用于人体手臂活动表面肌电信号的拉普拉斯高斯混合模型

Durgesh Kusuru，IEEE 学生会员；Anish C. Turlapaty，IEEE 会员；Mainak Thakur，IEEE 会员

arXiv:2301.01080v1 [eess.SP] 3 Jan 2023

摘要——表面肌电信号的概率密度函数（概率密度函数）符合以下单一标准分布之一：高斯分布或拉普拉斯分布。此外，模型的选择取决于肌肉收缩力（MCF）水平。因此，本文提出一种统一模型，用于描述不同 MCF 水平下表面肌电信号的统计特性。本文提出用于上肢记录信号的拉普拉斯高斯混合（LGM）模型。该模型能够解释不同活动中对应不同 MCF 水平的表面肌电信号。我们在不同的基准表面肌电数据集上测试了该模型，并从定性和定量两个角度对其进行了验证。结果表明，对于低和中等收缩力水平，所提出的混合模型比拉普拉斯模型和高斯模型都更准确；而在高收缩力水平下，LGM模型的表现与高斯模型相同。我们分析了LGM模型的混合权重，发现低和中等 MCF 水平下，LGM模型的两个混合权重均有贡献；而在高 MCF 水平下，拉普拉斯权重会减弱。所提出的上肢表面肌电信号LGM模型能够解释不同 MCF 水平下的表面肌电信号。该模型有助于加深对表面肌电信号统计特性的理解，并改善分类问题中的特征表示。

索引词——表面肌电图（表面肌电）、统计模型、概率密度函数（概率密度函数）、混合模型、肌肉收缩力、参数估计、期望最大化算法。

I. 引言

A. 背景

表面肌电图（sEMG）信号建模有多种应用，例如：1）深入了解由构成运动单位动作电位（MUAPs）生成表面肌电信号的过程，为表面肌电信号合成 [1] 和仿真研究 [2], 2奠定基础；2）改进临床环境中表面肌电信号的解读，例如用于诊断神经肌肉疾病 [3], 3；3）分析表面肌电信号与源肌群之间的相互关系，例如用于运动科学研究 [4], [5], [6], ；以及 4）构建可视化工具，支持运动科学 [7], 中的肌肉生理学检查和运动科学教育。表面肌电信号模型可依据以下原理分类：1）生物电原理；2）统计原理；3）机器学习原理。早期模型基于肌纤维和运动单位的生理特征及其电活动例如，在 [8] 和[9], 中，表面肌电信号被表示为运动单位动作电位的线性组合，其中动作电位被建模为电流三极子，从神经肌肉接头传播至肌纤维-肌腱末端。在 [10], 中，使用多尺度生理肌肉模型，根据自主运动对应的表面肌电信号估算肌肉力。

在统计方法中，表面肌电信号被视为随机信号，通常建模的特征包括信号强度（样本）、信号的时间演化、单个通道的自相关，以及多个通道之间的空间互相关。正如下一节所述，过去几十年来，表面肌电信号强度的概率模型已有显著发展。在时间模型中，表面肌电信号通常表示为线性自回归过程 [11] [12]。为估计运动单位动作电位，等长收缩时采集的表面肌电信号被建模为线性时不变系统的输出，其输入为非高斯白噪声 [13]。在基于方差的模型中，表面肌电信号被视为复合随机过程。例如，在尺度混合模型 [14], 中，信号强度被建模为在给定方差条件下的高斯过程，而该方差则被建模为逆伽马变量。

表面肌电信号的模式分类在矫形外骨骼控制[15], 、人体运动分析 [16], 以及神经肌肉疾病诊断等应用中发挥着关键作用。例如，这些信号可以提供运动控制参数等合适的输入，以驱动肢体外骨骼。在机器学习方法中，可以基于表面肌电信号的概率密度函数（概率密度函数） [17]提取合适的特征。在人体运动分析中，表面肌电信号可用于区分不同动作，例如手势与物体抓取[18]。在神经肌肉疾病诊断中，这些信号可用于研究肌病等病症。肌病与骨骼肌有关，会导致肌肉变弱，并引发肌肉疼痛、无力、疲劳等症状 [19]。解码表面肌电信号中包含的信息至关重要，需要可靠且精确的解决方案。在人机交互应用中，深度学习方法发挥着关键作用，可用于提升运动分类、关节角度预测和力/力矩估计等任务的性能 [20]–[22]。本文重点研究表面肌电信号强度的统计建模。

B. 表面肌电强度概率密度函数的现有模型

表面肌电信号统计信号模型的典型应用

本研究由印度政府下属的 SERB 通过项目资助编号 CRG/2019/003801 资助。作者来自印度安得拉邦奇图尔 517646 的印度信息技术学院斯里城校区生物信号分析组。（电子邮箱：durgesh.k@iiits.in、anish.turlapaty@iiits.in 和 mainak.thakur@iiits.in

包括：1）更深入地理解表面肌电信号的统计特性；2）改进分类问题中的特征表示；3）对信号进行定性分析。现有的表面肌电信号强度模型会根据肌肉收缩水平和肌肉类型，采用高斯或拉普拉斯概率密度函数等单一标准分布。以下根据自 $1970s$ 以来的研究，总结了现有的人体上肢不同肌群表面肌电信号模型。

1974年，Roesler[23] 进行了表面肌电测量，并提出在恒定力测量条件下，表面肌电信号服从高斯分布。Miler-Brown 等人 [24] 观察到，在较低力水平下，从第一背侧骨间肌（FDI，手背侧）记录的表面肌电信号，其分布在零值附近的峰比高斯分布更尖锐；随着力水平增加，零值附近的尖锐程度会减弱。在[25], 中，观察到从肱二头肌采集的表面肌电信号在低、中 MCF 水平下服从高斯分布。Hunter 等人 [26] 将恒定 MCF 下肱二头肌表面肌电信号的密度与高斯密度进行了比较，发现其在零值附近有一个窄峰。随后，Bilodeau 等人 [27] 观察到，在较低 MCF 水平下，肱二头肌表面肌电信号具有非高斯特性，且在零值附近有峰；在较高 MCF 水平下，其分布则趋向于高斯模型。Clancy 和 Hogan [28] 通过实验发现，恒定 MCF 下表面肌电信号的密度介于高斯概率密度函数和拉普拉斯概率密度函数之间。在 [29], 中发现，表面肌电信号的概率密度函数：1）在低、高 MCF 水平下，零值附近的峰比通常的高斯分布更尖锐，且尾部更长；2）在中等 MCF 水平下服从高斯模型。在 [30], 中发现，高 MCF 水平下的表面肌电信号分布为高斯分布。根据近期研究，较高 MCF 水平下，指浅屈肌 [31], [32] 和肱二头肌 [33], 的表面肌电信号服从高斯模型。综述这些研究可知，目前尚无一种独特的统计模型能够解释不同收缩力水平下的活动。在许多情况下，可能无法使用标准的单一密度模型来描述数据。此时，将数据建模为密度混合通常是适当的方法。贡献

- 针对表面肌电信号提出了一种统一混合模型，用于解释不同肌肉收缩力水平下信号的统计特性。
- 在多个基准 sEMG 数据集上对所提出的模型进行了测试，并采用定性和定量方法将其适用性与现有模型进行了比较。
- 分析了不同活动和强度下混合成分的权重，并展示了它们之间可能存在的相互关系。

II. 统计模型与问题描述

A. 拉普拉斯高斯混合模型

在 [34], 中，提出了一种拉普拉斯高斯混合（LGM）模型，并在单个 sEMG 数据集上进行了验证。在本研究中，我们进一步分析了 LGM模型，并评估其是否适用于对应不同 MCF 水平下不同上肢活动的各类基准数据集。以下将介绍所提出的模型。

设随机变量 Y 表示离散时间 sEMG 信号的强度。LGM模型可写为

$$f_Y(y;\Theta) = \lambda_1 f_1(y;\theta_1) + \lambda_2 f_2(y;\theta_2) \tag{1}$$

y 表示 Y的一个实现， $\Theta = [\lambda_1, \lambda_2, \theta_1, \theta_2]$ 是未知参数的集合。 λ_1 和 λ_2 是总和为 1 的混合权重。 θ_1 和 θ_2 是成分密度的参数。$f_1(y;\theta_1)$ 是定义如下的拉普拉斯密度

$$f_1(y;\theta_1) = \frac{1}{2\sigma_1}\exp\left(-\frac{|y-\mu_1|}{\sigma_1}\right) \quad -\infty < y < \infty \tag{2}$$

而 $f_2(y;\theta_2)$ 是如下给出的高斯密度

$$f_2(y;\theta_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right) \quad -\infty < y < \infty \tag{3}$$

注意， $\theta_1 = [\mu_1, \sigma_1]$ 和 $\theta_2 = [\mu_2, \sigma_2^2]$ 分别是相应密度的参数。如式 (1) 所示，混合权重 λ_1 和 λ_2 是隐含参数。利用期望最大化算法（EM 算法） [35], 从 sEMG 数据中估计 LGM 模型的未知参数。需要注意的是，EM 算法通常用于估计高斯混合模型的参数；基于此，本文为所提出的模型推导出类似的 EM 方法。

B. 参数估计问题

考虑一个数组 $\mathbf{y} = \{y_n\}_{n=0}^{N-1}$，其中 y_n 表示一个表面肌电信号的离散样本。根据高斯混合模型中使用的潜变量 [35], ，定义离散随机向量 $\mathbf{w}$ 如下

$$\mathbf{w} = \{w_n\}_{n=0}^{N-1} \tag{4}$$

这里 $w_n = [w_{n,1}, w_{n,2}]$ 有两个不同的状态，并分别对应相应的似然（混合权重）

$$\begin{aligned} p(w_{n,1}=1, w_{n,2}=0) &= \lambda_1 \\ p(w_{n,1}=0, w_{n,2}=1) &= \lambda_2 \end{aligned} \tag{5}$$

这些隐状态的边际似然为

$$p(w_n) = \lambda_1^{w_{n,1}}\lambda_2^{w_{n,2}} \tag{6}$$

y_n 在给定 w_n 和 Θ 条件下的条件概率密度函数为

$$f(y_n|w_n;\Theta) = \prod_{j=1}^{2}(f_j(y_n;\theta_j))^{w_{n,j}} \tag{7}$$

这里， y_n 是独立同分布的。数据、隐状态和未知参数的联合密度为

$$f(\mathbf{y},\mathbf{w};\Theta) = \prod_{n=0}^{N-1}\prod_{j=1}^{2}(\lambda_j f_j(y_n;\theta_j))^{w_{n,j}} \tag{8}$$

估计问题可表述如下：给定服从 LGM模型 (1) 的数据 $\mathbf{y}$, 目标是估计模型 (1) 中的参数 Θ 及相关统计量。下一节将介绍如何使用期望最大化算法估计 LGM模型的参数。

C. EM算法

完整数据对数似然为

$$L(\mathbf{y},\mathbf{w};\Theta)=\sum_{n=0}^{N-1}\sum_{j=1}^{2}w_{n,j}\ln(\lambda_j f_j(y_n;\theta_j)) \tag{9}$$

1) E步：给定数据 y 和 Θ的最新估计 $\Theta^{(i)}$， Λ(y Θ, $\Theta^{(i)}$) 是针对隐变量的条件似然计算得到的完整数据对数似然的期望。

$$\Lambda(\mathbf{y},\Theta,\Theta^{(i)})=E_{\mathbf{w}|\mathbf{y},\Theta^{(i)}}\left\{L(\mathbf{y},\mathbf{w};\Theta)\right\} \tag{10}$$

使用贝叶斯定理计算 w_n 的后验概率，得到

$$P(w_{n,j}=1|y_n;\Theta^{(i)})=\frac{f(y_n|w_{n,j}=1;\theta_j^{(i)})P(w_{n,j}=1)}{\sum_{l=1}^{2}f(y_n|w_{n,l}=1;\theta_j^{(i)})P(w_{n,l}=1)} \tag{11}$$

注意， w_n 的贝叶斯估计为

$$E(w_n|y_n,\Theta^{(i)})=P(w_{n,j}=1|y_n,\theta_j^{(i)}) \tag{12}$$

根据 (7)，对于 $w_{n,\ j}=1$，条件概率密度函数 $f(y_n|w_n;\ \Theta)$ 可简化为分量密度。然后，估计值 (12) 记为 $\gamma_n^{(i)}{}_{,\ j}$，可写为

$$\gamma_{n,j}^{(i)}=\frac{\lambda_j f_j(y_n;\theta_j^{(i)})}{\sum_{i=1}^{2}\lambda_i f_i(y_n;\theta_i^{(i)})} \tag{13}$$

因此，完整数据对数似然的期望变为

$$\Lambda(\mathbf{y},\Theta,\gamma^{(i)})=\sum_{i=1}^{n}\sum_{j=1}^{2}\gamma_{n,j}^{(i)}\ln(\lambda_j f_j(y_n;\theta_j)) \tag{14}$$

其中

$$\gamma^{(i)}=\{\gamma_{0,1}^{(i)},\gamma_{2,1}^{(i)},...,\gamma_{N-1,1}^{(i)},\gamma_{0,2}^{(i)},\gamma_{1,2}^{(i)},...,\gamma_{N-1,2}^{(i)}\} \tag{15}$$

2）M步：将拉普拉斯概率密度函数（3）和高斯概率密度函数（2）代入（14），得到

$$\begin{aligned}\Lambda(\mathbf{y},\Theta,\gamma^{(i)}) &= \sum_{n=0}^{N-1}\gamma_{n,j}^{(i)}\Bigg\{\ln\lambda_1-\ln\sigma_1-\frac{|y_n-\mu_1|}{\sigma_1}\\ &\quad \ln\lambda_2-\frac{1}{2}\ln\sigma_2^2-\frac{(y_n-\mu_2)^2}{2\sigma_2^2}\Bigg\}\end{aligned} \tag{16}$$

根据下面给出的优化问题，通过迭代估计参数。

$$\Theta^{(i+1)}=\max_{\Theta}\ \Lambda(\mathbf{y},\Theta,\gamma^{(i)}) \tag{17}$$

令式（16）中 Λ(y Θ,$\gamma^{(i)}$) 的偏导数等于零，并求解相应方程，可得到参数估计值如下。

$$\begin{aligned}\lambda_1^{(i+1)} &= \frac{N_1}{N}\\ \lambda_2^{(i+1)} &= \frac{N_2}{N}\\ \mu_1^{(i+1)} &= Median\left[\left\{\frac{\gamma_{n,1}^{(i)}}{N_1},y_n\right\}_{n=0}^{N-1}\right]\\ (\sigma_1)^{(i+1)} &= \frac{1}{N_1}\sum_{n=0}^{N-1}\gamma_{n,1}^{(i)}\left|(y_n-\mu_1^{(i)})\right|\\ \mu_2^{(i+1)} &= \frac{1}{N_2}\sum_{n=0}^{N-1}\gamma_{n,2}^{(i)}y_n\\ (\sigma_2^2)^{(i+1)} &= \frac{1}{N_2}\sum_{n=0}^{N-1}\gamma_{n,2}^{(i)}(y_n-\mu_2^{(i)})^2\end{aligned} \tag{18}$$

其中 $N_1=\sum_{n=0}^{N-1}\gamma_n^{(i)}{}_{,\ 1}$ 和 $N_1+N_2=N$。重复执行E步和M步，直到两个连续估计值 $\Theta^{(i)}$ 和 $\Theta^{(i+1)}$ 之间的平方差收敛。

D. 评估方法

使用期望最大化算法（18）得到的参数估计值，为表面肌电样本拟合LGM概率密度函数，方法如下

$$f(y;\hat{\Theta})=\hat{\lambda}_1 f_1(y;\hat{\mu}_1,\hat{\sigma}_1)+\hat{\lambda}_2 f_2(y;\hat{\mu}_2,\hat{\sigma}_2^2) \tag{19}$$

这里，ˆ λ_1、ˆμ_1、ˆσ_1、ˆλ_2、ˆμ_2、ˆσ_2^2 是收敛时由（18）得到的估计值。经验概率密度函数（mpdf）由信号样本的直方图构建。模型适用性的评估标准如下所述

视觉检查：通过视觉比较基于模型的概率密度函数（即根据模型拟合得到的近似概率密度函数）与 mpdf，以了解二者的一致程度 [36]。

库尔贝克-莱布勒散度：库尔贝克-莱布勒散度（KLD）[37] 是一种用于衡量两个概率密度函数之间差异的统计度量。设 p_1 和 p_2 为两个概率密度，则它们之间的 KLD 为

$$D_{KL}(p_1||p_2)=\sum_x p_1(x)\ln\left(\frac{p_1(x)}{p_2(x)}\right) \tag{20}$$

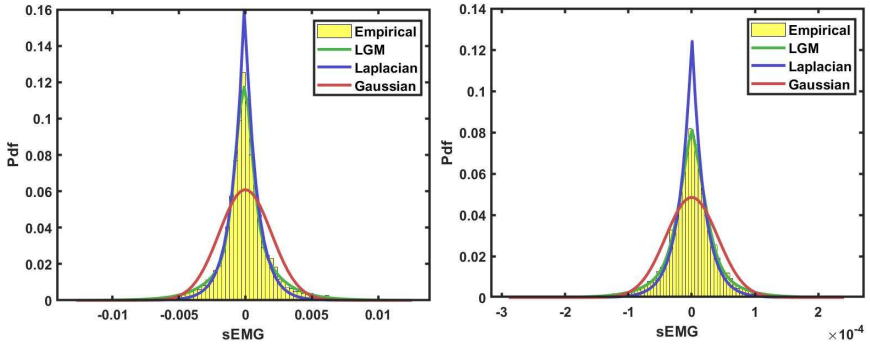
在本文中， p_1 是经验分布， p_2 是基于模型的近似概率密度函数。如果这两个分布相匹配，则 D_{KL}（$p_1||p_2$）等于 0。 D_{KL}（$p_1||p_2$）越低，近似结果就越接近 mpdf。

拟合优度图与决定系数 **[38]**：使用拟合优度图分析 sEMG 数据与基于模型的值之间的关系。数据点越接近相等线，模型拟合效果越好。另一方面，决定系数（R-squared）衡量观测到的因变量方差中有多少可由自变量解释。该值越接近 1，两个变量之间的相关性越强。

似然比检验（LRT）：似然比检验是一种用于比较两个不同模型的统计检验。为了确定

表 I：四个基准 sEMG 数据集的基本特征

	Ninapro DB2	Ninapro DB4	Rami-khushaba DB6	Intense 数据集
受试者数量	40	10	11	15
活动总数	练习-1- 17 练习-2- 23 练习-3- 09 总计 49	练习-1- 12 练习-2- 17 练习-3- 23 总计 52	40	1
考虑的活动数量	23	17	40	1
重复次数	6	6	6	1
通道数	12	12	7	8
电极类型	Delsys	Cometa Mini Wave	Delsys	Myo 腕带
采样率	2000 个样本/秒	2000 个样本/秒	4000 个样本/秒	200 个样本/秒



(a) 手势 (b) 抓握

(c) 手臂活动 (d) 剧烈活动

图1：手势、抓握、手臂活动和剧烈活动的多变量概率密度函数与模型估计概率密度函数的直观比较：LGM（绿色）、拉普拉斯（蓝色）和高斯（红色）；受试者分别为 10、3、1 和 10，对应的活动分别为 7、18、3 和 1

哪个模型具有统计显著性，需要评估两个模型的似然值。似然比检验定义为 [39]

$$T = 2(log(L_p) - log(L_e)) \tag{21}$$

其中， L_p 和 L_e 分别是LGM模型和任意现有模型的似然。

III. 数据描述

请注意，本研究分析的所有数据集均可通过公共资源获取。其简要介绍如下。

- Ninapro数据集：
  在 NinaproDB2 (NPDB2) [40] 和 DB4 (NPDB4) [41]数据集中，分别从 40 名和 10 名受试者群体中采集了 3 个训练项目。训练项目 1 和 2 与手势和抓握等活动相关。训练项目-3 对应不同力度下的手指动作，包括拇指外展和内收。本研究分析了 DB2 和 DB4 中与训练项目-2 对应的肌电信号；它们分别包含 23 种抓握动作和 17 种手势动作。该数据集中的表面肌电信号包含 12 个通道，对应于放置在手臂特定肌肉位置上的 12 个电极 [41]。在该数据集中，一项活动中的典型表面肌电信号持续 8 秒，其中休息 3 秒、活动 5 秒。每个试次重复六次。

- Rami-khushaba DB6 (RKDB6) [42]:
  该数据集包含从 11 名健全受试者（9 名男性和 2 名女性）在 5 种肢体姿势下完成 8 种不同动作时采集的表面肌电信号。肢体姿势的选择方式是

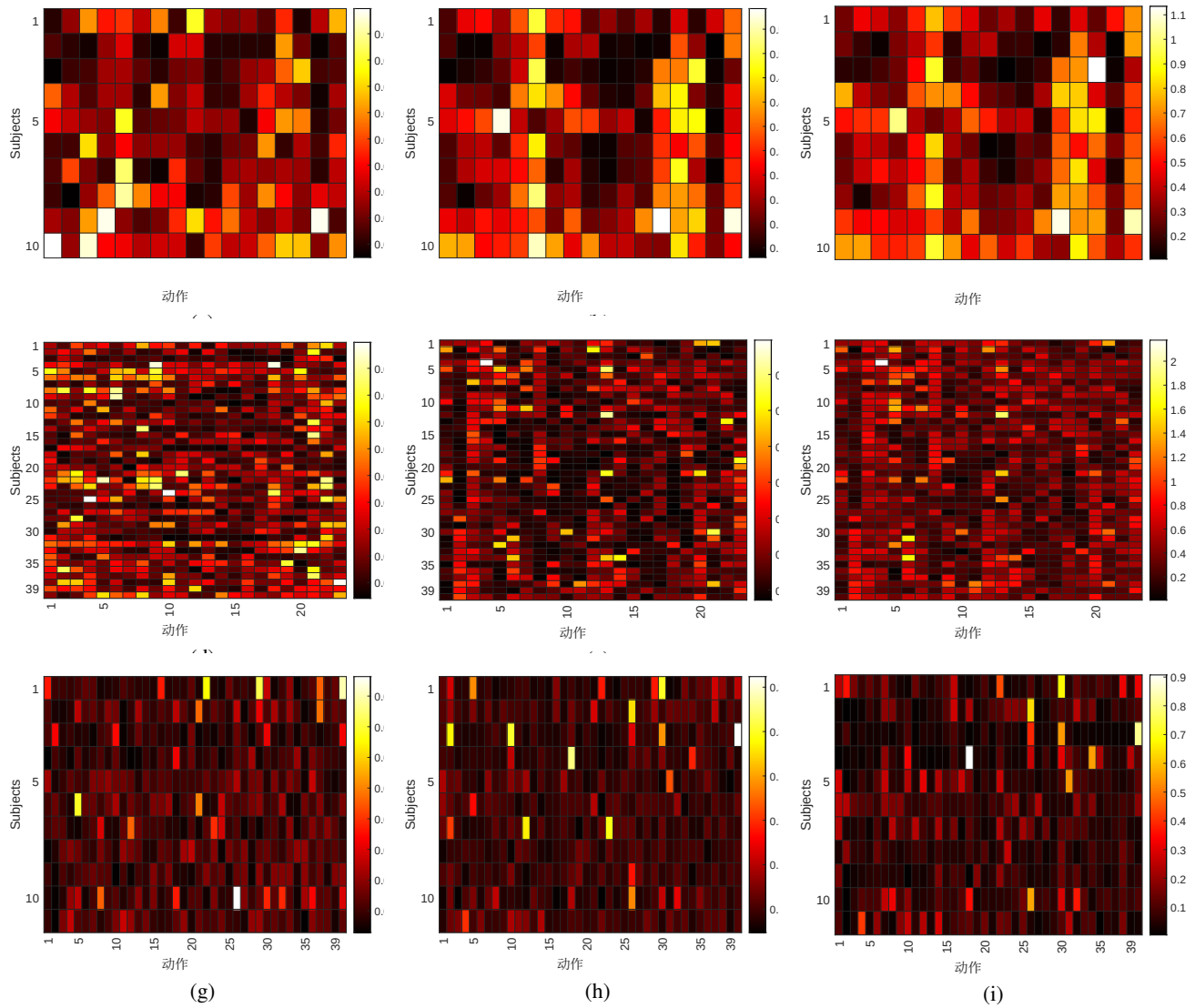


图 2：3 种模型的 KLD 热图：(a) LGM、(b) 拉普拉斯和 (c) 高斯，对应 Ninapro-DB4；(d) LGM、(e) 拉普拉斯和 (f) 高斯，来自 Ninapro-DB2；以及 (g) LGM、(h) 拉普拉斯和 (i) 高斯，来自 Rami-khushaba-DB6

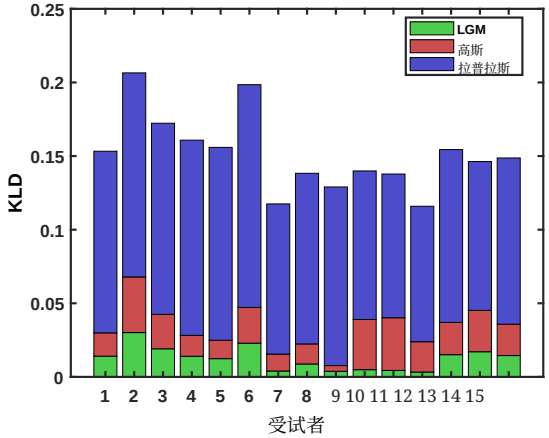


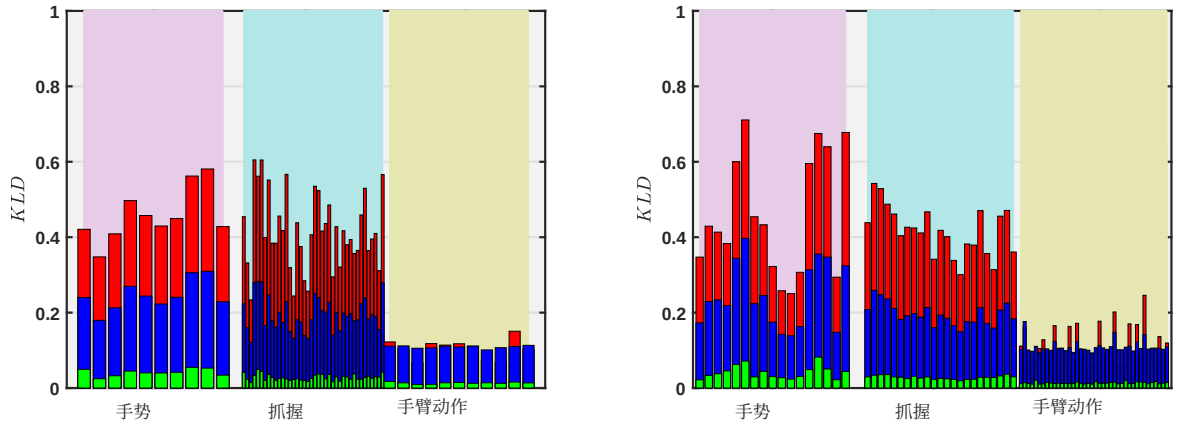
图3：LGM、拉普拉斯和高斯模型对高强度活动数据的KLD值

使每位受试者都能模仿日常活动。每项活动重复六次。一个表面肌电信号阵列由七个通道组成，分别对应七个 Delsys DE $2.x$ EMG 传感器，这些传感器沿前臂周向放置

- 高强度动作数据集（IAD） [43]:
该数据集包含从15名健康受试者在进行单次剧烈活动时采集的表面肌电信号，即要求每位受试者用右手握住一个6千克的哑铃，持续120秒。这些表面肌电信号包含8个通道，分别对应8个肌电电极，且每项活动仅进行一次。表I列出了这些基准数据集的基本特征。

IV. 结果与分析

对于上述每个数据集，使用这三种模型分析每位受试者在每项活动中的每个试次所对应的表面肌电信号。具体而言，从多个



(a) (b)

图 4：三个模型（LGM-绿色、拉普拉斯-蓝色、高斯-红色）在各条轨迹上的平均 KLD：（a）不同受试者在不同动作中的平均值；（b）不同动作在不同受试者中的平均值

通道中选取能量最高的通道，并使用基于以下评估方法的模型对其信号进行考察。

- 基于视觉检查的定性分析
- 定量分析：1）KL散度分析；2）拟合优度图，其中包含决定系数及其置信区间；3）似然比检验

A. 视觉检查

图 1 展示了 mpdf（黄色）与 LGM（绿色）、拉普拉斯（蓝色）和高斯（红色）模型拟合得到的概率密度函数之间的视觉比较。这些概率密度函数对应于以下不同活动的肌电信号：图 1(a)：活动 7，即受试者 10 指食指；图 1(b)：活动 18，即受试者 3 的四指抓握；图 1(c)：活动 3，即受试者 1 的腕关节旋后；图 1(d)：活动 1，即受试者 10 举起哑铃。图 1(a)、(b) 和 (c) 对应的概率密度函数分别是与手势、抓握和正常手臂活动相对应的表面肌电信号的概率密度函数。从中可以看出，与单独的拉普拉斯模型和高斯模型相比，mpdf 与 LGM 模型的重叠度更高。而图 1(d) 表示与剧烈活动相对应的表面肌电信号的概率密度函数，可以看到 LGM 模型与 mpdf 的重叠度与单独的高斯模型和 mpdf 的重叠度相近。相比之下，单独的拉普拉斯模型与 mpdf 的重叠度较低。

B. 定量分析

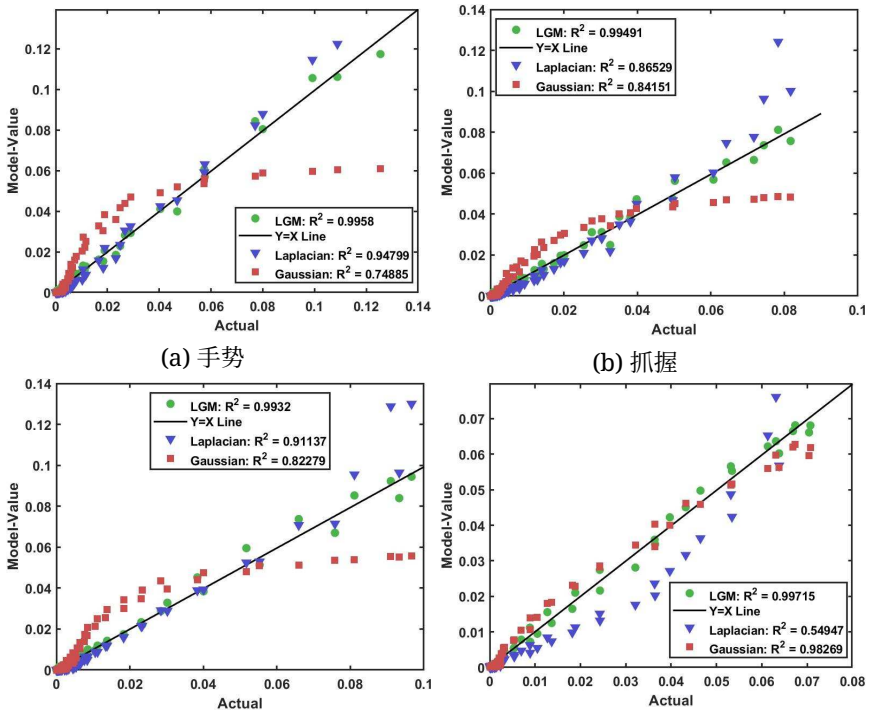
1) KL散度：对于所考虑的每个数据集，均计算 LGM 概率密度函数与 mpdf 之间的 KLD。为便于比较，也计算了高斯概率密度函数和拉普拉斯概率密度函数相对于 mpdf 的 KLD。相应结果如图 2 至图 4 所示。具体而言，图 2 展示了以受试者和动作作为变量的 KLD 热图。热图中的每个单元格表示特定受试者执行某项活动时，某个模型对应的 KLD。此外，此处所示的 KLD 是该活动各试次的平均值。图 2(a)–(c) 对应 Ninapro-DB4 的 KLD，图 2(d)–(f) 展示 Ninapro-DB2 的 KLD，图 2(g)–(i) 则表示 Rami-khushaba-DB6 的 KLD。对于这三个数据集，可以观察到这些热图中 LGM 模型的 KLD 最低。图 2 中热图的 KLD 下界和上界列于表 II。值得注意的是，LGM 模型的最高 KLD 值仍低于拉普拉斯模型和高斯模型的最低 KLD 值。

表 II：所提出模型与独立模型在不同数据集上的 KLD 下界和上界

数据集	LGM	拉普拉斯	高斯
NPDB4	[0.01 0.1]	[0.1 0.6]	[0.1 1.1]
NPDB2	[0.01 0.1]	[0.1 0.8]	[0.1 2.2]
RKDB6	[0.01 0.1]	[0.1 0.35]	[0.1 0.9]

剧烈活动情况下不同模型的 KLD 值如图3所示。值得注意的是，对于剧烈活动数据集，LGM模型的 KLD 值同样最低，其次是高斯模型，随后是拉普拉斯模型。对应于这三个模型的最小和最大 KLD 值分别为：LGM{0.0033、0.0301}；高斯模型 {0.0039、0.0378} ；以及拉普拉斯模型 {0.0920、0.1513}。

图4(a)展示了按受试者统计的各动作平均 KLD。图 4(b)展示了反向情况。LGM、拉普拉斯和高斯模型的 KLD 分别以绿色、蓝色和红色表示。从图4(a)和(b)可以看出，对于手势、抓握和手臂动作等活动，与其他模型相比，LGM模型在动作和受试者上的平均 KLD 值最低。



(a) 手势 (b) 抓握

(c) 正常活动 (d) 剧烈活动

图 5：受试者 10、3、1 和 10 在手势、抓握、正常及剧烈活动中的 LGM（绿色）、拉普拉斯（蓝色）和高斯（红色）模型的拟合优度图，以及对应的活动 7、18、3 和 1

2) 拟合优度图：图5展示了三种模型的估计值与实际数据之间的拟合优度图。具体而言，图5(a)至(d)分别对应手势、抓握、正常手臂活动和剧烈活动数据的结果。LGM模型、拉普拉斯模型和高斯模型分别以绿色、蓝色和红色数据点表示。从图5(a)、(b)和(c)的散点图可以看出，LGM模型的模型值靠近等值线，这意味着该模型的预测值接近表面肌电信号的实际值。而对于图5(d)所示的剧烈活动，LGM模型和高斯模型对应的模型值相近，且均靠近等值线。从图5可以得出，对于手势、抓握和正常手臂活动，LGM模型优于其他模型。不过，对于剧烈活动，LGM模型和高斯模型都能很好地拟合肌电数据。平均决定系数值见表 III。根据这些指标，LGM模型在前三类活动中表现更优。此外，对于剧烈活动，LGM模型和高斯模型的表现仍然相近。图5中各图对应的决定系数 95% 置信区间 [44] [45] 见表 IV。

表 III：模型评估所得四个数据集的决定系数

数据集	LGM	拉普拉斯	高斯
NPDB4	0.9958	0.94799	0.74885
NPDB2	0.99491	0.86529	0.84151
RKDB6	0.9932	0.91137	0.82279
IAD	0.99715	0.54947	0.98269

表 IV：LGM、拉普拉斯和高斯模型的决定系数置信区间

数据集	LGM	拉普拉斯	高斯
NPDB4	[0.9946 0.9970]	[0.9439 0.9520]	[0.7413 0.7564]
NPDB2	[0.9937 0.9961]	[0.8598 0.8708]	[0.8357 0.8473]
RKDB6	[0.9918 0.9946]	[0.9067 0.9160]	[0.8167 0.8289]
IAD	[0.9960 0.9983]	[0.5402 0.5587]	[0.9800 0.9854]

3) 似然比检验：如下所示，对 LGM模型和拉普拉斯模型进行式 (21) 中给出的似然比检验。

H_0 ： 拉普拉斯模型拟合数据H_1:
LGM模型拟合数据

此检验针对99%置信区间进行。值得注意的是，p值小于0.01，这意味着 H_0 被拒绝，H_1 被接受。对于 H_0，将拉普拉斯模型替换为高斯模型后重复此检验。该检验中的p值也小于0.01，因此 H_1 被接受。

C. 混合权重

不同数据集中的不同活动所对应的 LGM模型拉普拉斯分量的混合系数如图6所示。对于低和中等 MCF 水平，例如手势、抓握和正常活动，可以发现拉普拉斯

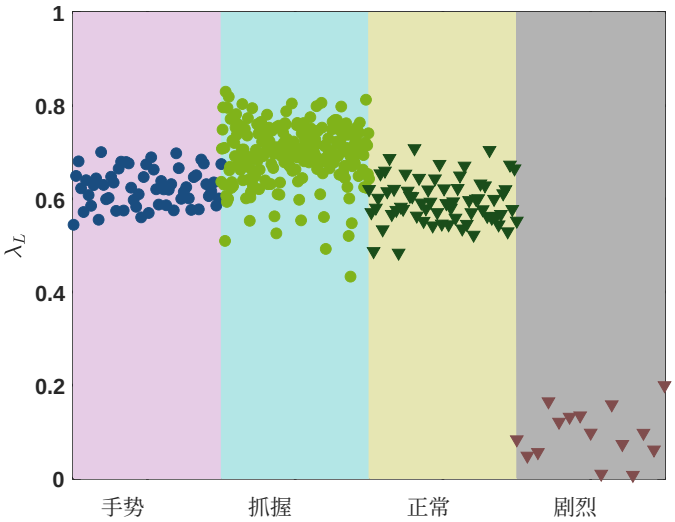
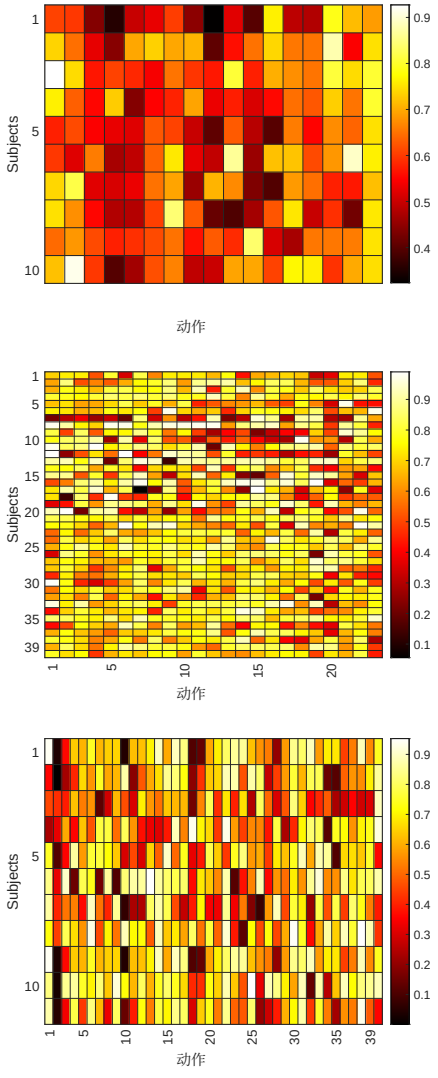


图6：不同受试者在手势、抓握、正常手臂动作和剧烈活动下各轨迹的拉普拉斯系数

分量的权重比高斯分量更大。而在剧烈活动的情况下，即MCF 水平较高时，高斯分量对应的混合权重较大，拉普拉斯分量的权重则较小。因此，从图6可以看出，随着活动强度的增加，也就是说，完成某个动作所需的能量增加，拉普拉斯分量的权重会降低。图7展示了所研究的前三种肢体活动对应的 LGM模型的拉普拉斯权重。热图中的横轴和纵轴分别表示动作数量和受试者。热图中的每个单元格表示某一受试者在某项活动中的 LGM模型拉普拉斯权重。可以发现，在大多数情况下，拉普拉斯权重 λ_1 高于高斯权重 λ_2。某些情况下，拉普拉斯权重低于高斯权重。例如，在图7(a)中，受试者1的活动4和9；在图7(b)中，受试者18的活动2；以及在图7(c)中，受试者1的活动2和10，高斯权重更大。



(c)

图7：LGM模型的拉普拉斯权重在各试次上的平均值，以受试者和动作划分：(a) 手势，(b) 抓握，以及 (C) 手臂活动

V. 讨论

根据第 IV 节所呈现的结果，对于对应于低水平和中等水平肌肉募集的肌电信号，即手势、抓握和正常手臂动作等活动，LGM 被发现比独立模型更适合。这一点通过以下方面得到验证：1）模型概率密度函数与 mpdf 的视觉检查；2）最低的 KLD；3）拟合优度图中模型值与真实值相符；4）较高的决定系数；以及 5）似然比检验接受备择假设。然而，对于剧烈活动，根据上述四种评估方法，LGM 和高斯模型的表现似乎相当。因此，在肌肉募集水平较高时，所提出的 LGM模型表现得类似于独立高斯模型。以下观察进一步证实了这一结果

：在 LGM模型中，拉普拉斯分量的混合系数与高斯分量相比变得非常小。

在对混合权重的分析中可以发现，对于前几种动作，拉普拉斯分量和高斯分量都对模型有显著贡献。然而，对于剧烈活动，高斯分量的作用要强得多。因此，根据这些发现，可以推断 LGM模型的权重可能与活动期间的MCF水平以及被激活的运动单位有关。例如，对于前几种活动，拉普拉斯权重 λ_1较高，这对应较低的 MCF水平和较少的激活运动单位。而在剧烈活动中，高斯权重 λ_2 较高，这对应较高的 MCF水平和较多的激活运动单位。这些发现与 I-B节中关于概率密度函数的文献报道一致，其中指出，在较低和中等 MCF 水平下，概率密度函数的中心峰值更尖锐，暗示

拉普拉斯结构；而在较高 MCF 水平下，则呈现出明显的高斯结构。

VI. 结论

本文针对来自上肢的表面肌电信号提出了一种拉普拉斯高斯混合模型。所提出的模型在多个基准 sEMG 数据集上进行了测试，并与现有的独立模型进行了比较。通过以下方法验证了该模型的适用性：（1）定性分析，例如与经验概率密度函数（mpdf）进行直观比较，结果显示 LGM模型具有最佳的一致性；（2）计算模型概率密度函数与 mpdf 之间的 KL散度，同样，LGM模型的 KLD 最低；（3）拟合优度图，对决定系数（CFD）- R^2 和置信区间进行比较，其中对于 R^2，可以观察到，在 LGM模型的情况下， R^2 最接近 1；（4）似然比检验（LRT），其结果同样支持 LGM模型。最后需要指出的是，在低和中等肌肉收缩力水平下，拉普拉斯权重比高斯权重更大；而在较高的肌肉收缩力水平下，拉普拉斯权重较低。未来，我们将扩展所提出的模型，以研究来自不同肌肉位置的表面肌电信号之间的相关性。

致谢

本研究由 SERB 根据印度政府项目资助，项目编号：CRG/2019/003801。

参考文献

[1] W. Wang、A. D. Stefano 和 R. Allen，“用于分析步态周期中肌肉活动的表面肌电信号仿真模型”，《生物与医学计算机》，第 36 卷，第 6 期，第 601–618 页，2006 年。[2] D. W. Stashuk，“肌电信号仿真”，《电肌电图与运动学杂志》，第 3 卷，第 3 期，第 157–173 页，1993 年。[3] P. A. Cuddon，“神经肌肉疾病中的电生理学”，《兽医诊所：小动物实践》，第 32 卷，第 1 期，第 31–62 页，2002 年。[4] A. D. Vigotsky、I. Halperin、G. J. Lehman、G. S. Trajano 和 T. M. Vieira，“解读运动与康复科学研究中的表面肌电信号幅值”，《生理学前沿》，第 8 卷，第 985 页，2018 年。[5] C. J. De Luca，“表面肌电图在生物力学中的应用”，《应用生物力学杂志》，第 13 卷，第 2 期，第 135–163 页，1997 年。[6] G. L. Soderberg 和 T. M. Cook，“生物力学中的肌电图”，《物理治疗》，第 64 卷，第 12 期，第 1813–1820 页，1984 年。[7] E. Hasanbelliu，“用于理解肌电信号在头部运动预判中作用的多维可视化工具”，载于 ACM SIGGRAPH 2004 Posters，2004 年，第 109 页。[8] R. Merletti、L. L. Conte、E. Avignone 和 P. Guglielminotti，“表面肌电信号建模。I. 模型实现”，IEEE生物医学工程汇刊，第 46 卷，820，1999 年 7 月，810 第。– 页。[9] J. Rodriguez-Falces、J. Navallas 和 A. Malanda，“肌电建模”，《肌电图分析中的计算智能——当前应用与未来挑战展望》，第 3–36 页，2012 年。[10] M. Hayashibe 和 D. Guiraud，“基于多尺度生理肌肉模型的自主肌电到力估计”，《生物医学工程在线》，第 12 卷，第 1 期，第 1–18 页，2013 年。[11] O. Paiss 和 G. F. Inbar，“表面肌电及其频谱的自回归建模及其在疲劳研究中的应用”，IEEE生物医学工程汇刊，第 10 期，第 761–770 页，1987 年。[12] T. Kiryu、C. J. De Luca 和 Y. Saitoh，“斜坡收缩期间肌电干扰信号的自回归建模”，IEEE生物医学工程汇刊，第 41 卷，第 11 期，第 1031–1038 页，1994 年。[13] S. Shahid、J. Walker、G. M. Lyons、C. A. Byrne 和 A. V. Nene，“将高阶统计技术应用于肌电信号以表征运动单位动作电位”，IEEE生物医学工程汇刊，第 52 卷，第 7 期，第 1195–1209 页，2005 年。

[14] A. Furui、H. Hayashi 和 T. Tsuji，《一种具有可变方差的表面肌电信号尺度混合随机模型》，IEEE生物医学工程汇刊，第66卷，第10期，第2780–2788页，2019年。[15] Z. O. Khokhar、Z. G. Xiao 和 C. Menon，《用于腕部外骨骼实时控制的表面肌电模式识别》，《生物医学工程在线》，第9卷，第1期，第1–17页，2010年。[16] H. Huang、T. A. Kuiken、R. D. Lipschutz 等，《利用表面肌电图识别运动模式的策略》，IEEE生物医学工程汇刊，第56卷，73，2008 1，65第期，第–页。[17] A. D. Chan 和 K. Englehart，《使用高斯混合模型对动力假肢的肌电信号进行连续分类》，载于《IEEE生物医学工程学会第25届年度国际会议论文集》（IEEE目录号：03CH37439），第3卷。IEEE，2003年，第2841–2844页。[18] A. Furui、T. Igaue 和 T. Tsuji，《基于尺度混合随机生成模型的贝叶斯推断肌电模式识别》，《专家系统与应用》，第185卷，文章号115644，2021年。[19] P. S. Giacomini，《肌电图与神经肌肉疾病：临床电生理相关性》，《麦吉尔医学杂志：MJM》，第9卷，第2期，第173页，2006年。[20] D. Xiong、D. Z hang、X. Zhao 和 Y. Zhao，《基于肌电的人机交互深度学习：综述》，《IEEE/中国自动化学会自动化学报》，第8卷，第3期，第512–533页，2021年。[21] W. Wei、Q. Dai、Y. Wong、Y. Hu、M. Kankanhalli 和 W. Geng，《基于多视图深度学习的表面肌电手势识别》，IEEE生物医学工程汇刊，第66卷，第10期，第2964–2973页，2019年。[22] B. Hudgins、P. Parker 和 R. N. Scott，《多功能肌电控制的新策略》，IEEE生物医学工程汇刊，第40卷，第1期，第82–94页，1993年。[23] H. Roesler，《用于比例控制的肌电信号统计分析与评估》，《上肢假肢与矫形器控制》，第44–53页，1974年。[24] H. Milner-Brown 和 R. Stein，《表面肌电图与肌力之间的关系》，《生理学杂志》，第246卷，第3期，第549–569页，1975年。[25] P. A. Parker、J. A. Stuller 和 R. N. Scott，《多状态肌电通道的信号处理》，《IEEE会议录》，第65卷，第5期，第662–674页，1977年。[26] I. Hunter、R. Kearney 和 L. Jones，《使用相位和脉冲响应函数技术估算肌肉动作电位的传导速度》，《医学与生物工程及计算》，第25卷，第2期，第121–126页，1987年。[27] M. Bilodeau、M. Cincera、A. B. Arsenault 和 D. Gravel，《斜坡和阶跃等长收缩期间肘屈肌肌电信号的正态性和平稳性》，《电肌电图与运动学杂志》，第7卷，第2期，第87–96页，1997年。[28] E. A. Clancy 和 N. Hogan，《表面肌电图的概率密度及其与幅度检测器的关系》，IEEE生物医学工程汇刊，第46卷，第6期，第730–739页，1999年。[29] P. Kaplanis、C. S. Pattichis、L. Hadjileontiadis 和 S. M. Panas，《表面肌电双谱分析》，载于《2000年第10届地中海电工技术会议：地中海国家的信息技术与电工技术，会议论文集，MeleCon 2000》（目录号：00CH37099），第2卷。IEEE，2000年，第770–773页。[30] K. Nazarpour、A. Sharafat 和 S. Firoozabadi，《表面肌电信号的负熵分析》，载于《IEEE/SP第13届统计信号处理研讨会》，2005年。IEEE，2005年，第974–977页。[31] G. R. Naik 和 D. K. Kumar，《不同力水平下多通道表面肌电的高阶统计量参数评估》，载于《IEEE生物医学工程学会2011年度国际会议》。IEEE，2011年，第3869–3872页。[32] G. R. Naik、D. K. Kumar 和 S. P. Arjunan，《不同最大自主收缩期间肌电信号的峰度与负熵研究》，载于《ISSNIP 2011年生物信号与仿生机器人会议》。IEEE，2011年，第1–4页。[33] K. Nazarpour、A. H. Al-Timemy、G. Bugmann 和 A. Jackson，《关于表面肌电信号概率分布函数的说明》，《脑研究通报》，第90卷，第88–91页，2013年。[34] D. Kusuru、A. C. Turlapaty 和 M. Thakur，《上肢表面肌电信号的拉普拉斯-高斯混合模型》，载于《IEEE生物医学与工程学会第43届年度国际会议》（EMBC），2021年。IEEE，2021年，第681–685页。[35] C. M. Bishop 和 N. M. Nasrabadi，《模式识别与机器学习》。施普林格，2006年，第4卷，第4期。[36] A. Spanos，《概率论与统计推断：基于观测数据的经验建模》。剑桥大学出版社，2019年。

[37] S. Kullback, 《信息论与统计学》。Courier Corporation，1997年。[38] J. Cohen、P. Cohen、S. G. West 和 L. S. Aiken，《应用多元回归》，《行为科学相关分析》，第2卷，1983年。[39] G. King，《统一政治方法论：统计推断的似然理论》。密歇根大学出版社，1998年。[40] M. Atzori、A. Gijsberts、C. Castellini、B. Caputo、A.-G. M. Hager、S. Elsig、G. Giatsidis、F. Bassetto 和 H. M¨uller，《用于非侵入式自然控制机器人手部假肢的肌电图数据》，Scientific Data，第1卷，第1期，第1–13页，2014年。[41] S. Pizzolato、L. Tagliapietra、M. Cognolato、M. Reggiani、H. M¨uller 和 M. Atzori，《六种肌电采集配置在手部运动分类任务中的比较》，PLOS ONE，第12卷，第10期，文章号e0186132，2017年。[42] R. N. Khushaba、M. Takruri、J. V. Miro 和 S. Kodagoda，《利用时变谱特征实现肢体位置不变的肌电模式识别》，《神经网络》，第55卷，第42–58页，2014年。[43] A. Ebied、A. M. Awadallah、M. A. Abbass 和 Y. El-Sharkawy，《利用多通道表面肌电进行上肢肌肉疲劳分析》，载于《2020年第2届新型智能与前沿新兴科学会议》（NILES）。IEEE，2020年，第423–427页。[44] J. D. Finn 等，《相关性再探》，《心理学通报》，第118卷，第1期，第155–164页，1995年。[45] P. Cohen、S. G. West 和 L. S. Aiken，《行为科学应用多元回归/相关分析》。Psychology Press，2014年。