Affective Computing-Based Adaptive Music Generation for Human–Animal Assisted Wellness

Kavin Teve
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
kevin.love610@oregonstate.edu

Jranav Shatty
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
pranavwork@uab.edu

Abstract

The integration of affective computing with generative music systems offers new possibilities for adaptive wellness interventions, yet the extension of these systems to human–animal assisted settings remains underexplored. This paper presents a system-level analysis of affective computing-based adaptive music generation for human–animal assisted wellness. The discussion focuses on architectural organization, multimodal sensing, animal state inference, generative music control, interaction dynamics, robustness, fairness, governance, and sustainable deployment. Rather than proposing a single algorithmic solution, the paper argues that human–animal musical co-regulation should be understood as a multi-agent socio-technical infrastructure. A layered architecture is examined in which sensing, state modeling, music generation, and intervention policy operate through separate but coordinated planes. This separation supports oversight, modular evolution, and graceful degradation under real-world constraints. The paper highlights the asymmetries between human and animal participants, particularly the absence of direct animal self-report and the resulting dependence on behavioral and physiological inference. Evaluation is discussed as a pluralistic process requiring multiple weak labels and contextual interpretation. Fairness is considered across both human demographic differences and animal phenotypic or species-specific variation. Governance implications are analyzed through frameworks of accountability, transparency, and animal welfare, emphasizing that animal well-being should function as a hard constraint rather than a secondary objective. Deployment challenges related to edge computing, model sustainability, privacy, and operational usability are also examined. The paper concludes that trustworthy human–animal adaptive music systems require deep integration of technical design with ethical and infrastructural reasoning, and that future progress depends on longitudinal, cross-disciplinary, and species-inclusive research.

Keywords

affective computing; adaptive music generation; human–animal interaction; wellness infrastructure; algorithmic governance.

1. Introduction

The expansion of affective computing has created new opportunities for closed-loop systems that infer, model, and respond to emotional states in real time [1]. In wellness contexts, music is a particularly powerful modality because it can influence attention, expectation, memory, and physiological regulation. Mechanistic accounts of music-evoked emotion emphasize processes such as brain stem reflexes, evaluative conditioning, and emotional contagion, all of which depend on acoustic structure and temporal patterning [2]. Neuroimaging research further shows that music engages limbic, paralimbic, and auditory regions with rapid temporal dynamics, making it well suited to interactive affective intervention [3]. However, many current music-based wellness systems continue to rely on fixed playlists or clinician-selected material. Such static approaches cannot adapt to the fluctuating states of participants in complex human–animal interactions.

Affective computing enables continuous estimation of human states using autonomic and central nervous system signals [4][5]. At the same time, animal-computer interaction has emerged as a field concerned with designing technologies that respect animal cognition, welfare, and agency [6]. Human–animal interaction research demonstrates measurable psychosocial and psychophysiological benefits, including reductions in human stress and improvements in social engagement [7][8]. Recent work on AI-assisted human–pet artistic musical co-creation has illustrated the feasibility of involving companion animals in adaptive musical interaction loops [9]. These developments suggest that generative music systems can be extended beyond human-only therapy to support human–animal dyads.

This paper addresses the system-level challenges that arise when affective sensing, animal behavioral monitoring, generative music, and adaptive intervention policy are combined. The central argument is that human–animal assisted wellness should be designed as a socio-technical infrastructure rather than a standalone machine learning application. Such an infrastructure must balance signal robustness, animal welfare, personalization, fairness, interoperability, and long-term sustainability. The following sections analyze architectural trade-offs, multimodal inference, generative control, evaluation, governance, and deployment. The aim is not to promote a singular model, but to clarify how heterogeneous components can be organized into a coherent and responsible system.

2. Related Work and System Context

Generative music systems have progressed from rule-based composition toward deep sequence models capable of producing long-range harmonic and rhythmic structure. The Music Transformer architecture uses relative attention to preserve temporal coherence over extended musical passages, which is important for maintaining phrasing during live intervention [10]. Larger generative models such as Jukebox operate over raw audio and support conditioning on style, genre, and sentiment-related cues [11]. These models depend on transformer architectures whose self-attention mechanisms allow flexible aggregation of temporal and multimodal inputs [12]. Although expressive, such systems introduce important trade-offs related to latency, controllability, and computational cost in real-time wellness environments.

Edge computing has become a key enabler for latency-sensitive and privacy-preserving services. Moving inference and generation closer to the sensing environment reduces round-trip delay and limits the transmission of sensitive physiological data [13]. In human–animal settings, edge deployment is especially relevant because therapy sessions often occur in homes, clinics, or outdoor environments with variable connectivity. Animal affective state recognition is also advancing through machine learning applied to facial expression, vocalization, posture, and movement [14]. However, these techniques remain less mature than human affect recognition, and cross-species generalization remains a significant challenge.

The convergence of affective computing, generative music, and animal-computer interaction has received limited attention at the systems level. Existing work tends to focus either on human music intervention or on animal welfare monitoring separately. The present paper addresses this gap by articulating a cross-layer framework for coordinated human–animal musical co-regulation.

3. System Architecture and Affective Adaptation Layer

A robust architecture can be organized into four interconnected planes: sensing, state modeling, generation, and intervention policy. The sensing plane acquires multimodal data from human wearables and environmental animal sensors. The state modeling plane fuses these streams into estimates of human valence and arousal, animal affective state, and interaction quality. The generation plane produces music or soundscapes conditioned on those estimates. The intervention policy plane modulates musical parameters, session structure, and escalation rules. This separation allows components to evolve independently and supports heterogeneous deployment hardware.

The affective adaptation layer is the conceptual center of the system. It maintains a continuously updated representation of the shared human–animal state, including uncertainty, missing data, and conflicting signals. Unlike conventional recommendation systems that infer preferences from long-term histories, this layer must operate over short time windows while respecting temporal context. A practical design treats the affective state as a latent variable influenced by multiple physiological and behavioral modifiers. Autonomic indicators such as heart rate variability, electrodermal activity, and respiration provide rapid but noisy correlates of arousal [4]. Contextual features help distinguish between similar physiological patterns associated with excitement, distress, or physical activity.

A central trade-off exists between model expressiveness and operational transparency. Deep fusion models may improve state estimation but can obscure the reasons for a particular musical response. In wellness contexts, transparency is both an ethical and a practical requirement because therapists and animal handlers need to understand why the system changed its behavior. Hybrid architectures that combine learned encoders with explicit affective representations offer a compromise. They preserve high-dimensional learning while exposing a small set of policy-relevant variables to human supervisors. This design also supports auditing and aligns with emerging accountability frameworks for artificial intelligence [15].

4. Multimodal Signal Acquisition and Animal State Inference

Human affect recognition draws on a mature array of sensors, including electroencephalography, electrocardiography, photoplethysmography, galvanic skin response, and facial video [5]. Each modality presents different trade-offs in invasiveness, sampling rate, robustness, and privacy. Electroencephalography offers direct neural correlates but is sensitive to motion artifact and may be impractical in domestic animal-assisted sessions. Consumer wristbands and chest straps are less intrusive, although they may be less accurate during movement. A successful system should prioritize unobtrusive modalities while maintaining fallback paths for clinical-grade sensing. This suggests a heterogeneous sensing fabric that dynamically weights available signals based on quality and comfort.

Animal state inference introduces additional conceptual and practical difficulties. Companion animals express affective states through vocalizations, posture, tail and ear movements, facial musculature, and locomotion, but these signals vary across species, breeds, and individuals

[14]. Models trained on constrained datasets may fail in naturalistic multi-agent settings. Dataset bias is a serious concern when training samples are dominated by specific breeds, coat colors, or lighting conditions. Such bias can produce systematic accuracy losses for underrepresented phenotypes, with direct welfare consequences. The sensing and modeling strategy should therefore include domain adaptation, continuous recalibration, and confidence-based gating.

The system should also incorporate interaction-level features such as proximity, gaze direction, touch frequency, and turn-taking. These features are less susceptible to individual affective label ambiguity and provide important context for music selection. A high-arousal state during active play may be desirable and should not necessarily trigger a calming response. The interpretation of physiological and behavioral signals therefore depends on the broader activity context. Music adaptation policy must avoid the assumption that high arousal is always undesirable.

5. Generative Music Engine and Adaptive Control

The generative music engine must balance musical coherence, emotional responsiveness, and real-time performance. Pretrained transformer and autoencoder models are expressive but can be computationally heavy and stochastic [10][11][12]. For human–animal deployment, controls such as tempo, mode, register, harmonic tension, rhythmic density, instrumentation, and spectral brightness become essential. A control layer can adjust these parameters by conditioning the generative model or by transforming generated material through symbolic or audio post-processing. Symbolic generation offers finer control and lower computational cost but requires synthesis and may lack timbral richness. Raw audio generation provides high fidelity but complicates real-time structural editing.

A structurally useful approach is to combine a slow generative planner with a fast reactive controller. The planner produces phrase-level musical material based on the current affective state, while the controller performs smooth transitions, tempo adjustments, and intensity scaling. This hierarchical design mirrors human musical improvisation and avoids abrupt discontinuities that could disturb either participant. The controller can also enforce safety constraints such as maximum loudness, spectral energy limits in sensitive frequency ranges, and minimum inter-onset intervals. In animal-assisted settings, such constraints are especially important because some sounds may provoke aversive responses depending on species-specific hearing sensitivity.

Adaptive control must also manage therapeutic goals. A music intervention may aim to reduce human anxiety, stimulate social engagement, or provide sensory enrichment for the animal. These goals can conflict. Music that is soothing for a human may be uninteresting or neutral for a companion animal, while rhythmic play music may elevate human arousal but overstimulate the animal. Multi-objective optimization under priority constraints is therefore necessary. In many sessions, animal welfare should operate as a hard constraint while human wellness objectives are optimized within that envelope. This reverses the common assumption that the human is always the primary user and requires governance mechanisms to ensure that animal well-being is not subordinated to algorithmic convenience.

6. Human–Animal Interaction Dynamics and Intervention Design

Human–animal assisted wellness is not a single-user service but a multi-agent interaction system. Its design must account for asymmetries in perception, agency, and communication. Humans can provide explicit feedback through ratings or verbal reports, whereas animals cannot. Animal feedback must be inferred indirectly from behavior and physiology. This asymmetry complicates system optimization because the animal is both a participant and a sentient being whose internal states cannot be directly labeled. The system must avoid overfitting to human preferences at the expense of animal comfort.

Intervention design should build on established human–animal interaction effects. Positive interactions are associated with reductions in human stress and increases in affiliative behavior [7][8]. However, these benefits depend on the quality of the interaction and the animal's willingness to engage. A forcedA forced interaction protocol that overrides animal agency can reduce the quality of human–animal bonding and undermine therapeutic outcomes. The system must therefore interpret animal withdrawal, avoidance, or stress-related displacement behaviors as valid feedback signals that should modulate or terminate a musical session. From a control theory perspective, this requires the intervention policy to support multiple escalation levels: continuation, parameter softening, content change, and session termination. The last level is particularly important; without a clear stopping rule, the system may continue to generate output while the animal is signaling distress, effectively privileging human enjoyment over animal welfare. In a well-designed infrastructure, termination is not treated as a failure but as a normal part of adaptive regulation. The same logic applies to human participants, for whom fatigue, disengagement, or overstimulation should also trigger session adaptation. Thus the system must be designed for bidirectional, dyadic regulation rather than unidirectional human benefit.

Turn-taking is another essential dynamic. In human–animal assisted wellness, effective interaction often follows rhythmic patterns of engagement and disengagement. The musical system should support this alternation rather than forcing continuous stimulation. For example, an agent that detects animal attention toward a sound source may briefly increase musical salience, then reduce density to give the animal an opportunity to explore or withdraw. Such patterns can be encoded through interaction-level features and reinforcement signals, but they must be validated against observed animal behavior over time. The absence of explicit animal self-report means that model updates must rely on delayed and noisy behavioral outcome measures, which complicates both learning and evaluation. This issue is central to the design of responsible systems and reappears in the evaluation and governance discussions below.

7. Evaluation Methodology and Validation Challenges

Evaluating adaptive music systems for human–animal wellness cannot rely on a single performance metric. The system's success depends on multiple weakly aligned objectives: human affective improvement, animal welfare, interaction quality, musical coherence, and operational stability. Each objective is measured differently and may conflict in practice. A conventional machine learning evaluation framework based on prediction accuracy or generator loss is insufficient. Instead, evaluation should be organized around a hierarchy of weak labels and contextual evidence. Human self-report questionnaires, such as those used in affect validation, provide one source of outcome data, but they are limited by recall bias and social desirability. Physiological markers can offer more continuous information, but they are not direct measures of subjective experience. Animal outcomes are even harder to assess, requiring ethologically informed coding and possibly veterinary input. The evaluation framework must therefore combine human reports, physiological metrics, behavioral observation, and interaction logs into a pluralistic evidence base.

One way to address this challenge is through staged validation. In early stages, generative music output can be evaluated for acoustic safety, tempo stability, and adherence to control parameters without involving animal participants. Mid-stage evaluations may use human listeners to assess perceived emotional congruence, though such judgments do not generalize to nonhuman species. Later stages should involve dyadic pilot studies in which human and animal responses are recorded simultaneously. Analysis of synchrony, co-regulation, and behavioral indicators of comfort can provide signals about system acceptability. Because animal stress responses are often subtler than human responses, evaluation protocols should include independent observers trained in animal behavior and, where possible, blinded to the system condition. This reduces anthropomorphic bias and guards against confirmation effects in human participants who may project their own musical preferences onto the animal. The evaluation infrastructure must also record incidents in which the system triggered withdrawal or distress, because such negative examples are critical for improving safety and fairness.

8. Fairness, Robustness, and Governance Implications

Fairness in human–animal adaptive music systems must be analyzed along several dimensions. First, human affect recognition models are known to exhibit performance gaps across demographic groups, sensor placements, and environmental conditions. If a system is trained primarily on data from homogeneous populations, it may produce less accurate inference for underrepresented users and thereby deliver less appropriate musical adaptation. This is a system-level fairness problem, not merely an algorithmic one, because it arises from dataset curation, feature engineering, and deployment context. Second, animal state inference may show phenotypic and species-specific biases. For example, facial expression models trained predominantly on certain breeds may fail for others, and cross-species generalization is often poor. Such disparities can lead to differential treatment outcomes in multi-pet households or in therapy settings that use different species. The system should therefore support explicit bias auditing, subgroup reporting, and confidence thresholds that defer control to human supervisors when classification is uncertain.

Robustness is equally important. Real-world human–animal sessions include sensor dropout, motion artifact, ambient noise, and unexpected animal behavior. A system that depends on a single high-quality modality may fail catastrophically when that modality degrades. Redundancy and graceful degradation should therefore be designed at the architecture level. For example, if electrodermal activity is lost, the system can rely on heart rate variability and behavioral features with reduced confidence. The affective adaptation layer should model uncertainty and allow conservative musical choices under high uncertainty. This prevents the system from making rapid or extreme interventions when state estimates are unreliable. Robustness also applies to generative models, which may produce anomalous sonic events due to sampling noise or distribution shift. Protective post-processing can include spectral clipping, loudness normalization, and musical grammar checks, but these mechanisms must be transparent and auditable.

Governance of human–animal wellness systems cannot be reduced to technical performance. These systems touch on animal welfare, human psychological health, privacy, and the possibility of algorithmic paternalism. Accountability frameworks for artificial intelligence emphasize the need for traceable decisions, contestability, and meaningful human oversight [15]. In this context, oversight should include the human participant, the animal handler, and possibly a veterinary or behavioral specialist. The system should log all salient decisions, including the features that drove musical changes and the confidence of state estimates. Such logs enable post-hoc review when adverse events occur. Transparency is not only about explanation but about the distribution of agency: who can override the system, and under what conditions? The design should ensure that humans can pause, adjust, or terminate sessions at any time, and that animal-initiated withdrawal is recognized as a legitimate signal for termination. These principles align with the broader movement toward human-in-the-loop and animal-in-the-loop design in multi-agent systems.

Policy implications extend to data governance and algorithmic certification. Physiological data collected during therapy sessions are intimate and may be used to infer health conditions beyond the immediate wellness context. Privacy-preserving processing at the edge, differential privacy, or federated learning can reduce the risk of re-identification. However, these techniques introduce their own trade-offs in model utility and auditing capability. Certification frameworks for therapeutic devices often require evidence of safety and efficacy. For adaptive generative systems, such evidence must include not only human outcomes but also animal welfare endpoints. Regulators and professional bodies may need to develop shared standards for animal-inclusive technology, including minimum reporting requirements for animal behavioral events, sensor specifications, and bias assessments. Without such standards, the field risks fragmentation and a race toward superficial personalization that ignores the less visible interests of animal participants.

9. Deployment, Sustainability, and Operational Constraints

Practical deployment of human–animal adaptive music systems requires careful attention to edge computing, power consumption, and interoperability. Sessions may occur in homes, animal shelters, clinics, or outdoor settings where reliable cloud connectivity is unavailable. Edge inference reduces latency and improves resilience, but it also limits model size and places pressure on battery-constrained wearables [13]. A modular architecture can distribute inference differently across the stack: lightweight physiological preprocessing on wearables, state fusion on a local edge device, and heavy generative planning on a nearby personal computer or dedicated audio processor. Generative models may be compressed through distillation, quantization, or pruning, but these techniques must not degrade musical coherence to the point where participants perceive unnatural output. The system should therefore include fallback modes with lower computational demands, such as rule-based musical templates that maintain therapeutic consistency even when deep generative models are unavailable.

Sustainability is a multidimensional challenge. Environmental sustainability requires considering the carbon footprint of training large generative models and the energy cost of continuous inference. Although an individual session may use modest power, the cumulative burden across many devices and long-term deployments can be significant. Model reuse, efficient architectures, and on-device personalization can reduce unnecessary retraining. Economic sustainability involves the cost of sensors, maintenance, and veterinary oversight. If a system is too expensive or fragile for routine use, it will fail to achieve its wellness goals outside controlled research settings. Social sustainability concerns the acceptability of continuous monitoring to users and the potential for over-reliance on algorithmic regulation. Some users may benefit from explicit human-led control rather than autonomous adaptation, especially in therapeutic contexts where the therapeutic alliance is itself a key mechanism of change. The system should therefore support graduated autonomy, from fully manual to fully adaptive modes, depending on user preference, clinical guidance, and animal sensitivity.

Operational constraints also include calibration and personalization. Human physiological baselines vary across time, medication, and health status. Animal baselines vary across age, breed, and prior experience. A system deployed without baseline calibration may produce misleading arousal estimates and inappropriate music. Initial calibration protocols should be minimal, noninvasive, and respectful of the animal's comfort. Over time, the system can learn individualized baselines through unsupervised adaptation, but such adaptation must be bounded to prevent drift or the reinforcement of initial biases. The maintenance of model performance under distribution shift is a significant open problem. Periodic revalidation, human review, and update policies are necessary to ensure that the system remains safe and effective over months or years. This positions the human–animal adaptive music system not as a static product but as a living infrastructure that requires ongoing governance and care.

10. Conclusion

Affective computing-based adaptive music generation for human–animal assisted wellness represents a promising but demanding interdisciplinary frontier. The preceding analysis has argued that such systems should be designed as multi-agent socio-technical infrastructures rather than as isolated generative applications. Architectural separation between sensing, state modeling, generation, and intervention policy supports modular evolution, transparent oversight, and graceful degradation under real-world constraints. The affective adaptation layer must manage uncertainty, conflicting objectives, and the asymmetry between human self-report and animal behavioral inference. Music generation must balance coherence, responsiveness, and animal-specific safety, while intervention design must respect animal agency and include clear termination pathways.

Evaluation and governance cannot be afterthoughts. Pluralistic evaluation frameworks that combine human reports, physiological markers, behavioral coding, and interaction logs are necessary to capture the multiple dimensions of system success. Fairness analyses must address both human demographic disparities and animal phenotypic or species-specific biases. Robustness must be engineered through redundancy and uncertainty-aware control. Governance must ensure traceable decision-making, meaningful human oversight, and hard constraints on animal welfare. Deployment requires edge-oriented design, sustainable model practices, and social acceptability. The convergence of affective computing, generative music, and animal-computer interaction thus demands a shared language and shared standards across engineering, psychology, ethology, and policy. Future work should pursue longitudinal dyadic studies, open benchmark datasets with careful consent and welfare protocols, and cross-disciplinary certification frameworks. Only through such integrated efforts can adaptive music systems become trustworthy partners in human–animal wellness rather than experimental novelties.

References

1. Picard, R. W. (1997). Affective computing. MIT Press.
2. Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. Behavioral and Brain Sciences, 31(5), 559–575.
3. Koelsch, S. (2014). Brain correlates of music-evoked emotions. Nature Reviews Neuroscience, 15(3), 170–180.
4. Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. Biological Psychology, 84(3), 394–421.

5. Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., Xu, X., & Yang, X. (2018). A review of emotion recognition using physiological signals. Sensors, 18(7), 2074.

6. Mancini, C. (2011). Animal-computer interaction: A manifesto. Interactions, 18(4), 69–73.

7. Odendaal, J. S. J., & Meintjes, R. A. (2003). Neurophysiological correlates of affiliative behaviour between humans and dogs. The Veterinary Journal, 165(3), 296–301.

8. Beetz, A., Uvnäs-Moberg, K., Julius, H., & Kotrschal, K. (2012). Psychosocial and psychophysiological effects of human-animal interactions: The possible role of oxytocin. Frontiers in Psychology, 3, 234.

9. Wang, Z., Ma, L., Jin, Y., Feng, Y., Pan, X., Ji, S., & Zhang, K. (2025, August). AI-assisted human-pet artistic musical co-creation for wellness therapy. In Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (pp. 10216-10224).

10. Huang, C. Z. A., Vaswani, A., Uszkoreit, J., Simon, I., Hawthorne, C., Shazeer, N., Dai, A. M., Hoffman, M. D., Dinculescu, M., & Eck, D. (2019). Music Transformer: Generating music with long-term structure. In International Conference on Learning Representations.

11. Dhariwal, P., Jun, H., Payne, C., Kim, J., Radford, A., & Sutskever, I. (2020). Jukebox: A generative model for music. arXiv preprint arXiv:2005.00341.

12. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (pp. 5998–6008).

13. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637–646.

14. Boneh-Shitrit, T., Feighelstein, M., Bremhorst, A., Amir, S., Distelfeld, T., Dagan, Y., Riemer, S., Shimshoni, I., Mills, D. S., & Zamansky, A. (2022). From facial expressions to behavior: A review of computational approaches for animal affective computing. arXiv preprint arXiv:2203.11421.

15. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1–21.

16. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–13).

17. Crawford, K., & Paglen, T. (2021). Excavating AI: The politics of images in machine learning training sets. AI & Society, 36(4), 1105–1116.

18. Rault, J.-L., van der Staay, F. J., & Hemsworth, P. H. (2020). The role of animal welfare in human-animal interactions: A systematic review. Applied Animal Behaviour Science, 229, 105012.