ is the environment force at the end-effector. The terms $\boldsymbol{K}_t^P$ and $\boldsymbol{K}_t^V$ are the stiffness and damping matrices at time t. Since the Geomagic Touch haptic device has no built-in force sensors to measure the environment forces, we perform an estimation via a nonlinear disturbance observer proposed in [32] together with the device kinematic and dynamic parameters from [33].

Reproduction: To adapt the model to a novel condition with the new reference frames of $\left\{\tilde{\boldsymbol{A}}_j, \tilde{\boldsymbol{b}}_j\right\}_{j=1}^{F}$, we linearly transform the frame-specific Gaussians back to global coordinates and compute the product of Gaussians. We determine the most likely state sequence, z_t^*, at each time t using the forward variable of HSMM. Then a smooth attractor trajectory, $\boldsymbol{y}_t^*$, is obtained with linear quadratic tracking (LQT). The reader is directed to [20] for more details. For reproduction, the control command is obtained using:

$$\boldsymbol{F} = \boldsymbol{K}_{z_t^*}^{\mathcal{P}}\left(\boldsymbol{y}_t^* - \boldsymbol{x}_t\right) + \boldsymbol{K}_t^{v}\dot{\boldsymbol{x}}_t, \tag{13}$$

where $\boldsymbol{K}_{z_t^*}^{\mathcal{P}}$ represents the estimated stiffness matrices for each state z_{t*} and $\boldsymbol{K}_t^{\mathcal{V}}$ is the damping matrix. The estimated stiffness reveals that high stiffness is necessary for precise picking from and placing onto the pillars of the rubber band, while more compliance is needed for reaching and moving between pillars during contact with the rubber band.

First, to understand the achievable success rate in the reproductions, we collected four demonstrations from each participant without any delay in the network. We used TP-HSMM for learning the attractor as in [20]. During reproduction, the robot should detach the rubber band from the start pillar and stretch it between the target and the anchor pillars to complete the task successfully. When there is no delay during demonstrations, the average success rate of the reproductions is 75% (shown in Fig. 13 with the left-most box plot), which we interpret as the highest achievable performance due to the inherent difficulty of providing clean and smooth demonstrations using the teleoperation setup.

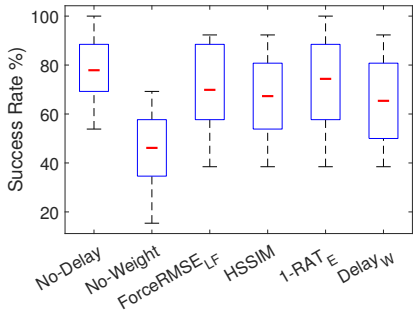


Fig. 13: Rubber band task success rate vs. weight metrics box plots.

Time-Varying Delay Experiments: We evaluated NetLfD for the rubber band task under a time-varying delay that alters between 0.15s to 0.45s with a sinusoidal profile of 15s period and at a DBP of 15%. As in the first task, we started each demo delay with a phase shift of 0.7π relative to the start time. We computed the reliability weights within a window length of 1s to capture the demonstration quality variations. For learning, we only considered successful demonstrations. If a demonstration failed at picking or placing the rubber band or if the rubber band lost contact with the pillars, we asked the participants to repeat the demonstration. We observed that the time-varying delay resulted in inaccurate movements, especially around the start and target pillars.

To evaluate the performance, we aggregated the results from all participants. Fig. 13 shows the average success rate in a boxplot per weighting strategy. We compare the weight metrics against the work in [20] as the baseline LfD of attractor learning with TP-HSMM which is denoted as No-Weight. The reproductions with No-Weight had the lowest success rate of 45.6%. A closer look reveals that the delay-based approach works well but is inferior to other weight metrics, while 1-RAT$_{\mathrm{E}}$ achieved the highest mean success ratio of 74.3%. This is mainly because the task success does not solely need force tracking accuracy but also trajectory precision, which 1-RAT$_{\mathrm{E}}$ successfully captured as an energy-based metric. Overall, our

approach improved the average success rate of the No-Weight ([20]) by up to 63%. With the ANOVA test, a significant difference was observed in the success rate results with $F(4, 85) = 6.71, p < 0.001$. Pairwise comparisons showed a significant difference between the No-Weight baseline and all metrics, whereas no statistically significant difference was detected among the weight metrics.

VI. Conclusions

In this work, we presented NetLfD to enhance the learning performance by accounting for unreliabilities in teleoperated demonstrations due to communication delay. Our analysis showed that delay could decrease demonstration quality and reproduction accuracy. NetLfD aims to improve learning and reproduction quality with a reliability-weighted likelihood formulation with weights linked to teleoperation transparency. We tested the proposed approach for two different in-contact skills. NetLfD outperformed the No-Weight baseline LfDs in force and position tracking accuracy in the letter-writing task by up to 41% and 30%, respectively, and increased the mean success rate by up to 63% in the rubber band placement task under time-varying delay. ANOVA tests verified that NetLfD significantly outperformed No-Weight baselines; however, no statistically significant difference was detected among the weight metrics. Considering the averaged results, HSSIM and $\text{ForceRMSE}_{\text{LF}}$ showed higher performance in the letter-writing in force tracking, while 1-RAT_{E} performed slightly better in position tracking. However, one drawback of HSSIM and $\text{ForceRMSE}_{\text{LF}}$ is the exchange of leader- and follower-side forces after each demonstration. In the rubber band placement, where both force and position are critical for success, 1-RAT_{E} slightly outperformed. Although Delay_{W} showed substantial improvements, it is inferior to other metrics. We foresee that NetLfD is not only applicable for delay but can also improve performance under communication interruptions and packet loss.

Acknowledgment

This work is funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop"(CeTI) of TU Dresden.

References

[1] G. P. Fettweis, "The Tactile Internet: Applications and Challenges," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 64–70, 2014.

[2] F. Fitzek, S. C. Li, S. Speidel, T. Strufe, M. Simsek, and M. Reisslein, *Tactile Internet with Human-in-the-Loop*. Cambridge, MA, USA: Academic Press, 2021.

[3] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robot. Auton. Syst.*, vol. 57, no. 5, 2009.

[4] M. Sakr, Z. J. Li, H. M. Van der Loos, D. Kulić, and E. A. Croft, "Quantifying demonstration quality for robot learning and generalization," *IEEE Robot. and Autom. Letters*, vol. 7, 2022.

[5] W. Si, N. Wang, and C. A. Yang, "A review on manipulation skill acquisition through teleoperation-based learning from demonstration," *Cogn. Comput. Syst.*, vol. 3, pp. 1–16, 2021.

[6] D. A. Lawrence, "Stability and transparency in bilateral teleoperation," *IEEE Transactions on Robotics and Automation*, vol. 9, no. 5, 1993.

[7] P. Hinterseer, S. Hirche, S. Chaudhuri, E. Steinbach, and M. Buss, "Perception-based data reduction and transmission of haptic data in telepresence and teleaction systems," *IEEE Trans. Sig. Proc.*, 2008.

[8] M. Panzirsch, J.-H. Ryu, and M. Ferre, "Reducing the conservatism of the time domain passivity approach through consideration of energy reflection in delayed coupled network systems," *Mechatronics*, vol. 58, 2019.

[9] X. Xu, M. Panzirsch, Q. Liu, and E. Steinbach, "Integrating haptic data reduction with energy reflection-based passivity control for time-delayed teleoperation," in *2020 IEEE Haptics Symposium*, 2020.

[10] S. M. Khansari-Zadeh and A. Billard, "Learning stable nonlinear dynamical systems with Gaussian mixture models," *IEEE Transactions on Robotics*, vol. 27, no. 5, pp. 943–957, 2011.

[11] S. Schaal, *Dynamic Movement Primitives -A Framework for Motor Control in Humans and Humanoid Robotics*. Springer Tokyo, 2006.

[12] S. Calinon, "A tutorial on task-parameterized movement learning and retrieval," *Intelligent service robotics*, vol. 9, no. 1, pp. 1–29, 2016.

[13] A. K. Tanwani *et al.*, "Generalizing robot imitation learning with invariant hidden semi-Markov models," in *International workshop on the algorithmic foundations of robotics*. Springer, 2018.

[14] P. Praveena, G. Subramani, B. Mutlu, and M. Gleicher, "Characterizing input methods for human-to-robot demonstrations," in *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2019.

[15] A. Pervez, A. Ali, J. Ryu, and D. Lee, "Novel learning from demonstration approach for repetitive teleoperation tasks," in *IEEE World Haptics Conference (WHC)*, 2017.

[16] F. Steinmetz, A. Montebelli, and V. Kyrki, "Simultaneous kinesthetic teaching of positional and force requirements for sequential in-contact tasks," in *IEEE-RAS International Conf. on Humanoid Robots*, 2015.

[17] M. Racca, J. Pajarinen, A. Montebelli, and V. Kyrki, "Learning in-contact control strategies from demonstration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016.

[18] C. Zeng, C. Yang, J. Zhong, and J. Zhang, "Encoding multiple sensor data for robotic learning skills from multimodal demonstration," *IEEE Access*, vol. 7, 2019.

[19] L. Rozo, S. Calinon, D. G. Caldwell, P. Jiménez, and C. Torras, "Learning physical collaborative robot behaviors from human demonstrations," *IEEE Transactions on Robotics*, vol. 32, no. 3, 2016.

[20] A. T. Le *et al.*, "Learning forceful manipulation skills from multi-modal human demonstrations," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.

[21] A. Sasagawa, S. Sakaino, and T. Tsuji, "Motion generation using bilateral control-based imitation learning with autoregressive learning," *IEEE Access*, vol. 9, 2021.

[22] M. A. Newton and A. E. Raftery, "Approximate Bayesian inference with the weighted likelihood bootstrap," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 56, no. 1, pp. 3–26, 1994.

[23] M. Markatou, "Mixture models, robustness, and the weighted likelihood methodology," *Biometrics*, vol. 56, no. 2, pp. 483–486, 2000.

[24] I. D. Gebru, X. Alameda-Pineda, F. Forbes, and R. Horaud, "EM algorithms for weighted-data clustering with application to audio-visual scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 12, pp. 2402–2415, 2016.

[25] E. Steinbach *et al.*, "Haptic codecs for the tactile internet," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 447–470, 2019.

[26] J.-H. Ryu, J. Artigas, and C. Preusche, "A passive bilateral control scheme for a teleoperator with time-varying communication delay," *Mechatronics*, vol. 20, no. 7, pp. 812–823, 2010.

[27] S. Calinon, A. Pistillo, and D. G. Caldwell, "Encoding the time and space constraints of a task in explicit-duration hidden Markov model," in *IEEE/RSJ Conf. on Intelligent Robots and Systems*, 2011.

[28] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. of the IEEE*, vol. 77, 1989.

[29] R. Hassen and E. Steinbach, "HSSIM: An objective haptic quality assessment measure for force-feedback signals," in *International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2018.

[30] F. Conti *et al.*, "The CHAI libraries," in *Proc. of Eurohaptics*, 2003.

[31] K. Lynch and F. Park, *Modern Robotics: Mechanics, Planning, and Control*. Cambridge Univeristy Press, 2017.

[32] A. Mohammadi, M. Tavakoli, H. Marquez, and F. Hashemzadeh, "Nonlinear disturbance observer design for robotic manipulators," *Control Engineering Practice*, vol. 21, no. 3, pp. 253–267, 2013.

[33] M. A. Arteaga, A. Gutiérrez-Giles, and J. Pliego-Jiménez, *The Geomagic Touch Haptic Device*. Cham: Springer International Publishing, 2022, pp. 361–374.