HUMANOID MUSICAL ROBOTS AS EXPERIMENTAL INTERFACES FOR MUSIC-EVOKED EMOTION

Vincent K.M. CHEUNG (cheung@csl.sony.co.jp)[1] and **Jia-Yeu LIN** (erin@aoni.waseda.jp)[2]

[1]**Sony Computer Science Laboratories**, Tokyo, Japan
[2]*Faculty of Science and Engineering*, **Waseda University**, Tokyo, Japan

ABSTRACT

Advances in technology have led to increasingly sophisticated musical humanoid robots. However, their use has largely been limited to performance and related research in human–robot interaction. In this position paper, we propose a novel perspective: musical humanoid robots as experimental interfaces for investigating music-evoked emotions. We argue that current research is constrained by paradigms relying on pre-recorded auditory stimuli, which fail to capture the multimodal, embodied, and interactive nature of real-world musical experience. Building on existing theories of music cognition and emotion, we identify mechanisms that require controlled manipulation of both acoustic and non-acoustic variables. We show that humanoid robots are well-suited as they enable parametric control of performance variables, reproducibility across trials, and the decoupling and recombination of auditory, visual, and interactive components. We illustrate the technical feasibility of this perspective through a case study of the WAseda Saxophonist Robot 5 (WAS-5), demonstrating reproducible control of acoustic and interaction variables that are prerequisites for future music-emotion experiments. Our work positions musical humanoid robots as a methodological platform that enables future controlled investigations of music-evoked emotions.

1. INTRODUCTION

Music is a powerful medium of communication that modulates human emotions [1–3]. A central goal in music cognition research is to understand how musical structures are perceived, integrated, and give rise to affective responses [4, 5].

Despite substantial progress over the past decades, much of this knowledge has been derived from laboratory paradigms relying on preselected, pre-recorded musical stimuli [6]. Such approaches implicitly assume that auditory information alone sufficiently captures the expressive content of musical performance and the mechanisms underlying music-evoked emotions. However, this assumption overlooks the embodied, multimodal, and interactive nature of real-world musical experience, where visual, motor, and social cues contribute to how music is perceived and felt [1, 3, 7].

In parallel, advances in sensors, actuators, and computation have led to increasingly sophisticated musical robots (see [8, 9] for an overview). Existing systems span percussion robots such as Cog [10] and Shimon [11], string robots including Violin MUBOT [12] and the Toyota Violin Robot, and wind robots such as the Waseda Flute and Saxophone Robots [13–16] and Ob [17]. These systems replicate key aspects of instrument sound production and have been primarily developed for performance or for studying human–robot interaction. As a result, their potential as experimental tools for investigating fundamental questions in music cognition remains underexplored.

In this position paper, we propose a novel use of musical humanoid robots as experimental interfaces for hypothesis-driven research in music-evoked emotions. Unlike prior work in robotic musicianship that focuses on performance or interaction design, this work positions humanoid robots as platforms that translate abstract musical structures and expressive intent into physically instantiated, controllable actions. This perspective enables the systematic manipulation of embodied and interactive aspects of musical performance that are otherwise difficult to isolate in conventional paradigms.

This paper makes three contributions:

(i) we introduce humanoid musical robots as experimental interfaces for investigating music-evoked emotions

(ii) we present a case study of the anthropomorphic saxophonist robot WAS-5 as technical feasibility evidence, demonstrating reproducible acoustic control and closed-loop interaction capabilities required for future music-emotion experiments

(iii) we present a set of theoretically grounded experimental paradigms that leverage capabilities afforded by musical humanoid robots.

Our work establishes the technical feasibility of musical humanoid robots as experimental interfaces capable of supporting future controlled investigations of music-evoked emotion.

2. ROBOTS AS EXPERIMENTAL INTERFACES

Current research on music-evoked emotions has largely relied on preselected, pre-recorded auditory stimuli, delivered over speakers or headphones and presented under controlled laboratory conditions [18, 19]. While this ap-

Copyright: © 2026. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

proach has yielded important insights into perceptual and cognitive mechanisms underlying music-evoked emotions, it implicitly assumes that auditory information alone is sufficient in eliciting affective responses during music perception. However, growing evidence suggests that music perception is inherently multimodal, involving not only sound but also visual [20], somatosensory [21], and social [1] cues. As a result, current paradigms are limited by their ability to systematically manipulate embodied and interactive aspects of musical performance while maintaining experimental control.

We propose that humanoid musical robots should not only be understood as performance systems, but also as experimental interfaces for studying music-evoked emotion. By *experimental interface*, we refer to a setup that allows the observation of changes in user behaviour or responses following the controlled manipulation of experimental variables [22]. In this view, humanoid robots translate musical expressive parameters into physically instantiated, controllable actions that can be systematically manipulated. Notably, the role of the robot is not to replicate human performance *per se*, but to provide a structured and reproducible medium through which the contribution of musical variables on emotional responses can be isolated.

Humanoid robots offer three key properties that make them well-suited for this role. First, they enable parametric control over embodied aspects of performance, such as movement dynamics, gesture timing, and expressive cues, while maintaining precise control over acoustic output. Second, they provide reproducibility, allowing performances to be repeated across trials with minimal variability, which is difficult to achieve with human performers. Third, they support the decoupling and recombination of modalities, which enable the independent manipulation of multimodal and interaction-related variables. Together, these properties allow researchers to clarify underlying mechanisms and move towards controlled investigation of how different components of musical performance contribute to emotional experience.

Importantly, we do not claim that all musical experiences are emotional, nor that humanoid robots can fully replicate the richness of human performance. Rather, we propose that in contexts where affective responses are of interest, robots provide a complementary methodological tool that enables controlled investigation of mechanisms that are otherwise difficult to isolate. This complements, rather than replaces, studies using human musicians. Importantly, humanoid robots are not intended to replicate the full richness or ecological validity of live human performances. Rather, their value lies in providing experimentally controllable and reproducible embodiments that complement studies using human musicians, particularly when isolating specific multimodal mechanisms of music-evoked emotion.

In the following sections, we ground our position with respect to our current understanding of music-evoked emotions, present a case study of a humanoid musical robot, and outline research directions that become accessible under this framework.

3. COGNITIVE MECHANISMS OF MUSIC-EVOKED EMOTIONS

3.1 From perception to evoked emotions

Existing research suggests that incoming sensory information from music is actively abstracted by the brain via two distinct systems of musical representations: tonal and metrical [23]. Tonal and metrical perception depend on sensory-acoustic constraints and statistical regularities learnt from extended exposure to music of a given style [2, 24]. The tonal system processes pitches not in isolation, but based on grouping mechanisms in terms of proximity of their relative pitches, as well as tonal hierarchies, where some pitches in a musical scale are considered more important with respect to the tonal centre established by its preceding context [5, 23]. Similarly, beats belong to a metrical hierarchy, such that beats placed or omitted at certain locations in a meter (recurrent patterns of pulses) are considered more important than others [25].

While the interaction of tonal and metrical representations engender expectations that are confirmed or violated, which in turn elicit an emotional response [26, 27], how this is combined with other musical information such as timbre, semantic information from lyrics, and multimodal associations such as body movements, remain unclear [28]. This calls for experimental paradigms where such elements can be systematically dissociated.

3.2 Models of music-evoked emotions

Current models of music-evoked emotions characterise affective responses elicited by music in different levels of abstraction [1, 6]: At the most rudimentary level, affective responses are described in terms of core affects (valence and arousal) that can be represented by a two-dimensional circumplex model. A higher level description uses categorical emotions (e.g., happy, sad) that are *sensed* and often regarded as basic. At the top of the hierarchy, elicited responses are explained in terms of complex emotions that are *experienced* and described using abstract concepts such as wonder and transcendence. Music-evoked emotions thus emerge from the interaction between perceptual, cognitive, and embodied mechanisms.

One influential model that explains the mechanisms through which music can induce emotions is the BRECVEMA framework [7, 29] (see [6] for a review of other models). These include lower-level *mapping* mechanisms that directly relate sound to affect, such as brain stem reflex, expectancy, entrainment, as well as embodied and socially grounded processes such as visual imagery and emotional contagion—the internal mirroring of emotions in the audience as expressed by the performer. Higher-level *evaluative* mechanisms further involve evaluative conditioning, episodic memory, and aesthetic appraisal.

That these mechanisms require the integration of auditory, visual, motor, and social information speaks for the

need of a multimodal and embodied approach to investigate music-evoked emotions. By enabling reproducible, dissociable, and controlled manipulation of movement dynamics, gesture, and perceived agency alongside acoustic output, embodied music systems such as humanoid musical robots allow specific components of emotion induction (e.g., entrainment, contagion, and social interaction) to be systematically investigated.

4. WHAT HUMANOID ROBOTS ENABLE

Here, we explain how musical humanoid robots can serve as an invaluable experimental interface towards understanding music-evoked emotions, particularly in three directions: (i) the multimodal nature of music perception, (ii) the limits of emotion contagion, and (iii) the role of social dynamics.

4.1 Multimodal dissociation

While existing knowledge on music-evoked emotions was built on music as audio alone, increasing studies show that our musical experience is multimodal. For example, music presented in both auditory and visual modalities were rated as having a higher aesthetical experience and show different physiological responses compared to auditory alone [30], and participants appear to evaluate musical performances primarily using visual instead of auditory information [20]. Furthermore, motor synchrony to dance music has shown to be related to imperceptible very low-frequency sounds at 8-37 Hz [21]. Listeners also showed enhanced activation of the mesolimbic reward network to live music compared to pre-recorded music [19]. Central to these studies is the claim that the sound produced via the physical interaction of acoustic, vibrotactile, and visual cues in music performed using actual instruments cannot be adequately reproduced by speakers alone. Although live music performance in itself does not necessitate humanoid robots, that the performance can be exactly controlled and replicable alleviates confounds that arise from variability in each human performance. That is, musical humanoid robots enable the dissociation of sound production and other emotional cues in a parametrically controlled manner.

4.2 Controlled emotion contagion

Another direction is to describe the precise boundary and parameters through which musical expression is transferable to the listener. Given that emotion contagion involves the internal mimicry of the performers' emotional expression, then what is the degree of anthropomorphisation (similarity to a human) required for this mimicry to occur? Current generative models of music assume that musical expression is encoded in terms of slight deviations in note onsets, velocities, and timbre [31], suggesting that none at all is required. This implies that it is only the *belief* of the performers' emotional expression that matters in inferring intentionality, for which the physical robot acts as a strong, modifiable prior.

Furthermore, humanoid robots allow us to clarify which actions are relevant to expression. This is important as recent research show that pianists' expressive touch is demonstrated via differences in piano key motion [32]. Humanoid robots offer a controllable framework to categorise and generalise these findings. Likewise, conductors' hand movements contain gesture information that signals differences in musical expression [33]. This can be explicitly tested by systematically examining actions that convey such expressive information. Examples of such controllable parameters include smoothness/jerkiness of movements, joint velocity, head-orientation, gaze direction and latency, breathing motion, degrees of freedom in a limb, as well as microtiming deviations.

4.3 Social dynamics and co-regulation

While the social function of music is apparent, how social dynamics modify music-evoked emotions is not well understood. Nevertheless, early studies suggest that synchronised tapping to the beat is related to increased trust and cooperation in adults and children [3], as well as differential responses when listening in a group versus alone [1]. Musical humanoid robots can fill in this gap by enabling a social environment with multiple robots actively performing or reacting to music with the listener in a controllable, co-creative set-up. Degrees of anthropomorphisation and mimicry to listeners' behaviour could further be tunable parameters that test their relevance on the social aspects of music-evoked emotions. In other words, synchrony between robot and human could be leveraged as a variable for real-time adaptation in musical co-creation paradigms.

5. CASE STUDY: WAS-5 AS TECHNICAL FEASIBILITY EVIDENCE

In this section, we introduce the Waseda Anthropomorphic Saxophonist (WAS-5) robot (see Figure 1) as a case study to show its feasibility in enabling parametric control over acoustic and interaction-level variables relevant to music-evoked emotion research. WAS-5 is the fifth generation of the Waseda Anthropomorphic Saxophonist robot [14–16], a humanoid robot designed to play a standard alto saxophone by mimicking human respiratory airflow mechanisms, as well as finger, eye, eyelid, and eyebrow movements with 31 degrees of freedom.

5.1 Advanced lip mechanism for fine-grained sound pressure control

To simulate the human respiratory tract, WAS-5 utilises an air pump and proportional valves to deliver precise airflow to the mouthpiece. In addition, its tongue and lips are made from a styrene thermoplastic elastomer to control airflow within its oral cavity with high biomimetic accuracy.

In human performance, the embouchure involves coordinated muscle activation around the lips that enable the performer to apply distributed and directionally controlled forces to the reed. A new lip mechanism implemented on WAS-5 employs an 8-directional actuation structure to better approximate the circumferential contact and radial

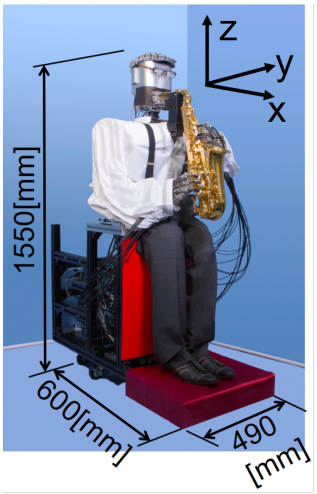


Figure 1. The Waseda Anthropomorphic Saxophonist (WAS-5) musical humanoid robot [15, 16].

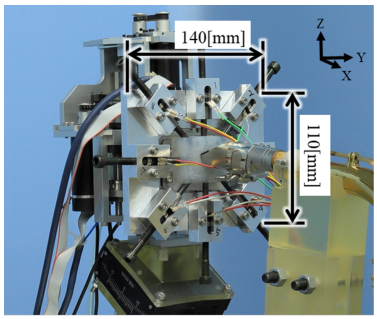


Figure 2. The 8-directional lip mechanism in WAS-5 for precise radial airflow control.

force application observed in human embouchure formation (see Figure 2).

Measurements show that this 8-directional lip mechanism increased the achievable sound pressure range by up to 33.8% compared to its previous configuration that was limited to two opposing directions, reaching a dynamic range of 24.3 dB when sounding pitch E4 and 26.5 dB when sounding pitch F4. This stable yet flexible interface can therefore support fine and reproducible control over vibration characteristics and airflow, enabling the robot to sustain the wide dynamic range essential for emotion expression.

From the perspective of music-evoked emotion research, these results provide technical feasibility evidence that the robot can reproducibly manipulate embodied acoustic parameters relevant to future music-emotion studies. Previous work has suggested that physical parameters of sound production such as embouchure adjustment, mouthpiece bite force, and reed depth contribute to acoustic features associated with emotion expression [34]. The WAS-5 robot thus provides the capability required for future parametric investigations of how embodied acoustic features influence music-evoked emotions via explicit mechanical control.

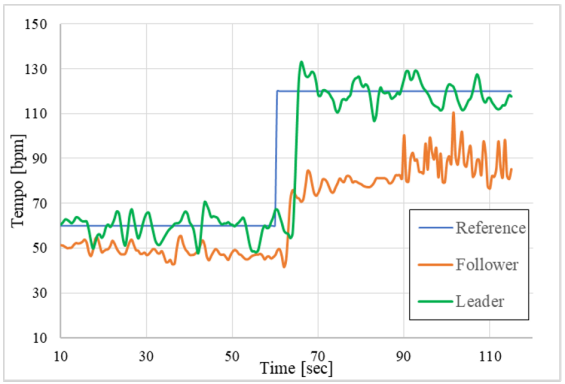


Figure 3. Feet-tapping dynamics of a representative subject, guided by WAS-5 in the Leader condition and followed by the robot in the Follower condition.

5.2 Robot-Assisted Music Interaction

While precise control of sound production is crucial, investigating music-evoked emotions also requires understanding how listeners and performers engage in dynamic contexts. To this end, WAS-5 is also capable of interacting with the listener in a closed-loop design, enabling real-time coupling between musical output and human behaviour.

In one experiment, participants were asked to tap their feet to a simple six-note melody as performed by WAS-5 under three experimental conditions. In the Control condition, the robot performed the melody at fixed tempi of 60 bpm and 120 bpm in 60 s intervals. In the Follower condition, the robot assumed a passive role and continuously adjusted its tempo according to participants' self-designated fast and slow tapping rates. In the Leader condition, the robot assumed an active role and modulated its tempo to guide the participant towards tapping at the predefined target tempi (60 or 120 bpm). In both adaptive conditions, tempo modulation was implemented via real-time tempo estimation based on video tracking of participants' movements.

Results (see Figure 3) showed that WAS-5 could reliably adapt its musical output to participants' behaviour in the follower condition, and systematically influence participants' tapping behaviour in the leader condition. These findings demonstrate that WAS-5 can reliably manipulate agency and interpersonal synchronization under controlled conditions. Importantly, this experiment evaluates behavioral synchronization rather than music-evoked emotion itself. Instead, it provides technical feasibility evidence that the platform supports closed-loop interaction paradigms which can subsequently be extended with subjective emotion ratings and physiological measures to investigate music-evoked emotional responses directly.

Together, these results show WAS-5 as an example humanoid musical robot that provides (i) precise acoustic control, (ii) reproducible performance, and (iii) closed-loop interaction capabilities, which are key to experimentally manipulating multimodal and social variables in music-evoked emotion research.

6. EXPERIMENTAL PARADIGMS ENABLED BY THE SYSTEM

In this section, we present paradigms as experimentally testable research directions. Their purpose is to illustrate how the demonstrated capabilities of musical humanoid robots may be leveraged to investigate specific mechanisms underlying music-evoked emotion in future empirical work. Building on the capabilities demonstrated above, we now outline some concrete experimental paradigms that operationalise the dimensions of multimodal integration, emotion contagion, and social interaction for research in music-evoked emotions. In each case, the use of robots is crucial as we require exact reproduction of a musical performance, direct manipulation of parameters relating to anthropomorphism, and precise movement synchronisation across multiple agents. We also summarise the mechanisms and main dependent variables explored in each paradigm in Table 1.

6.1 Embodiment and emotions

Here we test the effects of different levels of robotic embodiment on music-evoked emotions. The participant is seated behind a veil for visual isolation. We first compare live music performed by a musical humanoid robot versus the same performance as presented via speakers, matched as closely in loudness and proximity as possible. Next, we repeat the same presentation with the veil removed, except there will now be two presentations via speakers, with the robot either remaining static or moving synchronously to the music. After each presentation, we measure the participants' engagement, enjoyment, and trust in the robot. We hypothesise that participants will give a higher engagement and enjoyment rating to music performed live by the robot in both veiled and unveiled conditions compared to music reproduced via speakers, with the highest scores in the former. Findings from this experiment will enable a deeper understanding on the role of embodiment on music interaction towards a richer musical cocreation experience between human and robots.

6.2 Controlled emotion contagion

This experiment examines the extent visual cues signalled by the robot influences emotion contagion. The participant observes multiple repetitions of the same live music performance by a humanoid robot. At each repetition, features related to robot anthropomorphism, such as body sway, eyebrow/lid movement, facial expression, or breathing-related torso movement, are parametrically varied at predefined step sizes (i.e., constant auditory but varied visual output). After each presentation, the participant provides ratings on the perceived and felt intensity and intentionality of the emotions expressed. Given the importance of body sway in expressive music performance and rhythmic entrainment [35], we expect this feature to be most relevant in emotional expression. Results allow for the identification and quantification of visual features most salient for conveying music emotions and could be incorporated in future musical humanoid robot designs.

6.3 Social co-regulation

In this experiment, the participant is seated in between two humanoid robots in the audience. On stage is another musical humanoid robot performing live. At each repetition, the audience robots sway together in phase with the performer robot, together out of phase with the performer, or out of phase with each other and the performer. We measure participants' own body movement and electrodermal activity during the music, as well as collect behavioural ratings on their engagement and arousal, and ratings on how much they trust their neighbouring robots. Another variant would be to also vary the number of robots in the audience in a scaled paradigm. We hypothesise that synchronised body sway with the most audience robots would result in the highest engagement, arousal, and trust in the neighbouring robot, as well as synchronous movements by the participant. Findings from this experiment could inform the effects of social entrainment on music-evoked emotions and contribute towards developing more engaging creative systems.

7. LIMITATIONS

Our proposal on using musical humanoid robots as experimental interfaces to explore music-evoked emotions is not without limitations. First, the viable experimental state-space is constrained by current technological capabilities of a robot. For example, limitations in materials, sensors, and actuators could limit the performance of the robot and consequently restrict the questions that could be realistically explored. Nevertheless, although the paradigms presented previously were conceptual, they lie within the capabilities of existing musical humanoid robots. Second, constructing and calibrating the robot may be a challenge in itself and could introduce additional technical and logistical hurdles that must be overcome beforehand [8]. A trade-off between the benefits of building physical humanoid robots versus the design freedoms offered by virtual reality should thus be considered (see [36] for a relevant discussion). Third, the receptivity and attitude towards robots are individually and culturally dependent [37]. This means that results only reveal contemporary attitudes and may not generalise across all listeners geographically and over time. A multicultural approach with the potential to study their effects longitudinally may be required. Fourth, the musical humanoid robots should be designed to avoid the uncanny valley effect induced by perceptual mismatch arising from them having an almost-human resemblance. Fifth, despite the lack of physical grounding, virtual musician robots may also offer additional flexibility in research design [36]. Finally, the present work demonstrates only technical feasibility rather than empirical evidence that humanoid robots modulate music-evoked emotion. Establishing such evidence will require future experiments incorporating validated subjective emotion questionnaires together with physiological and behavioural measures.

Table 1. Summary of proposed paradigms enabled by humanoid robots for music emotion research, as well the underlying mechanisms targeted, and main psychological outcome variables that could be measured.

Paradigm	Target mechanism	Main dependent variables
Embodiment	Multimodal integration	Engagement, enjoyment, trust
Emotion contagion	Emotional mirroring	Perceived emotion, intentionality
Social co-regulation	Synchrony	Trust, engagement, arousal

8. CONCLUSION

In this position paper, we introduced musical humanoid robots as experimental interfaces for investigating music-evoked emotions. We demonstrated through a case study of the WAS-5 robot that such systems can achieve reproducible control of embodied acoustic parameters and closed-loop interaction with human behaviour, allowing humanoid robots to systematically manipulate multimodal and social aspects of musical performance. We further outlined experimentally testable paradigms targeting embodiment, emotion contagion, and social co-regulation. Findings from these research directions will prove invaluable for designing future creative systems with deepened understanding of agency and transmission of expressive intentionality, as well as further knowledge on how music is often regarded one of life's most important pleasures.

Acknowledgments

This work was supported by Sony Computer Science Laboratories and JSPS Grant-in-Aid for Early-Career Scientists [23K12755]. The authors would also like to thank Tokyo Women's Medical University/ Waseda University Joint Institution for Advanced Biomedical Sciences (TWIns) and Humanoid Robotics Institute (HRI).

9. REFERENCES

[1] T. Eerola, "Music and emotions," in *Springer handbook of systematic musicology*. Springer, 2018, pp. 539–554.

[2] S. A. Mehr, M. Singh, D. Knox, D. M. Ketter, D. Pickens-Jones, S. Atwood, C. Lucas, N. Jacoby, A. A. Egner, E. J. Hopkins *et al.*, "Universality and diversity in human song," *Science*, vol. 366, no. 6468, p. eaax0868, 2019.

[3] S. Koelsch, "Brain correlates of music-evoked emotions," *Nature reviews neuroscience*, vol. 15, no. 3, pp. 170–180, 2014.

[4] ——, "Toward a neural basis of music perception–a review and updated model," *Frontier in Psychology*, vol. 2, p. 110, 2011.

[5] M. A. Rohrmeier and S. Koelsch, "Predictive information processing in music cognition. a critical review," *International Journal of Psychophysiology*, vol. 83, no. 2, pp. 164–175, 2012.

[6] T. Eerola and J. K. Vuoskoski, "A review of music and emotion studies: Approaches, emotion models, and stimuli," *Music Perception: An Interdisciplinary Journal*, vol. 30, no. 3, pp. 307–340, 2012.

[7] P. N. Juslin and D. Västfjäll, "Emotional responses to music: The need to consider underlying mechanisms," *Behavioral and brain sciences*, vol. 31, no. 5, pp. 559–575, 2008.

[8] M. Bretan and G. Weinberg, "A survey of robotic musicianship," *Communications of the ACM*, vol. 59, no. 5, pp. 100–109, 2016.

[9] G. Weinberg, M. Bretan, G. Hoffman, and S. Driscoll, *Robotic musicianship: embodied artificial creativity and mechatronic musical expression*. Springer Nature, 2020, vol. 8.

[10] M. M. Williamson, "Rhythmic robot arm control using oscillators," in *Proceedings. 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems. Innovations in Theory, Practice and Applications (Cat. No. 98CH36190)*, vol. 1. IEEE, 1998, pp. 77–83.

[11] G. Weinberg and S. Driscoll, "The design of a robotic marimba player: introducing pitch into robotic musicianship," in *Proceedings of the 7th international conference on New interfaces for musical expression*, 2007, pp. 228–233.

[12] M. Kajitani, "Development of musician robots in japan," in *Proceedings of the Australian Conference on Robotics and Automation*, 1999.

[13] J. Solis, K. Petersen, T. Ninomiya, M. Takeuchi, and A. Takanishi, "Development of anthropomorphic musical performance robots: From understanding the nature of music performance to its application to entertainment robotics," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 2309–2314.

[14] J. Solis, A. Takanishi, and K. Hashimoto, "Development of an anthropomorphic saxophone-playing robot," in *Brain, Body and Machine: Proceedings of an International Symposium on the Occasion of the 25th Anniversary of the McGill University Centre for Intelligent Machines*. Springer, 2010, pp. 175–186.

[15] J.-Y. Lin, M. Kawai, Y. Nishio, S. Cosentino, and A. Takanishi, "Development of performance system with musical dynamics expression on humanoid saxophonist robot," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1684–1690, 2019.

[16] J.-Y. Lin, Y. Nishio, M. Kawai, K. Matsuki, S. Sessa, S. Cosentino, and A. Takanishi, "Design and evaluation of a robot limb for table key playing on humanoid saxophonist robot," in *2019 IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 2019, pp. 799–804.

[17] L. Maes, G.-W. Raes, and T. Rogers, "The man and machine robot orchestra at logos," *Computer Music Journal*, vol. 35, no. 4, pp. 28–48, 2011.

[18] L. A. Warrenburg, "Choosing the right tune: A review of music stimuli used in emotion research," *Music perception*, vol. 37, no. 3, pp. 240–258, 2020.

[19] W. Trost, C. Trevor, N. Fernandez, F. Steiner, and S. Frühholz, "Live music stimulates the affective brain and emotionally entrains listeners in real time," *Proceedings of the National Academy of Sciences*, vol. 121, no. 10, p. e2316306121, 2024.

[20] C.-J. Tsay, "Sight over sound in the judgment of music performance," *Proceedings of the National Academy of Sciences*, vol. 110, no. 36, pp. 14 580–14 585, 2013.

[21] D. J. Cameron, D. Dotov, E. Flaten, D. Bosnyak, M. J. Hove, and L. J. Trainor, "Undetectable very-low frequency sound increases dancing at a live concert," *Current Biology*, vol. 32, no. 21, pp. R1222–R1223, 2022.

[22] I. S. MacKenzie, "Designing hci experiments," *Human-computer interaction: an empirical research perspective*, vol. 176, 2013.

[23] S. A. Mehr, "Core systems of music perception," *Trends in Cognitive Sciences*, 2025.

[24] V. K. Cheung, P. Harrison, S. Koelsch, M. T. Pearce, A. D. Friederici, and L. Meyer, "Cognitive and sensory expectations independently shape musical expectancy and pleasure," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 379, no. 1895, 2024.

[25] P. Vuust and M. A. Witek, "Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music," *Frontiers in psychology*, vol. 5, p. 1111, 2014.

[26] V. N. Salimpoor, D. H. Zald, R. J. Zatorre, A. Dagher, and A. R. McIntosh, "Predictions and the brain: how musical sounds become rewarding," *Trends in cognitive sciences*, vol. 19, no. 2, pp. 86–91, 2015.

[27] V. K. Cheung and S. Sakamoto, "Separating uncertainty from surprise in auditory processing with neurocomputational models: implications for music perception," *Journal of Neuroscience*, vol. 42, no. 29, pp. 5657–5659, 2022.

[28] V. K. Cheung, P. M. Harrison, L. Meyer, M. T. Pearce, J.-D. Haynes, and S. Koelsch, "Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity," *Current Biology*, vol. 29, no. 23, pp. 4084–4092, 2019.

[29] P. N. Juslin, "From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions," *Physics of life reviews*, vol. 10, no. 3, pp. 235–266, 2013.

[30] A. Czepiel, L. K. Fink, C. Seibert, M. Scharinger, and S. A. Kotz, "Aesthetic and physiological effects of naturalistic multimodal music listening," *Cognition*, vol. 239, p. 105537, 2023.

[31] C. E. Cancino-Chacón, M. Grachten, W. Goebl, and G. Widmer, "Computational models of expressive music performance: A comprehensive and critical review," *Frontiers in Digital Humanities*, vol. 5, p. 25, 2018.

[32] K. Kuromiya, Y. Kobayashi, M. Hirano, and S. Furuya, "Motor origins of timbre in piano performance," *Proceedings of the National Academy of Sciences*, vol. 122, no. 39, p. e2425073122, 2025.

[33] G. Luck, P. Toiviainen, and M. R. Thompson, "Perception of expression in conductors' gestures: A continuous response study," *Music Perception*, vol. 28, no. 1, pp. 47–57, 2010.

[34] A. Almeida, W. Li, E. Schubert, J. Smith, and J. Wolfe, "Expressive goals for performing musicians: The case of clarinetists," *Musicae Scientiae*, vol. 27, no. 3, pp. 655–671, 2023.

[35] A. Chang, H. E. Kragness, S. R. Livingstone, D. J. Bosnyak, and L. J. Trainor, "Body sway reflects joint emotional expression in music ensemble performance," *Scientific reports*, vol. 9, no. 1, p. 205, 2019.

[36] T. R. P. Pessanha, H. Camporez, J. Manzolli, B. S. Masiero, L. Costalonga, and T. F. Tavares, "Virtual robotic musicianship: Challenges and opportunities," in *Proceedings of the Sound and Music Computing Conference (SMC'21). Sound and Music Computing Network*, 2021.

[37] S.-Y. Chien, M. Lewis, K. Sycara, J.-S. Liu, and A. Kumru, "The effect of culture on trust in automation: reliability and workload," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 8, no. 4, pp. 1–31, 2018.