ANTLER 1: A Multimodal Co-Creative Tutor for Stylistic Exploration in Tambour Embroidery

Anne Toomey[1], **Zowie Broach**[1], **Mojtaba Kargar**[1], **Samira Valorian**[1], **Henrietta Dent**[1], **Aemika Zou**[2], **Chaonan Lin**[2]

[1]The Royal College of Art (RCA), UK
[2]Laboratory for Artificial Intelligence in Design (AiDLab), Hong Kong

anne.toomey@rca.ac.uk, zowie.broach@rca.ac.uk, mojtaba.kargar@rca.ac.uk, samira.valorian@rca.ac.uk
henrietta.dent@rca.ac.uk,aemika.zou@connect.polyu.hk,chaonan.lin@rca.ac.uk

Abstract

Tambour embroidery is an embodied practice where expert knowledge lives in hand rhythm, needle angle, and material control. We present ANTLER 1, a style-aware co-creative tutoring *framework* for embodied craft, instantiated as a working multimodal prototype. The prototype captures live tambour stitching through synchronised video, Inertial Measurement Unit (IMU) telemetry, and hand tracking, transforming segmented stitch behaviour into profile-conditioned coaching feedback. We describe the framework design and present illustrative use-case scenarios showing how the same observed sequence is interpreted differently under precision, expressive, and hybrid profiles — contributing an early-stage design pattern for grounded, style-aware co-creative tutoring in embodied craft.

Introduction

Tambour embroidery emerged from Indian Ari hook work and later entered Ottoman and French couture traditions, yet its expertise remains embodied in wrist rhythm, hook angle, and material control rather than verbal instruction alone (Mora and others 2024; Salter and others 2010). As master artisans age, this tacit knowledge faces irreversible loss (Mora and others 2024; Boden 1992). Existing digital tools orient toward correctness and do not help practitioners explore how the same movement carries different creative meanings under different stylistic intentions.

ANTLER 1 frames AI tutoring as a co-creative, style-aware interaction grounded in observed stitch behaviour. The system acts as a computationally grounded coaching partner: given a target design context, observed stitch features, and practical constraints, it generates *what-if alternatives* that let the practitioner compare efficiency-oriented, expressive, and hybrid execution paths. Prior work addresses multimodal performance analysis and Large Language Model (LLM)-based tutoring, but does not address style-contingent feedback grounded in embodied craft action; ANTLER 1 bridges this gap (Boden 1992; Ritchie 2007; Yannakakis, Liapis, and Alexopoulos 2018; Davis and others 2013).

Three research questions structure this paper:

- **RQ1:** How can multimodal segmentation of tambour stitching support style-aware feedback in embodied creative practice?
- **RQ2:** How does grounding LLM feedback in observed performance features change perceived usefulness compared to ungrounded generation?
- **RQ3:** How can a co-creative tutor vary not only its recommendations but also its tutoring stance across precision, expressive, and hybrid profiles?

The contribution is a *design framework and early-stage prototype*, not a fully validated end-to-end system; claims are scoped accordingly throughout.

A corrective system applies a single target state; a co-creative system maintains multiple valid states and helps the practitioner navigate between them (Yannakakis, Liapis, and Alexopoulos 2018; Boden 1992). ANTLER 1 operationalises this through style profiles that redefine what counts as deviation: the same IMU signature that triggers a correction under precision may invite expressive exploration under an expressive profile (Franceschelli and Musolesi 2024; Anonymous 2024).

Related Work

Multimodal skill-analysis systems in surgery and motion training integrate video and kinematic traces to assess performance, but they optimise for correctness or safety rather than style-sensitive interpretation in creative craft practice (Pan, Wang, and others 2023; Chen et al. 2025). They demonstrate the feasibility of multimodal fusion for performance feedback, but their single-target optimisation is fundamentally incompatible with a domain where the same action may simultaneously be a technical deviation and a valid stylistic gesture.

LLM-based tutoring systems operate over textual or symbolic inputs and cannot distinguish practitioner-specific rhythmic signatures from generic craft description (Jain and others 2025; Mills and others 2025). Within computational creativity, co-creative systems are understood as interactive partners rather than autonomous generators, evaluated for situated interaction over output quality alone (Yannakakis, Liapis, and Alexopoulos 2018; Davis and others 2013; Mora and others 2024; Yannakakis and Togelius 2014).

The paper's two conceptual baselines are literature-grounded: a rule-based correction system aligned with prior motion-assessment work, and a vanilla LLM receiving style descriptions without multimodal evidence, aligned with text-

Figure 1: ANTLER 1 capture setup during a live tambour embroidery session. Overhead and underside cameras record stitch trajectory, fabric interaction, and embodied hand movement in real time. A sensor-integrated tambour tool captures fine-grained needle-motion dynamics used for segmentation and downstream co-creative coaching feedback.

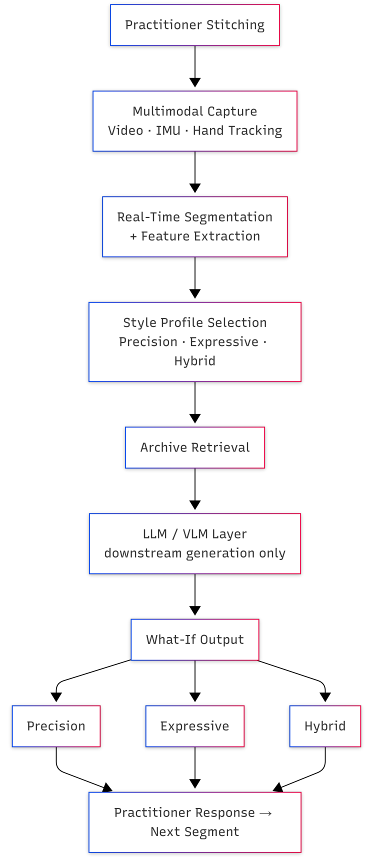

Figure 2: ANTLER 1 system pipeline. IMU wrist-velocity drives both feature extraction and real-time stitch boundary detection. Style profile selection modulates how the same feature is interpreted before retrieval and generation.

centric coaching systems. ANTLER 1 sits between and beyond these baselines by combining sensor grounding with style-conditioned stance variation. Recent surveys highlight explainability and process-transparency as core properties distinguishing genuine co-creative AI from post-hoc description systems (Llano et al. 2020; Bryan-Kinns and others 2023; Franceschelli and Musolesi 2024; Anonymous 2024); the style-profile architecture and what-if framing of ANTLER 1 are direct responses to these design requirements.

System Overview

Capture. Input: live tambour session. Output: synchronised video stream, IMU telemetry (100 Hz), and hand-skeleton joint positions at 30 fps. Figure 2 illustrates how three sensor streams combine. Video captures stitch trajectory and spatial layout; IMU captures wrist dynamics and rhythm; hand skeleton captures grip posture and needle-hand coordination. Table 2 maps each modality to extracted features and craft-relevant interpretation.

Segmentation. A stitch boundary is defined as the moment of minimum wrist velocity between a completed pull-through and initiation of the next insertion. Boundaries were established by manual annotation (Label Studio); a segmentation model trained on these labels performs real-time boundary detection at runtime via IMU-derived wrist-velocity drops, covering Pull, Chain, and Continuous Bead Insertion stitch types.

Feature Extraction. Spacing compression: percentage deviation from mean inter-bead distance. Tempo variance: coefficient of variation of inter-boundary durations. Wrist rotation direction is derived from gyroscope data; pull-through micro-adjustments are short-duration velocity spikes within a segment.

Style profiles. Style profiles are operational parameter sets, not prompt themes. Each encodes target regularity, variation tolerance, tempo elasticity, and tutoring stance. Table 1 summarises the three profiles. The key profile-dependent feature is error type G (Rhythm Arrhythmia): under Precision it is penalised; under Expressive the same rhythmic pattern may be reinterpreted as creative hesitation; under Hybrid it is permitted only at designated focal points.

Retrieval. Archive entries are ranked by joint match on `error_type`, `style_profile_tag`, and `stitch_type`; ties broken by recency in session history. Each entry stores stitch type, error labels [A–G], feature summary, technique reference, and stance tag.

Generation. The LLM receives a structured prompt containing the top-ranked archive entry, the current feature summary, and the active style profile; it renders these into coaching text and performs no independent classification.

Grounding example. For a Chain segment with high tempo

Modality	Features	Craft relevance
Video	spacing, curvature, trajectory density	regularity, motif geometry
IMU	tempo elasticity, motion drops, wrist rotation direction, pull-through micro-adjustments	rhythm, directional control, gestures associated with tension management
Skeleton	grip openness, wrist angle, coordination	technique type, needle control

Table 1: Modality-to-feature mapping in the current prototype. The IMU row reports motion patterns associated with tension management rather than direct measurement of thread tension.

Profile	Interpretation	Stance
Precision	variation reduced; type G penalised	corrective
Expressive	type G reframed; intentionality sought	reflective
Hybrid	controlled base; focal variation OK	curatorial

Table 2: Style profiles and tutoring stances. Error types A–G form a curated taxonomy orthogonal to style profile tags.

variance, 8% spacing compression, and clockwise wrist rotation drift: under the Precision profile the system generates *"Your chain is tightening and rotating on the pull-through. Stabilise the wrist earlier to keep spacing even."* Under the Expressive profile the same features yield *"The directional drift gives the chain organic movement. Repeat it at the motif edge to make it read as deliberate style."* This single-input, multi-output behaviour is the operational definition of style-aware co-creative tutoring in ANTLER 1.

Iterative loop. Each completed stitch triggers boundary detection, feature extraction, profile-conditioned retrieval, and feedback generation without intervention. The practitioner's response informs the next segment; an active learning component flags uncertain segments for expert review over multiple sessions (Yannakakis, Liapis, and Alexopoulos 2018; Davis and others 2013).

Use Cases and What-If Analysis

The scenarios below are illustrative mechanism examples constructed from a representative observed sequence to demonstrate the framework's profile-dependent behaviour. They draw on outputs generated during the pilot session but are presented as design demonstrations; practitioner responses are reported in Section .

All three scenarios share one observed sequence: a floral transition showing spacing compression, directional drift during pull-through, and uneven rhythmic control. A corrective system flags this as deviation; ANTLER 1 surfaces profile-dependent alternatives showing the same evidence admits multiple valid interpretations.

Technique background. *Lunéville crochet* works from the reverse of the fabric in a continuous chain, suited to repetitive, linear, high-volume bead application where consistency and speed are priorities. *Ari embroidery* works from the surface with a fine awl-hook, giving finer control over direction, spacing, and ornamental density at the cost of speed — supporting organic, non-linear motifs. The choice reflects an atelier judgement about where efficiency should yield to expressive intent: precisely the trade-off ANTLER 1's what-if layer is designed to surface.

Scenario A — Precision. Under this profile the system interprets the hesitation as a signal to restore base rhythm and spacing regularity. It surfaces Lunéville crochet as an efficiency-oriented alternative: continuous thread construction reduces bead-placement pauses and supports consistent chain spacing. The tutoring tone is corrective but supportive — recommending tightened tempo while preserving the practitioner's grip angle.

Scenario B — Expressive. The same hesitation is reframed as a possible stylistic breath, provided it is repeated intentionally. The system endorses Ari technique, whose individual pick-and-place method supports organic, non-linear bead orientation and richer ornamental surfaces. The tutoring tone is reflective, asking whether the pause marks compositional accent rather than mechanical difficulty.

Scenario C — Hybrid. The system preserves the controlled base chain (Lunéville register) while allowing expressive hesitation at focal floral nodes. This scenario most clearly demonstrates co-creative partnership: the system presents a conditional trade-off rather than prescribing a technique (Llano et al. 2020; Yannakakis, Liapis, and Alexopoulos 2018).

Figure 3: Left: Ari technique — surface-worked, individual bead placement, organic forms. Right: Lunéville crochet — reverse-worked, continuous thread, linear bead lines. The same IMU hesitation pattern can support recommending either technique depending on the active style profile.

What-if generation draws on four grounded inputs: observed motion features (tempo elasticity, directional drift, spacing, pull-through micro-adjustments); the active style profile; a technique archive mapping Lunéville and Ari characteristics to design intent; and a time map derived from the practitioner's own session rhythm. Time estimates are indicative and user-specific, not general benchmarks (Jain and others 2025; Mills and others 2025).

Table 3 operationalises the "same evidence, different meaning" claim by showing how identical observed features yield different technique recommendations under each profile.

Feature	Precision	Expressive	Hybrid
Hesitation (type G)	Correct: restore tempo	Reframe: stylistic breath	Allow at focal nodes
Spacing compression	Reduce: regularise	Preserve if intentional	Tighten base chain
Technique suggestion	Lunéville (efficiency)	Ari (expression)	Lunéville base + Ari focal
Tutoring tone	Corrective	Reflective	Curatorial

Table 3: Same observed sequence interpreted across three style profiles. This table operationalises style-contingent interpretation as the system's core CC contribution.

Preliminary Evaluation

The segmentation model was trained on data from 50 recorded sessions covering three stitch types: Pull, Chain, and Continuous Stitch with Black Bead insertion. Boundaries and error types were annotated manually before model training.

A single-session evaluation was then conducted with one expert embroiderer (5+ years atelier experience) performing the same three stitch types. Three conditions were compared: (1) ANTLER 1 with style profiles active; (2) a corrective-only baseline using the same sensor setup with a fixed single target; and (3) an ungrounded LLM receiving a plain-text session description without sensor data.

The five evaluation dimensions — *plausibility*, *usefulness*, *style fidelity*, *stance appropriateness*, and *perceived specificity* — were selected to distinguish ANTLER 1 from both baselines: a corrective system may remain plausible without style fidelity, while an ungrounded LLM may lack specificity. A structured interview was used to suit the single-participant design, at the cost of inter-rater reliability. Findings are reported as directional indicators; the single-participant design does not support statistical inference.

Dimension	ANTLER 1	Corrective	LLM only
Plausibility	high	high	medium
Usefulness	high	medium	low
Style fidelity	high	low	low
Stance appropriateness	high	low	medium
Perceived specificity	high	low	low

Table 4: Directional pilot ratings across three conditions (single practitioner; no statistical inference claimed).

All three profiles were rated plausible and distinguishable. Expressive and Hybrid profiles surfaced options not visible through correctness-only feedback. In one instance the Expressive profile over-interpreted a hesitation caused by thread resistance as stylistic intent — illustrating the core disambiguation challenge between expressive choice and mechanical difficulty. A structured participant study across skill levels is planned as the next evaluation phase (Mora and others 2024; Yannakakis and Togelius 2014).

Discussion and Limitations

Returning to the research questions: RQ1 is addressed by multimodal stitch-event segmentation, RQ2 directionally by improved perceived specificity and usefulness over the ungrounded LLM baseline, and RQ3 by the distinguishable three-profile tutoring architecture observed in the pilot.

ANTLER 1 presents a style-aware co-creative tutoring *framework* instantiated as a multimodal prototype, realising a four-stage pattern: segmentation → feature grounding → style parameterisation → stance-conditioned generation. Claims are scoped to this early-stage instantiation and do not extend to large-scale validation (Boden 1992; Yannakakis, Liapis, and Alexopoulos 2018; Davis and others 2013). The concrete contribution to the computational creativity literature is threefold: (i) a design framework for style-parameterised co-creative tutoring in embodied craft; (ii) an operational definition of style-contingent interpretation as a CC design pattern (same observed evidence, multiple valid tutor stances); and (iii) an early-stage demonstration that sensor-grounded feedback can be simultaneously explainable and style-sensitive, extending the XAI-for-arts agenda (Llano et al. 2020; Bryan-Kinns and others 2023) into the embodied creative-practice domain.

In one observed sequence the system over-interpreted a hesitation pause as stylistic intent when the cause was thread tension management — illustrating the core challenge of disambiguating expressive choice from mechanical difficulty, and motivating future work on multi-hypothesis intent modelling.

Recent surveys and theoretical work in computational creativity and human–AI co-creativity position systems like ANTLER 1 within a broader shift toward explainable, process-focused creative AI (Llano et al. 2020; Bryan-Kinns and others 2023; Franceschelli and Musolesi 2024; Anonymous 2024). ANTLER 1 aligns with these principles by exposing style parameters, grounding feedback in observable features, and offering profile-conditioned what-if scenarios rather than a single optimised output.

The style taxonomy was developed with one practitioner and reflects a specific couture lineage; other traditions may require different profile structures. Profiles are curated rather than learned, and outputs depend on practitioner-specific rhythm data, limiting population-level validity.

Future work will expand studies across skill levels, learn profile parameters from larger annotated datasets, compare more formally against both baselines, and investigate transfer to other embodied crafts — soutache embroidery, kantha stitching, and ceramic surface decoration — where tacit knowledge is similarly at risk (Mora and others 2024; Jain and others 2025; Mills and others 2025).

References

Anonymous. 2024. Human-AI co-creativity: Exploring synergies across levels of interaction. arXiv preprint arXiv:2411.12527.

Boden, M. A. 1992. *The Creative Mind: Myths and Mechanisms*. London: Routledge.

Bryan-Kinns, N., et al. 2023. Explainable AI for the arts: XAIxArts. In *Extended Abstracts of the ACM CHI Conference on Human Factors in Computing Systems*.

Chen, K.; Zhang, Y.; Weng, Y.; Wang, B.; and Liu, Z. 2025. Multimodal learning for automated robotic surgical skill assessment: Integrating video and kinematic data with ensemble model. *International Journal of Surgery*. Online ahead of print.

Davis, N., et al. 2013. An enactive model of creativity for computational co-creative systems. In *Proceedings of the International Conference on Computational Creativity*.

Franceschelli, G., and Musolesi, M. 2024. Creativity and machine learning: A survey. *ACM Computing Surveys* 56(11):283:1–283:41.

Jain, A., et al. 2025. An llm-guided tutoring system for social skills training. arXiv preprint arXiv:2501.09870.

Llano, M. T.; d'Inverno, M.; Yee-King, M.; McCormack, J.; Ilsar, A.; Pease, A.; and Colton, S. 2020. Explainable computational creativity. In *Proceedings of the International Conference on Computational Creativity*, 334–341.

Mills, C., et al. 2025. Llm-generated feedback supports learning if learners choose to use it. arXiv preprint arXiv:2506.17006.

Mora, M., et al. 2024. An embodied socio-cognitive lens. In *Proceedings of the International Conference on Computational Creativity*.

Pan, Y.; Wang, J.; et al. 2023. An automated skill assessment framework based on visual motion descriptors for robotic surgery training. *Sensors*.

Ritchie, G. 2007. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines* 17(1):67–93.

Salter, C., et al. 2010. Embodied interactive performance and the digital arts. *Digital Creativity*.

Yannakakis, G. N., and Togelius, J. 2014. Experience-driven procedural content generation. In *Affective Computing and Intelligent Interaction*.

Yannakakis, G. N.; Liapis, A.; and Alexopoulos, C. 2018. Evaluating creativity in computational co-creative systems. In *Proceedings of the Ninth International Conference on Computational Creativity*.