ALASCA: Asset Leakage with Acoustic Side-Channel Attacks in Multi-Joint Collaborative Robots

Ayan Joshi, Adharsh Jaikumar, Dr. Sabur Baidya, Saveer Jain

Abstract— **Collaborative robots are increasingly deployed in manufacturing environments where motion trajectories encode proprietary processes. We investigate whether an attacker can reconstruct these trajectories from acoustic emissions alone. While prior work has demonstrated acoustic side-channel attacks against additive manufacturing systems using stepper motors, no study has addressed collaborative robots equipped with brushless DC servo motors and harmonic drives, which produce fundamentally different acoustic signatures. We present the first acoustic side-channel attack targeting a collaborative robot (the Franka Emika Panda), extracting joint-level motion parameters from a single Ambisonic microphone recording. Using spectral, spatial, and temporal features derived from four-channel spatial audio, we train Random Forest classifiers on continuous sinusoidal sweeps and evaluate generalization to unseen discrete movement profiles. In multi-joint experiments, the base joint achieves 78.4% direction classification accuracy and 94.8% adjacent speed accuracy ($\pm$1 RPM), while a joint identification classifier distinguishes active joints at 95.9%. We further demonstrate that the underlying control algorithm (bang-bang, PID, or predictive) can be identified at 99.99% accuracy from acoustic emissions alone. An ablation study confirms that Ambisonic spatial features provide a significant accuracy improvement over single-channel recording. These results establish that collaborative robots are susceptible to acoustic side-channel reconnaissance and motivate investigation of acoustic countermeasures for security-critical deployments.**

I. INTRODUCTION

Collaborative robots (cobots) are specifically designed to work safely alongside humans within shared environments [1]. The integration of advanced sensing systems, force control mechanisms, and intelligent algorithms into cobots has revolutionized safe physical interaction and flexible task execution. *However, such a rich ecosystem in cobots can often be considered a "double-edged sword" since they can be used to steal assets.* To understand what level of leakage is appropriate, the key question boils down to: *What can trigger information leaks and how much information can be inferred from a given cobot without any intrusive mechanism?*

To answer the above question, we want to point out that cobots' motion trajectories often encode information on proprietary manufacturing procedures (e.g., aerospace composite layup paths) or sensitive experimental protocols (e.g., pharmaceutical compounding trajectories) that constitute valuable intellectual property (IP). The threat to such IP is well-documented. For example, recent corporate espionage cases, such as the theft of Micron's DRAM fabrication parameters [2] and the "Turbine Panda" campaign targeting aviation composite manufacturing [3], demonstrate that *manufacturing process knowledge or IPs* is a highly lucrative target. Therefore, for the first time, this paper demonstrates that *manufacturing process or IPs are correlated with robotic motion trajectories and the acoustic emission from the robotic motion can be used to nonintrusively leak confidential IPs without attacking the network security entirely [4].*

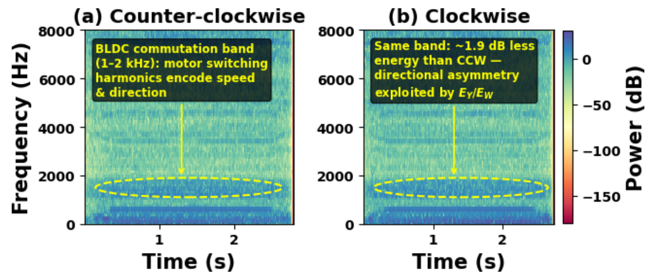

Fig. 1. Y-channel spectrograms of Joint 0 at 5.0 RPM, (a) CCW and (b) CW. Directional differences are sub-perceptual, motivating the spatial energy ratio features in Section IV-D.

Recent works [5]–[8] already give the impression that the acoustic emission can work as side channels to reveal sensitive information [5], [6], [9]. The seminal work by Al Faruque et al. [9], [10] demonstrated that stepper motor acoustics in 3D printers leak sufficient information to reconstruct printed geometries. Gatlin et al. [4] extended this threat to the power side channel, showing that encryption of digital design files is insufficient when physical emanations leak process information. However, these studies focus exclusively on additive manufacturing with stepper motors, which produce distinctive step-frequency acoustic signatures due to their discrete angular increments. On the contrary, cobot joints are driven by brushless DC (BLDC) servo motors with integrated torque sensors and harmonic drive gearboxes, producing acoustic emissions that differ substantially from stepper motors in both spectral content and spatial radiation patterns. Therefore, the solution proposed in [4], [9] may not work effectively in our novel cobot scenario with real-world industrial noisy conditions.

To motivate acoustic side-channel attacks on cobots, Fig. 1 presents Y-channel spectrograms of Joint 0 at 5.0 RPM for clockwise (CW) and counter-clockwise (CCW) rotation. The two panels appear nearly identical, sharing the same harmonic drive fundamental, BLDC commutation band (1–2 kHz), and broadband noise floor. Direction information is not recoverable by visual inspection, which explains why prior stepper-motor techniques that rely on salient step-frequency peaks do not transfer to servo-driven cobots.

In a novel cobot scenario, this paper uses acoustic emission from robotic motions by following two distinct strategies.

Strategy 1: We divide acoustic emissions from BLDC servo motors of cobots into non-overlapping temporal analysis windows to extract spectral, spatial, and temporal features from all four Ambisonic channels: W (omnidirectional sound pressure), Y (left-right particle velocity), Z (up-down particle velocity), and X (front-back particle velocity). This spatial decomposition allows analysis of how motor acoustic energy radiates directionally, which varies with joint rotation direction and joint position on the robot arm. We calculate 13 Mel-frequency Cepstral Coefficients (MFCCs) using 128 Mel bands, 12 chroma bins, 10 spatial energies (i.e., 4 channel energies, 3 energy ratios, 2 spatial differences, and 1 directional ratio), 26 Delta MFCCs, 13 Delta-Delta MFCCs, and 4 statistical energy dynamics, such as energy slope, mean, standard deviation, and range of sub-frame RMS energies. We use these features to classify the direction of rotation and the angular velocity of each active joint, and do the temporal sequencing of multi-joint movements, enabling full trajectory reconstruction of the robotic motions.

Strategy 2: Furthermore, cobots operate with sophisticated closed-loop control algorithms, such as PID controllers, predictive (acceleration-constrained) controllers, and time-optimal (bang-bang) control, each of which produces distinct velocity profiles and, consequently, distinct acoustic signatures. We classify the control algorithm from acoustic emission alone to reveal the sophistication of the underlying automation, a valuable intelligence target beyond trajectory reconstruction. Moreover, a realistic acoustic attack further requires that classifiers trained on one motion profile generalize to unseen motion profiles because an attacker who trains on known reference motions must infer trajectories from unknown victim movements. We explicitly evaluate this cross-profile generalization in this paper.

In this paper, we present the first investigation of acoustic side-channel attacks targeting cobots. Our contributions are:

1) We establish cobots as a viable target for acoustic side-channel attack. We demonstrate that joint rotation direction and angular velocity of a 7-DOF (Degree of Freedom) robot can be inferred from acoustic emissions alone on a real-world test bed comprising the Franka Emika Panda. *We consider the noisy scenario in our real-world test bed and prove the efficacy of our model in noisy conditions.*
2) We exploit four-channel Ambisonic spatial audio features (inter-channel energy ratios and directional components) to improve direction classification by 14.4 percentage points over single-channel baselines, achieving 78.4% direction accuracy and 94.8% speed accuracy within ± 1 RPM across six bins.
3) We evaluate cross-profile generalization, training on continuous sinusoidal motion and testing on discrete movements, showing that spectral speed features transfer across motion types while direction features depend on matched temporal dynamics.
4) We extend the attack to joint identification across multiple joints and to control algorithm classification (bang-bang, PID, predictive), demonstrating that acoustic emissions reveal not only kinematic parameters but also the control strategy governing the motion.

II. RELATED WORK

Acoustic emanations have been exploited as side channels across multiple domains. Genkin et al. [5] demonstrated extraction of RSA keys from CPU acoustic emissions during cryptographic operations. Backes et al. [11] showed that printers leak textual content through acoustic signatures, and Asonov and Agrawal [6] exploited keystroke acoustics to infer typed content.

In the manufacturing domain, Al Faruque et al. [9] presented the first acoustic side-channel attack on additive manufacturing systems, extracting Mel-frequency cepstral coefficient (MFCC) features from stepper motor sounds to predict axis of movement and travel distance. Subsequent work by Chhetri et al. [12] extended this to G-code reconstruction, while Song et al. [13] demonstrated similar attacks using smartphone microphones. Yang et al. [14] proposed detection of cyber-incidents in 3D printing by monitoring audio spectrograms via Wasserstein distance metrics. Shah et al. [15] extended acoustic attacks to a robotic arm (uArm Swift Pro), using smartphone recordings to fingerprint individual stepper-motor movements at approximately 75% accuracy and reconstruct warehousing workflows at 62%. However, their platform relies on stepper motors and their approach categorizes predefined movement patterns rather than decomposing per-joint motion parameters from multi-joint servo motor emissions.

All prior acoustic attacks on manufacturing and robotic systems target stepper motors, which produce characteristic acoustic signatures from discrete magnetic stepping. Servo motors, by contrast, operate with continuous rotation and produce broadband emissions modulated by speed, load, and control dynamics [16]. Our work bridges this gap by demonstrating that servo motor acoustics in cobots carry sufficient information for motion parameter inference.

III. THREAT MODEL

A. Motivation

Cobots present a uniquely vulnerable target for acoustic side-channel attacks for three reasons. *First*, cobots are designed for fenceless, human-proximate operation, making covert microphone placement trivial: an attacker needs only to carry a smartphone or place a compact recorder within the shared workspace. *Second*, cobots are increasingly deployed in high-value contexts where motion trajectories directly encode proprietary processes. For example, in aerospace composite layup, a robotic arm's trajectory encodes fiber orientation sequences and resin deposition paths that determine structural properties of flight-critical components. In pharmaceutical compounding, the trajectory encodes precise ingredient dispensing sequences, volumes, and mixing protocols. Reconstructing these trajectories from acoustic emissions would expose trade secrets protected under the Defend Trade Secrets Act (DTSA) and, in defense contexts, controlled under the International Traffic in Arms Regulations (ITAR). *Third*, unlike traditional industrial robots that operate in high-noise environments where hearing conservation programs are required above 85 dBA [19], cobot work cells are designed

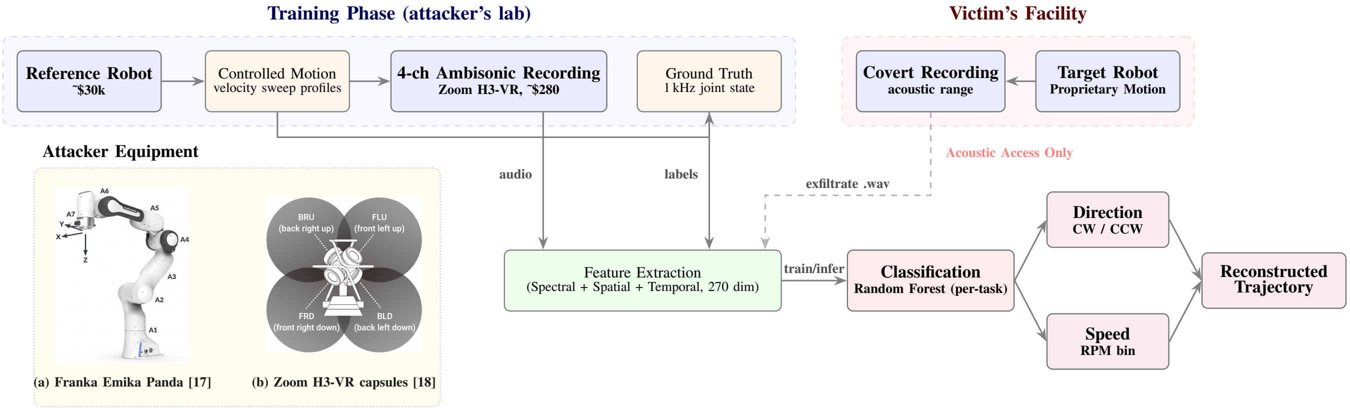

Fig. 2. Attack workflow. The attacker records a reference robot (a) executing controlled velocity sweeps, extracting 4-channel Ambisonic audio (b) and ground-truth joint states in parallel. Spectral, spatial, and temporal features (270 dimensions) train direction and speed classifiers. A covert recording of the target robot (right) is exfiltrated and processed offline through the same pipeline, producing per-window direction and speed predictions that are chained into a reconstructed trajectory. Only the recording step requires acoustic-range proximity to the target; all analysis occurs at the attacker's location.

for direct human proximity and verbal communication, resulting in substantially lower ambient noise that improves the acoustic signal-to-noise ratio available to an attacker.

B. Attack Scenario

We consider an adversary whose goal is to infer proprietary motion parameters of a cobot operating in a sensitive manufacturing or research environment. Concretely, the attacker seeks to recover four categories of information from acoustic emissions alone:

1) **Direction of rotation** for each active joint (clockwise vs. counter-clockwise),
2) **Angular velocity** (speed) of each joint, discretized into RPM bins,
3) **Temporal sequencing** of multi-joint movements, enabling full trajectory reconstruction, and
4) **Control algorithm identity** (e.g., bang-bang, PID, or predictive control), revealing the sophistication of the underlying automation system.

C. Attacker Capabilities and Cost

We assume a resourceful but non-privileged attacker with the following capabilities:

Access level. The attacker can position a recording device within acoustic range (approximately 0.5–2 m) of the target robot. The attacker has *no* network access to the robot's control system, *no* physical contact with the robot, and *no* access to proprietary control software, sensor telemetry, or joint-state data. This access model is realistic: cobot work cells are open by design, and recording devices can be concealed as personal electronics, modified IoT sensors, or left behind during a facility visit.

Resources. The attacker has access to the same or a similar robot model for offline training data collection. This assumption is reasonable given that platforms like the Franka Emika Panda are commercially available (~$30,000 USD) and widely deployed in university robotics laboratories. The recording hardware required is a consumer-grade spatial microphone (e.g., Zoom H3-VR, ~$280 USD), bringing total equipment cost below $31,000. *For comparison, the IP at stake in aerospace or pharmaceutical manufacturing can be valued in the millions of dollars, making this a low-cost, high-reward attack.*

TABLE I

COBOT ACTUATOR SURVEY.

Robot	DOF	Motor Type	Transmission
Franka Panda[†]	7	BLDC[a]	Harmonic drive[a]
UR5e	6	BLDC/PMSM[b]	Harmonic drive[b]
KUKA LBR iiwa 14	7	BLDC/PMSM[c]	Harmonic drive[c]
Fanuc CRX-10iA	6	PMSM[d]	N/D[d]

[†] Platform used in this study. N/D = not disclosed.
Sources: [a] [22], [23]; [b] [24], [25]; [c] [26], [27]; [d] [28], [29] (transmission inferred).

Moreover, the acoustic attack vector is not specific to a single platform. Table I surveys four widely deployed cobots: all platforms with disclosed specifications use brushless permanent-magnet motors with harmonic drive transmissions, the two components that dominate the acoustic signature. Other cobots (ABB YuMi, Techman TM5) do not disclose actuator specifications but are expected to follow similar designs based on industry conventions [20], [21]. This shared actuator architecture suggests that classifiers trained on one platform's emissions may transfer to other cobots with minimal retraining, though cross-platform validation remains future work.

Knowledge. The attacker knows the robot model and general joint configuration but does not know the specific motion trajectories, speeds, or control algorithms used in the victim's operation.

D. Attack Phases

The attack proceeds in two phases, illustrated in Fig. 2:

Training Phase. The attacker records acoustic emissions from a reference robot (which may be the same unit, accessed during a legitimate visit, or a separate unit of the same model) executing known motion profiles. In our experiments, the training motion is a continuous sinusoidal velocity sweep that covers the full range of directions and speeds the robot can execute. Ground-truth labels (joint velocities at 1 kHz) are obtained from the robot's control interface. The attacker extracts spectral and spatial audio features from these recordings and trains classifiers offline.

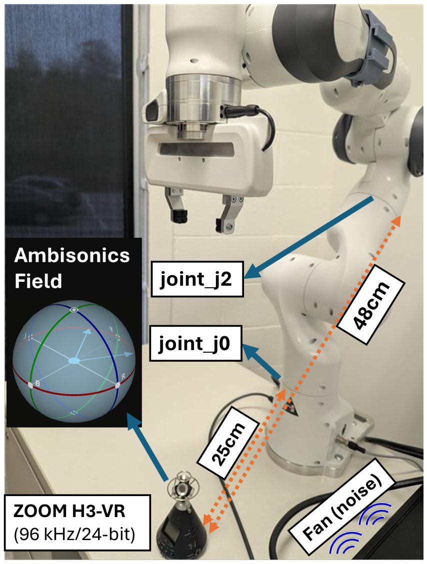

Fig. 3. Experimental setup: Franka Emika Panda with Zoom H3-VR Ambisonic recorder positioned near the robot base.

Inference Phase. The attacker deploys the trained classifiers against covert recordings of the victim robot executing unknown (proprietary) motion sequences. For each temporal analysis window in the recording, the model predicts direction and speed. Post-processing chains these per-window predictions into a reconstructed trajectory.

A critical aspect of this attack model is *cross-profile generalization*: classifiers trained on one motion profile (continuous sinusoidal sweeps) must generalize to fundamentally different motion profiles (discrete start-stop movements) encountered during the attack. An attacker cannot know *a priori* what motion profiles the victim robot will execute, so the classifier must learn acoustic features that transfer across motion types. We explicitly evaluate this cross-profile generalization as a central contribution of this work. Our experiments evaluate within-session cross-profile generalization; cross-session and cross-unit transferability remain open questions for future work.

IV. METHODOLOGY

A. Experimental Platform

We use a Franka Emika Panda, a 7-DOF cobot with torque-controlled joints driven by brushless DC (BLDC) servo motors coupled through harmonic drive gearboxes. The robot is controlled via the `libfranka` real-time interface at 1 kHz, providing direct access to joint positions, velocities, and torques. Acoustic emissions are captured using a Zoom H3-VR recorder in four-channel Ambisonic-A (AmbiX) format at 96 kHz with 24-bit resolution, positioned approximately 22 cm from the robot base on the workbench surface (Fig. 3).

B. Data Collection Protocol

Data is collected in two recording sessions, each executed by a single unified `libfranka` C++ program to ensure acoustic continuity between training and test phases within a session.

Session 1 (Multi-joint direction and speed). Training data consists of continuous sinusoidal velocity sweeps (0.4–1.0$\times$ maximum velocity, 0.55 rad/s) organized into three 3-minute sub-phases: Joint 0 only, Joint 2 only, and both joints simultaneously. A 10-second stationary pause separates training from the test phase and serves as the primary anchor for audio-to-CSV synchronization. The test phase consists of discrete point-to-point movements: 12 single-joint movements per joint at commanded speeds from $\pm$2.5 to $\pm$5.0 RPM (1.0 RPM increments), plus 10 simultaneous two-joint movements. Test movements follow trapezoidal velocity profiles (0.5 rad/s^2 acceleration, 1-second hold, symmetric deceleration) with 2-second pauses.

Session 2 (Controller comparison). Three velocity-level controllers (bang-bang, PID, and predictive) each execute 10 identical point-to-point movements on Joint 0, with target offsets of $\pm$0.3, $\pm$0.4, and $\pm$0.5 rad from the home position (with selected magnitudes repeated to reach 10 movements per controller). Controllers are executed sequentially in a single recording, yielding 60 movement groups (10 movements $\times$ 2 phases [to-target and return-to-home] $\times$ 3 controllers). We note that sequential execution introduces a potential temporal confound (the classifier could partially exploit recording-time position rather than acoustic content); we address this limitation in Section V-C.

The bang-bang controller commands constant velocity with abrupt transitions; the PID controller produces exponential decay toward the target; the predictive controller minimizes jerk, yielding a bell-shaped velocity profile. These three strategies produce fundamentally different mechanical excitation patterns.

Audio and joint-state streams are synchronized post-hoc via normalized cross-correlation between the W-channel RMS energy envelope and the sum-of-absolute-velocities envelope, both at 50 Hz, exploiting the 10-second silence as an alignment anchor (20 ms precision).

C. Audio Preprocessing

Raw recordings contain stationary interference (HVAC, building vibration, electronic hum) and transient events unrelated to servo activity. We isolate motor-active segments and suppress noise prior to feature extraction. The pipeline operates on the raw four-channel recording with Short-Time Fourier Transform (STFT) parameters $N_{\text{FFT}} = 2048$ and hop size $H = 512$ at $f_s = 96$ kHz, giving a frame rate of $\sim$187.5 frames/s and a frequency resolution of $\sim$46.9 Hz/bin.

*1) **Motion Detection via Spectral Flux:*** Motion detection operates on the omnidirectional W channel, which provides the highest broadband SNR. We compute spectral flux within the *servo whine band* (2–9 kHz), the dominant harmonic range of the Panda's BLDC motors and harmonic drive gearboxes. Spectral flux measures the frame-to-frame increase in spectral energy, providing a reliable indicator of motor engagement onset that is robust to stationary background noise:

$$\Phi(t) = \sum_{f \in \mathcal{B}} \max(0, |Z(f,t)| - |Z(f,t-1)|) \tag{1}$$

where $\mathcal{B}$ denotes the servo band and $Z(f,t)$ is the STFT coefficient at frequency f and frame t. Half-wave rectification focuses detection on energy onset events characteristic of motor engagement, suppressing post-movement ring-down that would otherwise extend segment boundaries.

The detection threshold is self-tuning: $\tau = \Phi_{15}+(1.1-\alpha)\cdot 12.0\cdot\sigma_{\text{low}}$, where Φ_{15} is the 15th-percentile flux (noise floor estimate), σ_{low} is the standard deviation of sub-median flux values, and α is a sensitivity parameter. The binary mask is smoothed with a median filter spanning $\sim$0.4 s to close brief intra-movement gaps, and edges are softened with a 50 ms fade window to prevent amplitude discontinuities that would introduce broadband spectral artifacts into downstream features.

*2) **Noise Suppression**:* Frames identified as non-motion provide a per-frequency noise profile $\hat{N}(f)$ via median spectral magnitude across quiet frames. This profile drives a soft Wiener filter that attenuates frequency bins dominated by stationary noise while preserving motor-generated spectral content, applied independently to each Ambisonic channel:

$$G(f,t) = \text{clip}\left(1-\beta\cdot\frac{\hat{N}(f)}{|Z(f,t)|+\epsilon},\ 0.2,\ 1.0\right) \tag{2}$$

where $\beta \in [0,1]$ controls denoise strength and $\epsilon = 10^{-9}$ ensures numerical stability. The gain floor of 0.2 preserves low-energy spectral components that may carry directional motor information in the spatial channels, avoiding the musical noise artifacts produced by hard spectral subtraction. Each channel is reconstructed via ISTFT with matching window parameters.

*3) **Head-Mute and Output**:* A 2-second head mute zeros the cleaned audio at recording onset to suppress servo drive initialization transients. The pipeline outputs cleaned four-channel audio alongside a CSV of movement segment timestamps derived from motion masks.

D. Feature Extraction

For each temporal analysis window (250 ms, non-overlapping), we extract spectral, spatial, and temporal features from all four Ambisonic channels. The 250 ms window balances frequency resolution for MFCC computation with temporal sensitivity to speed transitions in the 0–6 RPM operating range.

The four Ambisonic channels encode the sound field in AmbiX format [30]: W (omnidirectional sound pressure), Y (left-right particle velocity), Z (up-down particle velocity), and X (front-back particle velocity). This spatial decomposition allows analysis of how motor acoustic energy radiates directionally, which varies with joint rotation direction and joint position on the robot arm.

1) Spectral Features (per channel): From each channel we compute:

MFCCs. 13 Mel-frequency cepstral coefficients using 128 Mel bands [31]. We retain the first 13 coefficients, following convention in both speech [32] and mechanical sound classification [9]: the lower-order coefficients capture the gross spectral envelope (overall energy distribution across frequency), while higher-order coefficients encode fine spectral detail that is typically noise-dominated at the SNR levels present in our recordings.

Chroma features. 12 chroma bins capturing the distribution of spectral energy across pitch classes, encoding harmonic content related to motor resonance frequencies.

Extracting 25 spectral features (13 MFCC + 12 chroma) from each of four channels yields 100 spectral features per window.

2) Spatial Features: We compute 10 features that capture the directional distribution of acoustic energy. For each channel $c \in \{W,Y,Z,X\}$, the RMS energy is: $E_c = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x_c[n]^2}$, where $x_c[n]$ denotes the n-th sample in channel c and N is the window length. From these energies we derive:

- Channel energies: E_W, E_Y, E_Z, E_X (4 features)
- Energy ratios: E_Y/E_W, E_X/E_W, E_Z/E_W (3 features)
- Spatial differences: $E_Y - E_X$ and $(E_Y - E_X)/(E_W+\epsilon)$ (2 features)
- Directional ratio: $E_Y/(E_X+\epsilon)$ (1 feature)

where $\epsilon = 10^{-10}$ ensures numerical stability. The Y-channel features are particularly informative for base joint (J0) direction classification because J0 rotates horizontally, producing asymmetric left-right acoustic radiation that modulates Y-channel energy. The total feature vector is 110 dimensions (100 spectral + 10 spatial).

3) Temporal Features: To capture motor acceleration and deceleration dynamics, we augment the feature set with temporal derivatives computed from the cached MFCC matrices (no recomputation required):

Delta MFCCs (Δ). First-order temporal derivatives of MFCC coefficients within each 250 ms window, summarized as mean and standard deviation across frames (26 features per channel).

Delta-Delta MFCCs ($\Delta\Delta$). Second-order derivatives (13 mean features per channel).

Energy dynamics. Four features computed from 8 sub-frame energy measurements within each window: energy slope (linear regression coefficient), mean, standard deviation, and range of sub-frame RMS energies.

This yields 160 temporal features (39×4 channels + 4 energy dynamics), extending the direction feature vector to 270 dimensions and the speed feature vector to 260 dimensions. Speed features omit spatial features, which encode direction-specific information irrelevant to velocity estimation.

E. Classification

1) Direction Classification: We train a Random Forest classifier [33] (200 estimators, max depth 20) for binary classification of clockwise (CW) vs. counter-clockwise (CCW) rotation. Direction labels map to sign of joint velocity at each window's center timestamp: negative velocity indicates CW, positive indicates CCW. Windows where the absolute velocity falls below 0.05 rad/s are excluded, filtering out stationary and near-zero-velocity samples. Training data is balanced at 7,500 samples per direction class (15,000 total).

2) Speed Classification: Speed is discretized into bins of width $\Delta v = 1$ RPM, producing 6 classes spanning the 0–6 RPM operating range. Separate Random Forest classifiers are trained for CW and CCW directions. This direction-specific approach accounts for asymmetries in motor loading and acoustic coupling between rotation directions: for example, clockwise rotation of the base joint loads the cable harness differently than counter-clockwise rotation, producing subtly different acoustic profiles at the same angular velocity. Speed labels are derived from the absolute joint velocity converted to RPM: $\omega_{\text{RPM}} = |v_{\text{rad/s}}| \times 60/(2\pi)$.

We report both exact bin accuracy and adjacent bin accuracy (± 1 bin), measuring the model's ability to predict speed within ± 1 RPM of the true value.

F. Multi-Joint Source Separation

For multi-joint analysis, individual joint contributions must be distinguished from the composite acoustic signal. We exploit the spatial selectivity of Ambisonic recording: joints at different locations radiate sound from different directions, producing distinct spatial signatures across channels.

We evaluate two experiments. **Experiment A** (single-joint baselines) trains and tests direction and speed classifiers independently for Joint 0 (base rotation) and Joint 2 (elbow), establishing whether joints at different positions along the arm produce classifiable acoustic signatures. **Experiment B** (joint identification) trains a binary classifier to distinguish which joint is the dominant acoustic source in each window, using spatial features. Joint 0 (base, near microphone) produces high W-channel energy, while Joint 2 (elbow, elevated) produces relatively higher Z-channel energy.

V. Results

Direction and speed classifiers are evaluated under a cross-profile generalization protocol: training data consists of continuous sinusoidal velocity sweeps, while test data consists of discrete trapezoidal point-to-point movements that the classifier has never seen during training. This mimics the realistic attack scenario in which the attacker's training motion profile differs from the victim's operational motions. Controller classification (Section V-C) uses 5-fold grouped cross-validation on identical motion types.

A. Feature Ablation

To quantify the contribution of each feature group, we perform an incremental ablation study on Joint 0 direction classification (Table II). Starting from a single omnidirectional channel (W-only, 57.20%), expanding to all four Ambisonic channels yields 71.61%, a 14.4-percentage-point gain from the directional particle-velocity channels (Y, Z, X) that carry spectral information absent from the omnidirectional channel alone. Adding the 10 spatial features yields 74.20%, a modest but efficient 2.6 pp gain from only 10 additional dimensions, confirming that inter-channel energy ratios encode directional information beyond what per-channel spectral features capture independently. Notably, speed accuracy is unchanged by spatial features (92.2% adjacent in both configurations), confirming their contribution

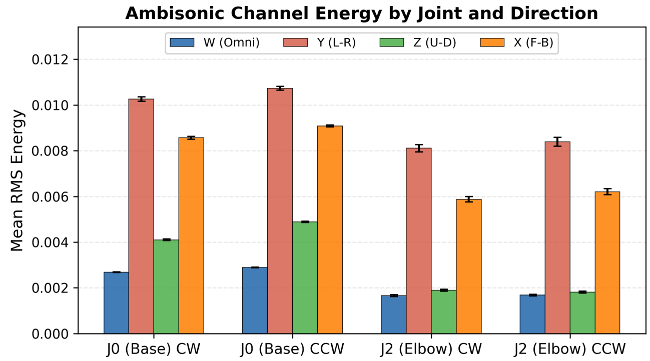

Fig. 4. Mean Ambisonic channel RMS energy by joint and direction (error bars: ± 1 SE). Joint 0 shows Y-channel asymmetry between CW and CCW rotation; Joint 2 shows none. This confirms that horizontal base rotation produces laterally asymmetric radiation, explaining the spatial feature gain in Table II and the inter-joint gap in Table III.

TABLE II

Feature ablation for J0 direction classification. Each row adds a feature group to the previous configuration.

Feature Set	Dim.	Accuracy (%)
W-channel only (mono)	25	57.20
All 4 channels (spectral)	100	71.61
+ Spatial features	110	74.20
+ Temporal features (full)	270	78.41

is direction-specific. The full 270-dimensional feature set achieves 78.41% direction accuracy.

Fig. 4 illustrates the underlying mechanism: Joint 0's horizontal rotation displaces the source laterally, modulating Y-channel energy asymmetrically between directions, while Joint 2's vertical rotation axis produces symmetric radiation.

B. Multi-Joint Direction and Speed

Table III summarizes direction, speed, and joint identification results for both joints under cross-profile evaluation.

Direction. Joint 0 (base) achieves 78.41% direction accuracy, substantially above the 50% chance level expected from random binary classification (Fig. 6). Clockwise and counter-clockwise classes are balanced: CW precision/recall = 0.78/0.80, CCW precision/recall = 0.79/0.77, indicating that the classifier does not exploit class imbalance. Joint 2 (elbow) achieves only 57.32%, with a pronounced recall asymmetry (CW: 0.76, CCW: 0.39). This disparity reflects the spatial geometry discussed in Section V-A: Joint 0 rotates horizontally at the robot base near the microphone, producing a strong left-right acoustic asymmetry, while Joint 2 rotates about a vertical axis farther from the microphone, generating weaker and less directionally distinct emissions.

Speed. Speed classification shows a different pattern. Joint 0 achieves 53.15% exact bin accuracy (6 classes, chance = 16.7%) and 94.82% adjacent accuracy (± 1 RPM). Joint 2 achieves 48.54% exact and 89.67% adjacent. The smaller gap between joints for speed (4.6 pp) compared to direction (21.1 pp) supports a key finding: speed classification depends primarily on motor commutation frequency and spectral energy distribution, which are physics-based properties that generalize across joint positions. Direction

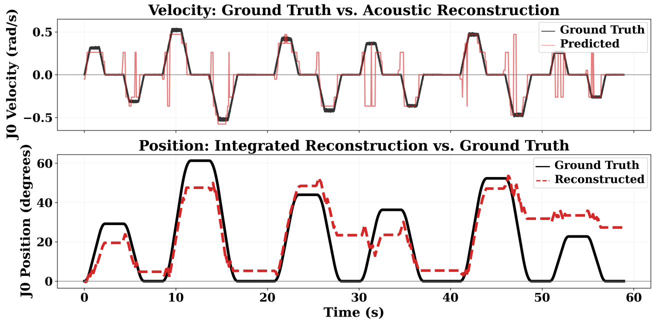

Fig. 5. Trajectory reconstruction: velocity predictions (top) and integrated position (bottom) vs. ground truth for 12 test movements (13.1° RMSE).

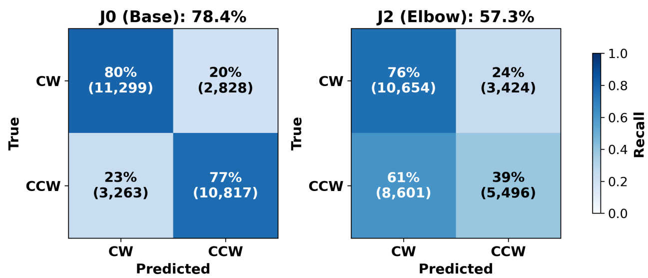

Fig. 6. Direction confusion matrices under cross-profile evaluation. Joint 0 shows balanced discrimination; Joint 2 exhibits CW bias (CCW recall: 39%).

classification depends on spatial radiation patterns that vary with joint geometry.

Joint identification. A binary classifier distinguishing Joint 0 from Joint 2 acoustic windows achieves 95.90% accuracy. Feature importance analysis reveals that spectral timbre features (MFCCs) dominate over spatial features: the top-ranked spatial feature (Y/X energy ratio) has importance 0.0031, an order of magnitude below the leading MFCCs. This indicates that different joints produce distinct spectral signatures due to differences in motor size, harmonic drive ratio, and mechanical load, rather than relying solely on spatial position relative to the microphone.

Together, these three capabilities enable end-to-end trajectory reconstruction. For each 250 ms window (50 ms step), the direction classifier outputs a sign ($+1$ for CCW, -1 for CW) and the speed classifier outputs an RPM bin, yielding a signed angular velocity estimate $\hat{\omega}(t)$. The joint position is then recovered by discrete integration: $\hat{\theta}(t) = \sum_{k=0}^{t} \hat{\omega}(k)\,\Delta t$, with energy-gated pauses held at zero velocity to prevent drift accumulation. Applied to 12 unseen test movements, this recovers the Joint 0 trajectory with 13.1° RMSE relative to ground truth (Fig. 5).

C. Controller Classification

Beyond motion parameters, we evaluate whether the control algorithm generating the robot's motion can be identified from acoustic emissions alone. The three controllers described in Section IV-B, i.e., bang-bang, PID, and predictive, each execute identical point-to-point movements on Joint 0 but produce qualitatively distinct velocity profiles (Fig. 7): bang-bang exhibits trapezoidal profiles with abrupt onset and offset transitions, PID shows exponential decay toward the target with occasional settling oscillations, and predictive produces smooth S-curve ramps with gradual transitions.

TABLE III

MULTI-JOINT CLASSIFICATION UNDER CROSS-PROFILE EVALUATION (TRAIN: SINUSOIDAL SWEEPS, TEST: DISCRETE MOVEMENTS).

Task	Joint	Accuracy (%)	Adj. Acc. (%)
Direction	J0 (base)	78.41	—
Direction	J2 (elbow)	57.32	—
Speed	J0 (base)	53.15	94.82
Speed	J2 (elbow)	48.54	89.67
Joint ID	J0 vs. J2	95.90	—

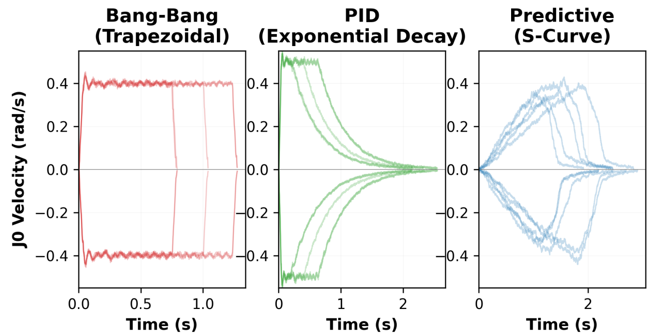

Fig. 7. Joint 0 velocity profiles for all recorded movements under each controller: bang-bang (left), PID (center), predictive (right).

Using 5-fold grouped cross-validation (CV) (grouped by movement to prevent data leakage between windows of the same movement), the Random Forest classifier achieves 99.99% ± 0.02% accuracy across 80,169 windows spanning 60 movement groups (Table IV). Only 10 windows are misclassified, all predictive windows labeled as PID. This is consistent with the velocity profiles: the predictive controller's smooth trajectories occasionally resemble PID's exponential settling during deceleration phases.

Feature importance analysis confirms that spectral features dominate (87.6%), with MFCC-4 across multiple channels ranking highest. This aligns with the physical interpretation: the controllers produce fundamentally different temporal patterns of motor loading, which creates distinct spectral energy distributions.

From the attacker's perspective, identifying the control algorithm enables controller-specific motion models for subsequent trajectory reconstruction, improving velocity estimation accuracy by matching the expected profile shape.

Caveat. The three controllers were executed sequentially within a single recording session, so the classifier could partially exploit slow temporal drift in environmental acoustics. Feature importance analysis argues against this: MFCC-4 (mid-frequency spectral shape) dominates across all four channels, consistent with motor-driven discrimination rather than broadband environmental drift. Nevertheless, interleaved controller ordering would strengthen this result. The effective sample size is 60 movement groups (the unit of grouped cross-validation), not the 80,169 feature windows.

VI. CONCLUSION

We presented the first acoustic side-channel attack on a collaborative robot, demonstrating that an attacker with

TABLE IV
CONTROLLER CONFUSION MATRIX (5-FOLD GROUPED CV)

True	**Predicted** Bang-bang	PID	Predictive
Bang-bang	20,198	0	0
PID	0	27,372	0
Predictive	0	10	32,589

a single Ambisonic microphone can infer joint direction (78.4%), speed (94.8% within ± 1 RPM), active joint identity (95.9%), and control algorithm (99.99%) from the Franka Emika Panda's acoustic emissions. Feature ablation confirms that four-channel spatial audio provides a 14.4 pp gain over monaural recording for direction, while speed depends on physics-based spectral properties that generalize across joints. End-to-end trajectory reconstruction recovers joint position with 13.1° RMSE. Current limitations include single-unit evaluation at close range; cross-session generalization remains an open challenge. Potential countermeasures include acoustic enclosures, active masking in motor frequency bands, and trajectory obfuscation [9]. These results establish cobots as viable targets for acoustic reconnaissance and motivate countermeasures for security-critical deployments.

ACKNOWLEDGMENT

Gemini (Google) and Claude (Anthropic) were used to assist with manuscript editing, code debugging, and figure refinement. All technical content, experimental design, and scientific interpretation are the sole work of the authors.

REFERENCES

[1] V. Villani, F. Pini, F. Leali, and C. Secchi, "Survey on human-robot collaboration in industrial settings: Safety, intuitive interfaces and applications," *Mechatronics*, vol. 55, pp. 248–266, 2018.

[2] U.S. Department of Justice, "PRC state-owned company, Taiwan company, and three individuals charged with economic espionage," DOJ Office of Public Affairs, Press Release No. 18-1404, Nov. 2018, available: https://www.justice.gov/opa/pr/prc-state-owned-company-taiwan-company-and-three-individuals-charged-economic-espionage.

[3] ——, "Chinese intelligence officer charged with economic espionage involving theft of trade secrets from leading U.S. aviation companies," DOJ Office of Public Affairs, Press Release No. 18-1336, Oct. 2018, available: https://www.justice.gov/opa/pr/chinese-intelligence-officer-charged-economic-espionage-involving-theft-trade-secrets-leading.

[4] J. Gatlin, S. Belikovetsky, Y. Elovici, A. Skjellum, J. Lubell, P. Witherell, and M. Yampolskiy, "Encryption is futile: Reconstructing 3D-printed models using the power side-channel," in *24th International Symposium on Research in Attacks, Intrusions and Defenses (RAID)*. ACM, 2021.

[5] D. Genkin, A. Shamir, and E. Tromer, "RSA key extraction via low-bandwidth acoustic cryptanalysis," in *Advances in Cryptology – CRYPTO 2014*. Springer, 2014, pp. 444–461.

[6] D. Asonov and R. Agrawal, "Keyboard acoustic emanations," in *Proceedings of the IEEE Symposium on Security and Privacy*, 2004.

[7] T. I. Tamiti, B. Joshi, R. Hasan, and A. Barua, "Hvac-ear: Eavesdropping human speech using hvac systems," 2025. [Online]. Available: https://arxiv.org/abs/2510.01082

[8] ——, "Reconstructing intelligible speech from the pressure sensor data in hvacs," 2025. [Online]. Available: https://arxiv.org/abs/2506.22311

[9] M. A. Al Faruque, S. R. Chhetri, A. Canedo, and J. Wan, "Acoustic side-channel attacks on additive manufacturing systems," in *Proceedings of the ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS)*, 2016, pp. 1–10.

[10] S. R. Chhetri, A. Barua, S. Faezi, F. Regazzoni, A. Canedo, and M. A. Al Faruque, "Tool of spies: Leaking your ip by altering the 3d printer compiler," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 2, pp. 667–678, 2019.

[11] M. Backes, M. Dürmuth, S. Gerling, M. Pinkal, and C. Sporleder, "Acoustic side-channel attacks on printers," in *Proceedings of the 19th USENIX Security Symposium*, 2010, pp. 307–322.

[12] S. R. Chhetri, A. Canedo, and M. A. Al Faruque, "KCAD: Kinetic cyber-attack detection method for cyber-physical additive manufacturing systems," in *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2016, pp. 1–8.

[13] C. Song, F. Lin, Z. Ba, K. Ren, C. Zhou, and W. Xu, "My smartphone knows what you print: Exploring smartphone-based side-channel attacks against 3D printers," in *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, 2016.

[14] W. Yang, J. Chen, C. Zhang, and K. Paynabar, "Online detection of cyber-incidents in additive manufacturing systems via analyzing multimedia signals," *Quality and Reliability Engineering International*, vol. 38, no. 3, pp. 1340–1356, 2021.

[15] R. Shah, C. M. Ahmed, and S. Nagaraja, "Fingerprinting robot movements via acoustic side channel," *arXiv preprint*, 2022. [Online]. Available: https://arxiv.org/abs/2209.10240

[16] P. Vijayraghavan and R. Krishnan, "Noise in electric machines: A review," in *Proceedings of the IEEE Industry Applications Conference*, vol. 1, 1999, pp. 251–258.

[17] Franka Emika GmbH, *Franka Emika Robot Product Manual*, 2021. [Online]. Available: https://download.franka.de/documents/100010_Product%20Manual%20Franka%20Emika%20Robot_10.21_EN.pdf

[18] Zoom Corporation, *H3-VR Handy Recorder Operation Manual*, 2018. [Online]. Available: https://zoomcorp.com/media/documents/E_H3-VR_2.pdf

[19] Occupational Safety and Health Administration, "Occupational noise exposure standard," 29 CFR 1910.95, 1983, amended 2008. Available: https://www.osha.gov/laws-regs/regulations/standardnumber/1910/1910.95.

[20] Z. Dong *et al.*, "Impedance control of collaborative robots with harmonic drive transmissions," *Asian Journal of Control*, 2024.

[21] E. Madsen, O. S. Rosenlund, D. Brandt, and X. Zhang, "Comprehensive modeling and identification of nonlinear joint dynamics for collaborative industrial robot manipulators," *Control Engineering Practice*, vol. 101, p. 104462, 2020.

[22] S. Haddadin, S. Parusel, L. Johannsmeier, S. Golz, S. Gabl, F. Walch, M. Sabaghian, C. Jähne, L. Hausperger, and S. Haddadin, "The Franka Emika robot: A reference platform for robotics research and education," *IEEE Robotics & Automation Magazine*, 2022.

[23] C. Gaz, M. Cognetti, A. Oliva, P. Robuffo Giordano, and A. De Luca, "Dynamic identification of the Franka Emika Panda robot with retrieval of feasible parameters using penalty-based optimization," *IEEE Robotics and Automation Letters*, 2019.

[24] D. De Santis, A. M. Giordano, and M. Garabini, "Electro-mechanical modeling and identification of the UR5 e-series robot," *Robotica*, 2024.

[25] Kollmorgen, "Kollmorgen drives the lightweight helpers of Universal Robots," Automate.org, 2015, available: https://www.automate.org/news/kollmorgen-drives-the-lightweight-helpers-of-universal-robots.

[26] German Aerospace Center (DLR), "Mechatronic design of the LWR III," DLR Institute of Robotics and Mechatronics, 2023, available: https://www.dlr.de/en/rm/research/robotic-systems/arms/lwr-iii/mechatronic-design-of-the-lwr-iii.

[27] A. Albu-Schäffer, S. Haddadin, C. Ott, A. Stemmer, T. Wimböck, and G. Hirzinger, "The DLR lightweight robot: Design and control concepts for robots in human environments," in *Industrial Robot: An International Journal*. Emerald, 2007, vol. 34, no. 5, pp. 376–385.

[28] FANUC Corporation, *AC Servo Motor αi Series Descriptions*, 2020, available: https://www.mroelectric.com/static/app/product/pdfs/A860-2001-T301.pdf.

[29] Premier Automation, "FANUC CRX10iA cobot," Premier Automation Blog, 2021, available: https://info.premierautomation.com/blog/fanuc-crx10ia-cobot.

[30] M. A. Gerzon, "Periphony: With-height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973.

[31] B. McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "librosa: Audio and music signal analysis in Python," in *14th Python in Science Conference*, 2015, pp. 18–24.

[32] S. Farhadi, M. Corrado, O. Borla, and G. Ventura, "Prestressing wire breakage monitoring using sound event detection," *Computer-Aided Civil and Infrastructure Engineering*, 2023.

[33] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.