Current Trends and Confounding Factors in Myoelectric Control: Limb Position and Contraction Intensity

Evan Campbell 1,2,*, Angkoon Phinyomark 2 and Erik Scheme 1,2,*

1 Department of Electrical and Computer Engineering, University of New Brunswick, Canada; Evan.Campbell1@unb.ca
2 Institute of Biomedical Engineering, University of New Brunswick, Canada; aphinyom@unb.ca
* Correspondence: escheme@unb.ca; Tel.: +1-506-453-4966

Received: 5 February 2020; Accepted: 9 March 2020; Published: 13 March 2020

Abstract: This manuscript presents a hybrid study of a comprehensive review and a systematic (research) analysis. Myoelectric control is the cornerstone of many assistive technologies used in clinical practice, such as prosthetics and orthoses, and human-computer interaction, such as virtual reality control. Although the classification accuracy of such devices exceeds 90% in a controlled laboratory setting, myoelectric devices still face challenges in robustness to variability of daily living conditions. The intrinsic physiological mechanisms limiting practical implementations of myoelectric devices were explored: the limb position effect and the contraction intensity effect. The degradation of electromyography (EMG) pattern recognition in the presence of these factors was demonstrated on six datasets, where classification performance was 13% and 20% lower than the controlled setting for the limb position and contraction intensity effect, respectively. The experimental designs of limb position and contraction intensity literature were surveyed. Current state-of-the-art training strategies and robust algorithms for both effects were compiled and presented. Recommendations for future limb position effect studies include: the collection protocol providing exemplars of at least 6 positions (four limb positions and three forearm orientations), three-dimensional space experimental designs, transfer learning approaches, and multi-modal sensor configurations. Recommendations for future contraction intensity effect studies include: the collection of dynamic contractions, nonlinear complexity features, and proportional control.

Keywords: electromyography; EMG; feature extraction; feature selection; myoelectric control; classification; pattern recognition; prosthetics; wearables; amputee

1. Introduction

Myoelectric control refers to the decoding of motor intent from electrophysiological properties of muscles for use as a control input for some external interface. Despite being a complex and non-stationary signal, surface electromyography (EMG) is leveraged in myoelectric control as a rich source of information. With adequate spatial resolution and a proper EMG pattern recognition pipeline, motions can be deciphered with remarkably high accuracy (>90% accuracy). Myoelectric control has been used in a variety of human-computer interfaces such as upper-limb prostheses or orthoses [1,2], electric wheelchairs [3], muscle-derived speech decoding devices [4,5], virtual reality control devices [6] and other clinical and consumer device designs [7]. While myoelectric control has been touted for decades as an intuitive means of control for assistive-devices, performance of these devices in daily living conditions has been notably inferior to benchmarks achieved in controlled laboratory environments.
The current challenges commonly associated with this lack of reliability in practical conditions can be roughly categorized into four confounding factors.

(a) **Limb position factor.** The muscular activity that maintains limb positions against gravitational force is dependent on the position of the limb. To an even greater degree, during limb motions even larger amounts of supplemental muscle activity is necessary. Additionally, while in different positions the underlying topography of the muscle fibers may shift relative to the electrodes changing the EMG signal measured substantially. These position-dependent muscular activations manifest as artefacts within the EMG signal which lead to confusion between expected muscle activation patterns and observed patterns.

(b) **Contraction intensity factor.** The contraction intensity of motions is subconsciously regulated according to the effort expected to act on the target load. The EMG signal amplitude is directly influenced by this intensity variability, as demonstrated by linear and nonlinear relationships in different works. Additionally, frequency characteristics have been shown to vary according to the contraction intensity elicited. These natural variations in intensity produce largely different muscle activation patterns for the same motion for different loads that may lead to differences between anticipated and observed patterns within a given motion.

(c) **Electrode shift factor.** When the position of an electrode shifts, as is common in socket-based prosthetic devices or circumference-style electrode placement, the underlying musculature changes relative to those electrodes. Subsequently, even in the event the same fibers are underneath the electrode after a shift occurs, a tissue filter effect results in the measured signal being different between electrode locations. Additionally, the change in contact impedance between the electrodes and the skin after a shift occurs introduces further changes in signal properties. As a consequence of these changes, the boundaries defined by a classifier are ill-suited to the properties of the new electrode location.

(d) **Within/between day factor.** The EMG signal is sensitive to many time-varying physiological, biochemical, or anatomical mechanisms [8], for instance blood flow [9]. An additional component of within/between day factor can encompass the change in electrode position when donning and doffing between sessions of myoelectric device use, or a change in perceived contraction intensities between separate uses of the device. These changes greatly influence the applicability of a trained classifier’s decision boundary. This factor has been similarly referred to as the effect of time and is largely responsible for the usability degradation found after prolonged use of myoelectric devices.

Past reviews have provided excellent motivation to investigate and alleviate confounding factors [8,10–13]; however, a comprehensive overview of all state-of-the-art solutions for confounding factors has not yet been presented. The necessary depth in surveying each confounding factor was accommodated within this survey by being split into two papers. This first paper aims to provide a comprehensive review of the two intrinsic confounding factors: limb position and contraction intensity. The limb position factor introduces a non-conservative domain adaptation between positions that limits operation of myoelectric devices in an unconstrained environment. The contraction intensity factor likewise introduces practical limitations through amplitude variation and frequency shift that standard single intensity level models neglect. The second paper aims to provide a mirrored survey of the current state-of-the-art solutions for the two confounding factors unaddressed in this manuscript: electrode shift and within/between day. The effect on classification performance, experimental design considerations, training procedures, and robust methods of the limb position and contraction intensity factors is the primary focus of this manuscript.

Finally, as an overview, the objective of this article is to illustrate the profound performance degradation caused by signal variability during real world use of myoelectric devices. Within the second section, the traditional topography of EMG pattern recognition architectures is presented. The third and fourth sections provide comprehensive analysis on the performance...
degradation and current solutions for the limb position and contraction intensity factors. The fifth section provides a summary of critical findings within the analyses and surveyed literature, current limitations of the state-of-the-art, and directions motivating future work.

2. EMG Pattern Recognition

The various stages of an EMG pattern recognition system are consistent across the literature, and are comprised of: data acquisition, pre-processing and segmentation, feature extraction, dimensionality reduction, and classification [10,14], as shown in Figure 1. While each of these stages influences the performance of the control system, the quality of the data itself is paramount in order for true motor intent to be decoded from the EMG signal.

2.1. Data Acquisition

EMG data are typically acquired through a guided protocol that prompts the end-user to elicit a set of known contractions that will later be recognized for control. During those prompts, patterns of muscle signals are typically recorded from multiple bipolar electrodes placed on the surface of the skin. The SENIAM (Surface ElectroMyoGraphy for the Non-Invasive Assessment of Muscles) guidelines [15] suggest that optimal electrode size and separation are 10 mm and 20 mm, respectively, with electrode pairs oriented parallel to the muscle fiber [16]. In the case of prosthetic devices, electrodes are embedded into the socket of the prosthetic limb. The number and placement of the electrodes are dependent on the classes of motion to be recognized, however, most studies have generally employed either muscle specific placement (determined through palpation) or general placement (circumferential around the forearm [17], for example, or in a grid arrangement). The use of other types of electrode configurations, such as monopolar and Laplacian, have been less successfully employed in myoelectric devices to date. Although data acquisition conventionally occurs only once during the training of pattern recognition-based myoelectric systems, adaptive classifiers which incorporate data over time have also been proposed [18].

2.1.1. Datasets

Table 1 summarizes the characteristics of these datasets. Although all data were recorded from intact-limbed subjects, the trends in performance should generalize to clinical populations with some consideration and caution [19,20]. When validating model architectures with amputee populations that have been developed with intact-limbed subject data, a bias is expected (i.e., a relatively constant decrease in accuracy compared to intact-limbed subjects), however, improvements to state-of-the-art models are expected to translate to amputee populations as long as migratory features are not required. For all datasets, the representation of all motion classes was balanced. Preparation of these data sets consisted of removal of power-line interference with
notch filters at 50 Hz or 60 Hz, motion-artefact removal with band-pass filters at 20 Hz to 450 Hz, down-sampling of the time-series to 1000 Hz (when relevant), and window segmentation using a window size of 150 ms and increment of 50 ms.

### Table 1. Datasets adopted in this work. N, C, M, R, Fs, and A/D indicate the number of subjects, channels, motions, repetitions, sampling frequency, and resolution of the analog-to-digital converter in bits, respectively. MVC refers to maximal voluntary contraction. Position abbreviations are outlined in Figure 4.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Dataset Description</th>
<th>N</th>
<th>C</th>
<th>M</th>
<th>R</th>
<th>Fs</th>
<th>A/D</th>
<th>Confounding Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>[21]</td>
<td>D1: 5 Limb Position</td>
<td>17</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>1000</td>
<td>16</td>
<td>P2, P3, P4, P9, P14</td>
</tr>
<tr>
<td>[22]</td>
<td>D2: 16 Limb Position</td>
<td>10</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>1000</td>
<td>n/a</td>
<td>16 Saggital and Humeral</td>
</tr>
<tr>
<td>[23]</td>
<td>D3: 3 Forearm Orientation</td>
<td>10</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>4000</td>
<td>12</td>
<td>P2s, P2, P2p</td>
</tr>
<tr>
<td>[24]</td>
<td>D4: 7% MVC Level</td>
<td>10</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>1000</td>
<td>16</td>
<td>20:10:80% MVC</td>
</tr>
<tr>
<td>[25]</td>
<td>D5: 3 Subjective Level (A)</td>
<td>10</td>
<td>8</td>
<td>7</td>
<td>4</td>
<td>1000</td>
<td>16</td>
<td>Low, Medium, High</td>
</tr>
<tr>
<td>[26]</td>
<td>D6: 3 Subjective Level (B)</td>
<td>10</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>4000</td>
<td>12</td>
<td>Low, Medium, High</td>
</tr>
</tbody>
</table>

#### 2.2. Data Pre-Processing

Pre-processing is used to improve the robustness of the data with respect to potential contaminants [8,25] by increasing the signal-to-noise ratio (SNR) to improve the distinguishing characteristics of the EMG signal. Prior to amplification, the EMG signal is typically between ±5 mV [8]; however, contaminants may corrupt the signal that are several times larger. Examples of contaminants include electrode shift, where the displacement of electrodes results in the surveyed muscle fibers being changed; power-line interference, where the fundamental frequency and harmonics of grid power are introduced into the recorded signal; motion artefact, where electrode-skin interface impedance changes or cable motion introduces low-frequency noise; and electronic noise, where noise is introduced by mismatched electrical components, heating of temperature sensitive components (Johnson–Nyquist noise), or inherent noise in semiconductors (Schottky noise). Among common contaminants, power-line interference is often removed using notch filters at key harmonic frequencies; motion artefacts are removed using a high-pass Butterworth filter with a typical cutoff frequency of 10 or 20 Hz [26]; ECG interference is sometimes removed using high-pass filters with cutoff of 100 Hz [27,28]; and electrical noise may be minimized by using high-performance components and/or a low-pass filter with cutoff-frequency of 450 or 500 Hz (the active energy of the EMG signal during contraction is considered negligible above 500 Hz). Other advanced signal processing techniques employed for EMG denoising include wavelet transform [29–31], wavelet packet transform [32,33], empirical mode decomposition [34–36], one-dimensional (1D) local binary pattern [37–39], and adaptive Wiener filtering [40].

#### 2.3. Data Segmentation

Because EMG is a stochastic process, the signal varies naturally over time. This variability results in the signal exhibiting non-stationarity, violating a common assumption among feature extraction methods. Consequently, data segmentation is commonly used to organize the EMG signal into frames of information from which properties can be considered to be weakly or wide-sense stationary. Two main techniques are commonly applied for windowing: adjacent or overlapping window segmentation [41]. Adjacent window segmentation extracts contiguous windows along the time series by incrementing an index by an amount equal to the size of the window. More commonly, overlapping window segmentation is used to increase the density of the decision stream by incrementing neighboring windows by a duration shorter than the window length, resulting in neighboring windows sharing common elements. When smaller increments are used, post-processing techniques can more readily be leveraged to improve the control outputs [42–47].
Regardless of the approach, the windowing operation extracts constant duration frames of EMG data that are necessary for feature extraction. Windows are typically enframed into short segments (100–300 ms) [48]. An upper-limit on window length is enforced in real-time myoelectric control, where the time between successive windows and computation time (update rate) cannot exceed 300 ms to avoid perceived delay [14,49].

2.4. Feature Extraction

Feature extraction is used to increase the information density in the signal by using representative properties of the segmented windows for classification, rather than the raw samples themselves. Extracting high quality features that possess good class separability, minimal complexity, and are robust to confounding factors is the most important contributor to myoelectric control system performance [50,51]. Features may be categorized according to two criteria: (1) the domain the property is extracted from, or (2) the modality of information the property captures.

Feature extraction is typically performed in three domain representations of the EMG signal: the time domain (TD), the frequency domain (FD), and time-frequency representation (TFR). TD feature extraction refers to procedures that extract properties from the EMG time series in its original form. These are commonly employed features in myoelectric pattern recognition systems due to their high accuracy in low-noise environments and low computational complexity, and provide intuitive information about muscle motor unit recruitment [52–57]. The mean absolute value (MAV) feature or root mean square (RMS) feature, for example, is a TD feature that represents the average energy of the EMG signal within a window [58,59]. FD features are extracted from the Fourier transform of the EMG signal to capture information about motor unit recruitment rates and muscle fatigue. The median frequency (MDF) feature, for example, is an FD feature that represents the frequency that divides the total power of the signal into equal upper and lower halves [60,61]. TFR refers to any transformation applied to the EMG signal that incorporates both time and frequency information [33]. The most commonly employed TFR is the wavelet transformation, where the cross-correlation between the EMG signal and a wavelet function is taken across multiple scale/shift configurations. Further dimensionality reduction is commonly required to interpret TFR into meaningful properties, however, this complexity is often justified by demonstrated performance gains [62]. Table 2 contains examples of commonly used features in EMG pattern recognition, with particular emphasis on robustness to confounding factors (see Sections 3 and 4).
Table 2. Examples of feature extraction methods employed in myoelectric control.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Feature Extraction Method</th>
<th>Name</th>
<th>Ref</th>
<th>Feature Extraction Method</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>[63]</td>
<td>Amplitude of the First Burst</td>
<td>AFB</td>
<td>[64,65]</td>
<td>Sample Entropy</td>
<td>SampEn</td>
</tr>
<tr>
<td>[63,66,69]</td>
<td>Difference Log Detector</td>
<td>DLD</td>
<td>[63,64]</td>
<td>Myopulse Percentage Rate</td>
<td>MYOP</td>
</tr>
<tr>
<td>[63,66]</td>
<td>Difference Temporal Moment</td>
<td>DTM</td>
<td>[64,70]</td>
<td>Bow-Counting Fractal Dimension</td>
<td>BC</td>
</tr>
<tr>
<td>[129]</td>
<td>L-Score</td>
<td>LS</td>
<td>[74]</td>
<td>Integral Square Descriptor</td>
<td>ISD</td>
</tr>
<tr>
<td>[74]</td>
<td>Coefficient of Regularization</td>
<td>COR</td>
<td>[74]</td>
<td>Modified Absolute Square Sum</td>
<td>MDIFF</td>
</tr>
<tr>
<td>[74]</td>
<td>Mean Difference Derivative</td>
<td>MDIFF</td>
<td>[75]</td>
<td>Activity</td>
<td>ACT</td>
</tr>
<tr>
<td>[75]</td>
<td>Mobility</td>
<td>MOB</td>
<td>[75]</td>
<td>Complexity</td>
<td>COMP</td>
</tr>
<tr>
<td>[63,67,69]</td>
<td>Integral of Electromyogram</td>
<td>IEMG</td>
<td>[64,72]</td>
<td>Maximum Fractal Length</td>
<td>MFL</td>
</tr>
<tr>
<td>[77]</td>
<td>Mean Absolute Value First Difference</td>
<td>MAVFS</td>
<td>[129]</td>
<td>Mean Squared Ratio</td>
<td>MSR</td>
</tr>
<tr>
<td>[63,78]</td>
<td>Modified Mean Absolute Value 2</td>
<td>MMAV2</td>
<td>[66,67]</td>
<td>Difference Cepstral Coefficients</td>
<td>DCC</td>
</tr>
<tr>
<td>[69,79,80]</td>
<td>Mean Absolute Value</td>
<td>MAV</td>
<td>[71,81]</td>
<td>Detrend Fluctuation Analysis</td>
<td>DFA</td>
</tr>
<tr>
<td>[64,82]</td>
<td>Maximum</td>
<td>MAX</td>
<td>[63,83]</td>
<td>Power Spectrum Ratio</td>
<td>PSR</td>
</tr>
<tr>
<td>[63,84]</td>
<td>Multiple Hamming Windows</td>
<td>MHW</td>
<td>[65,86]</td>
<td>Signal to Noise Ratio</td>
<td>SNR</td>
</tr>
<tr>
<td>[63,84]</td>
<td>Mean Power</td>
<td>MNP</td>
<td>[87,88]</td>
<td>Critical Exponent</td>
<td>CE</td>
</tr>
<tr>
<td>[63,79,80]</td>
<td>Root Mean Squared</td>
<td>RMS</td>
<td>[63,69]</td>
<td>Histogram</td>
<td>HIST</td>
</tr>
<tr>
<td>[63,84]</td>
<td>Spectral Moment</td>
<td>SM</td>
<td>[89,90]</td>
<td>Kurtosis</td>
<td>KURT</td>
</tr>
<tr>
<td>[63,84]</td>
<td>Sum of Squared Integral</td>
<td>SSI</td>
<td>[63]</td>
<td>Mean Absolute Value Slope</td>
<td>MAS</td>
</tr>
<tr>
<td>[63,64,84]</td>
<td>Total Power</td>
<td>TTP</td>
<td>[63,64]</td>
<td>Peak Frequency</td>
<td>PKF</td>
</tr>
<tr>
<td>[63,66,69]</td>
<td>v-Order</td>
<td>V</td>
<td>[89,90]</td>
<td>Skewness</td>
<td>SKEW</td>
</tr>
<tr>
<td>[63]</td>
<td>Waveform Length</td>
<td>WL</td>
<td>[85,86]</td>
<td>Signal to Motion Artefact Ratio</td>
<td>SMR</td>
</tr>
<tr>
<td>[93,94]</td>
<td>Median Frequency</td>
<td>MDF</td>
<td>[63]</td>
<td>Variance of Central Frequency</td>
<td>VCF</td>
</tr>
<tr>
<td>[93,94]</td>
<td>Mean Frequency</td>
<td>MNF</td>
<td>[64,95]</td>
<td>Variance Fractal Dimension</td>
<td>VFD</td>
</tr>
<tr>
<td>[63]</td>
<td>Slope Sign Change</td>
<td>SSC</td>
<td>[96]</td>
<td>Fused Time Domain Descriptors</td>
<td>FTDD</td>
</tr>
<tr>
<td>[99]</td>
<td>Graph Laplacian</td>
<td>GL</td>
<td>[100]</td>
<td>Relative Wavelet Packet Energy</td>
<td>RWPE</td>
</tr>
<tr>
<td>[74]</td>
<td>Mean Logarithmic Kernel</td>
<td>MLK</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Alternatively, features may be defined according to the properties targeted by their mathematical definitions [63]. The specific modalities of information of these theoretical properties include amplitude, variability, stationarity, entropy, linearity, similarity, and frequency. Regardless of designed motivation, characteristics of the EMG signal may result in some features being unable to capture these theoretical characteristics. Such is the case with variance (VAR) which was originally formulated to capture the variability of the signal; however, the zero-mean characteristic of the EMG signal results in VAR being virtually synonymous with other amplitude features. Instead, recent studies have characterized features according to a smaller set of domain-specific categories that have been defined through empirical observation. The specific modalities of information these empirical characteristics include are signal amplitude and power, nonlinear complexity and frequency information, time-series modeling, and unique [77].

The selection of appropriate features has a tremendous impact on the performance of any pattern recognition system, and the ideal feature set is heavily dependent on the classification task. Furthermore, the collection of features, and specifically the complementary information that they provide, has led the field to define and adopt known feature sets. For example, the Hudgins’s TD feature set is a widely adopted feature set used for myoelectric control studies with both able-bodied and amputee users. This feature set is comprised of four simple TD features: MAV, ZC, SSC, and WL [58]. Another example, the topologically selected TD feature set (TSTD) is comprised of twelve TD features: MAVFD, DASDV, WAMP, ZC, MFL, SampEn, and TDPSD (6 features) [77]. For a complete depiction of the feature sets contained within this survey, see Table 3.
Table 3. Datasets adopted or referenced in this work to investigate confounding factors in myoelectric control. AR4, and AR9 refer to autoregressive coefficients of the fourth and ninth order, respectively.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Abbrev.</th>
<th>Feature Set</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>[58]</td>
<td>TD</td>
<td>Hudgin’s Time Domain</td>
<td>MAV, ZC, SSC, WL</td>
</tr>
<tr>
<td>[101]</td>
<td>TDAR</td>
<td>Time Domain Autoregressive</td>
<td>MAV, ZC, SSC, WL, AR4</td>
</tr>
<tr>
<td>[77]</td>
<td>TSTD</td>
<td>Topologically Selected Time Domain</td>
<td>MAVFD, DASDV, WAMP, ZC, MFL, SampEn, TDPSD</td>
</tr>
<tr>
<td>[129]</td>
<td>LSF4</td>
<td>Low Sampling Frequency 4</td>
<td>LS, MFL, MSR, WAMP</td>
</tr>
<tr>
<td>[129]</td>
<td>LSF9</td>
<td>Low Sampling Frequency 9</td>
<td>LS, MFL, MSR, WAMP, ZC, RMS, IEMG, DASDV, VAR</td>
</tr>
<tr>
<td>[102]</td>
<td>TDPSD</td>
<td>Time Domain Power Spectral Descriptors</td>
<td>TDPSD</td>
</tr>
<tr>
<td>[103]</td>
<td>COMB</td>
<td>Combined</td>
<td>WL, SSC, LD, AR9</td>
</tr>
<tr>
<td>[104]</td>
<td>DFTR</td>
<td>Discrete Fourier Transform Representation</td>
<td>DFTR</td>
</tr>
<tr>
<td>[74]</td>
<td>invTDF</td>
<td>Inverse Time Domain Features</td>
<td>ISD, COR, MASS, MDIFF, MLK</td>
</tr>
<tr>
<td>[75]</td>
<td>Hjorth</td>
<td>Hjorth Parameters</td>
<td>ACT, MOB, COMP</td>
</tr>
</tbody>
</table>

2.5. Dimensionality Reduction

The creation of an appropriate feature set for myoelectric control requires three key considerations: (1) inclusion of features that have high-class discriminatory information, (2) exclusion of features that are heavily correlated with one another, and (3) minimization of the number of features included so as to combat the overfitting influence of the “curse of dimensionality”. Together, these aspects are addressed using dimensionality reduction techniques to improve the robustness of the classification algorithm by removing non-essential features. Although not strictly required for myoelectric control, dimensionality reduction can greatly improve classification accuracy and reduce computational cost.

Dimensionality reduction is primarily implemented using one of two methods: feature selection, which throws out redundant or less discriminative features, or feature projection, which transforms and reduces the dimensionality of the original feature space. Examples of feature selection techniques include sequential forward selection (SFS) and maximum-relevance minimum-redundancy (MRMR) [105]. Examples of feature projection include the unsupervised principle components analysis (PCA) and the supervised uncorrelated linear discriminant analysis (ULDA) [106,107]. Table 4 summarizes various dimensionality reduction techniques, organized by reduction method and whether class labels are required or not (i.e., supervised or unsupervised).

Table 4. Organization of dimensionality reduction techniques by reduction type, supervision. SR, NMF, CCA, t-SNE, and MDS refer to spectral regression, non-negative matrix factorization, canonical correlation analysis, t-stochastic neighborhood embedding, and multi-dimensional scaling.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Reduction</th>
<th>Supervision</th>
<th>Linearity</th>
<th>Algorithm Name</th>
<th>Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>[108]</td>
<td>Selection</td>
<td>Supervised</td>
<td>–</td>
<td>SFS (Wrapper)</td>
<td>Maximum Classification Accuracy</td>
</tr>
<tr>
<td>[108]</td>
<td>Selection</td>
<td>Either</td>
<td>–</td>
<td>SFS (Filter)</td>
<td>Optimal Statistical Measure</td>
</tr>
<tr>
<td>[105]</td>
<td>Selection</td>
<td>Either</td>
<td>–</td>
<td>MRMR</td>
<td>Class Separability - Feature Correlation</td>
</tr>
<tr>
<td>[107]</td>
<td>Projection</td>
<td>Supervised</td>
<td>Linear</td>
<td>ULDA</td>
<td>Maximum Separability</td>
</tr>
<tr>
<td>[109]</td>
<td>Projection</td>
<td>Linear</td>
<td>SR</td>
<td>Maximum Separability</td>
<td></td>
</tr>
<tr>
<td>[110]</td>
<td>Projection</td>
<td>Linear</td>
<td>NMF</td>
<td>Non-negative Factors</td>
<td></td>
</tr>
<tr>
<td>[111]</td>
<td>Projection</td>
<td>Linear</td>
<td>CCA</td>
<td>Maximize Domain Correlation</td>
<td></td>
</tr>
<tr>
<td>[112]</td>
<td>Projection</td>
<td>Linear</td>
<td>PCA</td>
<td>Preserve Variability</td>
<td></td>
</tr>
<tr>
<td>[113]</td>
<td>Unsupervised</td>
<td>Nonlinear</td>
<td>Isomap</td>
<td>Preserve Neighborhood Relationships</td>
<td></td>
</tr>
<tr>
<td>[114]</td>
<td>Unsupervised</td>
<td>Nonlinear</td>
<td>t-SNE</td>
<td>Preserve Neighborhood Relationships</td>
<td></td>
</tr>
<tr>
<td>[115]</td>
<td>Unsupervised</td>
<td>Nonlinear</td>
<td>Autoencoder</td>
<td>Reduce Reconstruction Loss</td>
<td></td>
</tr>
<tr>
<td>[116]</td>
<td>Unsupervised</td>
<td>Nonlinear</td>
<td>MDS</td>
<td>Preserve Pairwise Distance</td>
<td></td>
</tr>
</tbody>
</table>
2.6. Classification

Classification is the process of inferring the class of an unknown observation using a predictive model trained with earlier observations. Classification algorithms can be divided into two categories: generative classifiers, and discriminative classifiers. Generative classifiers assume that the training data obey some general distribution, and thus can be represented using the parameters of the distribution. This assumption results in decision boundaries that are more robust to minor perturbations, have lower computational complexity, and require less data to reach an optimal performance level. Examples of generative classifiers used in myoelectric control include the naïve Bayes classifier (NB), linear discriminant analysis classifier (LDA), quadratic discriminant analysis classifier (QDA), and Gaussian mixture model (GMM). Discriminative classifiers make no such assumption, and instead directly use training observations to model their discriminant function. These classifiers can have higher specificity than generative classifiers, but typically incur higher computational complexity, require more data to form appropriate discriminant functions, and are therefore subject to over-fitting. Examples of discriminative classifiers used in myoelectric control include \( k \)-nearest neighbor (kNN), random forest (RF), support vector machine (SVM), and sparse representation classification (SRC). Deep learning classifiers are a subset of discriminative classifiers that automatically encode relevant distribution characteristics directly from the raw training data into the bias and weight terms of a neural network architecture by optimizing a loss function. Deep learning classifiers have the highest degree of specificity of all classifiers but require vast amounts of data to reach optimal performance and ensure good generalizability. Examples used in myoelectric control include the recurrent neural network (RNN), long-short term memory (LSTM), probabilistic neural network (PNN), convolutional neural network (CNN), adaptive domain adversarial neural network (ADANN) [117], recurrent convolutional neural network (RCNN) [118].

2.7. Performance Evaluation

The evaluation of myoelectric pattern recognition systems may be performed at different stages of the architecture, namely: (1) at the feature level, (2) after the classification stage, and (3) during feed-forward use, in an online scenario. First, the evaluation of the feature extraction stage quantifies the available features based on the feature space representation of the training data. Considerations for high-quality features include (a) high separability, where clusters of features from each class occupy different regions of feature space; (b) high robustness, where perturbations to the input data result in negligible difference in the feature distributions; and (c) low complexity, where computational overhead of the extraction process does not exceed hardware or real-time requirements [119]. The Davies-Bouldin index (DBI) is often used to quantify these aspects of feature space prior to the actual classification problem. For each cluster, the multi-dimensional standard deviation, \( S_i \), is computed to describe the magnitude of within-cluster similarity. Additionally, the Euclidean distance between each pair of clusters, \( D_{ij} \), is computed to describe the magnitude of between-cluster similarity (Equation (2)). Within- and between-cluster similarity measures are combined to form a singular measure of overlap, \( R_{ij} \). Finally, the highest magnitude of overlap for each cluster is averaged to quantify the worst-case separability of neighboring class clusters in feature space. Equations (1)–(4) demonstrate how to compute the DBI value. The parameters \( x_j, \mu_i, N_i \), and \( K \) are the \( i \)th feature vector, the mean feature vector of the \( i \)th cluster, the number of feature vectors in the \( i \)th cluster, and the number of \( N \) choose 2 cluster combinations, respectively.

\[
S_i = \sqrt{\frac{1}{N_i} \sum_{j=1}^{N_i} (x_j - \mu_i)^T (x_j - \mu_i)}, \quad x_j \in C_i \tag{1}
\]

\[
D_{ij} = \sqrt{(\mu_i - \mu_j)^T (\mu_i - \mu_j)} \tag{2}
\]
Classifier evaluation quantifies the performance of a myoelectric system according to three measures: (a) the total error rate (TER), (b) active error rate (AER), and (c) robustness. TER denotes the total prevalence of all misclassifications in a test set. In contrast to TER, which is a general descriptive measure, AER quantifies the prevalence of misclassifications that result in incorrect prediction of an active motion class (anything but no motion class) of the control system. This distinction, motivated by the higher cost incurred by moving erroneously than by not moving at all, leads to AER ignoring false negative predictions (incorrect predictions of no motion class). Equations (5) and (6) show the computation of TER and AER, respectively. Parameters $n$, $N$, $p_n$, and $l_n$ are the index of the test feature vector, the total number of test feature vectors, the class prediction of the classifier for feature vector $n$, and the true class label of feature vector $n$, respectively.

\[
TER = 100\% \times \left(1 - \frac{\sum_{n=1}^{N} p_n = = l_n}{N}\right)
\]

\[
AER = 100\% \times \frac{\sum_{n=1}^{N} p_n \neq l_n \cap p_n \neq NM}{\sum_{n=1}^{N} p_n \neq 1}
\]

The robustness of a classifier is evaluated using strategic cross-validation of the dataset to highlight sensitivity to factors that may degrade motion classification. Cross-validation frameworks can be categorized into one of four forms: $x$ vs. $x$, $x$ vs. $y$, $x$ vs. all, and $N$ vs. all. The $x$ vs. $x$ validation framework provides the upper-limit of performance that can be expected when the training conditions identically match the testing conditions, and reflects the ability of the system to recognize testing similar to the training data. The $x$ vs. $y$ validation framework provides the lower-limit of performance expected when the training and testing conditions are completely disjoint. The difference between the performance of $x$ vs. $x$ and $x$ vs. $y$ validations indicate the sensitivity of the myoelectric device to the factor being assessed. The $x$ vs. all validation framework provides a realistic measure of performance expected when the device is trained in a single condition and tested in multiple other conditions. The $N$ vs. all validation framework illustrates the trend in performance when increasing the variability in the training set to better match the variability of the testing set. All possible configurations of training and testing conditions are explored for $x$ vs. $x$, $x$ vs. $y$, $x$ vs. all, and $N$ vs. all forms of validation. A demonstration of $x$ vs. $x$, $x$ vs. $y$, $x$ vs. all, and $N$ vs. all forms of validation is included using in Figure 2, using the limb position effect as an example factor.
Online validation is considered to be the most direct measure of the usability of a myoelectric pattern recognition system (other than clinical trials of specific use cases, such as prostheses). In myoelectric control, usability is most often assessed using a Fitts’ law test, which quantifies the information conveyed through a human-computer interface [120]. During this test, trained motions are mapped to cursor movements in a virtual target acquisition environment. Individual target trials are given an index of difficulty (ID) based on the distance to be traveled to acquire the target, $D$, and the size/width of the target, $W$, allowing different configurations to be tested and compared. The range of IDs appropriate for myoelectric control assessment is typically beneath 4 bits, as trial completion rates (within some allowable time) drastically decrease beyond that level [120]. Over multiple repeated trials, the ID is varied and a range of performance metrics (throughput, path efficiency, average speed, overshoot, completion rate) are computed [121]. Throughput is the primary summary metric of Fitts’ law, combining the time taken to acquire the target and the ID to quantify the information transfer rate of the control scheme. Using such a 2 degree of freedom Fitts’ law test, Wurth et al. [122] concluded that both sequential and simultaneous pattern recognition control schemes outperform conventional myoelectric control. Similarly, Robertson et al. [123] used a Fitts’ law test to conclude that rejection of non-confident decisions (to a no motion state) significantly improves the throughput of myoelectric control. Although the Fitts’ law test provides a reasonable measure of usability, any interactions between usability and confounding factor variability have thus far been largely unexplored.
3. The Limb Position Effect

3.1. Background and Theory

While myoelectric control performance is well-understood in controlled settings, variability in the patterns of EMG can be introduced by a number of different factors. For one, electrodes placed over or adhered to the skin cannot necessarily record the activity of the same underlying musculature under all conditions. Muscle fibers may change positions relative to the sensors as muscle is displaced to provide the tension required during contraction. The shifting of the muscle and subcutaneous tissues then alters the tissue filter effect which modulates the recorded signal. Separately, muscles associated with limb stabilization or locomotion may contract, leading to EMG activity that is unrelated to the motion of interest. This is particularly deleterious when these muscles share common space with intended recording sites. One example of this is the brachioradialis muscle in the forearm which supports both forearm and elbow flexion. In some prosthesis applications, the contact between the electrode and skin may also vary with residual limb position, subject to gravity or device loading.

During motion classification, these factors add ambiguity to the known patterns of muscle co-activation [124,125]. Enforcing rigid constraints on limb position and forearm orientation during data collection manages but also ignores these important sources of signal variability which are present during natural use conditions. Despite improved offline performance, this has been widely reported to lead to a decrease in classification accuracy and robustness during online testing [21,23,124,126,127].

In controlled experiments, the limb position effect is minimized for intra-position classification, when classifiers are trained in the same position as tested. When inter-position classification is performed, however, the limb position effect introduces large amounts of variation that degrade both accuracy and precision [21,124]. When classification is performed across a range of daily conditions, accuracy and precision tend to lie between those of intra- and inter-position because of the presence of some testing positions during training [22]. As the training set grows and includes a greater coverage of testing positions, the accuracy and precision increases towards the upper, intra-position, performance limit. This, however, requires substantial and often not viable collection of vast amounts of training data.

The degradation from variability in limb position has been demonstrated in both able-bodied and amputee populations. In amputee populations the position effect is thought to be less profound than able-bodied populations, possibly due to shorter muscle lengths and sometimes fixed attachment points on the bone [125]. When a prosthesis is donned, however, the weight of the device compresses muscle fibers, altering the patterns exhibited during muscle contraction. These patterns change as a function of limb position on account of gravitational forces on the device providing relief of pressure, shifting of pressure, or increased pressure on the musculature. Additionally, modern surgical interventions (i.e., agonist-antagonist myoneural interface) avoid anchoring individual muscles to enhance control of the residual muscles [128]. The development of solutions to the limb position effect are therefore necessary for robust operation of myoelectric control systems for both amputees and able-bodied individuals.

3.2. Investigating the Limb Position Effect

In this work, a comprehensive investigation was performed to evaluate the limb position effect and serve as a discussion point for comparison to the related literature. Three datasets: (D1) 5 Limb Position [21], (D2) 16 Limb Position [22], and (D3) 3 Forearm Orientation [23], were adopted to demonstrate the effect in different conditions. The high-level signal characteristics of the included datasets can be found in Table 1, with further details available in the original works. The TD [58], LSF4 [129], LSF9 [129], TDPSD [102], and TSTD [77] feature sets outlined in Table 3 were prepared using window lengths of 150 ms and increments of 50 ms.
Results were compared for various frameworks found within the limb position literature; (a) \( x \) vs. \( x \) (intra-position): training and testing in the same position; (b) \( x \) vs. \( y \) (inter-position): training and testing in different positions; (c) \( x \) vs. all (single-position): training in a single position and testing in all positions; and (d) \( N \) vs. all (multi-position): training in multiple positions and testing in all positions. Each testing framework was assessed using leave-one-trial-out cross-validation, within-subject, using \( k \)NN, LDA, QDA, RF, and linear SVM classifiers. Hyperparameters (SVM: regularization factor, RF: number of trees, and \( k \)NN: number of neighbors) were selected in preliminary analysis by a grid search algorithm on a subject of each dataset that was excluded from subsequent analysis. This naive selection greatly reduced computation time for analysis. For the sake of clarity, only results from the TD and TSTD feature sets using LDA and SVM classifiers are shown within this section. For the complete set of results, refer to Supplementary Table S1. The results of these testing frameworks are summarized in Table 5 and Figure 3.

### Table 5. Classification accuracies (%) of the different testing frameworks for the limb position effect.

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>Classifier</th>
<th>Test Framework</th>
<th>5 Limb Position</th>
<th>16 Limb Position</th>
<th>3 Forearm Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean±SD (Min, Max)</td>
<td>Mean±SD (Min, Max)</td>
<td>Mean±SD (Min, Max)</td>
</tr>
<tr>
<td>TD</td>
<td>LDA</td>
<td>Intra-Position</td>
<td>94.5 ± 1.2 (93.1, 96.0)</td>
<td>86.9 ± 1.9 (83.7, 90.2)</td>
<td>96.1 ± 1.4 (94.6, 97.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inter-Position</td>
<td>80.2 ± 7.1 (67.7, 92.0)</td>
<td>75.4 ± 6.6 (58.2, 89.4)</td>
<td>36.8 ± 8.6 (27.8, 47.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single-Position</td>
<td>83.1 ± 2.2 (80.6, 86.6)</td>
<td>76.1 ± 2.9 (70.9, 80.1)</td>
<td>56.6 ± 1.1 (55.5, 57.6)</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>Intra-Position</td>
<td>95.2 ± 0.9 (93.7, 96.1)</td>
<td>85.9 ± 1.8 (83.0, 88.8)</td>
<td>92.3 ± 1.7 (90.7, 94.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inter-Position</td>
<td>86.5 ± 4.7 (77.5, 93.8)</td>
<td>74.2 ± 6.4 (57.2, 86.4)</td>
<td>36.1 ± 8.8 (25.7, 47.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single-Position</td>
<td>88.3 ± 0.8 (86.9, 89.1)</td>
<td>74.9 ± 2.6 (70.3, 78.4)</td>
<td>54.8 ± 2.2 (53.1, 57.3)</td>
</tr>
<tr>
<td>TSTD</td>
<td>LDA</td>
<td>Intra-Position</td>
<td>96.9 ± 0.6 (96.1, 97.7)</td>
<td>91.2 ± 1.8 (88.7, 94.6)</td>
<td>97.0 ± 1.1 (95.9, 97.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inter-Position</td>
<td>86.3 ± 5.9 (75.9, 95.6)</td>
<td>79.9 ± 7.0 (57.1, 92.5)</td>
<td>38.2 ± 7.7 (31.1, 50.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single-Position</td>
<td>88.4 ± 0.9 (87.3, 89.7)</td>
<td>80.6 ± 2.6 (76.1, 83.6)</td>
<td>57.8 ± 1.9 (55.7, 59.4)</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>Intra-Position</td>
<td>94.9 ± 0.6 (93.9, 95.3)</td>
<td>86.7 ± 2.1 (83.5, 90.1)</td>
<td>94.4 ± 1.4 (93.2, 95.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inter-Position</td>
<td>85.6 ± 5.1 (76.6, 92.7)</td>
<td>75.1 ± 6.5 (59.4, 88.6)</td>
<td>45.8 ± 7.4 (35.6, 54.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single-Position</td>
<td>87.4 ± 1.0 (86.0, 88.7)</td>
<td>75.9 ± 2.6 (71.6, 79.3)</td>
<td>62.0 ± 1.2 (60.7, 63.1)</td>
</tr>
</tbody>
</table>

x vs. x (intra-position): The results of the x vs. x testing framework illustrated the conditions under which the majority of myoelectric control experiments have been conducted; those that avoid the limb position effect. In this testing framework, any variability in limb position between training and testing data is minimized, resulting in classification accuracies that are substantially greater than those encountered in clinical practice. As shown in Table 5, this trend of highest accuracy was consistent across feature sets, classifiers, and datasets, often with mean accuracies exceeding 95%. Additionally, the performance was consistently high across different positions, as demonstrated by average standard deviations of 0.6–2.1%. This signifies that repeatability and separability of feature space is consistent across isolated positions, regardless of which position is assessed, as long as single positions are used. This assessment highlights that performance degradation due to the limb position effect is not due to differences in feature separation across different positions. Instead, degradation is more likely influenced by a non-conservative domain adaptation when classification is performed inter-position.
Figure 3. Mean classification accuracies (%) when training with \( N \) positions in the \( N \) vs. all testing framework using the 16 limb position dataset, the TD feature set and the LDA classifier. Error bars indicate standard error measurements across subjects. Significant differences were denoted by * when the \( p \)-value was less than 0.05, whereas n.s. indicates no significant difference.

\( x \) vs. \( y \) (inter-position): The results of the \( x \) vs. \( y \) testing framework illustrated the worst-case performance of all. This framework was used in the original demonstration of the limb position effect [124] because the differences due to limb position are maximized between training and test data. As shown in Table 5, the mean accuracy was substantially lower when training and testing positions were different, as demonstrated by a 12.0%, 11.4%, and 54.9% decrease in accuracy compared to intra-position classification on the 5 limb position, 16 limb position, and 3 forearm orientation datasets, respectively. Moreover, the inter-position classification performance on the 3 forearm orientation dataset was substantially lower than that of the 5 and 16 limb position datasets, signifying that the effect of forearm orientation is more profound than other changes in limb position (such as the elbow and shoulder joints). Performance was better when positions were in close proximity to each other, such as P9 with P14, and P2 with P3 position pairs (Figure 4) which averaged 91.8% and 91.6%, respectively, on the 5 limb position dataset. The similarity of feature space between proximal limb positions demonstrates that the domain adaptation between positions manifests gradually as a function of distance or joint space between limb positions. To train and inform a classifier with maximum position variability information with minimal training time, dissimilar positions (those that exhibit a large degree of domain adaptation) are therefore favoured. An example of such a configuration would be a total of six position conditions: P3, P4, P9, and P14 limb positions, while including supination, neutral, and pronation forearm orientations from a single limb position (Figure 4).

\( x \) vs. all (single-position): The results of the \( x \) vs. all testing framework illustrate the performance that might be expected during regular operating conditions when trained using a single position, as is commonly seen in clinical and laboratory settings. As expected, the \( x \) vs. all testing framework experienced a large decrease in accuracy and precision compared to \( x \) vs. \( x \), but significantly outperformed the \( x \) vs. \( y \) framework. A degradation in performance of 9.5%, 10.5%, and 18.4% was found compared to the intra-position case for the 5 limb position, 16 limb position, and 3 forearm orientations datasets, respectively.

Statistical analysis was conducted using a 4-way ANOVA to identify statistical differences between the various factors of the investigation. A subject-average measure of accuracy was computed for each subject that was the mean accuracy across-cross-validations and testing conditions. This procedure ensures that the dependent variable of interest (accuracy) is independently and identically distributed, an assumption of an ANOVA. A statistical difference was found between testing frameworks (\( p < 0.05 \)), datasets (\( p < 0.05 \)), and classifiers (\( p < 0.05 \)); however, no significant difference was found between feature sets (\( p = 0.18 \)). Of note, all testing frameworks were significantly different, with average accuracies of 88.8%, 76.0%, and 62.8% for the \( x \) vs. \( x \), \( x \) vs. all, and \( x \) vs. \( y \) cases, respectively. The 12.8%
difference between the x vs. x and x vs. all cases highlights the difference between constrained study results and those of more realistic conditions during active use of a device. Post-hoc analysis with Bonferroni correction of the datasets revealed significant differences between all pairs of datasets, with average accuracies of 85.8%, 75.1%, and 66.6% for the 5 limb position, 16 limb position, and 3 forearm orientation datasets, respectively.

The performance of these datasets indicates the amount of variability introduced across the positions surveyed. While the degradation in performance of the 16 limb position dataset was greater than the 5 limb position dataset, it should be noted that it also contained fewer channels of EMG (Table 1). Also, given that it included more positions, the proximity of neighboring positions to one another was closer than those of the 5 limb position dataset. Post-hoc analysis with Bonferroni correction of the classifier factor revealed significant differences between the $k$NN classifier and all other classifiers, with an average accuracy of 79.4%, 77.9%, 77.1%, 76.8%, and 68.0% for the LDA, RF, QDA, SVM, and $k$NN classifiers, respectively. Lorrain et al. [130] also determined that an LDA classifier was the best candidate using the TDAR feature set for motion recognition in the presence of confounding factors with more than 6% improvement when factor-specific variability was introduced to the training set. This is likely because the use of a common (pooled)
covariance promotes more general decision boundaries than other discriminative classifiers [42]. Conversely, the SVM classifier is known to perform well with the wavelet feature set [41,130], likely because it explicitly integrates structural risk minimization. Of note, post-hoc analysis revealed no significant differences between feature sets. This is not to say that features do not affect performance, but that the chosen feature sets have all been validated in myoelectric control and possess meaningful and robust discriminatory information.

**N vs. all (multi-position):** The first three frameworks evaluated accuracy using a single training position; whereas this framework uses multiple training positions. Accuracy was consistently saturated before all limb positions were included. Including more training positions improved accuracy (Figure 3 and Supplementary Figure S2), with no significant improvement found after adding more than 4 positions for the 5 limb position dataset, and more than 6 positions for the 16 limb position datasets. Conversely, no saturation point was reached for the forearm orientation dataset, indicating that all orientations are necessary for an optimal training strategy. Despite these improvements, even when all positions were included in the training set, classification performance of the N vs. all remained significantly worse than the x vs. x testing framework (Table 5 and Figure 3). The N vs. all testing framework accuracy when all positions were included remained 2.3%, 1.5%, and 7.9% worse than the x vs. x framework across the three datasets. The larger discrepancy between these cases for the 3 forearm orientation dataset suggests again that the perturbation of feature space across forearm orientations is more meaningful than those of limb positions. Inclusion of varied training exemplars can effectively overcome the variability of limb position; however, forearm orientation variability remains a current challenge.

### 3.3. Experimental Protocols

Researchers have predominantly adopted limb position experimental protocols in one of four forms: (1) the static forearm orientation protocol, (2) the static limb position protocol, (3) the dynamic two-dimensional (2D) space protocol, and (4) the dynamic three-dimensional (3D) space protocol. Within these, researchers have also proposed set positions in which to elicit hand and wrist motions. Compiled from the surveyed literature, Figure 4 contains a list of positions and unifies them under a single frame of reference. Figure 5 summarizes the frequency with which these various positions and experimental design protocols have been used in the literature.

![Figure 5](image.png)

**Figure 5.** Infographic of the prevalence positions in the literature across 44 articles. Outer ring: limb position and forearm orientation by article count (%). Inner ring: experimental protocol by article count (%).

#### 3.3.1. Static Forearm Orientation

Static forearm orientation experiments have been conducted by many researchers [23,131–137]. Within their protocols, shoulder and elbow joint angles are typically held constant throughout all
collections. Variability of position is introduced via articulation of the wrist joint using the supinator and pronator muscles. Positions within these experimental protocols are generally classified as wrist/forearm supination, wrist/forearm neutral, and wrist/forearm pronation. Note that the term forearm rotation is more appropriate, however the term wrist rotation is often used interchangeably.

The primary focus of static forearm orientation experiments is to determine the muscular co-activation patterns that remain consistent across changes in forearm orientation. Myoelectric control often uses electrodes placed within the housing of a prosthesis, or around the forearm in the case of intact-limbed individuals. As muscles involved in supination and pronation of the forearm are located underneath common collection sites, deviations in forearm orientation greatly influence the observed patterns of muscular activation, and hamper classification performance during online and offline classification [134]. Transradial amputees may or may not exhibit forearm orientation degradation depending on the length of the residual limb, where large residual lengths allow forearm pronation/supination and thus invite forearm orientation degradation. Importantly, hand and finger contractions use both intrinsic muscles within the hand and extrinsic muscles within the forearm that are in close spatial proximity to the pronator and supinator muscles that control forearm orientation, making them sensitive to variations in forearm orientation.

3.3.2. Static Limb Position

The most commonly reported limb position experimental protocol in the literature is one that uses static limb positions [21,22,102,124–127,131,133–135,138–150]. Among the articles surveyed, these make up 50% of all experiments. This protocol prompts subjects to conform to specified wrist, elbow, and shoulder joint angles to reach a characteristic position, commonly those purported to occur during daily activities. While remaining in the desired position, subjects perform repetitions of hand and wrist motion classes. Across the static limb position literature, the average number of positions collected is 8, with a maximum of 91. The most commonly employed positions were P4, P3, P9, P2, P1, and P14 (Figure 4), in descending order of frequency.

The primary focus of static limb position experiments is to determine muscular activation patterns that remain constant across changes in limb positions, or to highlight those that suffer the most. While finger activation patterns are sensitive to variations in forearm orientations due to the proximity of involved muscles, wrist motions are similarly sensitive to changes in limb position due to the use of forearm and shoulder muscles in limb stabilization.

3.3.3. Dynamic 2D Space

Contrary to static forearm orientation and static limb position experimental protocols, dynamic 2D space experimental protocols involve transitions between elbow and wrist joint angles within the collection window [22,127,134,142,144,151–154]. Conventionally, these experiments require subjects to sustain wrist or finger motions while transitioning between two or three limb positions. Figure 6a illustrates an example of dynamic 2D space limb position variation while transitioning between two positions. Figure 6a illustrates an alternative protocol where the subject is given visual guidance through many targets along a trajectory.
The primary focus of dynamic 2D space experiments is to overcome the unreliable performance of myoelectric control systems during online use. By introducing dynamic changes between limb positions, not only are limb stabilization synergies collected, but muscle contractions involved in wrist and elbow joint articulation are also collected. The measurement of these dynamic factors aids in differentiation between motion classes of interest and limb position changes [127].

3.3.4. Dynamic 3D Space

Dynamic 3D space experiments involve the transition between shoulder, elbow, and wrist joint angles while eliciting a target motion class [22, 155–160]. Generally, these protocols can be grouped into two subsets: reach-to-grasp, and activities of daily living (ADLs). Within the reach-to-grasp protocol, subjects sequentially transition between a series of phases: a rest phase, arm-extension to object, contact with object, grasp of object, unhand object, return to initial position, rest. Variability can be introduced by requiring different object-related grips (for instance, palmar, pinch, or power grip) or by modifying the elevation or lateral position of the object to introduce shoulder flexion and abduction variability [155–158]. Within the activities of daily living protocol, progressions between limb positions are designed to model functional tasks: for instance, moving a glass from a tabletop to drinking position or moving from a relaxed position to reaching in a cupboard [22].

The primary focus of dynamic 3D space experiments is to model compound muscle contractions across the entire limb as in real-world functional tasks. The muscle activation patterns of these compound motions have been found to be repeatable under the same conditions; however, when object position or grip style changes, the activation pattern is significantly different [157]. While limited information about shoulder muscle activations can be obtained using forearm electrodes, electrodes situated at the clavicular head of the pectoralis, medial deltoid, infraspinatus, teres major, and superior trapezius muscles have proven successful at distinguishing between shoulder articulations [158]. An alternative application of the reach-to-grasp protocol is the decoding of motor planning using a combination of EMG and electroencephalography (EEG) signals. Neural correlates are defined between the observed EEG signal and key landmarks during the reach-to-grasp motion. Current practices achieve 93.5% accuracy when using EEG to detect the onset of the motion; however, the recognition of the type of grasp using EEG alone has been shown to be a much more difficult task (65.9% for three classes) [161]. This alternative application is greatly motivated by the desire for intuitive neuroprosthetics to restore function to patients with high levels of motor lesion.

3.4. State-of-the-Art Approaches

Over the past decade, numerous studies have proposed methods to minimize the degradation caused by variation of limb position. These methods can be roughly categorized into four main
approaches: (1) developing feature extraction, dimensionality reduction, and classification algorithms that are less susceptible to the variations of signal patterns between positions; (2) including prior information about the variability of EMG signals by collecting motions in numerous pre-defined static and dynamic positions; (3) improving generalization of features across positions through transfer learning; and (4) incorporating limb position measurements taken from extra sensor modalities.

3.4.1. Robust Algorithms

The first method of minimizing the effect of limb position is the exploration of feature extraction, dimensionality reduction, and classification algorithms that are inherently less susceptible to this effect [23,102,127,142,147,152,162–164]. In an exploratory study conducted by Liu et al. [127], repeatability metrics between positions were computed for various commonly used feature sets. The results showed that current feature sets involving Hudgins’ TD features, auto regressive coefficients, and cepstrum coefficients underwent significant changes when limb position is changed. Khushaba et al. [102,164] expanded upon the Hjorth parameter feature set to extract further time-domain power spectral descriptors (TDPSD) from the time domain using more stable normalization processes. The TDPSD feature set proved to be less variable than other feature sets under the perturbations of limb position [23], and performed similarly in response to force levels [165]. The TDPSD feature set achieved 85%, 91% and 87% accuracy on low, medium, and high intensity contractions between three forearm orientations; whereas, the Hudgins’ set of TD features achieved 78%, 83%, and 85%, respectively. An alternative feature set was proposed by Gu et al. [152], where six band-limited frequency bands were defined using DFT, DWT, and WPT prior to extracting TD features. Between seven static limb positions, SVM classification accuracies of TD features in isolation improved on average by 30% accuracy when extracted from DFT, DWT, and WPT representations instead of from the raw data.

The use of novel classification algorithms has achieved strong performance in the presence of limb position variability. Yu et al. [148] validated the use of a mixed-LDA classifier that used stable representations of motions defined by taking the common-mode of class-specific covariance clusters across all positions. Across five static limb positions, the mixed-LDA classifier achieved 93.6% accuracy opposed to 82.5% accuracy using a standard LDA classifier. Liu et al. [142] found that conditional Gaussian mixture models (cGMM) consistently outperformed conventional LDA during inter-position classification. Across four static limb positions and three dynamic 2D space positions, cGMM had 2% higher accuracy than LDA across all six subjects. Mukhopadhyay et al. [166] achieved 98.9%, 98.7%, 90.6%, 91.8%, and 88.4% accuracy across 5 positions using a DNN, SVM, kNN, RF and DT with the TDPSD feature set. Power et al. [167] determined dynamic time warping (DTW) of the RMS value of the signal yielded higher accuracy and lower computational cost than the TDPSD feature set. Liu et al. [168] used a linear-nonlinear cascade regression to simultaneously estimate shoulder, elbow, and wrist joint angles accounting for 93%, 90%, and 84% of the variance in able-bodied subjects, and 85%, 91% and 85% of the variance in stroke subjects, respectively. Betthauser et al. [147] validated the use of a sparse representation classification (SRC), which had found prior success in image detection in cases of heavily occluded objects or missing pixels. The application of SRC yielded significantly better performance across all combinations of training and testing positions ($p < 0.001$). Additionally, the use of extreme learning machines (ELM) in adaptive sparse representation classification (EASRC) greatly reduced computational burden of SRC while maintaining stability under untrained limb positions during an online experiment [162].

These algorithmic approaches to reducing the impact of the limb position effect have indeed enhanced myoelectric control system usability. Nevertheless, the adoption of these more robust methods incurs a tradeoff between the otherwise highest achievable classification accuracy and reliability. In order for existing state-of-the-art configurations to approach the usability of these newer, more robust algorithms, variability between positions may be introduced through strategic training strategies.
3.4.2. Training Strategies

The simplest method to improve predictive power when extrapolating outside of the measured conditions is to collect exemplars of motions classes in the unknown condition [21,23,124,126]. Leveraging this technique, Scheme et al. [124] trained two LDA classifiers: the first was trained using a single limb position, while the second was trained with all eight limb positions. The single-position classifier had an average inter-position accuracy of 65%, whereas the eight position classifier performed substantially better with accuracies between 86–95% among all positions. The effect of forearm orientation was likewise examined by Khushaba et al. [23], where an SVM classifier was trained to detect six motions collected during wrist supination, wrist neutral, and wrist pronation. Although intra-orientation motion recognition had acceptable accuracy (94.2%), no single orientation achieved robustness across positions (67.3%, 60.6%, and 67.6% accuracy across forearm orientation testing frameworks trained in P2, P2s, and P2p, respectively). Further, the magnitude of variability introduced by limb position and by forearm orientation was investigated by Yang et al. [134], with findings that hand and finger gestures were more influenced by forearm orientation than limb position.

Although the benefit of the inclusion of exemplars from multiple positions is evident from these studies, a practical limitation of this approach is the requirement for longer training procedures to inform position tolerant decision boundaries. However, by instead allowing dynamic motions that span static positions, or free motion of the arm in the 3D space, a greater breadth of limb positions and forearm orientations can be measured in less time [22,169,170]. A second consideration of this method involves the possibility that inter-position variability, if in excess, may ambiguate decision boundaries, increasing inter-class error in a single stage classifier [162].

As demonstrated in Section 3.2, Table 5, motion recognition rates are optimal when all possible testing positions are recorded within the training period. While this solution is the conceptually simplest solution to mitigating the limb position effect, a collection period that requires repetition of the training protocol in multiple positions is extremely cumbersome and could result in low user adoption of the system, regardless of the usability once trained. To overcome the limitation of exhaustive training protocols, transfer learning has been proposed as a viable alternative.

3.4.3. Transfer Learning

A third approach to minimizing the limb position effect, while avoiding exhaustive training procedures, is the use of transfer learning. The fundamental aim with transfer learning is to leverage information learned in one domain to improve performance in another. Although now commonly associated with deep learning transfer learning can also be achieved using dimensionality reduction techniques. For example, canonical correlation analysis (CCA) has been applied to reduce the impact of variability between limb positions. The intention of transformations like CCA is to learn a linear projection between EMG feature space from each position to a reference position. This projection can be applied during device operation when the position is known to minimize the degradation caused by the limb position effect. Also using CCA, Cheng et al. [163] attained similar inter-position and intra-position accuracies and demonstrated that further benefits were gained through using CCA together with uncorrelated linear discriminant analysis (ULDA). The potential of CCA has been demonstrated for various sources of variability, particularly for minimizing the amount of training necessary for novel users by adapting the domain according to subject specific information [141].

In addition to CCA, other feature projection techniques have been explored for myoelectric control. For example, Ishii et al. [133] used a two-stage cascaded bilinear transform, building on the work of Matsubara et al. [171], in an effort to remove the subject and position effect from EMG signals. An emerging avenue of transfer learning for myoelectric control is the strategic retraining of existing neural networks with confounding factor specific data to specialize model weights against the factor’s variability. Although weight-adjustment transfer learning has not been explored as a solution for the limb position effect, supporting work exists for such methodologies.
applied to reduce inter-subject variability [172,173], and has been validated for use within-subject and between-sessions [174].

3.4.4. Extra Sensor Modalities

A final approach to minimizing the effect of limb position is the use of specialized sensors. Fougner et al. [21] defined two frameworks for sensor fusion between EMG and accelerometer/inertial measurement unit information: (1) a cascaded classifier that, first, classifies position using accelerometer information, then performs motion recognition on the appropriate position-specific classifier; and (2) a single-stage classifier trained using features from accelerometers and EMG signals. The first framework describes multiple position-specific classifiers, each of which require adequate data collection in their respective discrete positions. This framework is usually preceeded by position classification. The second framework describes a single position-inclusive classifier, where features from all positions are combined into a single-stage classification scheme. This framework requires features that distinguish position and motion specific information for optimal separability of classes; and thus increases the amount of training data to avoid the curse of dimensionality [144]. Masters et al. [143] explored such frameworks using only EMG to distinguish between positions and found significant improvement over training in the neutral position; however, no significant difference was found between the position-specific and position inclusive frameworks. Further, Shazad et al. [175] determined single-stage classifiers trained with continuous position data outperformed discrete position classifiers with 98.7% and 97.6% accuracies, respectively.

The use of sensors other than accelerometers has been explored to provide positional awareness. Geng et al. [139] added mechanomyography (MMG) signals to augment EMG, yielding improved position specificity. As a result, the MMG-informed classifiers outperformed EMG alone in both the single-stage and two-stage classifier frameworks. Interestingly, Khushaba et al. [99] explored graph laplacian (GL) based feature extraction for accelerometer and MMG modalities, and found that GL based feature extraction yielded 93.8% and 94.1% accuracy, respectively, across 40 motion classes. This is in stark contrast to Hudgins’ TD features based EMG classification, which obtained only 66% accuracy. This work suggests that further investigation is required to better understand the discriminative power of these alternative sensors. These results should be interpreted with caution, however, as the dataset used exploits positional variance to enhance separability of motion classes. This is a common trait among human-computer interaction studies, however, biomedical studies motivated by prosthetic control cannot exploit positional variance as motions are required to be elicited regardless of limb position [176].

4. Variation in Contraction Intensity

4.1. Background and Theory

As with limb position, neglecting variability introduced by muscle contraction intensity can also lead to differences in observed performance between controlled lab results and clinical usability. When repeated measurements of the same motion and contraction intensity are taken, the integrated EMG signal is highly reproducible, attaining correlation coefficients on the order of 0.98 [177–180]. Signal variability between contraction intensities, however, remains a current challenge for myoelectric control systems in clinical applications. This is particularly important given the widespread use of proportional myoelectric control, which necessitates variations in contraction intensity to control the speed of actuation of a terminal device or cursor [24,181–183].

Variability in muscle contraction intensity is modulated physiologically by two factors: motor unit (MU) recruitment threshold, and firing frequency. First, as the intended intensity of the motion increases, so does the neuromuscular signal potential. As the signal increases, greater number of MUs are fired as their recruitment threshold is surpassed, resulting in stronger muscle contractions. Generally, innervation of low-threshold type I (aerobic, fatigue-resistant) MUs persists across all
contraction intensities, while higher-threshold type II (anaerobic, fatigue-prone) MUs are additionally recruited when the neural drive is high. Second, the inter-pulse interval of all MUs decreases with contraction intensity, with refractory periods between contractions range from 0.1 s during moderate contraction intensity to 0.025 s during maximum intensity contractions. These two mechanisms introduce signal variability across contraction intensities within the same class of motion.

While the innervation and activation of individual MUs remains largely consistent across intensities, when excluding signal-dependent motor noise \[184,185\], force production is modulated macroscopically using combinations of these two factors. Additionally, as a consequence of larger signal amplitude, the multiplicative signal-dependent motor noise manifests with larger variance. Consequently, the generated EMG signal can vary widely even within the same motion class. This variation is often leveraged to facilitate proportional control in myoelectric classification, or as part of more continuous regression schemes. Notably, the commonly used MAV feature is used, paradoxically, for both classification and proportional control. Nevertheless, without due consideration, the introduced variability can greatly decrease classification performance and overall usability.

### 4.2. Investigating the Force Effect

As before, a comprehensive investigation was performed to evaluate the effect of contraction intensity and to serve as a discussion point for comparison to the related literature. Three datasets: (D4) 7 different\% MVC levels \[24\], (D5) 3 different subjective levels (A) \[24\], and (D6) 3 different subjective levels (B) \[23\], were adopted to demonstrate the effect in different collection conditions. The details and pre-processing of the included datasets can be found in Table 1, with further information found in the original works of each dataset. Feature extraction was performed to create TD \[58\], LSF4 \[129\], LSF9 \[129\], TDPSD \[102\], and TSTD \[77\] feature sets using a window length and increment of 150 ms, and 50 ms, respectively.

Testing frameworks were defined identically to those in the limb position investigation; however, groups were instead stratified using contraction intensity level. Consequently, motion recognition was performed in the following conditions: (a) \(x\) vs. \(x\) (intra-level): training and testing data in the same intensity level; (b) \(x\) vs. \(y\) (inter-level): training and testing data in different intensity levels; (c) \(x\) vs. all (single-level): training in a single level and testing in all levels; and (d) \(N\) vs. all (multi-level): training in an increasing number of levels and testing in all levels. As before, within-subject, leave-one-trial-out cross-validation was used to determine performance metrics across five classifiers: \(k\)NN, LDA, QDA, RF, and SVM. For the complete set of results, refer to supplementary Table Supplementary Table S2.

#### Table 6. Classification accuracies (%) of the different testing frameworks for the contraction intensity effect.

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>Classifier</th>
<th>Test Framework</th>
<th>7% MVC Level Mean±SD (Min, Max)</th>
<th>3 Subjective Level (A) Mean±SD (Min, Max)</th>
<th>3 Subjective Level (B) Mean±SD (Min, Max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD</td>
<td>LDA</td>
<td>Intra-Level</td>
<td>94.0 ± 5.8 (83.0, 99.2)</td>
<td>92.7 ± 2.9 (89.5, 95.2)</td>
<td>96.1 ± 2.0 (93.8, 97.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inter-Level</td>
<td>67.9 ± 26.0 (19.7, 99.0)</td>
<td>67.2 ± 19.2 (36.7, 90.2)</td>
<td>71.4 ± 11.9 (55.5, 81.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single-Level</td>
<td>71.7 ± 6.6 (61.2, 79.3)</td>
<td>75.7 ± 6.4 (68.4, 79.9)</td>
<td>79.6 ± 5.7 (76.0, 86.2)</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>Intra-Level</td>
<td>94.1 ± 6.3 (81.6, 99.2)</td>
<td>93.4 ± 3.2 (90.3, 96.7)</td>
<td>92.8 ± 1.8 (91.7, 94.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inter-Level</td>
<td>67.2 ± 25.1 (19.3, 98.4)</td>
<td>65.7 ± 17.6 (36.7, 85.8)</td>
<td>59.4 ± 12.0 (40.5, 71.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single-Level</td>
<td>71.0 ± 6.1 (63.0, 78.5)</td>
<td>75.0 ± 4.3 (70.2, 78.7)</td>
<td>70.6 ± 6.0 (64.1, 76.0)</td>
</tr>
<tr>
<td>TSTD</td>
<td>LDA</td>
<td>Intra-Level</td>
<td>95.0 ± 5.5 (85.1, 99.2)</td>
<td>94.1 ± 3.4 (91.1, 97.7)</td>
<td>96.9 ± 1.1 (95.7, 97.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inter-Level</td>
<td>71.2 ± 23.1 (25.7, 98.9)</td>
<td>68.4 ± 15.8 (49.6, 87.8)</td>
<td>69.0 ± 17.4 (40.4, 84.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single-Level</td>
<td>74.6 ± 7.9 (59.3, 82.5)</td>
<td>77.0 ± 6.0 (70.9, 82.9)</td>
<td>78.3 ± 8.6 (69.3, 86.6)</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>Intra-Level</td>
<td>91.8 ± 5.3 (81.6, 96.1)</td>
<td>91.0 ± 2.9 (88.2, 94.0)</td>
<td>93.8 ± 2.4 (91.3, 96.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inter-Level</td>
<td>66.3 ± 20.4 (28.4, 94.1)</td>
<td>62.7 ± 13.7 (46.8, 77.8)</td>
<td>62.7 ± 11.1 (45.1, 72.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single-Level</td>
<td>69.9 ± 8.1 (54.2, 77.8)</td>
<td>72.1 ± 5.2 (66.7, 77.0)</td>
<td>73.1 ± 4.5 (68.8, 77.8)</td>
</tr>
</tbody>
</table>
Figure 7. Classification accuracies (%) of the \(x\) vs. \(y\) testing framework using the 7\% MVC level dataset and the TD feature set.

Figure 8. Classification accuracies (%) of the \(N\) vs. all testing framework using the 7 intensity dataset, the TD feature set and the LDA classifier. Error bars indicate standard error measurements for each number of position across subject. Significant differences were denoted by \(*\) when the \(p\)-value was less than 0.05, whereas n.s. indicates no significant difference.

\(x\) vs. \(x\) (intra-level): The results of the \(x\) vs. \(x\) framework demonstrated the most controlled case (as often done in myoelectric control experiments), using only within-intensity classification. As with the limb position investigation, the accuracies obtained in this frameworks were greater than clinical benchmarks obtained in more practical conditions. As shown in Table 6, the mean intra-level classification accuracy across all subjects was often found to exceed 94\%. In contrast to the intra-position frameworks, the intra-level frameworks spanned a higher range of standard deviations in classification accuracy ranging from 1.1–6.3\%. These larger standard deviations result from challenges in the precise repeatability of contraction intensity and differences in the separability of feature space across the set of intra-level tests. Unlike the limb position effect, a secondary factor of contraction intensity variability (and thus induced performance degradation) is decreased performance when classifying lower intensity contractions. This difficulty is further magnified for motion classes such as forearm pronation and supination which are not typically associated with high force muscle activations.

\(x\) vs. \(y\) (inter-level): The results of the \(x\) vs. \(y\) framework illustrated the worst-case scenario, with maximum contraction intensity variability between training and testing data. As expected, the \(x\) vs. \(y\) framework experienced the worst performance of all testing frameworks. As in Table 6, the mean accuracy attained during motion recognition was substantially lower when training and testing intensity levels were different, demonstrated by a 25.6\%, 26.8\%, and 29.3\% decrease in mean accuracy when performing inter-level classification as opposed to intra-level classification on the 7 intensity, 3 intensity (A), and 3 intensity (B) datasets, respectively. The classification accuracy of the inter-level framework varied widely, with standard deviations in the range of 11.1–26.0\%. This variability supports that performance across levels is highly dependent on the selection of the training intensity, as outlined in [24]. The 7 intensity dataset, for instance, had a mean accuracy of 66.3\% when using the TSTD feature set with the SVM classifier; however, one train-test pair of intensity levels achieved 94.1\% accuracy while another achieved only 28.4\%. Additionally, performance was consistently worse
when testing on low intensities and training on high intensities when compared with the opposite arrangement of intensities (Figure 7). While the classifier trained on the median intensity level achieved the highest mean accuracy, a classifier based on any single intensity level resulted in likely unusable performance across all feature sets. The collection of multiple intensity levels is therefore required to avoid degrading usability due to muscle contraction intensity variability. Moreover, as the contraction intensity of the training set increases, the performance degradation when testing against even neighboring intensity levels decreased.

\textit{x vs. all (single-level):} The results of the x vs. all framework illustrated the impact of contraction intensity variability across a range of contraction intensities day when trained with a single intensity level. As expected, the x vs. all testing framework experienced a large decrease in accuracy and precision compared with the x vs. x framework. Despite this decrease, the x vs. all testing framework still significantly outperformed the x vs. y testing framework. A degradation in performance of 21.9\%, 17.9\%, and 19.5\% was observed when compared to the x vs. x testing framework for the 7 intensity, 3 intensity (A), and 3 intensity (B) datasets, respectively.

Statistical analysis of the accuracy of the three single training level testing frameworks was conducted using a 4-way ANOVA. A statistical difference was found for the testing framework (p < 0.05), classifier factor (p < 0.05), and feature set factor (p < 0.05); however, no significant difference was found between datasets (p = 0.138). All pairs of testing frameworks were significantly different during post-hoc analyses, where the mean accuracy of x vs. x, x vs. y, and x vs. all was 92.0\%, 63.2\%, and 71.1\%, respectively. As initially hypothesized, the x vs. x framework attained the best performance among testing frameworks, the x vs. y framework attained the worst-case performance, and the x vs. all framework attained better performance than the worst-case scenario when using the same dataset and pattern recognition pipeline. The choice of classifier was found to have an impact on the robustness to contraction intensities, with LDA, QDA, SVM, RF, and kNN classifiers attaining mean classification accuracies of 80.5\%, 77.2\%, 75.2\%, 72.2\%, and 72.2\%, respectively. Notably, LDA achieved the highest accuracy across all classifiers, significantly outperforming all other tested classifiers across all testing frameworks. In contrast to the limb position investigation, the feature set factor was found to be significant. The TD and LSF9 feature sets significantly outperformed the TDPSD feature set with mean classification accuracies of 76.1\% and 76.0\% compared to 73.8\%, respectively. Also unlike the limb position investigation, results across datasets were not found to be significantly different. During this investigation, the amount of variability introduced across the range of intensities of each dataset was relatively consistent, as this variation is more readily modelled experimentally. Therefore, when designing experiments to accommodate the contraction intensity effect, while there is a requirement to survey a range of intensities, the number of levels elicited is of less importance than with limb position.

Of note, this outcome only corresponds to differences in intensity levels in able-bodied subjects. Due to diminished proprioception, muscular atrophy, or surgical outcomes, amputees often report difficulty producing low- and high-intensity contractions [186]. When purposeful regulation is required, the use of visual feedback (such as the overall mean MAV) as an accompanying modality during the training procedure may have a meaningful impact for these users [186]. Likewise, similar difficulties have been reported for other user groups, such as spinal cord injury (SCI) populations, in regulating a targeted intensity level, and in particular, high levels. For these groups, the use of biofeedback with operant conditioning has been used for rehabilitation to increase maximum contractions, and during collection to augment the precision of the recorded intensities [187].

\textit{N vs. all (multi-level):} In contrast to the first three testing frameworks which used a single training intensity level, the effect of multiple training intensity levels was determined using the N vs. all testing framework. As shown in Figure 8, accuracy and precision improved on the 7\% MVC level dataset as the number of training intensity levels increased (similar relationships on other datasets shown in Supplementary Figure S2). Across all datasets, the minimum number of levels required to obtain near-best performance was two (p < 0.05); 20\% MVC (low) and 70\% MVC (high). This
finding corroborates that classification performance can be optimized with few training repetitions as long as the main variability of contraction intensity is included. As expected, there was a positive relationship between the number of training levels included and classification accuracy; however, the mean accuracy when all positions were included remained significantly lower ($p<0.05$) than the $x$ vs. $x$ framework. This suggests that, despite including the contraction intensity variability from all levels, the classifiers have difficulty in combining this variability completely. The use of these common classifiers to create adequate decision boundaries appears to be limited by the increased entropy introduced by multiple intensities. As in the limb position effect, one solution to this problem is the use of cascaded classifiers, where a preliminary classifier identifies the % MVC and selects the appropriate single level classifier for motion classification.

4.3. Experimental Protocols

Across the literature, the experimental designs of contraction intensity works can be divided according to three properties: (1) types of muscle contractions, (2) normalization techniques, and (3) intensity levels selected within the experiment. Figure 9 illustrates the distribution of these three traits across the 18 applicable contraction intensity articles surveyed.

4.3.1. Types of Muscle Contractions

First, the type of contraction elicited during the experiment can have large implications on the interpretation of research outcomes. Isometric, or static, contractions are performed when muscles are under tension but no movement is generated. Isometric contractions are evident when a position is maintained outside of its relaxed state. During these contractions, MUs are sequentially fired at a sufficient rate to maintain tension within the same motion. These contractions model the steady-state phase of the motion, which is the fundamental target of real-time myoelectric methodologies. While experiments with this type of contraction provide a direct measure of the control scheme’s effectiveness in steady-state, natural motions have temporal characteristics and transitions that this format does not capture. Dynamic, isotonic, or ramp contractions, on the other hand, involve changing muscle lengths (shortening or lengthening) and thus movement of a body part is generated. In contrast to isometric contractions, dynamic ramp contractions capture the continuous temporal evolution of natural motions from rest to the desired intensity and sometimes back to no-motion. As such, these contractions are better suited as models for the usability of the device outside of laboratory settings. Training myoelectric devices with such contractions has demonstrated meaningful usability improvements over devices trained with static contractions, where training with dynamic ramps yielded 16.6% higher accuracy than their isometric counterparts [24]. From Figure 9a isometric contractions dominate experimental designs of contraction intensity literature, with only 3 of 18 applicable articles using a dynamic contraction profile.

4.3.2. Intensity Level Normalization Techniques

As in Figure 9b, subjective regulation of contraction intensity is the most ubiquitous regulation technique for contraction intensity studies. This subjective method is achieved by attempting intended
levels with no additional feedback provided to the user. This subjective estimation, however, lacks specificity to define precise intensity levels; therefore, subjects are typically told to produce “low” contractions as opposed to, say, 20% MVC. An advantage of subjective estimation, however, is that it is appropriate for gauging the usability of the myoelectric device when trained unaided by external intensity feedback, as would commonly be the case clinically and during everyday use.

In contrast to subjective regulation techniques, objective regulation techniques specify levels of intensity as a normalized percentage of some signal amplitude. The predominant normalization techniques employed include: (1) muscle-specific maximum voluntary isometric contraction (MSMVIC), (2) task-specific maximum voluntary isometric contraction (TSMVIC), and (3) dynamic-range maximum voluntary contraction (DRMVC) [188]. MSMVIC primarily uses a manual muscle test to achieve maximal activation of the target muscle. Manual muscle tests are motions that produce maximal neural activations of the chosen muscle; for an extensive survey of manual muscle tests for various muscles, see Halaki et al [188]. When subjects have appropriate coaching and familiarity with the desired motion, repeatability of contractions is high; however, contractions are often reported exceeding 100% of the maximum activation indicating that selection of manual test may not be appropriate for all subjects [189–191]. Studies that assess physiological properties of muscles gravitate towards MSMVIC normalization; however, myoelectric control scarcely employs this technique. Conversely, TSMVIC relies on the measurement of maximal muscle contractions during a target task. Myoelectric control predominantly employs TSMVIC to collect training data for intensity robust motion recognition systems, where the training procedure prompts users the elicit maximal contractions of each motion to establish objective reference intensity levels for further repetitions of each respective motion. In contrast to MSMVIC, TSMVIC observes fewer isometric contractions exceeding 100%; however, 100% is still exceeded under dynamic contractions and the relative activation of muscles is variable across-subjects. Finally, DRMVC is suggested as the only appropriate normalization technique for contractions accompanied by movement. In response to changing activation levels during stages of eccentric and concentric contractions [192], reference intensity levels are established using maximal isokinetic contractions. During future myoelectric control experiments where dynamic/ramp contractions are emphasized, it is recommended that DRMVC predominantly be used.

When determining the reference scale using any normalization technique, it is suggested to use at least 3 repetitions, each separated by 2 min, to improve robustness of measurements while limiting fatigue [188,193]. As an alternative to maximal contractions, the use of submaximal contractions (roughly 80% MVC) has been proposed when subjects experience difficulty in total motor recruitment [194]. As shown in Figure 9b subjective regulation dominates the experimental designs in the literature, with only 7 of 18 contraction intensity effect related articles using a motion specific MVC normalization regulation technique.

4.3.3. Intensity Levels

Although signal amplitude is commonly used as a proxy for contraction intensity, other characteristics are also influenced by the intensity of contraction, including the observed innervation zone, the low frequency peak, and the probability distribution function of the signal. As mentioned in Section 4.1, an increase of contraction intensity is manifested through higher MU recruitment and an increase in firing frequency. The increase in recruitment results in more frequent activation of type II MUs. As a consequence of the recruitment of other muscle fibers, the observed innervation zone of the muscles may shift; for example the biceps brachii has been observed to shift up to 2.4 cm over the range of 10%-70% MVC [195]. The increase in firing rate is directly observable via a low frequency peak in the power spectrum of the signal. As the intensity of contraction increases, the frequency of this peak can change from 10 Hz to 40 Hz. The consensus on the overall change in frequency, however, is still under debate. Mean and median frequency have been observed to increase [196–198], decrease [199–202], or remain consistent as intensity increases. The probability distribution function
of the EMG signal is typically described as a combination of Gaussian and Laplacian, with their weighting being dependent on the contraction intensity [203]. The distribution has been reported to tend towards Laplacian as intensity increases [204], Gaussian as intensity increases [205,206], or towards Laplacian as intensity deviates from 50% [207].

The distribution of intensity levels adopted across studies in the literature is consolidated in Figure 9c. Across all surveyed studies, the mean number of contraction intensity levels was $3.7 \pm 2.0$, ranging from 2 to 11 levels. The most commonly used number of intensities explored was 3, where those three intensities were low, moderate, and high intensities. Combining the statistical analysis of Section 4.2 with previous outcomes presented in the literature, an appropriate set of intensity levels for an experimental design can be chosen to include 20% MVC (low intensity) and 70% MVC (high intensity) for able-bodied subjects. Additionally, a small sample of current studies successfully incorporate additional contraction variability using ramp contractions. These ramps dynamically range between roughly 20% MVC and 80% MVC and incorporate much of the information without the need for individual repetitions or feedback. Nevertheless, the robustness of myoelectric control systems to dynamic contractions and intentional proportional control remains an active challenge.

4.4. State-of-the-Art Approaches

A variety of solutions have been proposed to minimize the impact of contraction intensity variability. Two main methodological approaches have been employed to minimize the variability observed across intensities: (1) training strategies, including prior information about the variability of EMG signals by collecting motions in numerous isometric or dynamic intensity levels; and (2) robust algorithms, including feature extraction, dimensionality reduction, and classification algorithms that are less susceptible to these variations. A third category focuses instead on using the intensity variability itself as an additional control scheme, so as to facilitate proportional control.

4.4.1. Training Strategies

Many experiments have been conducted to determine the optimal training intensity levels for myoelectric control, with little consensus or convergence to a single solution. Khushaba et al. [132] examined the effectiveness of seven feature sets (TDPSD, DFT, DWT, TD, TDAR, MFMF, TDAR with RMS) when trained on low, medium, or high contraction intensities. Classifiers trained with low intensity contractions performed significantly worse than classifiers trained with moderate and high intensity contractions. Across all feature sets, a difference of 10.0% and 4.4% was observed between low intensity trained classifiers and moderate and high intensity counterparts, respectively. Li et al. [208] investigated the robustness of classifiers trained exclusively on low, medium, or high intensity levels and achieved 69%, 81%, and 74% accuracies, respectively, on test sets containing all intensity levels not included in the training set. Similarly, He et al. [104] suggested using the median intensity level along with the range of anticipated test intensities. While the specific intensity level for an optimal single-intensity training procedure differs across the literature, there is agreement that the simplest method of improving robustness to intensity variability is training with multiple intensity levels. There is less research reported, however, on the best combination of intensities with which to train. Scheme et al. [181] proposed training with exemplars of low intensity (20% MVC) and high intensity (80% MVC) to ensure robustness under all intensity levels. Finally, Scheme et al. [10,24] also recommended including ramp contractions within the training set to both capture the natural variability of contraction intensity during regular use, and to expedite the collection process. This approach has since been adopted as best practice by many groups across the field [209,210].

Despite the robustness gained by collecting multiple intensity levels, or these dynamic training regimes, there is motivation to further improve robustness through algorithmic solutions. Consequently, the development of pattern recognition methodologies that are invariant to contraction intensity have been explored.
4.4.2. Robust Algorithms

In contrast to the above training strategies, the aim of robust algorithms is to be inherently less sensitive to variability introduced by different contraction intensities. These methods have typically fallen into three categories: (1) feature extraction, (2) feature projection, and (3) classification algorithms.

Robust feature extraction methods aim to extract features that are resilient to changes in contraction intensity while retaining discriminative motion specific information to facilitate classification. Tkach et al. [103] determined that the selection of appropriate time domain features can improve classification performance in the presence of intensity variability by 16%, with a particular emphasis on the robustness of time-series modeling features, such as autoregressive and cepstral coefficients. Similarly, nonlinear features have been found to outperform linear features (only AR features performed well inter-level), with power spectral entropy, detrended fluctuation analysis (DFA) [81], and maximum fractal length achieving the highest performances in isolation [211]. Nonlinear features were further explored by Iqbal et al. [212], with entropy features derived from the wavelet packet transform outperforming those from DFA.

In addition to evaluation of isolated features, sets of features have also been evaluated. A comparison between TD with KURT and AR with RMS feature sets determined that the former yielded the best performance for two amputee subjects [186]. The use of time domain (TD, TDPSD, TDAR, TDAR with RMS), frequency domain (DFT, MFMF), and time-frequency representation (wavelet) feature sets were compared across 3 intensity levels [23,132]. TDPSD significantly outperformed all other feature sets ($p < 0.01$), while DFT appeared consistently as the second best. In addition, Li et al. [213] explored the use of common spatial patterns extracted from the EMG signal, and found that they performed 5.3% better than the TD feature set. Through the use of an improved DFT, Lv et al. [214] demonstrated greater accuracy using FD features than TD features across intensities. Feature sets that harness nonlinear complexity information, such as fractal features and power spectral entropy with AR, demonstrated the possibility of motion recognition in the absence of reliance on amplitude features; however, this system under-performed compared to TDPSD [211]. Finally, Asogbon et al. [74] examined the combination of contraction intensity variability and subject mobility on myoelectric control. Their proposed feature set, invTDF, outperformed TD by 12% and reported considerably lower across-subject variability of performance.

Alternatively, feature projection techniques selectively retain information from a broader set of features. Motivated by the modularity of motor control for myoelectric control, Atoufi et al. [215] employed NMF to attempt to extract neural primitives from the EMG signal across a range of intensities. In cases of both known and unknown testing intensities, however, MAV outperformed the neural input based feature set. Feature normalization methods demonstrated a meaningful improvement of 5% and 11% for TD and DFTR feature sets, respectively, during the classification of unseen intensity levels [104]. Likewise, the TDPSD feature set has been employed in conjunction with SR [216] to improve robustness to contraction variability within both congenital and traumatic amputee populations. An improvement of 6–8% accuracy was obtained across intensities using TDPSD when compared against VDMOM, AR+RMS, TD, and wavelet feature sets that persisted across subjects and classifiers (LDA, RF, NB, kNN) [165].

Finally, classification algorithms have been validated for their robustness to contraction intensity. Al-Timemy et al. [165] determined that the LDA classifier was optimal compared to the RF, naive Bayes, and $k$NN classifier when all intensity levels were included in the training set. Moreover, the application of temporal classifiers like DTW was justified by Powar et al. [217], where DTW performed on RMS representations of amputee and able-bodied EMG signals yielded higher accuracy than conventional classifiers using the TDPSD feature set. Cascaded classification architectures have also been employed to alleviate the impact of changing contraction intensities, with a first stage selecting the appropriate force-specific second-stage motion classifier. Li et al. [218] demonstrated the cascaded classification architecture yielded a 18.1% increase in accuracy when using three second-stage classifiers.
4.4.3. Proportional Control

The variability introduced through muscle contraction intensities may degrade classification performance; however, this phenomenon has also been harnessed to improve control. These proportional control schemes modulate the amount of class-specific output (such as the velocity of a device or cursor) based on the contraction intensity. Conventional myoelectric control systems utilize this within-class variability by mapping an estimate of the EMG amplitude (traditionally MAV or RMS [219]) to the output velocity of the recognized class. Hoozemans et al. [220] predicted the intensity of prehensile motion ranging between 0 N and 300 N using a low-pass representation of the EMG signal from six forearm electrodes obtaining mean differences between predictions and true intensities of 24 N. Three amplitude reliant proportional control schemes were validated using Fitts’ Law tests by Scheme et al [181]: (1) class-specific proportional control (CSPC), (2) auto-calibrated class-specific proportional control (ACSPC), and (3) bounded proportional control between 10% MVC and 70% MVC (BPC). BPC yielded the best performance among the paradigms used, where improvements over CSPC of 40%, 22%, and 72% were found for able-bodied subjects and 21%, 10%, and 44% for amputee subjects in throughput, efficiency, and overshoot, respectively.

While amplitude estimates are widely used within proportional control schemes, other metrics have also been explored. Atoufi et al. [221] explored the use of neural weights produced through NMF to predict contraction intensities of 8 hand and wrist motions using an ANN classifier. Neural weights significantly outperformed MAV between-subjects on metrics of RMSE (0.90 ± 0.43 vs. 1.17 ± 0.64) and $R^2$ (0.76 ± 0.22 vs. 0.52 ± 0.49) during intensity estimates. The use of frequency features was explored by Bilodeau et al. [222] where MNF and MDF significantly increased ($p < 0.05$) with an increase in contraction intensity for three quadriceps muscles. In contrast, Throngpanja et al. [199] determined that MNF, MDF, and time-dependent MNF and MDF significantly decrease as contraction intensity increases during isotonic contractions of the biceps brachii. Additionally, Karlsson and Gerdle [223] explored the use of time-frequency representation metrics on the same quadriceps muscles, obtaining $R^2$ values of 0.30–0.43, and 0.91–0.93 for time-frequency measures and RMS, respectively.

5. Concluding Remarks and Future Recommendations

The primary focus of myoelectric control has historically been the development of human-computer interfaces that improve quality of life for motor-impaired populations. The use of a controlled environment in myoelectric laboratory experiments has led to reports of high performing systems; however, these systems have been found to be less robust in daily life. Accordingly, the field has sought to improve the robustness of myoelectric control under all daily-living scenarios. The primary confounding sources of variance that have been identified including the limb position effect, the contraction intensity effect, the electrode shift effect, and the within/between day effect. This hybrid survey and analysis highlighted the challenges and current solutions for the limb position and contraction intensity effects. The summary findings of this work are outlined as follows:

5.1. Limb Position Effect

- An expected 12.8% decrease on average in accuracy is expected when systems trained in a single position are tested with naturally varying limb position. The relative performance between various testing frameworks (found using the 16 static limb position data set) suggests that at most 6 positions are needed to capture the variability of limb positions across the humeral and transverse plane. Four positions are sufficient to achieve best performance in the 5 static limb position data set, which may have more closely modelled the range of positions used in everyday tasks. The impact of forearm orientation variation, however, can not be minimized using a subset of the training positions. The variance across forearm orientations is sufficiently high that the neutral, pronated, and supinated orientations all must be included in the training procedure.
to achieve best performance. The combination of these findings are unified to suggest that future myoelectric studies include repetitions in 4 limb positions and all orientations, for example using P3, P4, P9, P9s, P9p, and P14.

- The static limb position protocol serves as a strong environment to assess the robustness of novel algorithmic solutions. The usability challenge in the field, however, is the development of systems that are robust to dynamic motions that transpire while moving between positions. For this reason, a heavy focus on 2D and 3D space experimental protocols is recommended in future works.

- The review of state-of-the-art methods reveals an abundance of feature extraction techniques that minimize the positional variance and segment classes effectively, such as the TDPSD feature set. While the LDA classifier does not attain the highest accuracy in highly controlled settings, the classifier outperforms all other classifiers in the presence of the limb position effect. More meaningful improvements are demonstrated by novel classification architectures that yield significantly better performance than conventional architectures, such as SRC or cascade classification strategies. The design of novel feature extraction methods and classification strategies, however, remains a current research challenge to minimize the limb position effect.

- Transfer learning has the potential to reduce the variance across positions through algorithmic approaches such as CCA and the bilinear transform. Given the phenomenal impact transfer learning has had in other fields (e.g., object detection, deep learning), applications of unsupervised transfer learning may lead to drastic improvements in the state-of-the-art in the future.

- The incorporation of accelerometer and MMG signals into traditionally EMG only systems has lead to improvements in segmenting feature space locally according to position. Current practices mainly include single-stage position informed classifiers, or cascade classifiers that perform position and motion recognition independently. The treatment of the accompanying (non-EMG) modality, however, has been limited in most cases. Recent works suggested that the graph laplacian IMU feature provides better accuracy than an EMG feature set alone, revealing a potentially new avenue for feature extraction methods to incorporate into EMG pattern recognition pipelines.

5.2. Contraction Intensity Effect

- An expected 20.4% decrease on average in accuracy is expected when systems trained with a single intensity are tested in the context of naturally occurring contraction intensity variation or proportional control. Best performance was achieved when two intensities were included in the training set: 20% MVC and 70% MVC, or low and high subjective intensities. The best single training intensity level determined in the investigation, and across literature, is the median level across the expected range of intensities, however performance was maximized when including examples from across the expected range (statically or dynamically).

- Traditionally, isometric contractions have been used in myoelectric studies. The intensities of these contractions are commonly regulated using task-specific maximum voluntary isometric contraction normalized levels. A focus of future works should be the extension to dynamic contractions, either isokinetic or isotonic, which more accurately model natural motions. Accordingly, the dynamic-range maximum voluntary contraction normalization technique will be key in appropriately scaling intensity levels.

- The selection of an appropriate pattern recognition pipeline can drastically improve robustness to intensity variation. Nonlinear complexity features demonstrate greater robustness to contraction intensity than the traditional use of amplitude features. Feature projection techniques such as NMF and SR are used to improve resilience of features to intensity variability. Again, the LDA classifier outperformed other traditional classifiers in intensity robustness. Further improvements are found when using cascade classifiers.

- Auto-calibrated class-specific proportional control has been used in clinical applications; however, bounded proportional control appears to outperform this scheme by ensuring that users can leverage the full range of control.
The state of EMG pattern recognition robustness to confounding sources of variability has advanced drastically over the past decade. Challenges remain in striking a balance between the benefit to classification accuracy and decrease in user-acceptance caused by longer, repetitive, unengaging multi-condition training protocols. Algorithmic solutions provide opportunity for improved robustness without additional user-effort, however, novel approaches designed for one confounding source of variability could have unforeseen sensitivities to other confounding factors. The use of offline evaluation in most investigations surveyed here neglect the influence of sensory feedback, which shapes user behaviours during the use of myoelectric control. An online evaluation framework for EMG pattern recognition in the presence of confounding factors is therefore recommended in future works. The usability of myoelectric devices over long durations during activities of daily living (e.g. driving, office-work, exercise, and cooking) remains largely unaddressed and offers an exciting avenue for exploration. Despite the vast advancements summarized here, confounding sources of variability remain a current challenge for EMG pattern recognition usability in such daily living scenarios. Readers with interest in myoelectric control usability in such daily living scenarios are encouraged to consult a similarly structured future work for an extensive overview of effects of electrode shift and between/within day variations.
Table 7. Survey of experimental parameters across position effect myoelectric control experiments.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Subject Details</th>
<th>Motion Details</th>
<th>Electrode Details</th>
<th>Instrumentation Details</th>
<th>Pattern Recognition Architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEW [156]</td>
<td>N 4 23.75 ± 2.06 M 3H Dyn 3D</td>
<td>8</td>
<td>DD-arr, DT Sparse</td>
<td>2000</td>
<td>n/a</td>
</tr>
<tr>
<td>[150]</td>
<td>N 4 23.75 ± 2.06 M 3H Dyn 3D</td>
<td>8</td>
<td>DD-arr, DT Sparse</td>
<td>2000</td>
<td>n/a</td>
</tr>
<tr>
<td>[124]</td>
<td>N 8 23.75 ± 2.06 M 3H Dyn</td>
<td>8</td>
<td>DF, arr, DT Sparse</td>
<td>1000</td>
<td>16</td>
</tr>
<tr>
<td>[151]</td>
<td>N 2 8 23.75 ± 2.06 M 3H Dyn</td>
<td>4</td>
<td>DD-arr</td>
<td>512</td>
<td>n/a</td>
</tr>
<tr>
<td>[120]</td>
<td>N 5 20–45 4M 1F 6H + NM Iso P1 P2 P3 P4</td>
<td>8</td>
<td>F</td>
<td>Sparse</td>
<td>300/100</td>
</tr>
<tr>
<td>[127]</td>
<td>N 4 22–33 3M 7F 7H + NM Iso P1 P2 P3</td>
<td>4EMG</td>
<td>Sparse</td>
<td>400/150</td>
<td>TD</td>
</tr>
<tr>
<td>[157]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[158]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[139]</td>
<td>N 10 22–33 3M 7F 7H + NM Iso P1 P2 P3</td>
<td>4EMG</td>
<td>Sparse</td>
<td>150/50</td>
<td>TD</td>
</tr>
<tr>
<td>[146]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[141]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[139]</td>
<td>N 10 22–33 3M 7F 7H + NM Iso P1 P2 P3</td>
<td>4EMG</td>
<td>Sparse</td>
<td>150/50</td>
<td>TD</td>
</tr>
<tr>
<td>[146]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[141]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[139]</td>
<td>N 10 22–33 3M 7F 7H + NM Iso P1 P2 P3</td>
<td>4EMG</td>
<td>Sparse</td>
<td>150/50</td>
<td>TD</td>
</tr>
<tr>
<td>[146]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[141]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[139]</td>
<td>N 10 22–33 3M 7F 7H + NM Iso P1 P2 P3</td>
<td>4EMG</td>
<td>Sparse</td>
<td>150/50</td>
<td>TD</td>
</tr>
<tr>
<td>[146]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[141]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[139]</td>
<td>N 10 22–33 3M 7F 7H + NM Iso P1 P2 P3</td>
<td>4EMG</td>
<td>Sparse</td>
<td>150/50</td>
<td>TD</td>
</tr>
<tr>
<td>[146]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[141]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[139]</td>
<td>N 10 22–33 3M 7F 7H + NM Iso P1 P2 P3</td>
<td>4EMG</td>
<td>Sparse</td>
<td>150/50</td>
<td>TD</td>
</tr>
<tr>
<td>[146]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[141]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[139]</td>
<td>N 10 22–33 3M 7F 7H + NM Iso P1 P2 P3</td>
<td>4EMG</td>
<td>Sparse</td>
<td>150/50</td>
<td>TD</td>
</tr>
<tr>
<td>[146]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[141]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[139]</td>
<td>N 10 22–33 3M 7F 7H + NMIso P1 P2 P3</td>
<td>4EMG</td>
<td>Sparse</td>
<td>150/50</td>
<td>TD</td>
</tr>
<tr>
<td>[146]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[141]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[139]</td>
<td>N 10 22–33 3M 7F 7H + NM Iso P1 P2 P3</td>
<td>4EMG</td>
<td>Sparse</td>
<td>150/50</td>
<td>TD</td>
</tr>
<tr>
<td>[146]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[141]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[139]</td>
<td>N 10 22–33 3M 7F 7H + NM Iso P1 P2 P3</td>
<td>4EMG</td>
<td>Sparse</td>
<td>150/50</td>
<td>TD</td>
</tr>
<tr>
<td>[146]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
<tr>
<td>[141]</td>
<td>N 5 21–40 n/a 3 H2O</td>
<td>Dyn</td>
<td>16</td>
<td>DD DB DF Sparse</td>
<td>150/100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Subject Details</th>
<th>Motion Details</th>
<th>Electrode Details</th>
<th>Instrumentation Details</th>
<th>Pattern Recognition Architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. Age Sex No. Type Intensity No. Location Layout Fs Res. Gain Win Size/Inc Features Classifier</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[150]</td>
<td>2A 32,29 2M 4H Low Moderate High 12 A 2000 n/a 1000 160/40 TD + KLRT AR(4) RMS LDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[226]</td>
<td>3N 25.8 ± 0.8 3M 2F 7H + NM 30 60 90 8 DF 1000 10 0-4.5 V 128/50 TD LDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[24]</td>
<td>10N 25–50 7M 3F 6H + NM 20 30 40 50 60 70 80 ramp 8 DF 1000 16 n/a 160/16 TD LDA MV9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[221]</td>
<td>8N 23–53 7M 1F 8H Dyn 25 50 12 DF 1000 12 5000 200/50 NMF ANN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[219]</td>
<td>9N n/a n/a 6H + NM Dyn 20 30 40 50 60 70 80 n/a n/a n/a n/a 200/200 NMF LDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[152]</td>
<td>10N 20–30 M 6H Low Moderate High 6,8 DF 4000/2000 12,16 1000 150/75 TDPASD DFT WT TD DAR TDAR2 SVM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[227]</td>
<td>9N 20–30 M 8H + NM 20 50 80 8 DF Sparse 2000 n/a n/a 200/150 cnDFT TRDTR TD COMB LDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[218]</td>
<td>5N 25–26 n/a 4H + NM 20 50 80 4 DF Sparse 1000 n/a n/a 150/100 TD Parallel LDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[23]</td>
<td>12N 1A 20–33 12M 1F 4H Low Medium High 6 DF Sparse 4000/2000 12,16 1000 150/75 TDPASD DFT WT TD DAR TDAR2 SVM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[5]</td>
<td>3N 25–32 n/a 6 DF 4 Circ. 2 Sparse 1024 n/a n/a 150/100 invTDF TD4 AR NOV LDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[100]</td>
<td>8N 35 ± 15 4M 4F 4F + NM Iso 25 65 2 DB DT 4x4 Grid 2500 n/a 60 dB 150/75 MAV ZC SSL WAMP V-Order LD VAR HIST9(9) AR(9) CepC LDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[134]</td>
<td>6N 26.5 ± 3.1 M 8 DF Circ. 2000 n/a n/a 128/64 TD FD DFT DFT_TD WT_TD WPT_TD PNN LDA KNN SVM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[214]</td>
<td>8N 20–30 n/a 4EMG 4IMU DW Circ. 2000/148 n/a n/a 200/150 TD + DFTm ACC TDvACD DFT+ACC LDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[211]</td>
<td>9A 31.8 ± 10.6 7M 2F 6H Iso Low Medium High 8 A Circ. 2000 16 1000 250/125 FIS PSEAR COMB DFA MLTD TPASD PSE AR HG TD LDA SYM PNN QDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[12]</td>
<td>9A 31.8 ± 10.6 7M 2F 6H Iso Low Medium High 8 A Circ. 2000 16 1000 250/125 SEN PEN ShannonWPE LogEnergyWPE NormWPE SureWPE ThresholdWPE LDA SYM PNN QDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[217]</td>
<td>9A 31.8 ± 10.6 7M 2F 6H Iso Low Medium High 8 A DF Circ. 2000 16 1000 250 5000 250 RMS TPASD DFTW</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Materials: The following are available at http://www.mdpi.com/1424-8220/xx/1/1/s1, Figure S1: Multi-position accuracy for datasets D1-D3, Figure S2: Multi-level accuracy for datasets D4-D6, Table S1: Intra-, inter-, and single-position accuracy for datasets D1-D3, Table S2: Intra-, inter-, and single-level accuracy for datasets D4-D6.

Author Contributions: Conceptualization, E.C. A.P, E.S.; methodology, E.C. A.P, E.S.; software, E.C.; validation, E.C. A.P, E.S.; formal analysis, E.C. A.P, E.S.; investigation, E.C. A.P, E.S.; resources, E.S.; data curation, E.C.; writing–original draft preparation, E.C. A.P, E.S.; writing–review and editing, E.C. A.P, E.S.; visualization, E.C. A.P, E.S.; supervision, A.P, E.S.; project administration, A.P, E.S.; funding acquisition, E.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Natural Sciences and Engineering Research Council of Canada grant number 2014-04920, and New Brunswick Innovation Foundation.

Conflicts of Interest: The authors declare no conflict of interest.

References


72. In brief, another reference to...


188. Halaki, M.; Ginn, K. Normalization of EMG signals: To normalize or not to normalize and what to normalize to? In *Computational Intelligence in Electromyography Analysis—A Perspective on Current Applications and Future Challenges*; IntechOpen: London, UK, 2012.


© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).