Toward Long-Term FMG Model-Based Estimation of Applied Hand Force in Dynamic Motion During Human–Robot Interactions

Umme Zakia, Student Member, IEEE, and Carlo Menon, Member, IEEE

I. INTRODUCTION

FORCE myography (FMG) is a wearable technology used to monitor volumetric changes that occur in a limb during muscle contraction. It can be used as an indirect method of monitoring force or pressure present at the surface of the limb by reading the underlying musculotendinous contractions [1]. FMG is similar to the traditional and widely accepted noninvasive surface electromyography (sEMG) technology that detects electrical activity of underlying muscles [2]. Signals from both technologies are nonstationary, transient, subject-specific, time-series biosignals that can be measured continuously. The standard sEMG is a more prevailing technology used in medical research, habilitation, ergonomics, sports science, exoskeleton and prosthetic control, human–machine interactions (HMI), and human–robot interactions (HRI) [3]–[5]. As a new emerging technology, FMG has gained attention and been found effective in hand gesture recognition, motion recognition, prosthetic control, and rehabilitation applications [6], [7]. FMG biosignals can also be used as an intuitive and promising method in HMI or HRI using standard machine learning (ML) techniques [8] such as sEMG-based HRI applications. Although FMG is comparably new, it is wearable, inexpensive, generates reproducible signals, is easy to operate, and requires simple signal processing compared to sEMG [9]. Furthermore, the FMG technology achieved better results than sEMG in several recent research works [10], [11]. Therefore, this study focused on utilizing the FMG biosignal for an HRI application.

Determining human motions in physical HRIs (pHRI) can play an important role in real-time (RT) interactions and improving safety. Using image and sensory data, human intentions of posture, gesture, hand force, or motion can be predicted with ML algorithms [12], [13]. Common human–robot collaborative tasks such as object handling, transport, or handover require workers to use hand force. Since FMG signals can read muscle contractions during such activities, it can be a viable tool for predicting worker’s intended interactions or applied forces [14]. Therefore, a few studies were conducted to determine isometric hand forces, dynamic forces in motion via FMG signals [15], [16]. Although the study in [16] showed impressive results in recognizing applied forces in dynamic motion during HRI, it had a few shortcomings. First, a separate regression model was trained for predicting forces applied by each participant for one specific motion. Second, separate training datasets were needed if
motions were different because of variations in muscle contractions. This required collecting adequate training datasets each time an individual interacted using a different motion (task) with the robot. Additionally, retraining from scratch was necessary occasionally when FMG bands were taken off and put back on later due to the nonstationary, transient, and individual-specific nature of FMG signals. The collection of another training dataset was time-consuming and impractical for regular use.

Potential application of a wearable FMG band for human workers in industrial HRI workspaces would require general applicability to all workers for control and safety aspects. A trained model that can estimate dynamic hand force via FMG signals is preferable for all individuals in any intended motion during pHRI. However, obtaining such a trained model is feasible only with a bigger and more diverse population dataset. Hence, a novel calibrated FMG-based model is proposed in this study where population data (multiple source domains) were collected over a long period of time during RT interactions between several participants and a linear robot (a biaxial stage). The viability of the proposed long-term calibrated FMG (LCFMG) model was based on the ability to recognize RT unlearned “new input sample data” (target domain) that were out-of-distribution (OOD) compared to the “learned” population data. By “unlearned,” we meant estimating hand force either in “a new unlearned motion” or for “a new unlearned participant.” By “learned” we referred to the population data available to form a long-term “baseline dataset” (aggregated multiple source distributions) for generalization. The proposed model was evaluated in estimating forces in a) scenario 1: unlearned/ unseen motion, and b) scenario 2: unlearned/ unseen participant. As RT test data (target domain) in these scenarios were quite different from the baseline dataset, few calibration data were required for practical evaluation of the model, thereby allowing for the use of generalized zero-shot learning (GZSL) for the proposed model. Recognizing unseen scenarios using calibration data for domain adaptation is a relatively unexplored area in FMG-based HRI. These calibration data were distinct from the long-term population dataset and similar to test samples. Several combinations of population data and calibration data were used to train a few LCFMG-based models. Performances of these LCFMG models were evaluated in RT and compared to each other. A few models were also trained with “new input sample data” only (no population data), as mentioned in [16]. These models were termed as “specialized trained models (STMs)” hereinafter and were compared with the proposed LCFMG models for performance evaluation.

The rest of this article is organized as follows. Section II describes the experimental setup, and Section III presents the methodology used. A novel LCFMG-based model is proposed in Section IV while the protocol followed is explained in Section V. Performance evaluations of the LCFMG models are presented in Section VI and discussed in Section VII. Section VIII concludes this article.

II. EXPERIMENTAL SETUP

In this study, HRI was observed between a healthy participant and a biaxial stage/linear robot. A custom-made knob-like gripper/end-effector mounted on top of the stage allowed participants to grasp and apply force in an intended motion, as shown in Fig. 1. Two customized FMG bands, each consisting of 16 force-sensing resistors (TPE 502C, Tangio Printed Electronics, Canada), were donned on a participant’s dominant forearm and upper arm positions to read muscle contractions (32 feature spaces). The biaxial stage consisted of two identical linear stages of 450 mm in dimension (X-LSQ450B, Zaber Technologies, Vancouver, BC, Canada) and the total workspace was 450 mm x 450 mm. A force-torque (FT) sensor (Mini45, ATI Industrial Automation, Apex, NC) housed inside the gripper measured user-applied forces as a true label generator. Labeled FMG data were transmitted to a desktop computer (Intel core i7 7700k, GTX 1080) using data acquisition platforms (NI DAQs 6259, 6341, and 6210, National Instruments, Austin, TX, US). Data collection, training of the models, RT test phases, and control command were executed through a LabVIEW interface and MATLAB scripts.

For RT control design, the conventional admittance control scheme was implemented [26]. External forces applied to the gripper were translated into torques at each joint of the stage, and the stage moved to a new position based on the calculated displacement. Thus, the gripper slid easily in the XY planar workspace according to the applied hand force, ensuring “complaint collaboration.” Support vector regression (SVR), a popular supervised ML technique, was implemented for estimating applied hand force in an intended motion. Hyperparameters such as $L_2$ regularization, Gaussian kernel functions, and epsilon insensitive function as loss functions were used [17]. To fit real continuous numbers, best values for cost (C) and gamma (G) were obtained using a grid search. During the training/learning phase, the biaxial stage was controlled by the FT sensor readings of the applied hand force. While in the RT test/evaluation phase, displacements of the stage were controlled by estimated applied force via FMG signals for compliant collaboration.

III. METHODOLOGY

In ML, the ability of a trained model to fit unseen test data without compromising performance is essential for many real-world applications. To predict categories or regressing new, unlearned test samples, zero-shot learning (ZSL) can be implemented where the trained model transfers knowledge learned from the source (training) distribution to the target (test) distribution. Hence, in ZSL, a model trained with source distribution attempts to predict test samples from a target domain that it has never seen before (source distribution ≠ target distribution) [18]. This approach allows a model to perform better with lower generalization error when the “new input sample data” (test data) has some similarity to the model population data (training data) [19]. The generality to correctly predict such OOD samples (test data) is doable when training data include all possible distribution, which is practically not feasible. Specially, it is difficult to implement ZSL in practical FMG-based applications. As FMG is a transient, nonstationary time-series biosignal, and affected by arm posture and motion of limb movements or physiological changes (sweats, tiredness), variations are always
present in the streaming signal. Therefore, direct transfer using the population/baseline dataset (multiple source domain) to implement ZSL might fail in predicting RT, dynamic, unseen test data (target domain) via FMG signals.

A generalized approach to ZSL allows both seen/learned source samples and unseen/unlearned test samples available during evaluation, hence known as generalized ZSL [20]. It bridges the gap between the seen source domain and the unseen target domain by leveraging their semantic information. Since an FMG biosignal is subjective-specific and sensitive to sensor position changes/band displacement each time a band is donned on a participant, it is treated as a different domain. To overcome such performance and inherent challenges and to recognize unseen RT test data, domain adaptation is a favorable alternative as it allows a few test samples included in the source distribution. During RT test phase, this will enable the model to better predict unseen, new target data.

Interestingly, these approaches were investigated recently using sEMG. Like the FMG signal, the traditional sEMG signal is influenced by electrode shift, electrode placement on the limb, muscle contraction intensity, and transient changes in the signal [21]. To address these dynamic challenges, many recent studies showed that domain adaptation using transfer learning can be valuable for real-world sEMG-based applications [22], [23]. Some of these studies addressed electrode shifts and day-to-day variability through adaptive transfer learning [24], [25], inter-session gesture recognition using deep domain adaptation on unlabeled test data or fine-tuning labeled calibration data [26], and periodic recalibration for multiple days use for prosthetic control by applying transfer learning with few training data [27]. Furthermore, a recent study showed that aggregating source distributions from multiple users with deep transfer learning in gesture recognition enhanced model performance [28]. Since an FMG signal has similar characteristics, multiple domain adaptation was investigated with the traditional ML algorithm in this study.

The proposed framework utilized a fully supervised multiple domain adaptation with a modified GZSL method where both source and target distributions were somewhat different but had the same feature spaces. In this study, our objective was to train a model with multiple source distributions [baseline dataset aggregated from a full/subset of “reference dataset” only, and/or “learned participant dataset” only, (Section IV)] and evaluate new, unseen RT test samples [target domain: unseen motion (scenario 1)/unseen participant (scenario 2)], as shown in Fig. 2. For FMG-based RT interaction, instantaneous signals were required to represent muscle contractions. In recognizing unseen RT test samples, a modified GZSL in adapting knowledge transfer helped where a few test samples were included in the aggregated multiple source distributions. Therefore, fewer “new input sample data” or calibration data (applied force in an intended motion) were collected from a participant at the beginning of an evaluation period. These data were like the RT test samples (target domain) and were aggregated with baseline source distribution for training purposes, as described in Section IV. As the seen calibration data were like the unseen RT test data, the model could learn what was expected from the target domain and predicted better. Therefore, the essence of the proposed model to predict OOD data was evaluated in a) scenario 1 # unseen motion: a learned participant applying force in a “new, unlearned” motion, and b) scenario 2 # unseen participant: a “new, unlearned” participant applying force in a learned motion. In this context, “learned” referred to the aggregated multiple source distributions acquired from a “reference dataset” and/or a “learned participant dataset.” In the case of “unlearned,” no/negligible “new input sample data” were included in the training dataset and referred to as the “calibration dataset.”

In this study, a total of 15 participants (P1–P15) contributed voluntarily. Source domains collected from the first five participants (P1–P5) formed a “reference dataset” for generalization. The other 10 participants (P6–P15) contributed to evaluate the proposed framework in RT. Among them, three participants (P6–P8) contributed to collect “learned participant dataset” and to evaluate model performance in scenario 1; while the remaining...
seven participants (P0–P15) took part in scenario 2 for evaluation. To recognize unseen/unknown motion or unseen/unknown participant in the two scenarios (as described in Section IV-A and IV-B), a training dataset was formed using different source distributions.

### A. Long-Term Baseline Dataset

A “long-term baseline dataset” or simply “baseline dataset” (aggregated multiple source distributions) was accumulated from a “reference dataset (subset/full)” only, and/or a “learnt participant dataset (subset)” only.

1) **Reference dataset** $D_{R1}$: Multiple source domains were aggregated into a reference dataset. It was a collection of data from two FMG bands placed on the forearm and upper arm of five participants (P1, P2, P3, P4, and P5) capturing muscle contractions during interactions with the linear robot. Participants applied forces in dynamic motions, namely X direction ($M_1$), Y direction ($M_2$), diagonal ($M_3$), square ($M_4$), and diamond ($M_5$) in the XY plane demonstrating both simple and complex planar motions. These diverse pools of multiple source domains offered better generalization in domain adaptation. The reference dataset (full/subset) was used in both scenarios 1 and 2 in recognizing unseen test samples.

2) **Learnt participant dataset** $D_{R2}$: To recognize unseen motion in scenario 1, a few source distributions were collected from volunteering participants (P6–P8). For each participant, a separate “learnt participant dataset” was created during interactions in five different motions (M1, M2, M3, M4, and M5). This helped to generalize source domains collected from a participant for target domain adaptation. Participants (P6–P8) in scenario 1 were termed as “learned” because some known or seen data from the “learnt participant dataset” (subset of it) was used in training, as described in Section IV-A.

Both the “reference dataset” and “learnt participant dataset” were collected over a long period of time before the evaluation period, and hence “long-term” was used to describe this model.

### B. Calibration Dataset $D_{SRT}$

The “calibration dataset” was collected at the beginning of the RT evaluation period of the framework. It was called the “new input sample data” (0, 1, or 2 repetitions of 0, 400, or 800 samples data where a participant interacted with the robot for a certain time in each repetition). To evaluate the proposed models, force estimation in two intended motions: “X direction ($M_1$)” and “Diagonal ($M_3$)” were considered. So, either in scenarios 1 or 2, calibration data collected from a participant (P6–P15) in an intended motion determined the RT intended motion for interaction during evaluation.

To evaluate the proposed LCFMG-based model implementing GZSL multiple domain adaptation, baseline dataset and calibration dataset were aggregated in various combinations to train few models and predict unseen test samples, as shown in Fig. 3. Several cases were investigated to obtain a possible solution in real-world FMG-based HRI scenarios, as described below:

**1) Scenario 1:** Estimating Force in Unlearned Motion $M_1$:

In this scenario, the applied force in a new unlearned motion $M_U$ ($U = 13$) was attempted by a learned participant $P_L$ ($L = 67,8$) to interact with the biaxial stage, as shown schematically in Fig. 4. “New input sample data” from executions of such a motion were used to create a “calibration dataset.” Two cases were considered in scenario 1.

   a) **Case 1 # intra-participant:** The “baseline dataset 1” for intended motion $M_1$ was formed from a subset of a “learnt participant dataset” i.e., FMG data of applied force in four motions: $M_2, M_3, M_4, M_5$ by a participant, $P_L$ ($L = 67,8$). Similarly, the force exerted in $M_1, M_2, M_4, M_5$ motions by a participant $P_L$ ($L = 67,8$) contributed to the “baseline dataset 1” for intended motion $M_3$. The intra-participant training dataset was formed as follows:

   Training dataset $[M_U = 1]$
   \[
   \left\{M_2, M_3, M_4, M_5\right\}_{P_L} \cup \left\{M_1\right\}_{P_L} \quad (1)
   \]

   Training dataset $[M_U = 3]$
   \[
   \left\{M_1, M_2, M_4, M_5\right\}_{P_L} \cup \left\{M_3\right\}_{P_L} \quad (2)
   \]
b) Case ii # inter-participant: The “baseline dataset 2,” in this case, was formed by augmenting “baseline dataset 1” from intraparticipant with a subset of the reference dataset. Such as, for intended motion $M_1$, FMG data from five participants ($P_1$–$P_5$) interacting in $M_2$, $M_3$, $M_4$, $M_5$ motions were used. According to (3) and (4), inter-participant training datasets were formed.

Training dataset $[M_U = 1]$

\[
\left\{ M_2, M_3, M_4, M_5 \right\} p_1, p_2, p_3, p_4, p_5, p_6 \cup \{ M_1 \} p_L \quad \text{Calibration dataset}
\]

Training dataset $[M_U = 3]$

\[
\left\{ M_1, M_2, M_4, M_5 \right\} p_1, p_2, p_3, p_4, p_5, p_6 \cup \{ M_3 \} p_L \quad \text{Calibration dataset}
\]

In each case, no/few calibration data (“new input sample data”: $P_L$ interacting with the robot in unlearned motion $M_U$) of zero, one, or two repetitions (0, 400, 800 samples) augmented with baseline datasets resulted in the LCFMG-0, LCFMG-1, LCFMG-2 trained models, respectively. To compare the performance of these models, STMs (STM-1, STM-2, and STM-5) were generated with different repetitions (1, 2, and 5) of calibration datasets of an unlearned intended motion, $M_U (U = 13)$, performed by $P_L (L = 67,8)$.

2) Scenario 2: Estimating Force for Unlearned Participant $P_U$: In this scenario, a new unlearned participant, $P_U (U = 9, \ldots, 15)$, with no prior information, interacted with the biaxial stage in a learned intended motion, $M_U (L = 13)$. For generalization, the reference dataset (subset/full) was used. A “new input sample data” from the execution of the intended motion was used as the “calibration dataset.” To construct the training dataset, two cases were considered:

a) Case i # intra-motion: The “baseline dataset 1” included a subset of the reference dataset of FMG data collected from five participants ($P_1$–$P_5$) applying force in an intended motion $M_L (L = 13)$. The intra-motion training dataset was formed according to (5) and (6).

Training dataset $[M_L = 1]$

\[
\left\{ M_1 \right\} p_1, p_2, p_3, p_4, p_5 \cup \{ M_1 \} p_U \quad \text{Calibration dataset}
\]

Training dataset $[M_L = 3]$

\[
\left\{ M_3 \right\} p_1, p_2, p_3, p_4, p_5 \cup \{ M_3 \} p_U \quad \text{Calibration dataset}
\]

b) Case ii # inter-motion: The full reference dataset was used as the “baseline dataset 2.” According to (7) and (8), the inter-motion training dataset was formed.

Training dataset $[M_L = 1]$

\[
\left\{ M_1, M_2, M_3, M_4, M_5 \right\} p_1, p_2, p_3, p_4, p_5 \cup \{ M_1 \} p_U \quad \text{Calibration dataset}
\]

Training dataset $[M_L = 3]$

\[
\left\{ M_1, M_2, M_3, M_4, M_5 \right\} p_1, p_2, p_3, p_4, p_5 \cup \{ M_3 \} p_U \quad \text{Calibration dataset}
\]

LCFMG-based models for these cases were trained by including no/few calibration data (“new input sample data”: unlearned $P_U$ interacting with the robot in intended motion $M_U$) of zero, one, or two repetitions (0, 400, or 800 samples) augmented with the baseline dataset resulted in the LCFMG-0, LCFMG-1, LCFMG-2 trained models, respectively. A schematic of this scenario is illustrated in Fig. 5. For performance comparison, STMs (STM-1, STM-2, and STM-5) were generated using different repetitions (1, 2, and 5) of the calibration dataset in an intended motion $M_L (L = 13)$ performed by a new participant $P_U (U = 9, \ldots, 15)$.

V. STUDY PROTOCOL

Fifteen healthy participants (33±8 years, 12 males and 3 females, all right-handed) voluntarily participated in this study. All participants acknowledged the study protocol and gave their written consent as approved by the Office of Research Ethics at Simon Fraser University, Canada. Two FMG bands were donned on the forearm and upper arm positions of each participant during a PHRI session, as shown in Fig. 1. During interactions, the applied force was kept within 20–60 N with a maximum voluntary contraction of 30%–80%. This helped to avoid muscle fatigue and provided force applied within reasonable ranges. Observations showed that during interactions between participants and the linear robot, mean displacements in the intended motions were 400 mm (X motion) and 475 mm (Diagonal motion: ~320 mm in X-dimension, ~350 mm in Y-dimension) although maximum allowed displacements were higher [450 mm (X motion) and 630 mm (Diagonal motion)].

The study spanned over two separate periods of a) “long-term training data collection period,” followed by b) “real-time evaluation period” of the proposed models. Datasets collected in these two periods and the formation of the training dataset are shown in Table I for intended motion $M_I$. 

Additional figure showing the schematic of scenario 2: LCFMG-based model generation to recognize unlearned participant applying force in motion ($M_I$).
TABLE I

<table>
<thead>
<tr>
<th>Training Data Collection Period</th>
<th>Real-time Evaluation Period</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reference dataset</strong></td>
<td><strong>Baseline dataset</strong></td>
</tr>
<tr>
<td><strong>Scenario 1</strong></td>
<td><strong>Case 1: B = Q₁</strong></td>
</tr>
<tr>
<td>A. Long-Term Training Data Collection Period</td>
<td>Here, Q₁ = {P₃[M₃], ...} for each participant</td>
</tr>
<tr>
<td>5 participants (P₃–P₅)</td>
<td>8000 samples/sensor</td>
</tr>
<tr>
<td>5 motions (M₃–M₅)</td>
<td>5 motions * 2000 samples/motion</td>
</tr>
<tr>
<td>2000 samples/motion for each participant</td>
<td></td>
</tr>
<tr>
<td><strong>Scenario 2</strong></td>
<td><strong>Case 1: B = D ∈ R</strong></td>
</tr>
<tr>
<td>B. RT Evaluation Period</td>
<td>Here, D = M₉ {P₆[M₉], ...} for each participant</td>
</tr>
<tr>
<td>50000 samples/sensor [5 motions * 5 participants * 2000 samples/motion]</td>
<td></td>
</tr>
<tr>
<td>Tested participants</td>
<td>10000 samples [1 motion * 5 participants * 2000 samples/motion]</td>
</tr>
<tr>
<td>X</td>
<td></td>
</tr>
<tr>
<td>DIAGOAL</td>
<td>Fig. 6. Applied forces and displacements during the interaction between a contributing participant (P₆) and the linear stage in RT evaluation.</td>
</tr>
</tbody>
</table>

A. Long-Term Training Data Collection Period

The multiple source domains, i.e., the baseline datasets [“reference dataset” from P₁–P₅ and “learned participant dataset” from P₆–P₈] required for different scenarios were collected in this period. These datasets were collected in multiple sessions over a few days where several participants (P₁–P₈) interacted with the robot. The baseline dataset was considered as the offline training dataset used for generalization and domain transfer knowledge. In this period, only the data collection phase was executed. In each training data collection session, the participant applied force in an intended motion during interaction and was considered as a separate source domain.

At the beginning of a session, a participant wearing the two FMG bands sat comfortably on a chair in front of the biaxial stage with his/her shoulder and back straight on a chair locked in position. For compliant collaboration, the participant grasped the gripper and applied force in an intended motion and continued repeatedly until 400 data samples/sensor were collected; this was termed as one “repetition.” Five repetitions were performed to collect a total of 2000 samples of “source domain sample data” from each sensor. Collected data were labeled and saved for later use in training the models. This phase lasted for approximately 12–15 min.

B. RT Evaluation Period

Ten participants (P₆–P₁₅) contributed to the RT evaluation of the proposed LCFMG models. During this period, a separate session was conducted for each participant in investigating a single scenario (two cases) for two intended motions (M₁, M₃), as shown in Fig. 6. Six LCFMG models and three STM models were evaluated for one participant in a scenario (1/2) in one intended motion (M₁/M₃). This required around 90 min to complete the evaluation for each participant for the two intended motions. A calibration dataset was collected at the beginning of the evaluation period. To evaluate the proposed models in each motion, three phases were executed consecutively:

1) Calibration data collection phase in which participants interacted with the biaxial in an intended motion.
2) Model training phase in which several LCFMG-based models were trained by merging different combinations of baseline and calibration datasets of acquired FMG data as described in Section IV-A and -B.
3) Test phase in which the performance of the LCFMG-based models was evaluated in RT to estimate user-applied forces in dynamic motion.
TABLE II
SUMMARY OF TRAINING DATASETS USED FOR LCFMG-BASED MODEL

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Case</th>
<th>Contributing participant</th>
<th>Performed motion</th>
<th>Baseline dataset</th>
<th>Calibration dataset</th>
<th>Repetitions</th>
<th>Samples/sensor</th>
<th>LCFMG model</th>
<th>R²</th>
<th>NRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCFMG-0</td>
<td>P&lt;sub&gt;1&lt;/sub&gt; (L=67,8)</td>
<td>M&lt;sub&gt;0&lt;/sub&gt; (U=1,3)</td>
<td>In case of L=1:</td>
<td>[M&lt;sub&gt;1&lt;/sub&gt;, M&lt;sub&gt;2&lt;/sub&gt;, M&lt;sub&gt;3&lt;/sub&gt;, M&lt;sub&gt;4&lt;/sub&gt;, M&lt;sub&gt;5&lt;/sub&gt;]</td>
<td>(M&lt;sub&gt;1&lt;/sub&gt;)&lt;sub&gt;L&lt;/sub&gt;</td>
<td>0</td>
<td>8000</td>
<td>Intra-Participant-LCFMG-G0</td>
<td>&lt;0.45</td>
<td>&gt;0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>In case of L=3:</td>
<td>[M&lt;sub&gt;1&lt;/sub&gt;, M&lt;sub&gt;2&lt;/sub&gt;, M&lt;sub&gt;3&lt;/sub&gt;, M&lt;sub&gt;4&lt;/sub&gt;, M&lt;sub&gt;5&lt;/sub&gt;]</td>
<td>(M&lt;sub&gt;1&lt;/sub&gt;)&lt;sub&gt;L&lt;/sub&gt;</td>
<td>1</td>
<td>8400</td>
<td>Intra-Participant-LCFMG-G1</td>
<td>0.89</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>In case of L=3:</td>
<td>[M&lt;sub&gt;1&lt;/sub&gt;, M&lt;sub&gt;2&lt;/sub&gt;, M&lt;sub&gt;3&lt;/sub&gt;, M&lt;sub&gt;4&lt;/sub&gt;, M&lt;sub&gt;5&lt;/sub&gt;]</td>
<td>(M&lt;sub&gt;1&lt;/sub&gt;)&lt;sub&gt;L&lt;/sub&gt;</td>
<td>2</td>
<td>8800</td>
<td>Intra-Participant-LCFMG-G2</td>
<td>0.92</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>In case of L=3:</td>
<td>[M&lt;sub&gt;1&lt;/sub&gt;, M&lt;sub&gt;2&lt;/sub&gt;, M&lt;sub&gt;3&lt;/sub&gt;, M&lt;sub&gt;4&lt;/sub&gt;, M&lt;sub&gt;5&lt;/sub&gt;]</td>
<td>(M&lt;sub&gt;1&lt;/sub&gt;)&lt;sub&gt;L&lt;/sub&gt;</td>
<td>0</td>
<td>40000</td>
<td>Inter-Participant-LCFMG-G0</td>
<td>&lt;0.45</td>
<td>&gt;0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>In case of L=3:</td>
<td>[M&lt;sub&gt;1&lt;/sub&gt;, M&lt;sub&gt;2&lt;/sub&gt;, M&lt;sub&gt;3&lt;/sub&gt;, M&lt;sub&gt;4&lt;/sub&gt;, M&lt;sub&gt;5&lt;/sub&gt;]</td>
<td>(M&lt;sub&gt;1&lt;/sub&gt;)&lt;sub&gt;L&lt;/sub&gt;</td>
<td>1</td>
<td>40400</td>
<td>Inter-Participant-LCFMG-G1</td>
<td>0.91</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>In case of L=3:</td>
<td>[M&lt;sub&gt;1&lt;/sub&gt;, M&lt;sub&gt;2&lt;/sub&gt;, M&lt;sub&gt;3&lt;/sub&gt;, M&lt;sub&gt;4&lt;/sub&gt;, M&lt;sub&gt;5&lt;/sub&gt;]</td>
<td>(M&lt;sub&gt;1&lt;/sub&gt;)&lt;sub&gt;L&lt;/sub&gt;</td>
<td>2</td>
<td>40800</td>
<td>Inter-Participant-LCFMG-G2</td>
<td>0.94</td>
<td>0.16</td>
</tr>
</tbody>
</table>

1) Calibration Data Collection Phase: Labeled FMG data were collected the same way as described in Section V-A. For compliant collaboration, a participant sat comfortably with their arm parallel to the horizontal space, grasped the gripper of the robot, applied force in an intended motion (M<sub>1</sub> or M<sub>3</sub>), and continued interaction repeatedly until 400 samples were collected in one “repetition.” Five repetitions were executed to collect a total of 2000 samples of “new input sample data” from each sensor. Among these, repetitions 1 and 2 only were used as “calibration data.” Collecting two repetitions of calibration data required approximately 5 min while the whole session was conducted in 12–15 minutes.

2) Model Training Phase: During this phase, several models were trained while the participant sat comfortably and relaxed with the FMG bands still wrapped around his/her arm. The training dataset for each model augmented the baseline dataset with “calibration data” of 0, 400, or 800 samples of “new input sample data.” Three separate long-term models (LCFMG-0, LCFMG-1, LCFMG-2) were trained for each case in one scenario, as described in Section IV. For two cases in one scenario, six separate LCFMG-based models (Scenario 1: Intra-participant-LCFMG-0/1/2 and inter-participant-LCFMG-0/1/2, Scenario 2: Intra-motion-LCFMG-0/1/2 and inter-motion-LCFMG-0/1/2) were trained. Also, three STMs (STM-1, STM-2, STM-5) were generated using only “new input sample data” (1, 2, and 5 repetitions or 400, 800, and 2000 samples from calibration dataset) for performance evaluations. Table II lists a detailed description of baseline and calibration datasets, number of samples collected from each sensor, and the model generated for each scenario and case.

3) Test Phase: A block diagram of the RT test phase summarizing the procedure followed to evaluate the performance of the trained models is shown in Fig. 7 (also in Figs. 4 and 5). After the models were trained, each of them (six LCFMG models and three STM models for one motion in a scenario) was evaluated separately, as listed in Table II. During this phase, the

![Fig. 7. RT test phase evaluating an LCFMG model in recognizing unlearned motion or unlearned participant.](image-url)
robot was controlled by the estimated FMG-based applied force in motion predicted by the LCFMG model on incoming RT test data. The estimated force was mapped into displacements for the biaxial stage. This allowed compliant collaboration where the robot followed the same trajectory as the intended motion. Each trained model was evaluated approximately for 120 s.

VI. PERFORMANCE EVALUATION

Two outcome measures: co-efficient of determination ($R^2$) and normalized root mean square error (NRMSE) were used to evaluate the performance of the LCFMG-based trained models. Performance of the different models in RT test phases is reported in this section as box plots of $R^2$ and NRMSE with median values in Figs. 9–12 and listed in Table II.

To verify the capability of the proposed framework, a five-fold cross validation (CV) on the training dataset (baseline dataset augmented with calibration data) was conducted. CV was carried out for each participant (P6–P15) in each intended motion ($M_1$ and $M_3$). Inter-participant-2 and inter-motion-2 cases were considered where baseline datasets had larger labeled multiple source distributions (see Table II). Training data were shuffled for uniform distribution after aggregating baseline and calibration data. CV accuracies ($R^2$) in estimating force in each dynamic motion ($M_1$ and $M_3$) are reported separately in the boxplot shown in Fig. 8. In both scenarios, median values of CV accuracies were quite higher ($R^2\approx 90\%$) and approximately the same for both “X” and “Diagonal” motions across participants. This verified the applicability of generality and implementation of GZSL with domain adaptation in recognizing new, unseen, RT test data.

As explained in the previous sections, two different cases of intra/inter-participant and intra/inter-motion were considered in each scenario. During the RT evaluation, the model LCFMG-0 trained with no calibration data, failed to estimate user-applied forces (average $R^2<0.45$, average NRMSE>0.3) in each case. Therefore, the intra/inter-participant-LCFMG-0 and the intra/inter-motion-LCFMG-0 models are not discussed in the following sections.

A. Co-efficient of Determination ($R^2$)

A statistical tool, coefficient of determination ($R^2$), was used to assess the ability of the ML model to predict future values in regression analysis [19]. In the case of analyzing a continuous stream of data in RT, $R^2$ quantified how much variations of the dependent variable were due to the variations of independent variables that were explained by the model as

$$R^2 = \frac{\text{Variations of independent variables}}{\text{Total variance}}.$$  \hspace{1cm}(9)

$R^2$ values varied between 0 and 1, with higher values representing better estimation accuracy.

1) Scenario 1: Unlearned Motion $M_U$: RT evaluation of the long-term models in scenario 1 showed that inter-participant-LCFMG-2 performed better in force estimation for both unlearned motions ($M_1 = \text{X-direction, } M_3 = \text{Diagonal}$). The inclusion of two repetitions of calibration data boosted the performance (see Fig. 9); this was noticeable in the case of $M_3$ ($R^2 = 0.85$ with LCFMG-2 versus $R^2 = 0.67$ for LCFMG-1). Comparison of the performance of the long-term models with the STMs showed the competitive advantages of the proposed model. The inter-participant-LCFMG-2 outperformed STM-2 and was comparable with STM-5 in estimating user-applied forces in the “X-direction” motion ($M_1$). This LCFMG-based model also outperformed both STM-2 and STM-5 during “diagonal” motion ($M_3$) (highlighted in “gray” in Table II).
2) Scenario 2: Unlearned Participant PU:

RT evaluation results of Scenario 2 where unlearned participants interacted with the stage in the intended motions of M\textsubscript{1} and M\textsubscript{3} (X-direction, Diagonal) are illustrated in Fig. 10. In both intra- and inter-motion cases, LCFMG-1 and LCFMG-2 showed comparable performances, although LCFMG-1 did not perform well for all participants. In both intended motions, the inter-motion-LCFMG-2 model performed better and was comparable with STMs STM-2 and STM-5, as observed in Table II and Fig. 10.

B. Normalized Root-Mean-Square Error

Root-mean-square error was used to measure errors that the regression model showed in predicting the output. This measure was then divided by the mean of measured data to obtain NRMSE:

\[
\text{NRMSE} = \sqrt{\frac{1}{n} \sum_{i}(Y_{\text{est}} - Y_i)^2}{\text{mean}(Y)}
\]  

(10)

where \( Y \) was measured data, \( n \) was the number of samples, and \( Y_{\text{est}} \) was the prediction made by the regression model. This measure was particularly useful to compare model performance for datasets with different ranges [17]. Lower NRMSE values indicated a better fit between model estimations and true labels.

1) Scenario 1: Unlearned Motion M\textsubscript{U}:

Comparing the RT performance of different models in both cases of this scenario showed that in most cases, LCFMG-2 outperformed LCFMG-1 as observed by a smaller average NRMSE (see Table II) and lower range of error variations (see Fig. 11). The inter-participant-LCFMG-2 model was superior to STM-2 and STM-5 in estimating the user-applied forces during the two intended motions M\textsubscript{1} and M\textsubscript{3} (see Table II). In addition, the range of variations of NRMSE with inter-participant-LCFMG-2 was less than that of the STM-2 model (see Table II) during M\textsubscript{1}.

2) Scenario 2: Unlearned Participant PU:

Calculating NRMSE in the RT evaluation of this scenario showed similar results such as scenario 1. In most cases, LCFMG-2 outperformed LCFMG-1, STM-2, and STM-5 with a lower NRMSE. These errors are presented in boxplots of Fig. 12 and reported in Table II.

C. Comparative Analysis

A comparison of the performance of different models is illustrated in Fig. 13. The LCFMG-based models generally acquired better prediction accuracies and lower standard deviations in contrast to the STMs. This superior performance was clearly noticed when comparing LCFMG-2 models with the STM-2 and STM-5 models in Fig. 13. The inter-participant-LCFMG-2 model in scenario 1 (recognizing new, unseen motion) and
Fig. 12. RT evaluation of scenario 2: NRMSE values are reported for each trained model. (a) Unlearned participant performing X direction (M1) pattern. (b) Unlearned participant performing Diagonal (M3) pattern.

Intra-motion-LCFMG-2 model in scenario 2 (recognizing new, unseen participants) performed better among the models. It is worth noting that the LCFMG-2 model obtained high prediction accuracies with limited information, i.e., only two repetitions of calibration of “new input sample data.”

Statistical comparison of the performance among the proposed LCFMG-based models with the specialized model was conducted for the viability of generalizations. In this study, the two one-sided test (TOST) [29], an equivalence test, was conducted. The TOST equivalence test showed that inter-participant-LCFMG-2 was equivalent to STM-5 (scenario 1: “unlearned motion”) with a hypothesized mean difference of 0.71 at 5% significance level. Similarly, intra-motion-LCFMG-2 was found equivalent to STM-5 (scenario 2: “unlearned participant”) with a hypothesized mean difference of 0.707 at 5% significance level. However, it was difficult to be conclusive with a smaller population size, as in this study.

In both scenarios for “Diagonal” intended motion, the inter-participant/intra-motion-LCFMG-1 model did not perform well, as indicated in Fig. 14. To interact in a “Diagonal” motion pattern, a participant had to apply enough force to cause displacements in both X and Y dimensions. Recognizing applied force in a “Diagonal” motion in the planar space required simultaneous predictions from two models (Model X: applied force in X dimension, and Model Y: applied force in Y dimension) from the same RT observations. For compliant collaboration, substantially higher estimation accuracies were required in each dimension; otherwise, the participant might lose control of the gripper. With a lower calibration dataset (400 samples), recognition became relatively tricky during RT test phase because slight deviations in applied forces, arm motion speed, or posture affected the model performance due to uncertainties involved during interactions. Therefore, a model trained with too little calibration data might fail to learn possible motions. In scenario

Fig. 13. Comparison of performance of different models in estimating user-applied forces during different movement patterns. (a) Scenario1: unlearned motion M_U = X-direction. (b) Scenario1: unlearned motion M_U = Diagonal. (c) Scenario2: unlearned participant PU performing motion M_L = X. (d) Scenario2: unlearned participant PU performing motion M_L = DG.
The study in this article is focused on learning abilities of LCFMG-based models estimating applied forces in dynamic motions via FMG signals when a participant (P6–P15) interacted with a linear robot. Two main trends were considered for training the FMG-based regression models:

1) creating the proposed generalized long-term FMG-based calibrated regression models with minimum calibration data, LCFMGs;

2) administering extensive data collection sessions to record FMG data from each new participant/new motion performing several repetitions to create STMs, STMs.

It was worthy to acknowledge that the RT performance of an FMG-based model varied greatly due to its inherent subject-specific nature. During the RT evaluation period, continuous inbound test data were subject to individual variances such as sudden deviations in arm motion, posture, or changed applied forces. Also, it was noticed that physiological attributes (arm length, forearm and upper arm perimeter, sweats, skin hair, fatigue) affected each participant’s muscle contraction readings (better/poor observations in calibration and test data samples). Variations in the calibration data in the intended motion (“X,” “Diagonal”) along with these phenomena impacted model performance; hence, larger deviations were seen, as depicted in Fig. 13. Using a small calibration dataset, the model worked well for most unlearnt participants, while showed moderate performance for a few others. But when compared to STM-2, the inter-participant/inter-motion-LCFMG-2 model was superior.

In all scenarios, the inter-participant/inter-motion-2 model was more generalizable compared to the intra-participant/intra-motion-2 model. Aggregating calibration data with these diverse multiple sources provided a versatile and unique training dataset, applicable to the contributing participant only. Furthermore, calibration data in the “X direction” and “Diagonal” intended motions were quite different. It was the calibration data that determined which intended motion the model would recognize. In scenario 1 for the inter-participant-2 model, although the same reference dataset was used for all participants, training data became different because of the unique “learnt participant dataset” and calibration data. The “learnt participant dataset” provided enough seen data from a contributing participant. Similarly, during recognition of unseen participant motion, inter-motion-2 model used two repetitions of calibration data (800 samples) aggregated with a large, diversified reference dataset for training. Apart from the calibration data, this trained model had no data from a new, unseen participant. Although CV accuracies were ≈90% in “X” or “Diagonal” for both inter-participant-2 and inter-motion-2 models, the RT accuracies were 94%, 90% in “X” and 85%, 80% in “Diagonal” motions, respectively. These gaps between the offline model performances with RT evaluations (specifically in M3: “Diagonal”) were observed in another similar research project using sEMG biosignals [30]. It was interesting to note that for “X” motion, the generalized
trained model performed well because of a simpler computation in one dimension only. However, larger standard deviations were observed for “Diagonal” motion because more complex predictions were required simultaneously in X and Y dimensions (inter-participant-2: 85±5% and inter-motion-2: 80±7%) across all participants (P₀–P₁₅). Results indicated that adding “learnt participant dataset” in scenario 1 was worthy in recognizing unseen “Diagonal” motion, however, it was practically not easy to collect. On the contrary, inter-motion-2 model was more practical to implement.

A. Observations

This study revealed some useful observations that might benefit future FMG-based real-world applications.

1) Impact of Training Dataset: For generalization, the same multiple source distributions were used in training for recognition of the unseen test data, as described in Table II. For instance, in scenario 1 (recognizing an unseen motion M₁ or M₃), although the “learnt participant dataset” was distinct for each participant (P₀–P₈), the reference dataset was the same for all. Interestingly, in scenario 2, one baseline dataset (subset/full reference dataset) was used in recognizing unseen participants (P₀–P₁₅).

A model with the largest multiple source distributions added more generality, diversity, and contribution to supervised transfer knowledge. The training dataset for the LCFMG-2 model was approximately five times greater than the LCFMG-1 model. Thus, the model was more generalized, although it required almost twice the training time. Also, generalization errors were low and comparable with the STMs; thus, accomplishing generalized ZSL in regressing FMG signals was useful. It was observed that the STM-1 and STM-2 models trained with small training data (calibration data of 400, 800 samples only) could not provide better estimates of RT interactions. Whereas the LCFMG-based models estimated quite well due to the multiple source distributions contributing to the transfer of knowledge.

2) Impact of Calibration Data: It was observed that the generalized ZSL with few calibration data allowed the LCFMG-1 and LCFMG-2 models to perform adequately, but the LCFMG-0 model with no calibration data did not work. Due to the transient, nonstationary nature of FMG signals, muscle contractions during different activities were individual-specific, time-bound, and related to sensor placements. Therefore, any calibration data collected was valid during the period an individual continued wearing the FMG bands. Thus, removing the bands and redonning them, either by different users or by the same user, resulted in different FMG readings. Therefore, the need for collecting calibration data every time a participant wore the bands was essential and was revealed during testing with the LCFMG-0 model (average $R^2<0.45$). It was not surprising that models trained without any calibration data failed in RT testing, and hence, ZSL could not be implemented using FMG signals.

RT model performance was governed by the calibration data. Introducing few calibration data implemented generalized ZSL with multiple source domain adaptation technique where it tuned a model with the current state of FMG signals. These data simply transformed the trained model distinctively and applicable to the participant performing the action in a RT session. For effective applied force prediction in the test phase, motion pattern and interaction force were required to be similar to the calibration data. Therefore, the inclusion of calibration data during RT evaluation was vital for

1) involving few test samples in training data to recognize unseen motion or unseen participant using GZSL learning;
2) representing current states of muscle contraction and sensors positions;
3) selecting an intended simple motion;
4) reducing training time i.e., reducing fatigue;
5) allowing donning of (on/off) the band during longer time usage because new calibration data could be collected again.

With fewer calibration data, the model might lean toward multiple source distributions, predicting well for some participants while moderately for others. But evaluation of the proposed model across participants indicated that even with a bias toward multiple source domains, the proposed model predicted better when calibration data were used in training, thereby tuning the model toward the target domain.

3) Impact of Intended Motion: For an “X” intended motion, one trained model was required (biaxial stage moves along X dimension only), and in a “Diagonal” intended motion, two trained models performed simultaneous predictions in X and Y dimensions. For compliant collaboration, synchronous RT conversions between applied force in motion and displacements in the linear stages were required. Due to fundamental differences in arm posture, motion pattern, or applied forces in the two intended motions, a model with the same baseline dataset (inter-motion-2/inter-participant-2) performed differently because the calibration data were different. Therefore, reported accuracies in RT predictions [as shown in Figs. 9, 10, and 13] were distinct and influenced by the selection of the intended motion.

During interactions between participants and the linear robot in the “X” intended motion, applied forces: 20–30 N, mean displacement: 400 mm and average speed/motion: 70–80 mm/s were observed. While in “Diagonal” motion, applied forces: 20–60 N, mean displacement: 475 mm, and average speed/motion: around 60–90 mm/s were observed across participants. The intended motion used in the calibration phase allowed the model to predict applied force in that motion. So, applied force and motion speed in the calibration phase was expected to be maintained by a participant in the RT test phase.

4) Few Exceptions: Although LCFMG-0 failed in estimating user-applied forces, surprisingly it worked exceptionally well in one or two instances. In recognizing unseen X motion, the inter-participant-LCFMG-0 model obtained higher estimation accuracies ($R^2 = 70\%$) for a learned participant (P₇: P₈), while the intra-motion-LCFMG-0 model estimated well ($R^2 = 77\%$) for recognizing an unseen participant (P₁₀: P₁₃) applying force in a learned motion M₁. These two incidents indicated that a long-term FMG model might estimate user-applied forces during a simple straightforward motion even without “calibration data.” This would be achievable only when the population dataset
includes all possible force ranges in dynamic motions from more participants and requires further investigation.

B. Potential Applications

The study revealed the fact that an LCFMG-based model would be a better choice in practical applications of daily life, prosthetic control, HMI, or HRI. This framework would allow a new first-time user or repeated user to interact in either seen or unseen motions. Also, it is preferable for ease of operation with lower calibration data and faster training. During RT evaluation, calibration data were only collected from a contributing participant (P6–P15). Hence, in practical HRI scenarios, the framework will allow faster and fewer calibration data from a participant and avoid muscle fatigue. Thus, this framework will reduce the physiological effects of wearing the FMG band for a long time by retraining quickly when the band is taken off and put back again.

In real-world HRI applications, recognizing an unseen motion would be beneficial. Inter-participant-LCFMG-2 seemed a good choice in recognizing force in a new/unlearnt motion. Although, this required a “learnt participant dataset” from a participant. This might be possible only by repeated use where a human worker would interact with a robot on a regular basis and contribute source data (including calibration data from daily interaction).

On the contrary, scenario 2 used the approach where a new, unseen worker interacted with a robot without spending a longer training time in collecting large training data. The inter-motion-2 would be particularly useful in FMG-based applications for a first-time user. Therefore, the inter-motion-LCFMG-2 model is recommended for the real-world HRI environment. The model will allow any new worker to interact with a robot in an intended motion such as the “X direction” or “Diagonal.” This versatile model would work for other simple motions too. Interestingly, the model did not distinguish a repeated user from a new user, meaning that any worker (unseen or seen) would be treated as an unseen participant. This was realistic because even for a repeated user, new calibration data were required for domain adaptation between the source and target data. By aggregating calibration data from previous, repeated use, the inter-motion-LCFMG-2 can become personalized (converting gradually toward inter-participant-LCFMG-2) for an individual where human interactions with machines are desirable on a regular basis.

C. Limitations and Future Works

Although we have investigated multiple source domain adaptations to overcome similar challenges imposed on practical usage of the sEMG technology, it was not included in this study. Collecting sEMG signal and FMG signal simultaneously from the same forearm and upper arm positions during RT interactions could become difficult. This could complicate the control process and would require more time in evaluating both methods. Investigating different LCFMG models and a few STM models in RT required around 1.5 h for each participant. Due to this time-consuming nature of the study, only two motions were investigated to evaluate the proposed models. Also, to investigate unseen motion, the requirement of collecting “learnt participant dataset” limited the number of participants (P6–P15) in evaluating the framework. Also, in a few instances, it was observed that the force and motion were governed by the large baseline dataset, slightly limiting the behavior of the RT interaction.

In this study, all participants interacted with a linear robot, where labeled multiple source domains were collected. The proposed model was evaluated with the same linear robot. It would be interesting to study whether these multiple source domains can work with a different industrial robot using transfer learning. Also, one ML algorithm, i.e., SVR was implemented for the RT evaluation. In the future, deep transfer learning would be studied for better performance where a model can learn latent features from the source distributions. Repeated interaction of a participant in an intended motion could gradually contribute input source data (including calibration data from daily interaction). Further study can be done to implement incremental learning by augmenting new input data and increase a model’s knowledge.

VIII. Conclusion

In this study, estimating applied hand force for OOD sample data was investigated using multiple source domain adaptation. Using GZSL concept, models were trained with a long-term FMG baseline dataset (multiple source domains) and calibration data for practical reasons (few sample data during RT interaction). It was observed that the LCFMG-based models suitably estimated applied hand force during a new/unlearned motion, or with a new/unlearned participant manipulating the linear robot. With the primary interest to understand the usability of FMG-based model in a real-world application, this novel method investigated FMG-based pHRI for the first time in new, unseen scenarios and was proven effective. This study demonstrated that FMG-based ML models can be generalized to interpret new, unlearned data (a new user or a new executed motion) if some calibration data are added to the long-term training dataset. This can provide a viable solution for FMG-based pHRI, rehabilitation, and prosthetic control to interact with machines on daily basis.

ACKNOWLEDGMENT

The Principal Investigator, Carlo Menon, and members of his research team have a vested interest in commercializing the technology tested in this study, if it is proven to be successful and may benefit financially from its potential commercialization. The data are readily available upon request.

REFERENCES


