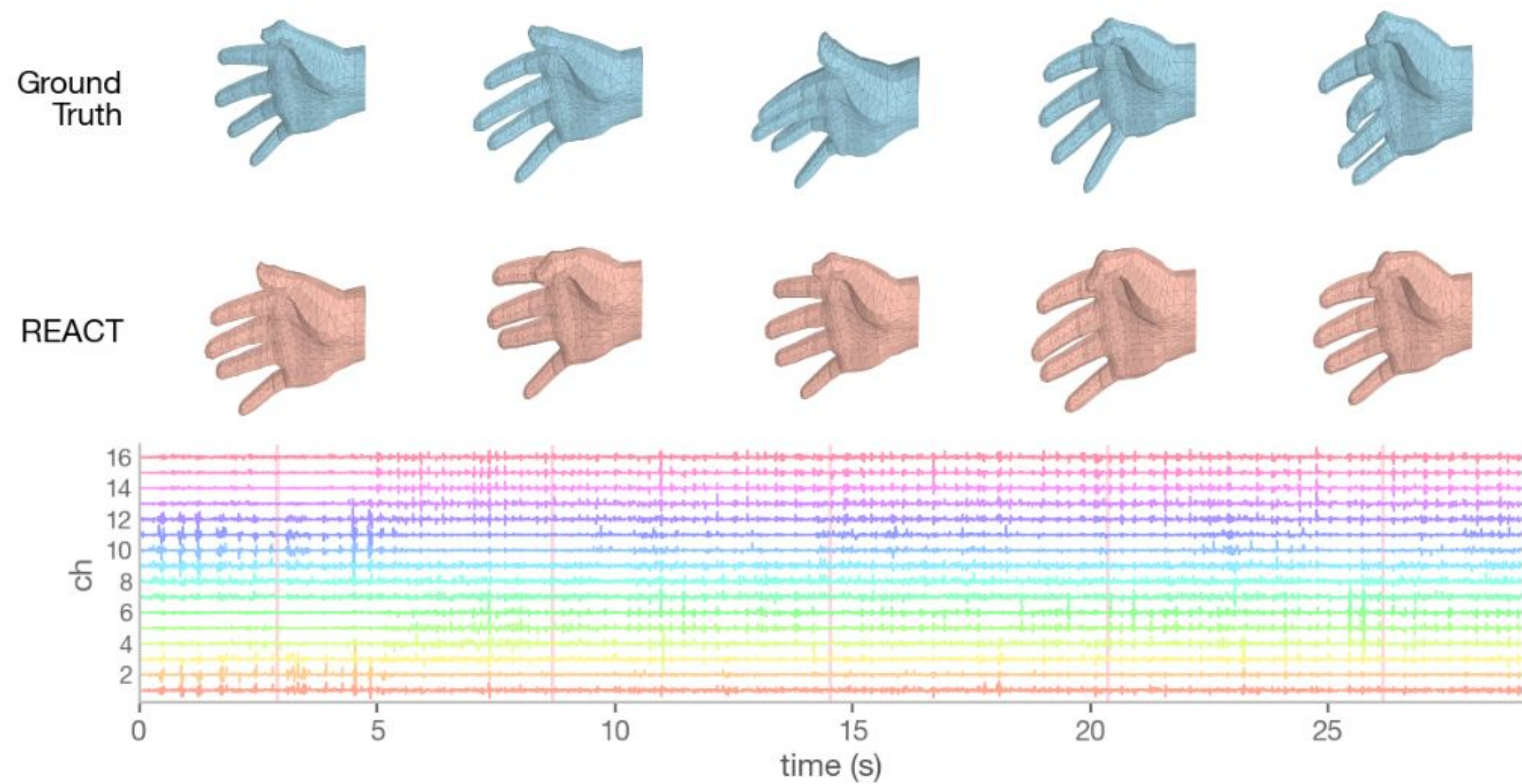


REACT: A Conditioning Framework for User-Adaptive sEMG Hand Pose Estimation

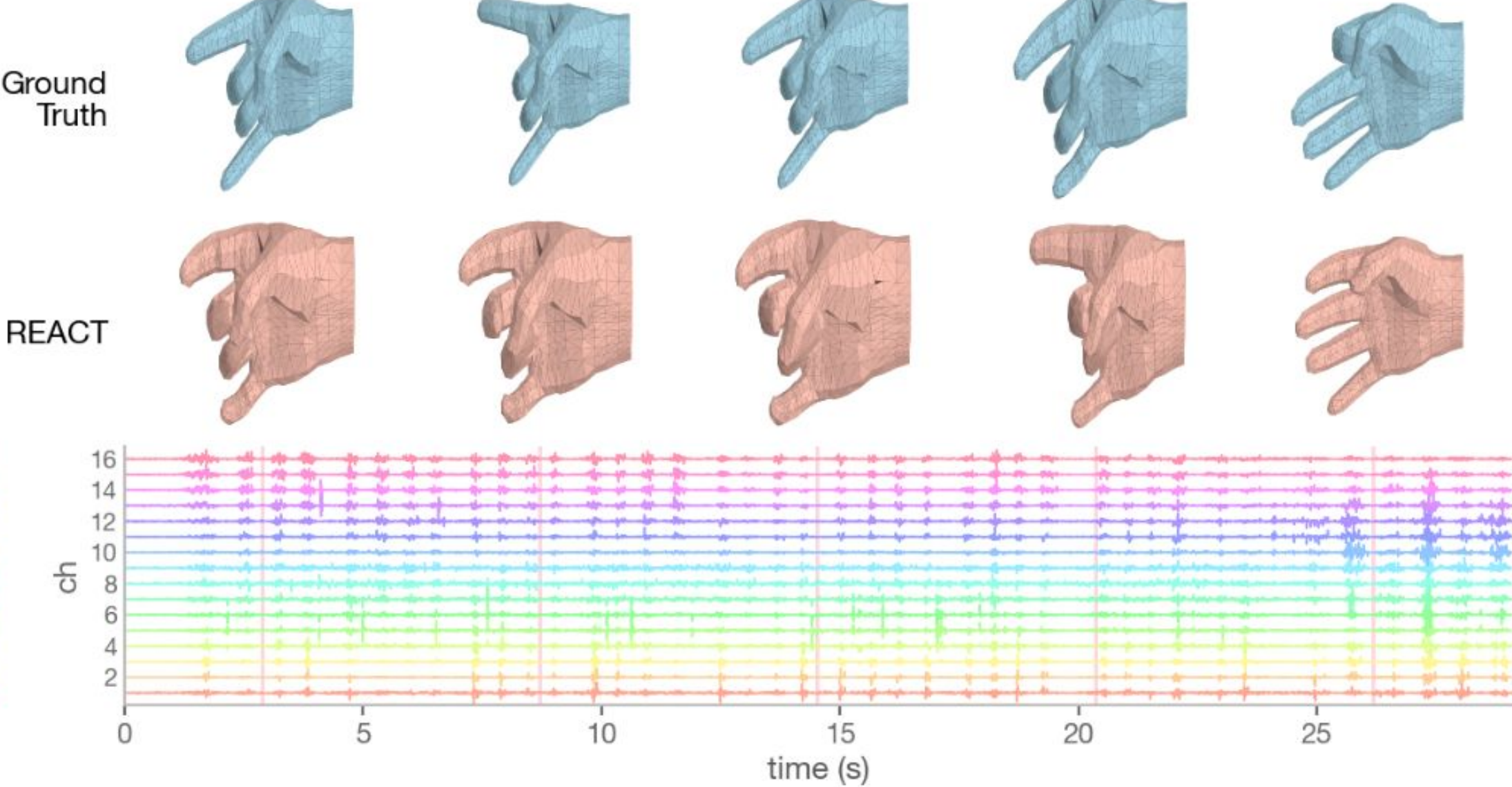
Eric Xie, Hayson Cheung



Finger Wiggling & Spreading



Hook'em, Horns, OK, Scissors



Motivation

Question:

- Can we overcome the cross-user generalization challenge in sEMG (surface electromyography) to hand pose estimation with a learned calibration module?

Requires:

- k samples of sEMG data for calibration

Produces:

- Conditioning parameters computed during calibration phase and applied to encodings at inference time

Challenges in Hand Pose Estimation

- Cross-user variability in muscle anatomy, skin impedance, and electrode placement causes significant signal differences
- Existing models degrade substantially on unseen users
- Fine-tuning require separate training procedure which is impractical for real-time use
- No prior work addresses gradient-free, continuous pose regression for calibrating new users

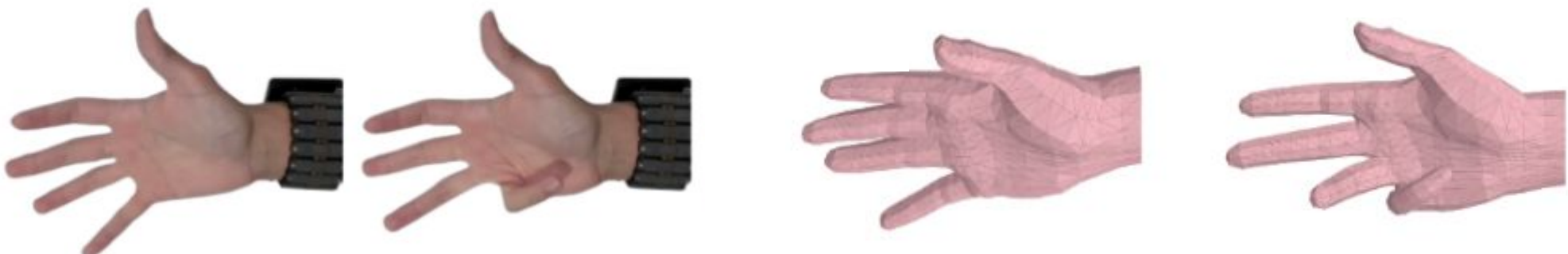
Public Dataset Benchmark

- Meta's emg2pose: 193 users, 25,253 recordings, 424 hours of synchronized sEMG + hand pose labels
- Three generalization axes: User (new users), Stage (new gestures), User+Stage (both)
- Two prediction modes: Regression (absolute angles) and Tracking (velocity integration from initial pose)
- REACT achieves lower error across all metrics

Split	Model	Regression		Tracking	
		Err (°)	Dist (mm)	Err (°)	Dist (mm)
User	VEMG2POSE	12.2	15.8	7.7	10.3
	REACT (ours)	12.0	15.6	7.4	10.0
Stage	VEMG2POSE	15.2	20.4	11.2	15.2
	REACT (ours)	14.6	19.9	10.9	14.9
User+Stage	VEMG2POSE	15.8	21.6	11.0	15.4
	REACT (ours)	15.3	21.2	10.8	15.3

Discussion

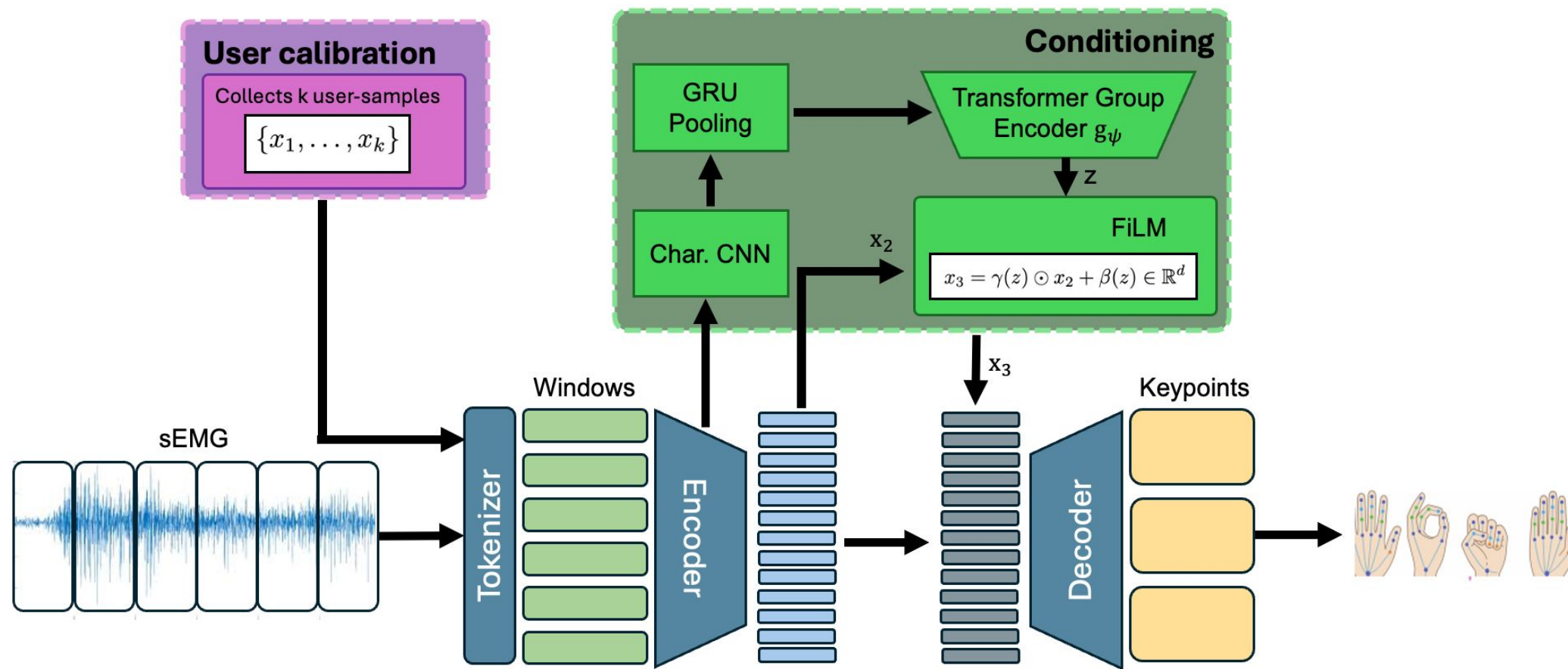
- Consistent improvements of 1.6–3.9% across all splits
- Why does conditioning work? We hypothesize that inter-user variability manifests in channel-wise muscle activation amplitudes that FiLM helps shift & scale
- Limitations:** moderate gains, single dataset
- Future work:** deployment on physical robotic hands using Meta wristbands to teleoperate in real-time. Experimentation on a broader dataset, even beyond EMG, in domains with user calibration potentials



User teleoperating with Meta wristband

Robotic hand

Model



- REACT inserts a lightweight conditioning pathway between the baseline vemg2pose encoder and decoder
- Calibration = four stages: a characteristic CNN extracts user-specific features, a bidirectional GRU pools each recording temporally, a transformer encoder aggregates k recordings into a single user embedding, and a FiLM projection produces per-channel (γ, β) applied to the encoder output
- Training:** two phases, frozen backbone first, then unfreeze for joint fine-tuning with LR decay
- Backbone agnostic:** not a standalone predictor. Can enhance a strong backbone but not fix a weak one. Flexible, can attach to any future SOTA models



UNIVERSITY OF
TORONTO

