Abstract—Numerous techniques have been used to minimize error in relating the surface electromyogram (EMG) to elbow joint torque. We compare the use of three techniques to further reduce error. First, most EMG-torque models only use estimates of EMG standard deviation as inputs. We studied the additional features of average waveform length, slope sign change rate and zero crossing rate. Second, multiple channels of EMG from the biceps, and separately from the triceps, have been combined to produce two low-variance model inputs. We contrasted this channel combination with using each EMG separately. Third, we previously modeled nonlinearity in the EMG-torque relationship via a polynomial. We contrasted our model vs. that of the classic exponential power law of Vredenbregt and Rau [1]. Results from 65 subjects performing constant-posture, force-varying contraction gave a “baseline” comparison error (i.e., error with none of the new techniques) of 5.5 ± 2.3% maximum flexion voluntary contraction (%MVC). Combining the techniques of multiple features with individual channels reduced error to 4.8 ± 2.2 %MVC, while combining individual channels with the power-law model reduced error to 4.7 ± 2.0 %MVC. The new techniques further reduced error from that of the baseline by ≈15%.

Index Terms—Biological system modeling, electromyogram, EMG-force, multiple-channel EMG

I. INTRODUCTION

Since at least the work of Inman et al. in 1952 [2], the surface electromyogram (EMG) has been investigated as an estimator of muscle force/joint torque. Much of the early work studied the linearity of the relation using agonist muscle EMG during constant-posture, quasi-constant force contractions (“quasi-static”) [1-5]. During the intervening years, numerous studies (see review in [6]) have expanded the experimental conditions and reduced the error in the EMG-torque relationship through various improvements, including: modeling both agonist and antagonist muscle activity [7-10], accounting for subject-to-subject differences in the relationship [4, 11], reducing variability in the estimate of EMG standard deviation (EMGσ) by whitening the EMG signal and/or (for large muscle groups) utilizing multiple-channel EMGσ–torque estimators [12-21], modeling EMG-torque dynamics [19, 22-24], incorporating a range of joint angles [25-29], and applying robust system identification methods [11, 19, 24, 30, 31]. The various techniques are relevant in several areas in which a noninvasive EMG-torque estimate is useful, such as prosthesis control [32, 33], clinical biomechanics [34, 35] and ergonomics assessment [36, 37].

In a related problem in EMG-based prosthetics control, multiple EMG features have been used as inputs to the task of classifying distinct movement classes. In particular, Hudgins et al. [38] (see [39] for a review) added to EMGσ the features of slope sign change rate (SSC), zero crossing rate (ZC) and average waveform length (WL). Only recently has the success of these “additional” features in the EMG classification problem led to their investigation as model inputs in the EMG-torque problem [40-45].

In this study, we report on three techniques for continuing performance improvement in the EMG-torque relationship. First, most past studies using dynamic models of EMG-torque have exclusively utilized EMGσ as the input EMG feature. Thus, we look at the applicability of adding the additional features of Hudgins et al., comparing models with and without their inclusion. Second, for large muscles, EMGσ variability has been reduced by combining the information from multiple electrodes into one EMGσ estimate [12-14, 16]. This method of channel combination is optimal assuming that the underlying muscle contains the same information across its multiple electrode locations, varying only due to the stochastic nature of motor unit firing times. However, there is evidence that large muscles—this research studies the biceps and triceps muscles—contain some degree of control based on neuromuscular compartments [46-48]. As such, combining EMG sites to produce a feature estimate would no longer be justified. Thus, we contrast combining EMG sites to estimate a feature vs. extracting features from each individual electrode. Third, our own dynamic EMG-torque models have incorporated the static power-law nonlinearity described by Vredenbregt and Rau [1] via the use of a polynomial relation [29, 31]. This method simplifies the mathematics, allowing the use of linear least squares estimation, but can require many parameters—which can have its own detrimental effects [49]. Other authors have captured a nonlinear relationship with other model forms, e.g., parallel-cascade models [24] and neural networks [40, 42, 44, 45, 50]. Therefore, we directly compared use of the power-law nonlinearity of [1] (requiring parameter estimation via nonlinear least squares) to that of the polynomial model. Finally, we examined if combining pairs of...
these various improvement techniques provides an additive benefit. We also varied the dynamic model order (i.e., number of time lags) and the tolerance value associated with the Moore-Penrose inverse method used to linear least squares fit model parameters. These parameters influence EMG-force errors [31] and thus should be optimized within each of the three primary techniques studied in this work.

II. METHODS

A. Experimental Subjects, Apparatus and Methods

Experimental data from 54 subjects acquired during three prior experiments [29, 51, 52] were combined with the data from 11 new subjects to form a pool of 65 total subjects. The new data collection and all analysis was approved and supervised by the WPI Institutional Review Board. Each of the 65 subjects provided written informed consent for their respective experiment. For the new data collection (similar methods were used in the prior experiments), subjects were seated and strapped via three belts into the custom-built straight-back chair shown in Fig. 1, with their right shoulder abducted 90°, the angle between their upper arm and forearm 90°, their forearm oriented in a parasagittal plane, and their supinated right wrist (palm perpendicular to the floor) tightly cuffed to a load cell (Vishay Tedea-Huntleigh Model 1042, 75 kg full scale). Skin above the biceps and triceps muscles was vigorously scrubbed with an alcohol wipe and a bead of electrode gel was massaged into the overlying skin. Four custom-built bipolar EMG electrode-amplifiers were applied in a transverse row across each of the biceps and triceps muscle groups, midway between the elbow and the midpoint of the upper arm (to avoid the innervation zone proximally and the tendon distally), centered along the muscle mid-line. Each electrode-amplifier had a pair of 8 mm diameter, stainless steel, hemispherical contacts separated 1 cm edge-to-edge, oriented along the muscle’s long axis. The distance between adjacent electrode-amplifiers was ~1.75 cm. A reference electrode was gelled and secured to the lateral aspect of the upper arm, between the flexion and extension electrodes. All electrodes were secured in place on the right arm with elastic bandages. Custom electronics amplified and filtered each EMG signal (CMRR > 90 dB at 60 Hz; 8th-order Butterworth highpass at 15 Hz; 4th-order Butterworth lowpass at 1800 Hz) before being sampled at 4096 Hz with 16-bit resolution. The RMS EMG signal level at rest (representing equipment noise plus ambient physiological activity) was on average 5.0 ± 7.3% of the RMS EMG at 50% maximum voluntary contraction (MVC) for these 11 new subjects. The load cell (torque) signal was synchronously sampled at 4096 Hz with 16-bit resolution.

All contractions were constant-posture. Subjects were provided a warm-up period. Separate extension and flexion MVCs were then measured in which subjects took 2–3 seconds to slowly ramp up to MVC and maintained that force for two seconds. The average load cell value during the contraction plateau was taken as the MVC. Five second duration, constant-force contractions at 50% MVC extension, 50% MVC flexion and at rest (arm removed from the wrist cuff) were next recorded. These contractions were used to calibrate whitening filters and to gain-normalize the EMG and force data [52, 53], as further described below. Then, three tracking trials of 30 s duration were recorded during which the subjects used the load cell as a feedback signal to track a computer-generated torque target. The target moved on the screen in the pattern of a bandlimited (1 Hz) uniform random process, spanning 50% MVC extension to 50% MVC flexion. Three minutes of rest were provided between trials to avoid cumulative fatigue.

Fig. 1. Subject seated in the experimental apparatus with right arm cuffed at the wrist to the load cell and electrodes applied over the biceps and triceps muscles. Inset shows six electrodes positioned transversely across the biceps muscles (with securing bandage removed for visualization). Middle four electrodes used for the analysis reported herein. Triceps electrodes were arranged similarly.

B. Methods of Analysis

Analysis was performed offline in MATLAB. All EMG filters were designed as specified below, then applied in the forward and reverse time directions to achieve zero phase and the square of the magnitude response. Each EMG channel was powerline notch filtered (2nd-order IIR notch filter at the fundamental and each harmonic, bandwidth ≤ 1.5 Hz), since whitening at high frequencies is particularly susceptible to powerline interference. These filters attenuate powerline noise with limited reduction in signal statistical bandwidth [54]. Each EMG channel was then highpass filtered to reject motion artifacts (5th-order Butterworth, 15 Hz cutoff) and whitened using the adaptive whitening algorithm of [52] and [53]. Features were next extracted from each of the eight whitened EMG signals. EMGr[n] was formed by rectifying each signal and WL[n] was computed as the absolute difference between adjacent samples [38, 39, 55], where n was the discrete-time sample index at the sampling rate of 4096 Hz. ZC[n] and SSC[n] [38, 39, 55] were formed by assigning a value of one to each sample corresponding to a thresholded zero crossing/slope sign change, and a value of zero otherwise. For each electrode, the noise threshold used for ZC and SSC was...
3% of the RMS of a rest contraction. Two different EMG channel selection options were studied: (1) features from the four biceps and, separately, triceps channels were each ensemble averaged to form four-channel feature estimates (for EMGσ and WL, the channels were gain normalized prior to doing so [16]), or (2) features from each of the eight individual EMGs were retained separately. EMG features and the torque measurement were next lowpass filtered at 16 Hz, then downsampled to 40.96 Hz. This rate is fast enough to capture the system dynamics while also eliminating high-frequency noise outside of the torque signal band that can confound the ensuing system identification [30, 49]. Note that the lowpass filter stage prevents aliasing when downsampling, while simultaneously smoothing (averaging) the EMG features. Further smoothing is inherently customized to each subject, provided by the dynamic models (described in the next paragraph). Hence, the dynamic models optimize the final lowpass cutoff frequency (and shape) in order to minimize EMG-torque error [56].

The decimated extension and flexion EMG features from each subject (inputs) were related to their respective elbow torque (output) via one of two dynamic models. The first “quadratic” model incorporated a second degree polynomial (based on prior optimization of a subset of these data [31], and consistent with the nonlinearity in the EMG-force curve at the elbow [1]):

\[
T[m] = \sum_{q=0}^{Q} \sum_{f=1}^{F} \sum_{e=1}^{E} c_{q,f,e} V_{f,e} [m-q] + c_{2,q,f,e} V_{f,e}^2 [m-q],
\]

where \(m\) was the decimated discrete-time sample index, \(T[m]\) the measured torque, \(Q\) the number of time lags in the model (to provide dynamics), \(F\) the number of EMG features included (EMGσ was always included; optionally either one or all of the remaining three features was included), \(E\) the number of EMG channels (\(E=2\) was used when the four biceps and four triceps channels were combined into two channels; \(E=8\) was used when eight individual channels were retained), \(c_{q}\) were the fit coefficients, and \(V[i]\) were the EMG feature values. The fit coefficients were estimated using regularized (Moore-Penrose inverse) linear least squares, in which singular values of the design matrix were discarded if their ratio to the largest singular value was less than a selected tolerance value (\(Tol\)) [31, 57]. Thus, the EMG features, and their squared values, were least squares fit to torque.

The second “power-law” model was:

\[
T[m] = \sum_{q=0}^{Q} \sum_{f=1}^{F} \sum_{e=1}^{E} c_{q,f,e} V_{f,e}^{r_f} [m-q],
\]

where \(r\) was also a fit parameter equal to a continuous-valued exponent applied to the feature value. This exponent directly implemented the EMG-force nonlinearity of Vredenbregt and Rau [1]. The fit coefficients \((c_i, r)\) were fit using nonlinear least squares. Anecdotally, initial solution guesses for \(r\) of 0.5, 1 and 2 were evaluated, with the \(c_i\) coefficients then initialized via linear least squares (using a pseudo-inverse tolerance of 0.005). Only the \(r = 1\) value converged rapidly for most subjects. When each of these three \(r\) values led to convergence, they arrived at the same minimum solution. Thus, \(r = 1\) was used as the initial guess value. This initial guess value happens to be the optimal linear solution.

Of the three available tracking trials, two were used for training and one for testing. Since the nonlinear minimizations were time-intensive and the sample size was already quite large for an EMG-torque study (65 subjects), cross-validation was not used. Error is reported as the test set RMS error between the actual and EMG-estimated torque, normalized to maximum flexion torque for each subject. The first and last 2 s of data were excluded to account for filter startup and tail transients. 1. We investigated combinations of: model orders between \(Q=5\) to 40, five distinct EMG feature selections (EMGσ only, EMGσ paired with each of the other three features and all four features), two EMG channel selections (a four-channel biceps EMGσ with a four-channel triceps EMGσ, or retaining all eight individual electrodes), two models (quadratic and power-law), and various pseudo-inverse linear least squares tolerance values (starting at \(Tol=0.1\) and decrementing by 0.002 to \(10^{-3}\), and \(10^{-4}\) and \(10^{-5}\)). Note that we did not investigate every combination of model order and pseudo-inverse tolerance vs. the other parameters, since doing so would have been prohibitively time-consuming and the influence of model order and tolerance has already been characterized in prior work [31]. Rather, tolerance was fixed at \(Tol=0.005\) while model order was varied; and model order was fixed at \(Q=15\) while tolerance was varied.

Finally, for comparison to conventional EMG-torque models, we also investigated cascade of a fixed, second-order Butterworth filter (cut-off frequency of 1.5 Hz, as optimized for a subset of these data in prior work [56]) after each of the extension and flexion EMGσ signals, as derived from single-channel unwhitened EMG (selecting one of the central electrodes on each muscle). The gains of both filters (i.e., the fit coefficients for the Butterworth model) were calibrated from the test data using linear least squares (\(Tol = 0.005\), two training trials, one test trial).

Statistical evaluation used multivariate ANOVA (significance level of \(p = 0.05\), with post hoc pair-wise comparisons conducted using Tukey’s honestly significant difference test (which adjusts for multiple comparisons).

III. RESULTS

Our strategy was to individually compare the three study techniques of EMG features, EMG channel combination and model vs. our “baseline” best prior technique [31] comprised of the EMGσ feature only, four-channel EMG processors and the quadratic polynomial model. As appropriate, we also assessed performance as a function of dynamic model order (\(Q\)) and pseudo-inverse tolerance (\(Tol\)). Then, we assessed

1 In real-time applications, all processing would be conducted using causal filters, eliminating the need to exclude any tail transients. (They would not exist.) The startup transient would occur at device power-up and thus not interfere with regular device operation.
improvement (beyond that found due to one study technique) when pairs of study techniques were combined. We do not report results from all three study techniques combined, since the nonlinear minimization frequency failed to converge—presumably due to the large number of features encountered when using all (five) features and eight individual EMG channels. Note that for several analyses, model order was fixed at $Q=15$ and the pseudo-inverse tolerance was fixed at $Tol=0.005$. These fixed values were determined based on prior analysis of a portion of these data [31] (and are consistent with our results reported herein). Example time-series EMG-torque estimates are shown in Fig. 2. For comparison, the conventional Butterworth model had average ± std. dev. RMS error of $8.9 \pm 3.0 \%\text{MVC}_F$.

At $Q=15$, the baseline technique error (mean ± std.) was $5.5 \pm 2.3 \%\text{MVC}_F$. Fig. 3, bottom, shows results vs. pseudo-inverse tolerance ($Tol$) with model order fixed at $Q=15$. Error reduced rapidly as tolerance decreased, and the full feature set showed the lowest error. To avoid the interactions at the larger tolerance values, a two-way ANOVA (Factors: tolerance, feature set) omitted tolerance values above 0.011. This comparison was only significant for the main effect of feature set [$F(4, 40)=3.1, p=0.011$], without interaction. Post hoc Tukey comparisons only found differences between the EMGσ-only feature vs. the full feature set. Overall, the full feature set generally produced lower errors.

**EMG Channel Selection:** Next, we compared results between the baseline technique vs. individual EMG channels. Fig. 4, top, shows results vs. model order $Q$ ($Tol$ fixed at 0.005). Error reduced as model order initially increased and the individual EMG channels had lower error. A two-way ANOVA (Factors: model order, EMG channel selection) was significant for both main effects [$F(35, 35)=7.4, p=10^{-6}$ for model order; $F(1, 35)=96, p=10^{-6}$ for channel selection], without interaction. Post hoc Tukey evaluation of model order found that lower orders had higher errors than the highest orders for orders $Q=5$ through 7. Results for model orders 8–40 did not differ. Post hoc Tukey evaluation of EMG channel selection found individual EMG channels to have lower error. Fig. 4, bottom, shows results vs. tolerance ($Q$ fixed at 15). For consistency, a two-way ANOVA (Factors: tolerance, EMG channel selection) omitted tolerance values above 0.011. This comparison was only significant for the main effect of EMG channel selection [$F(1, 10)=34, p=10^{-6}$; no interaction], with post hoc Tukey evaluation finding individual EMG channels to have lower error. Overall, using eight separate channels—as opposed to extension/flexion four-channel processors—consistently led to lower error.

**Power-Law Model:** Then, we compared results between the baseline technique and the power-law model. Fig. 5 shows these results vs. model order ($Q$), with $Tol=0.005$ selected for the quadratic model. Tolerance was not examined as a separate factor, as it is not varied with the power-law model. Error using the power-law model was consistently lower than that of the baseline model. A two-way ANOVA (Factors: model order, model type) was significant for both main effects [$F(35, 35)=9.4, p=10^{-6}$ for model order; $F(1, 35)=33, p=10^{-6}$ for model type], without interaction. Post hoc Tukey evaluation of model order found that lower orders exhibited higher errors than the highest orders for orders $Q=5$ through 7. Results for model orders 8–40 did not differ. Post hoc Tukey evaluation of model form found lower errors with the power-law model. Overall, the power-law model produced lower errors.

---

2 Note that a three-way ANOVA with factors model order, tolerance and feature set was not pursued since results from all combinations of model order and tolerance were not computed (see Methods). Instead, model order and tolerance were analyzed in separate two-way ANOVAs (here and also below).
Fig. 3. Baseline Model vs. Feature Set: Average RMS errors from 65 subjects, four-channel EMG. Legend refers to both plots. Single-sided standard error bars shown for two of five feature sets (standard errors were similar for the other three feature sets) for selected $Q$ values. Top: Results vs. quadratic model order ($Q$), using pseudo-inverse tolerance of $Tol=0.005$. Bottom: Results vs. pseudo-inverse tolerance, with quadratic model order $Q=15$.

Fig. 4. Baseline Model vs. EMG Channel Selection: Average RMS errors from 65 subjects, EMG$\sigma$-only feature set. Single-sided standard error bars shown for selected $Q$ values. Top: Results vs. quadratic model order ($Q$), using pseudo-inverse tolerance of $Tol=0.005$. Bottom: Results vs. pseudo-inverse tolerance, with quadratic model order $Q=15$.

Fig. 5. Baseline Model vs. Power-Law Model: Average RMS errors from 65 subjects, EMG$\sigma$-only feature set. Results vs. model order ($Q$). Quadratic model used pseudo-inverse tolerance of $Tol=0.005$. Single-sided standard error bars shown for selected $Q$ values.
**B. One Improvement Technique vs. Two**

We concluded from the above results that each of the three techniques improved EMG-torque performance individually. Thus we next evaluated pairs of techniques, comparing each pair to the individual improvements. For EMG feature sets, we only retained two options, EMGσ only and all features. The results above showed that the other feature set options had performance that fell between these two. Also, we eliminated the reporting of post hoc statistical evaluation for model order and tolerance, as their roles were well established by the results above and prior literature results [31]. Doing so placed our focus on the three improvement techniques.

**EMG Feature Set & EMG Channel Selection:** Above, Fig. 3 showed the error improvements from the baseline technique due to EMG feature set and Fig. 4 to EMG channel selection. Here, Fig. 6 repeats both of these individual results, then adds the results when these techniques are combined (quadratic model with all features and eight individual EMG channels). For the results shown in Fig. 6, top, a two-way ANOVA (Factors: model order, the three techniques) was significant for both main effects [F(35, 70)=8.7, \(p=10^{-6}\) for model order; \(F(2, 70)=23, p=10^{-6}\) for technique], without interaction. Post hoc Tukey evaluation of technique found that using all EMG features with four-channel EMG had higher error than the other two techniques (EMGσ only, eight individual channels; all features, eight individual channels). At \(Q=15\), the technique with all features and eight individual EMG channels had an error mean ± std. of 4.8 ± 2.2 %MVC\(_F\). The error results shown in Fig. 6, bottom, are high for all techniques for large tolerance values and climb when using all features and eight individual channels for tolerances ≤10\(^{-3}\). A two-way ANOVA (Factors: tolerance values ≤0.011, the three techniques) was only significant for technique [F(2, 20)=4.2, \(p=0.02\); no interaction]. Post hoc Tukey evaluation of technique found that using EMGσ only with eight individual channels exhibited lower error than the other two techniques. Nonetheless, Fig. 6 shows similar performance specifically in the region of the optimum tolerance value (e.g., \(Tol = 0.005\)), when comparing the technique of EMGσ only with eight individual channels to the technique of all features with eight individual channels.

**EMG Feature Set & Model Form:** Above, Fig. 3 showed the error improvements due to EMG feature set and Fig. 5 to model form. Here, Fig. 7 repeats both of these individual results, then adds the results when these techniques are combined (four-channel EMG with all features and the power-law model). A two-way ANOVA (Factors: model order, the three techniques) was significant for the main effect of model order [F(35, 70)=13, \(p=10^{-6}\) ], but not significant for the main effect of technique [F(2, 70)=2.8, \(p=0.06\)], without interaction. Thus, this paired set of improvements did not reduce error beyond that found from each individual technique. At \(Q=15\), each of the three techniques had an error mean ± std. of approximately 5.1 ± 2.1 %MVC\(_F\).

---

**Fig. 6.** Quadratic Model: All Features vs. Eight Individual EMG vs. Both. Average RMS errors from 65 subjects. Legend refers to both plots. Single-sided standard error bars shown for two of three feature sets (standard errors were similar for the third feature set) for selected \(Q\) values. Top: Results vs. quadratic model order \(Q\), using pseudo-inverse tolerance of \(Tol=0.005\). Bottom: Results vs. pseudo-inverse tolerance, with quadratic model order \(Q=15\).
This study evaluated three techniques to reduce error in the EMG-torque relationship about the elbow—additional EMG features, EMG channel selection and EMG-force model form. Figs. 3–5 (and their associated statistical analyses) show that each of these techniques individually improved upon a “baseline” model that used only the EMGσ feature, four-channel EMG for each of the biceps and triceps, and the quadratic polynomial model. Note that this baseline model already optimizes several processing steps, including using EMG signal whitening, selecting the degree of the polynomial model and selecting the pseudo-inverse tolerance [15, 16, 31]. Whitening has previously been shown to reduce the variance of EMGσ estimates [13-16], e.g. providing an ≈63% improvement in SNR for constant-posture, constant-force elbow contractions when using a 245 ms smoothing window [15]. EMG whitening leads to significant performance improvements in EMG-torque estimation [17, 31], e.g. a 14.1% reduction in RMS error during constant-posture, repetitive elbow exertions [17]. Whitening has also been shown to reduce the variance of WL and (to some extent) ZC estimates [55] (leading to performance improvements in multifunction prosthesis control [55]). The variance reduction is attributed to an increase in signal statistical bandwidth provided by whitening [13, 55]. Therefore, we would expect similar variance reduction in whitened estimates of the SSC feature.

Of the three techniques, Figs. 6 and 8 show that using eight individual EMGs (as opposed to a four-channel EMG for each of the biceps and triceps) provides the clearest advantage. The concept of combining the information from multiple electrodes sited over a large muscle assumes that the spatially diverse information represents different statistical samples of the same underlying stochastic process [12, 13]. The elbow contractions used herein were constrained to a single plane, reinforcing this assumption. Certainly, prior work has shown that four-channel EMG over the biceps and triceps leads to lower EMG-torque error than if only one biceps and one triceps EMG were used [12, 13, 30, 31], attributed largely to lower EMGσ variance [12, 13, 15-17]. However, our current results show that further error reduction is realized if the multiple EMG channels are used as separate inputs to the system identification model. Several concepts could explain this further improvement, all challenging the assumption that each electrode is stochastically sampling the same distribution. First, the individual electrodes could be sampling from distinct spatial regions with distinct neuromuscular control (i.e., neuromuscular compartments [46-48]). Second, we have anecdotaly noticed that electrodes placed further from the muscle midline are more prone to crosstalk from the antagonist muscles, and that the EMG from such electrodes leads to a poorer EMG-torque estimate. The use of individual electrodes would permit the system identification model to de-emphasize those EMG channels that contribute less to reducing the EMG-torque error. Third, the quality of the EMG signal (e.g., signal to noise ratio) can vary electrode-to-electrode. When used as individual channels, the system identification model can de-emphasize the noisy electrodes; but when combined into a four-channel EMG, the emphasis of individual channels is purposely equalized [12, 16]. Future research should examine which EMG channels are more heavily weighted in these identified models.

Because the decrease in EMG-force error due to eight individual EMG channels was robust to model form, it may be especially applicable to other nonlinear models used in the literature, such as parallel-cascade models [24] and neural networks [40, 42, 44, 45, 50]. All model forms, however, become increasingly ill-conditioned as more fit parameters are added, the relative importance of which may vary model form to model form.

Figs. 6–8 show that the remaining two improvement
techniques (EMG feature sets and EMG-force model form) each provide approximately the same error reduction—Fig. 6, top, shows that all features, eight individual EMG channels and the quadratic model ($Q=15, \text{Tol}=0.005$) had an error mean ± std. of 4.8 ± 2.2 %MVC$_F$; while Fig. 8 shows that the EMGσ-only feature, eight individual EMG channels and the power-law model ($Q=15$) had an error mean ± std. of 4.7 ± 2.0 %MVC$_F$. These error performances represent an ≈15% reduction in error compared to the baseline model error of 5.5 ± 2.3 %MVC$_C$. The power-law model has the advantage of fulfilling the static EMG-force nonlinearity found by Vredenbregt and Rau [1] with a single exponential parameter per EMG channel, but the disadvantage of requiring significantly more computation time for determining fit coefficients via nonlinear least squares. A concern with using additional EMG features is their effect on the conditioning of the linear/nonlinear least squares fit, since conditioning is inversely related to the number of fit coefficients [49]. In particular, Fig. 6, bottom, shows error increasing for tolerances below $10^{-3}$ when all four features for each of eight individual channels are fit using the quadratic model (64 coefficients in total). Tolerances below $10^{-3}$ provide progressively less regularization, the opposite of what is needed when the number of fit coefficients grows. Hence, model error grows, likely due to overfitting. As a result, the range of tolerance values over which error is minimum shrinks, making the modeling less reliable.

Considering systematic errors in the EMG-torque techniques, the use of multiple features expands the model shapes that can be fit (i.e., beyond the shapes that can be accommodated when only using EMGσ as an input). The performance of EMG-torque models also suffer from random errors due to the stochastic nature of EMG. The uncorrelated components of the four EMG features would tend to average and reduce variance errors. (E.g., when one feature value is randomly above its “true” value, another feature might be below.) Hence, both systematic and stochastic improvements can result. Of course, a challenge is to improve EMG-torque performance due to these advantages, in spite of the detrimental effects of overfitting (due to the increased number of parameters) and feature correlation (which, combined with overfitting, degrade the conditioning of the least squares fit). Future modeling might consider a compromise approach that only utilizes a sub-set of the additional features.

For the quadratic polynomial model, we focused our attention on a model order of $Q=15$ and a tolerance of $\text{Tol}=0.005$. Our ANOVA results showed statistical differences (reductions) in error as model order increased from $Q=5$ to $Q=7$ or 8 (depending on the condition). Nonetheless, all of our graphical results show continuing decline in error up to about $Q=15$. Although we had a large sample size of 65 subjects, it is likely that statistical power limited our ability to find statistical differences for orders above 8. In particular, paired statistical tests can be more powerful when assessing different treatments (i.e., EMG-force techniques) applied to the same data. For example, consider the technique that demonstrated the lowest average error: EMGσ-only feature, eight individual channels and the power-law model (Fig. 8). If we successively compute paired sign tests [58] between adjacent model orders at/above $Q=8$, we find statistical differences ($p<0.01$) until comparing orders $Q=12$ to 13. Such comparisons support our choice of $Q=15$ (and are more fully detailed in [31]). A similar argument supports our use of $\text{Tol}=0.005$.

Within the literature, it is difficult to directly compare EMG-torque results between studies, since error is a function of many variables, including the experimental conditions (e.g., constant-posture vs. freely moving) and experimental tasks (e.g., random, broadband torques vs. sinusoidal). Further, several different error measures are used within the literature. However, relative changes in performance in the same dataset, evaluated with the same error measure, should be more robust when compared. To that end, we have studied sub-portions of this data set in several published studies. The highest error of 19.2 ± 11.2 %MVC$_F$ was found when supplying single-channel, unwhitened EMGσ to a simple second-order Butterworth model, calibrated from 50% constant-force contractions [31]. Our results herein reduced the error to 8.9 ± 3.0 %MVC$_F$ when the single-channel, unwhitened EMGσ supplied to the Butterworth model was calibrated from two dynamic contractions. This error was reduced to 5.5 ± 2.3 %MVC$_F$ with our “baseline” method that used four-channel, whitened EMG and a quadratic nonlinearity (and FIR linear model). Finally, error was reduced to 4.7 ± 2.0 %MVC$_F$ (the primarily work reported herein) by substituting individual EMG channels (rather than grouping them, separately, from the biceps and triceps muscle groups) and the power-law model (or reduced to 4.8 ± 2.2 %MVC$_F$ by substituting individual EMG channels and all features). Thus, dramatic reduction in EMG-torque error has been achieved overall. In many applications in clinical biomechanics and ergonomics assessment, electrodes would be mounted on a subject, calibration data recorded and then a clinical/experimental task completed in a single session. Since appropriate EMG-torque calibration data is required for these scenarios and computation is readily available, all of the performance gains realized by these modeling techniques could be utilized. For prosthesis control, however, there is some evidence that improved off-line classification results do not always translate into improved on-line classification performance when assessed on standard prosthesis tasks [59, 60]. Although our research involved EMG-torque estimation and not EMG-based classification, similar concerns exist [61].

We limited this work to constant-posture contractions in order to reduce the complexity of a problem that already considers many modeling variables. In so doing, our results are directly relevant to prosthesis control when EMG is observed over remnant muscles whose posture is constrained (e.g., secured at both ends to the same bone), and in clinical/ergonomic assessments in which such postural constraint is appropriate. But, when joint angle is varied, additional study will be necessary. That said, the reduction in RMS error can be thought of as a reduction of two error components: a variance error and a bias error. Those
processing techniques that generally reduce variance (e.g., whitening, averaging due to multiple EMG channels, averaging due to multiple EMG features) should reduce EMG-torque error regardless of the experimental conditions. Techniques that reduce bias would likely need to be substituted with appropriate posture-varying models. The relative magnitude of variance vs. bias error can also change in posture-varying contractions. Nonetheless, our results should be informative to future studies of the reduction of both components of the RMS error in posture-varying contractions.

V. CONCLUSION

Our baseline technique for relating EMG to torque—EMGσ feature only, four-channel EMG from each of the biceps and triceps and a dynamic, quadratic nonlinear model—produced an error mean ± std. on this dataset of 5.5 ± 2.3 %MVC̄.

This baseline technique already includes several technique optimizations, including EMG signal whitening, multi-site EMG and the use of the quadratic nonlinearity [31]. Three technique improvements were individually applied. These improvements were: additional EMG features, the use of eight individual EMG channels and a power-law model. Each technique individually lowered EMG-torque fit error. Combining the techniques of additional features and individual channels reduced error to 4.8 ± 2.2 %MVC̄, while combining individual channels with the power-law model reduced error to 4.7 ± 2.0 %MVC̄. These error performances represent an ~15% reduction in error compared to the baseline model. Hence, these combined techniques represent a substantial improvement in performance. These results should be informative to application areas, including prosthesis control, clinical biomechanics and ergonomics assessment.

References


Edward (Ted) A. Clancy (S’83–M’91–SM’98) received the B.S. degree from Worcester Polytechnic Institute (WPI), and the S.M. and Ph.D. degrees from Massachusetts Institute of Technology (MIT), all in Electrical Engineering.

He has worked in industry for medical instrumentation and analysis companies interested in EMG, EEG, ECG and blood pressure, and the defense industry (aircraft instruments and radar). He is Professor of Electrical and Computer Engineering, and of Biomedical Engineering at WPI. He is interested in signal processing, stochastic estimation and system identification, particularly as applied to problems in medical engineering and human rehabilitation.