Adaptation and Perceived Creative Autonomy in Gesture-Controlled Interactive Music

Jason Brent Smith
jason.smith1@northwestern.edu
Northwestern University
Evanston, Illinois, USA

Jason Freeman
jason.freeman@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Abstract

With the variety and rapid pace of developments in Artificial Intelligence (AI), musicians can face difficulty when working with AI-based interfaces for musical expression as understanding and adaptation to AI behaviors takes time. In this paper, we explore the use of AI in an interactive music system designed to adapt to users as they learn to perform with it. We present *GestAlt*, an AI-based interactive music system that collaborates with a performer by analyzing their gestures and motion to generate audio changes. It uses computer vision, online machine learning, and reinforcement learning to adapt to a user's hand motion patterns and allow a user to communicate their musical goals to the system. It communicates its decision-making to the user through visualizations and its musical output. We conducted a study in which five musicians performed using this software over multiple sessions. Participants discussed how their preferences for the system's behavior were influenced by their experiences as musicians, how adaptive reinforcement learning affected their expectations for the system's autonomy, and how their perceptions of the system as a creatively autonomous, collaborative partner evolved as they learned how to perform with the system.

Keywords

Interactive Music, Artificial Intelligence, Machine Learning, Human-Computer Interaction

1 Introduction

Artificial intelligence (AI) has long been sought as a way to allow computers to assist with musical composition, performance, and theory [53]. Algorithmic composition is used to create unique sounds [40, 62], leading to the study of musical systems as autonomous composers and performers [56]. Composition and music generation systems [1, 7] have allowed AI to become part of a changing music creation landscape, but these tools present challenges to musicians unfamiliar with how to use and adapt to them. This paper explores how interactive AI-based systems can support a user's understanding by adapting to users.

Perception can deeply affect a user's satisfaction with an AI-based system. The user can reflect on whether or not they see the system exhibit creative autonomy [34], the act of independent decision-making that exhibits the AI's creativity rather than an extension of the musician's. Boden describes perception as an element of creativity alongside memory, conceptual thinking, and self-reflection for both humans and computers [4]. This work highlights two techniques that directly impact how users interact with an AI-based system and how they *perceive* that interaction: explanation and adaptive machine learning behaviors. Explainability [29] promotes users' understanding of the decisions made by an AI, allowing them to form a mental model of what the system maps to their actions [9]. Online machine learning (OML) is the paradigm of a machine learning system in which a model is updated with new data over time [25], and reinforcement learning (RL) is the use of a reward signal such as user feedback to train a model [12]. Tools such as these enable AI-based interactive systems to adjust to their users over time.

Gill [28] describes an ideal human-machine interaction as a collaboration that combines "the computational capacity of the machine and the knowledge of the human" in a "symbiotic and interactive relationship that valorizes human knowledge and computational resources." These symbiotic relationships form from complementary but independent decision-making, another aspect of creative autonomy [34]. Evaluating the AI components of AI-based interactive music systems in their ability to support *symbiotic* relationships with human performers is critical to inform AI-based systems that foster human creativity.

Gesture-controlled systems for music and dance connect human input motion to generated music through the exploration of latent spaces, or semantic representations of audio and visual concepts [47]. Interactive, dance-based performance systems explore the augmentations and limitations imposed by communication between performer and system [16, 17]. The ability for gesture-controlled systems to non-verbally communicate and react to decision-making is especially relevant to modern concerns about supporting users creatively using Large Language Models (LLMs) [37, 44] and generative audio models [36]. This paper explores how systems can display an understanding of a user's input through the creation of an AI-based interactive music system that recognizes and interprets gestures.

This paper introduces *GestAlt*, an AI-based interactive generative music system. *GestAlt* is designed for live musical performance as an improvisational tool that creates music alongside a user's motion and gestures while visualizing its decision-making. Users communicate with it through interactive training and demonstration non-verbally. *GestAlt* combines gesture recognition with adaptive musical changes to explore how users and an AI form a shared mental model *during* a creative performance task. AI predictions can be used to model behavior in real-time performance [45]. Continuous learning can act as an analog to the self-awareness of an AI system in a musical performance [54]. *GestAlt* is designed to demonstrate autonomy through continuous adaptations to user feedback.

2 Related Work

This work draws from frameworks as well as existing new interfaces for musical expression (NIMEs) in evaluating a system's agency in music-making. We also examine tools such as adaptive machine learning, sensors, and using visual programming for improvisatory performance with interactive music systems.

NIME '25, June 24–27, 2025, Canberra, Australia

Figure 1: The user view of *GestAlt*, with key commands (left) and motion (center) and gesture (right) visualizations.

2.1 Frameworks for Creativity and Understanding

Creative autonomy is the ability of a system to demonstrate independent decision-making [34]. This work focuses on *perceived creative autonomy,* or how users form an interpretation of a system's level of creative autonomy and role in the creative process, rather than *actual* creative autonomy to account for the limited scope of autonomous behaviors by systems such as *GestAlt.*

Theory of Mind represents the ability to read signals and understand expectations [41]. Mutual Theory of Mind is found when multiple parties build shared expectations, such as the shared goals between performers to create music. Interactions between users and a conversational agent were studied to obtain a sense of a community's perception through perceived anthropomorphism, intelligence, and likeability of the system [68].

Synchrony between humans is the psychological effect that shared movement can lead to greater cooperation between members of a group [71]. Synchronous movement has been attributed to increased senses of "altruism" and "friendship" toward artificial agents [8, 26]. Despite a minimal impact on the output of their task, these adoptions allow users to "recognize themselves in the actions of the AI" and establish trust in the AI's decisions [48]. An example of this interaction in AI-powered music is *Shimon*, the robotic marimba player, which performs head movements to match those of a human player to increase Synchrony when engaging with live improvisation [32, 58].

2.2 Musical Agents

A performer and a computer operate with a recurrent feedback loop with various modalities given the physical aspects of the system [5]. Machine learning can be integrated into "tools" and AI agents can function as "actors" in musical performance, due to the variety of human-AI controller paradigms and technical aspects of the learning algorithm [11, 57]. Musical agents are autonomous systems that can perform alongside a user [65]. These can range from "reactive," rule-based systems that display some autonomous behaviors, such as George Lewis's *Voyager* [40], to "completely autonomous agents" employing statistical sequence modeling and cognitive modeling [43]. Systems such as Somax2 reactively improvise using cognitive memory [2].

Rowe presents a paradigm of interactive music systems as "instruments" that analyze gestures to create musical outputs and "players" that act as separate performers that follow the user to varying degrees [55]. This dichotomy, as well as the role of AI in musical improvisation, is further explored by Fiorini in a case study [23] using the Somax2 [2] system. Fiorini relates the use of AI-supported improvisational tools to the contrast between composition to improvisation in terms of musical creativity [6] as well as a continuum of a spectrum of musical interactions based on cognitive models [50].

2.3 Adaptive Machine Learning

Online machine learning (OML) systems update a model with new data over time [25]. Interactive modeling training has been used to great effect in musical applications that allow a performer to incorporate the process as part of a live performance [20], but automated machine learning for doing so is relatively new [11].

Reinforcement learning (RL) is a learning paradigm in which an agent performs actions based on a reward, usually derived from its perception of its environment [12]. Learning from Demonstration is the process in which agents mimic the behavior demonstrated by an expert [51], with users demonstrating a task and approving the system when it imitates the task correctly. An example of this is the TAMER framework, or Training Agents Manually via Evaluative Reinforcement [39]. RL has been used in musical applications, such as *Co-Explorer*, a system for interactive exploration of musical parameters by sound designers [59].

2.4 Sensors and Visual Programming

The laptop computer contains a variety of sensors that make it a suitable platform for expressive musical performances [22], such as the camera. Achieving expressiveness in computer musical performance is possible through understanding mappings, visual feedback, the development of virtuosity through practice, and critical discourse of the system [15]. Participants in a study using a constrained musical instrument displayed various techniques in creating musical variety through problem-solving or exploratory mindsets [30]. Interactive machine learning can allow users of digital musical instruments to use (digital) mappings to create control paradigms, with many-to-many mappings and nondeterministic processes increasing the control users have over complex interactions with the system, promoting "partnerships" between user and system through the act of creating those mappings [20]. There are many modes of input in such a system, but gesture is linked to the user's intention in musical performance [3].

Visual programming languages have a long history in the field of interactive music and the development of NIMEs [21]. Hartley [31] provides a collection of Max/MSP[1] objects designed to help with prototyping musical interfaces and augmenting physical instruments, further illustrating the software's use in multimodal and modular approaches. Additionally, the use of visualization has supported tools for analyzing input to musical systems [49], as well as physical interfaces such as the Reactable that respond to physical interactions with audiovisual output [35].

3 System Design

GestAlt[2] is an interactive software-based generative music system that uses a laptop camera to capture a performer's motion for improvisatory musical performance. It is hosted in a modified application that demonstrates Google's MediaPipe gesture detection library [64, 73], combined with a musical output patch in the visual programming language Max/MSP. It uses MediaPipe to measure the position and orientation of a user's hands.

[1]https://cycling74.com/products/max
[2]Video demonstration: https://bit.ly/gestalt-demo

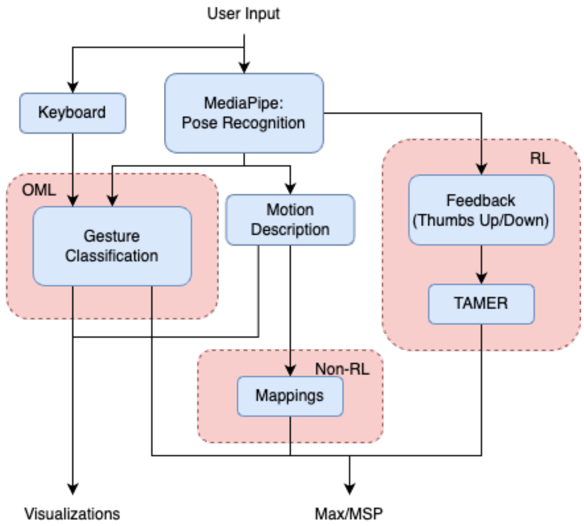

Figure 2: System flow of *GestAlt*. User input is processed by the MediaPipe model and the user's gesture is classified alongside keyboard commands for online machine learning (OML). The system has two configurations: "RL", which uses the TAMER reinforcement learning agent alongside the user's second hand to alter musical mappings, and "non-RL", which uses predefined musical mappings.

GestAlt then uses two separate neural networks to determine a gesture in the form of a hand sign (a static depiction of the shape of the hand) and a motion (a pattern of the user's positions throughout multiple frames). *GestAlt* generates visualizations of the user's input gesture, including classification labels, a bounding box, and a trail representing a user's motion (see Figure 1).

Figure 2 depicts the flow of information between user input and audiovisual output. The motion description model is a classification model trained on clockwise, counterclockwise, and straight motions. *GestAlt* contrasts previous interactive music systems by the authors [60, 61] through the use of Mediapipe for hand sign classification and the addition of adaptive behaviors in the form of online machine learning and reinforcement learning.

3.1 Online Machine Learning

GestAlt uses Online Machine Learning (OML) to allow user customization by adding new gestures to a hand sign classification model [64] and retraining the model during a performance. These changes embody direct communication between the user and AI as the user initiates them through keyboard commands on a dimension they can easily control (choice of gesture). *GestAlt* begins already trained to recognize an open hand, a pointed index finger, and a fist. Users can register new hand signs with a number and text label using the "L" key, pressing "K" to enable recording and holding the number key associated with the hand sign while performing it with their other hand before using the "R" key to retrain the gesture detection model (see Table 1). These keys were visible to users, as shown in Figure 1.

Hand detection (through MediaPipe) and hand sign classification (using OML) are directly mapped to filter parameters in a looping Max/MSP patch (see Figure 3). These parameters include low-pass filter cutoffs on looping drum and bass samples ("bass filter" and "drum filter"), the resonance and cutoff of a band-pass

Key	Function	Explanation
L	Label Gestures	Opens Terminal to name/number a gesture.
K	Log Keypoints	Allows users to hold 1-9 and record gestures.
R	Retrain Model	Retrains model and turns off gesture recording.
N	Return	Disables gesture recording without retraining.

Table 1: Keyboard commands for online machine learning with *GestAlt*.

filter ("roughness"), and the rate of a randomly-pitched synthesized melody ("rate") (see Figure 4). To maintain the system's performance for continuous gestural capture, the gestural recognition model is retrained in a separate Python thread. This allows users to continue performing gestures and those gestures to be transmitted to the musical output in the Max/MSP patch while the system retrains. As the user trains new gestures using OML, they are mapped to new melodic patterns (the "rate" parameter).

3.2 Reinforcement Learning

RL can allow users to directly communicate their desired behavioral changes to the system while maintaining system autonomy when implementing those changes. As collaborative music-making is open-ended and subjective, exploration-focused models that rely on human feedback over an environmental reward signal are most appropriate for *GestAlt*.

GestAlt uses RL to map motion patterns to changes in musical parameters. The model can be re-trained using the constant stream-of-motion classification outputs collected during a session and user feedback. Similar adaptation is used in systems for interactive audio parameter [59] and sound [67] design. *GestAlt* uses Deep TAMER, designed to allow rapid training of low-dimensional tasks using deep learning [69]. A two-layer, linear neural network takes in a representation of the current gesture and current musical output and evaluates possible changes the system can cause against the potential for user approval.

Inputs	Output
Gesture (1 - 9) Hand Position ([0 - 0.99, 0 - 0.99]) Hand Size/Distance (0 - 0.99) Motion (1 - 4) Musical Parameter (0 - 3) Parameter Value (0.00 - 0.99)	User Feedback (-1 - 1)

Table 2: Inputs and Output of the Deep TAMER reinforcement learning model used in *GestAlt*. The system conducts actions (Musical Parameter, Parameter Value) to maximize a predicted User Feedback output variable by training the above Inputs as a state variable.

The mappings are defined by two variables. The first, "slider," represents major Musical Parameter mappings (see Figure 4): "rate" of a pitched droplet melody (0), the "roughness" caused by band-pass filters throughout the audio loops (1), and the "bass filter" (2) and the "drum filter" (3) that are linear in the accompanying Max/MSP patch as well as a second number that correlates

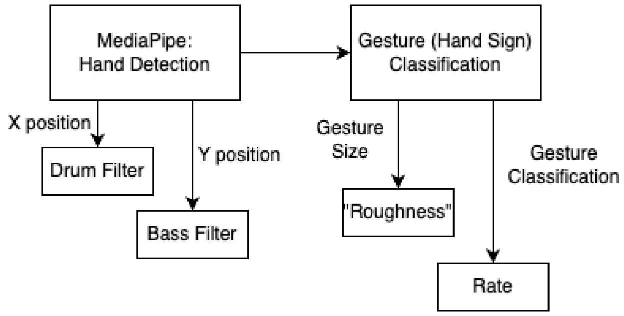

Figure 3: Mappings between model output and musical parameters in *GestAlt.* These mappings are altered during a performance using *GestAlt's* Online Machine Learning functionality, or performed by a separate model using Reinforcement Learning functionality.

to the Parameter Value of a "slider". This model outputs a value that represents user feedback, ranging from −1 to 1. With reinforcement learning enabled, the user provides feedback to the model by giving a "thumbs up" (1) or "thumbs down" gesture (-1) with their "non-dominant" hand, which is determined by the system as whichever hand is physically lower. The weights of the neural network continuously adjust to optimize positive feedback for a given motion-music state. The model retrains itself every 15 frames (roughly once per second), as long as any positive or negative feedback has been received within that window. Table 2 shows the inputs and outputs of the state variable representing performance with *GestAlt*.

After the model trains, *GestAlt* simulates the actions it could take by comparing the predicted user feedback for changing each Musical Parameter option in the model's input state variable. The combination of musical parameters and values that resulted in the model predicting the highest user feedback is then saved as the agent's action. Like the rule-based mappings, the values are sent to Max/MSP via an Open Sound Control (OSC) protocol.

4 Methodology

To evaluate *GestAlt* as an adaptive and interactive music system and to explore the role of perceived creative autonomy in interaction dynamics between human performers and an AI-based system, we designed a study to answer the following question:

- How can the perceived ability of an interactive generative music system to adapt to a user increase user ratings for anthropomorphism, perceived intelligence, and creative autonomy? How do these perceptions and ratings change across multiple usage sessions?

This study is designed to have participants interact with *GestAlt* in a way that supports reflection on the co-creative dynamics they share with the system. Subjects answered questions about how they perceived the AI component of the system's creative autonomy, which was analyzed alongside interviews and their interactions with the system.

4.1 Participant Information

This study adopted the model of a case study [18, 24] to focus on the perceptions of the participants while using *GestAlt* in depth. Case studies have often been used to assess interactive music systems and NIMEs [30, 38, 72].

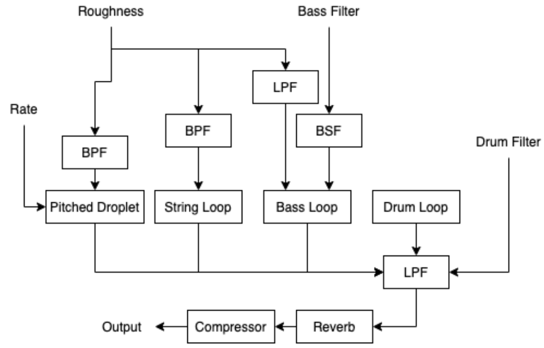

Figure 4: Mappings between musical parameter values and filter effects in *GestAlt's* Max/MSP patch.

This study has five participants (two women and three men, ages 24 to 70) with a minimum requirement of self-reported musical performance experience (a minimum of two years) and understanding of computer science, but a variety of skill levels among them. Participants completed a questionnaire before using the software to assess their level of musical experience, preconceived notions of human-AI musical interaction, and understanding of machine learning concepts. The responses encompassed the full range of 7-point Likert scale in experience performing with AI and the ability to understand the development of machine learning models. All participants reported at least 4 out of 7 regarding their interest in performing with AI. Table 3 lists the five participants and their self-reported experiences.

	Primary Instrument	Musical Experience	Experience Performing with AI	CS Experience
A	Synthesizer	High (6)	Low (1)	Medium (4)
B	Flute	High (7)	Low (1)	Low (1)
C	Percussion	High (7)	High (6)	Medium (4)
D	Percussion	High (7)	High (7)	Low (1)
E	Brass	Medium (5)	Low (1)	High (6)

Table 3: The five participants' experience levels related to music, performance with AI, and computer science.

4.2 Study Procedures

In this study, participants answered a series of Likert-scaled survey and open-ended interview questions as they completed a performance task with *GestAlt* (see Table 4 in the Appendix).

4.2.1 Pre-Questionnaire. Participants answered questions about their musical experience, including their experience with AI in music and their familiarity with AI and machine learning (see Appendix A). In repeated trials, we compared the participants' responses to these questions with their previous responses to detect a change in their interest in performing with AI.

4.2.2 Performance. Participants were randomly assigned to use one of the two versions of *GestAlt* first in the three sessions: one with (RL) or without reinforcement learning (non-RL). They practiced by improvising with *GestAlt* for five minutes while thinking aloud and talking freely. They then completed a two-minute recorded improvisatory performance and repeated the above steps for the other version.

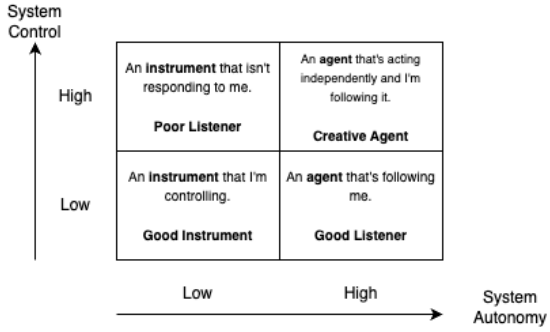

Figure 5: Expectations for participant interpretations of System Autonomy and System Control ratings.

Afterward, the participants and the researcher watched the recordings of their performances. They identified any musical interactions they saw as notable and rated *GestAlt* on whether they were leading the creative process or the system was, reflecting Rowe's "instrument" and "player" paradigms [55]. They rated the system by answering two questions on a 7-point Likert Scale while watching recordings of themselves playing with it (see Section C in the Appendix). The first, System Autonomy, asked participants to rate the system's independence on a scale of "instrument" (low) to "agent" (high) in how it responds to their gestures. The second, System Control, asked whether they felt the system was following them (low) or leading them (high) in the creative process. Figure 5 depicts expectations for how the participants might interpret combinations of low and high System Autonomy and System Control ratings based on how they perceive the *system's* level of autonomy and control.

4.2.3 Post-Questionnaire. After performances, participants compared the two versions of *GestAlt* in creativity support, satisfaction with visualizations as explanations, and trust [13, 33, 46] (Appendix B.1, B.2, B.3). Interview questions included how users perceived the system's ability to communicate with and understand their own musical goals (known as Mutual Theory of Mind [68]) as well as a sense of shared movement or "synchronization" with the system [71] (see Appendix D). Participants completed the study once every two weeks for four weeks throughout Fall 2023, and in subsequent sessions, compared their results to measure changes in perceptions over time.

4.3 Analysis Methods

This case study is designed to observe how participants' perceptions of *GestAlt* as a creatively autonomous creative partner develop over time. By performing a thorough qualitative analysis of a small number of participants, we compared their experiences and accounts with previous sessions.

4.3.1 Qualitative Data. Although participant ratings were measured as quantitative data, their main purpose was to facilitate discussion. From this experience, participants produced data from interview questions that can be analyzed qualitatively. This choice of qualitative analysis was motivated by a focus on participants' experiences recorded through verbal responses, and a case study's inclusion of only five participants would be insufficient for statistical tests. Creativity support ratings have also been used to contextualize interview responses in other Co-Creative AI studies [52] such as the evaluation of Co-Explorer, an RL-based tool for audio parameter manipulation [59].

4.3.2 Thematic Analysis. We performed deductive thematic analysis on participants' experiences to group them with specific areas of interest, such as perceived autonomy, changes in creative goals as a result of system behavior, or the effects of specific visual or auditory cues [19, 27, 66]. First, we recorded statements made by participants and combined them with their numerical ratings (for example, a participant's verbal comparison of the two systems' behaviors was paired with their preference for one version in terms of Explanation Satisfaction ratings). We grouped this data into themes related to our research question such as anthropomorphism and perceived adaptation. We further categorized participant data into frameworks such as expectancy violation [10], mutual theory of mind [68], or synchrony [71] (see Section 2.1). We also analyzed participant recordings to find performance-related patterns, such as variation in the density of effect parameter changes over time and whether or not these changes match the experiences or goals of participants.

5 Findings

System ratings, recorded performances, and interview responses share themes: First, participant preferences for behavior by an AI-based interactive music system were influenced by their experiences and changed through interaction with the AI. In addition, participants evaluated the versions of the system in terms of how their adaptive behaviors relate to anthropomorphism and changes in expectations. Finally, participants' perceptions about the system's actions relative to their own changed over time.

5.1 Preferences for AI Behavior

Participants reported preferences for AI behavior in a music system based on their previous music and computer science experience, and their preferences changed throughout the trials as they adapted to using *GestAlt*.

Participant A first indicated that they would prefer a mostly autonomous system that shows "response, not control." This preference was based on their experience using synthesizers, where they could fully control when sound starts and stops. When they could not force the RL version to return to a previously learned state, they perceived the system's autonomy as negatively impacting collaboration. However, once they understood the system better, they felt the "possibility of a conversation" and reported increased trust for the more autonomous RL. Participant B, an experienced instrumentalist with limited CS knowledge, initially expected AI systems to act fully as instruments. After their trials, they desired systems to display the "full spectrum" of autonomy in performance. Participant E, with the highest reported CS knowledge, stated that their preferred AI-based music system would be very autonomous but receptive to communication and to avoid "stubbornness," a display of autonomy that acts independently of the user in terms of creative leadership. As the RL version periodically surprised them, they stated that stubbornness should remain "low, but not zero," contrary to their initial negative description of stubbornness.

Changes in perception of *GestAlt*'s creative autonomy depended on the participants' stated preferences for AI behavior in music performance. Participant C increased their evaluation of the system after it showed autonomy and control levels they expected to dislike: it changed "rate" mapping based on changes set by the retrained gestural classification model (see Figure 3), adjusting the melody in a fashion the participant rated highly in terms of the system taking control in the improvisatory process.

5.2 Reinforcement Learning: Anthropomorphism and Expectancy Violation

The act of adding gestures in the non-RL or RL version greatly affected user perceptions of the system: for non-RL, participants were able to identify mappings by observing the mappings changed by new gestures, which increased their understanding and trust in the system; for RL, participants were able to create specific motion patterns that resulted in distinct musical decisions by the system, increasing their ratings of the AI's autonomy.

Adding a second adaptive behavior, reinforcement learning, affected the anthropomorphism participants applied to the system. When describing changes in their perception, participants referred to non-RL primarily with things they learned about the mappings created by the system. The addition of reinforcement learning influenced how the participants described their adaptation, using anthropomorphic language such as "musician" (Participants A, C, and D) and "duet" (Participant C).

The RL version proved to be more capable of surprising users. While this reduced trust for some users, it increased participants' sense of reward for exploratory behavior. For example, when Participant A adjusted to reinforcement learning with *GestAlt*, they stated that they would "want to be surprised" by future AI interactions. Participant B was surprised that AI was taking on the role of a "performer" rather than just an "instrument." Participant D was surprised when the system did not recognize their crossed-hand motion - they later attributed an increase in their rating of how well they understood the system to this experience, as it exposed a limit to the system's abilities. Whenever some behavior change was noticed, such as RL causing a new mapping between motion and a filter parameter for drums or bass (see Figure 4), the participants' perceptions of system autonomy increased. This represents a positive example of expectancy violation [10]. This also reflects the "dialogic creative process, emblematic of an improviser's way of working" caused by "momentary inspiration" and "immediate sonic realization," the goals of George Lewis' *Voyager* [40].

Participants' understanding of the RL version was based on learning how to control when reinforcement learning occurs. Participants A, B, and C provided sparse feedback to reward specific sonic combinations, ceasing their feedback once they heard a filter parameter mapping change make the drums, bass, or melody more prominent in the system's musical output. Participants D and E provided continuous feedback to force change at varying rates. Seeing a change caused by their input to the reinforcement learning agent allowed participants to learn when and how the agent manipulated the audio parameters. For example, Participants B, D, and E heard parts of the sound they had not noticed before filtered, such as the harmonic string loop, only heard during combinations of high "roughness" settings and low filter cutoffs for the bass and drum loops.

5.3 Evolving Perceptions: Mutual Theory of Mind and Synchrony

Participants increasingly relied on visualizations over time. For Participants A B and D, visualizations began as distractions but were eventually seen as necessary components to understand and perceive the system as autonomous. These visualizations impacted participants' actions while using the system, such as by encouraging Participant B to make circular motions to expand the area represented by a captured gesture or Participant D to avoid areas of the screen less likely to be accurately recognized. While participants adjusted to using *GestAlt* and learned how it adjusted to them, they developed creative goals and reflected on the system's behavior related to their movement.

In this study, participants' evolving perceptions of the system's autonomy reflected a sense of *perceived* Mutual Theory of Mind [68] with *GestAlt*, despite it lacking a model of the user's behavior or goals. For example, Participant E stated that the non-RL system version was an "instrument" lacking creative input: they felt fully in control creatively, as the system made predictable musical changes in response to gestures and only changed those mappings when they retrained the gesture classifier. RL was an AI they initially did not acknowledge the decision-making of, with Participants A, B, and E perceiving the actions of RL as random. Once they understood RL, Participants A and E had a "conversation" or "negotiation" with the system. Participant B stated that, after practicing with the system and reviewing their performance, their understanding and trust in it increased. They stated that the system is "trustworthy by definition because it is a machine... I trust that if I do it right, it will do what I ask it to do." In addition to these descriptions of dialogic interaction with the system, participants evaluated how well the system built an understanding of their motions through OML and RL. These factors demonstrate a developing mental model of the AI's behavior and "goals" (through practice with the system) alongside the user's increasing sense that the AI understands their goals (through adaptive machine learning).

As participants were improvising with *GestAlt*, they evaluated synchronization between their input motion, visualizations by the system, and the equivalent changes in musical parameters. This perceived sense of Synchrony [71] with the AI increased as the participants reflected on their actions, especially as they watched recordings of their motions and the musical results of those motions. A sense of synchronization with either version of *GestAlt* increased their trust and satisfaction with the system.

6 Discussion

This paper evaluates the interactive generative music system *GestAlt* as an AI-based creativity support tool. Musicians evaluated the system in terms of their preferences and expectations of AI behavior. Their sense of anthropomorphism was influenced by system adaptations and instances of surprise, and their perceptions of a system as a creative partner evolved alongside their understanding of the system and how to perform with it. These findings, while subject to several limitations due to system and study design, represent aspects that may be used to guide future research in designing AI-based music systems. We present the following observations as design considerations for creating music systems that support human-AI collaboration [14]. Future research exploring human perceptions during other, more varied forms of human-AI music-making may be able to develop more concrete principles for AI-based interactive music system design.

6.1 Moments of surprise can build trust and understanding in a musical AI.

The five participants each reported some positive surprise with the system or an experience that subverted their expectations while providing insight into how the system worked. For example, participants A and D both reported that they momentarily came to "some sort of understanding" with the system when it exhibited a sudden change in response to their feedback. Participants B, D, and E specifically noticed new parts of *GestAlt's* audio output when their RL feedback resulted in sudden filter changes that they had not triggered when using the non-RL version.

Furthermore, participant preferences for AI behavior were altered by these surprises. Participant B initially reported a preference for low System Control. Later, they articulated that they want systems that display autonomy and periodically take control creatively. They appreciated taking less of a "leadership" role over time. Similarly, participant C reacted positively to behaviors that contradicted their preference ratings. AI-based music systems can be designed to facilitate positive surprises, using constant change to differentiate themselves from static systems while teaching the user how to interact with them.

6.2 Visualizations can teach musicians how to watch and listen to an AI.

Depending on their backgrounds or creative goals, the participants behaved differently with information from visualizations either to learn more about the AI or learn how to perform with it. As with participants in a past study comparing the effects of visualizations on users' perceptions of creative autonomy [61], participants formed expectations for system behavior that correlated with their background and used visualizations to find methods of interacting with the system that matched those expectations. Participants with musical backgrounds used the pose recognition visualization to find stable, linear mappings that they could use expressively, and participants with computer science expertise used the motion description output to determine the limits of the system's ability to respond to them, allowing it to create output freely.

Participants' reactions to the visualizations changed over time. Participants A and B, the least experienced with AI, initially referred to visualizations as distracting but grew to rely on them. Watching their performance videos led them to reflect on the relationship between visuals and music. Participants B and E began to develop a pattern of training the system, reflecting on the visualizations, then using new movements to expand the musical output further. These changes brought about by learning through visualization demonstrate how visual explanation [29] can promote expression with an AI-based system.

6.3 Communication can support a musical system being treated co-creatively.

Although the two versions of *GestAlt* contained interactive adaptive behaviors, participants referred to the RL version that enabled constant real-time feedback with a higher degree of anthropomorphic language. The participants mainly referred to non-RL as an "instrument." Their descriptions of RL were more varied but consisted of comparisons to human-to-human collaboration with terms such as "musician" and "duet."

In particular, these terms changed as participants developed an understanding of the goals of the system or a perceived Mutual Theory of Mind with the agent, even though the system does not contain any internal modeling of goals. For non-RL, they noted between trials the mappings they had learned, but with RL, they noted what they felt the system had learned from them. The increased anthropomorphism did not lead to universally higher ratings for creativity support, satisfaction, or trust, but the RL version saw a more positive change in trust ratings between sessions than the non-RL version.

Participants were asked how "synchronized" they felt with *GestAlt* in open-ended interview questions. When discussing how "synchronized" with it they were, participants measured non-RL in terms of musical synchronization through reliability and latency. When talking about RL, they noted the "common ground" they came to with the AI regarding the intent of their gestures and the system's visualizations or musical output, relating to Synchrony [71]. By forming a sense of shared musical goals and movement, AI can provide tools with evolving creative decision-making and musical expression alongside a user.

7 Conclusion and Future Work

This paper investigated how adaptive AI behaviors, when built into an interactive generative music system, allow users to form an understanding of the system as an autonomous creative partner. Participants were more able to identify preferred behaviors as their understanding and trust of the system grew with repeated usage, and training and system adaptation allowed them to find unexpected behaviors that changed their initial perceptions. Participants assigned more anthropomorphism to the system with additional adaptive behaviors, but their creative goals did not universally align with the desire for a more autonomous system.

The participants in this study come from a limited group of people who are proficient musically and aware of basic computer science principles. As part of its case study design, this study had only five participants, representing only a specific, knowledgeable population. By analyzing trends in larger groups, or participants with more variety in experience, future research can determine more generalizable information about how musicians interact with an AI that exhibits adaptive behaviors.

Additionally, *GestAlt* represents a single kind of human-AI musical interaction: a sequential "wheelbarrow" topology in which the human musician performs gestures that are interpreted and transformed into musical output by the AI [70]. The findings and observations from using these systems may differ from those that use other models of interaction, such as compositional systems [42] and those that support symmetrical interactions with musical input and output [40]. Other performance scenarios, such as using *GestAlt* alongside another human musician, may reveal different insights about the AI component's role in facilitating human-to-human collaboration [63].

Non-improvisatory performance tasks, such as writing a piece for *GestAlt* performance, may demonstrate more effects of constrained AI behavior due to the system's limited variety in musical output changes by the AI. *GestAlt's* limited number of mappings to parameters and use of looping samples (see Figure 4) may also contribute to the ease with which participants identified system behavior, and more elaborate mappings or audio generation [1, 7] may reveal more about how adaptive behaviors or explanations affect users' learning how to interact with an AI-based music system.

8 Ethical Standards

This material is based upon work supported by the National Science Foundation Award No. 2300633. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Informed consent was received from all human participants compliant with Institutional Review Board requirements.

References

[1] Andrea Agostinelli, Timo I Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, et al. 2023. Musiclm: Generating music from text. *arXiv preprint arXiv:2301.11325* (2023).

[2] Gérard Assayag, Laurent Bonnasse-Gahot, and Joakim Borg. 2022. Cocreative Interaction: Somax2 and the REACH Project. *Computer Music Journal* 46, 4 (2022), 7–25.

[3] Frederic Bevilacqua, Fabrice Guédy, Norbert Schnell, Emmanuel Fléty, and Nicolas Leroy. 2007. Wireless sensor interface and gesture-follower for music pedagogy. In *Proceedings of the 7th international conference on New interfaces for musical expression*. 124–129.

[4] Margaret A Boden. 2007. Creativity in a nutshell. *Think* 5, 15 (2007), 83–96.

[5] Bert Bongers. 2000. Physical interfaces in the electronic arts. *Trends in gestural control of music* (2000), 41–70.

[6] David Borgo. 2006. Sync or swarm: Musical improvisation and the complex dynamics of group creativity. In *Algebra, Meaning, and Computation: Essays Dedicated to Joseph A. Goguen on the Occasion of His 65th Birthday*. Springer, 1–24.

[7] Zalán Borsos, Raphaël Marinier, Damien Vincent, Eugene Kharitonov, Olivier Pietquin, Matt Sharifi, Dominik Roblek, Olivier Teboul, David Grangier, Marco Tagliasacchi, et al. 2023. Audiolm: a language modeling approach to audio generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2023).

[8] Petter Bae Brandtzaeg, Marita Skjuve, and Asbjørn Følstad. 2022. My AI Friend: How Users of a Social Chatbot Understand Their Human–AI Friendship. *Human Communication Research* 48, 3 (04 2022), 404–429.

[9] Nick Bryan-Kinns, Berker Banar, Corey Ford, Courtney N Reed, Yixiao Zhang, Simon Colton, and Jack Armitage. 2021. Exploring XAI for the Arts: Explaining Latent Space in Generative Music. In *eXplainable AI approaches for debugging and diagnosis*.

[10] Judee K Burgoon. 2015. Expectancy violations theory. *The international encyclopedia of interpersonal communication* (2015), 1–9.

[11] Baptiste Caramiaux and Marco Donnarumma. 2021. Artificial Intelligence in Music and Performance: A Subjective Art-Research Inquiry. In *Handbook of Artificial Intelligence for Music*. Springer, 75–95.

[12] Sonia Chernova and Andrea L Thomaz. 2014. Robot learning from human teachers. *Synthesis lectures on artificial intelligence and machine learning* 8, 3 (2014), 1–121.

[13] Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 4 (2014), 1–25.

[14] Perry Cook. 2017. 2001: Principles for designing computer music controllers. *A NIME Reader: Fifteen years of new interfaces for musical expression* (2017), 1–13.

[15] Christopher Dobrian and Daniel Koppelman. 2006. The'E'in NIME: Musical Expression with New Computer Interfaces.. In *NIME*, Vol. 6. 277–282.

[16] Sarah Fdili Alaoui. 2019. Making an interactive dance piece: Tensions in integrating technology in art. In *Proceedings of the 2019 on designing interactive systems conference*. 1195–1208.

[17] Sarah Fdili Alaoui, Thecla Schiphorst, Shannon Cuykendall, Kristin Carlson, Karen Studd, and Karen Bradley. 2015. Strategies for embodied design: The value and challenges of observing movement. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*. 121–130.

[18] Joe R Feagin, Anthony M Orum, and Gideon Sjoberg. 1991. *A case for the case study*. UNC Press Books.

[19] Jennifer Fereday and Eimear Muir-Cochrane. 2006. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods* 5, 1 (2006), 80–92.

[20] Rebecca Fiebrink. 2017. Machine learning as meta-instrument: Human-machine partnerships shaping expressive instrumental creation. In *Musical instruments in the 21st century*. Springer, 137–151.

[21] Rebecca Fiebrink and Laetitia Sonami. 2020. Reflections on eight years of instrument creation with machine learning. (2020).

[22] Rebecca Fiebrink, Ge Wang, and Perry R Cook. 2007. Don't forget the laptop: using native input capabilities for expressive musical control. In *Proceedings of the 7th international conference on New interfaces for musical expression*. 164–167.

[23] Marco Fiorini. 2024. Being the Artificial Player: Good Practices in Collective Human-Machine Music Improvisation. In *EAI ArtsIT 2024*.

[24] Bent Flyvbjerg. 2011. Case study. *The Sage handbook of qualitative research* 4 (2011), 301–316.

[25] Óscar Fontenla-Romero, Bertha Guijarro-Berdiñas, David Martinez-Rego, Beatriz Pérez-Sánchez, and Diego Peteiro-Barral. 2013. Online machine learning. In *Efficiency and Scalability Methods for Computational Intellect*. IGI Global, 27–54.

[26] Ken Fujiwara, Rens Hoegen, Jonathan Gratch, and Norah E. Dunbar. 2022. Synchrony facilitates altruistic decision making for non-human avatars. *Computers in Human Behavior* 128 (2022), 107079. https://doi.org/10.1016/j.chb.2021.107079

[27] Graham R Gibbs. 2007. Thematic coding and categorizing. *Analyzing qualitative data* 703 (2007), 38–56.

[28] Karamjit S Gill. 2019. Designing AI Futures: A Symbiotic Vision. In *Creativity in Intelligent Technologies and Data Science: Third Conference, CIT&DS 2019, Volgograd, Russia, September 16–19, 2019, Proceedings, Part I 3*. Springer, 3–18.

[29] David Gunning. 2017. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web* 2, 2 (2017).

[30] Michael Gurevich, Adnan Marquez-Borbon, and Paul Stapleton. 2012. Playing with constraints: Stylistic variation with a simple electronic instrument. *Computer Music Journal* 36, 1 (2012), 23–41.

[31] Kobi Hartley, Steve Hodges, and Joe Finney. 2023. Jacdac-for-Max: Plug-and-Play Physical Prototyping of Musical Interfaces. In *New Interfaces for Musical Expression (NIME)*.

[32] Guy Hoffman and Gil Weinberg. 2010. Shimon: an interactive improvisational robotic marimba player. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. 3097–3102.

[33] Robert R Hoffman, Shane T Mueller, Gary Klein, and Jordan Litman. 2018. Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv:1812.04608* (2018).

[34] Kyle E Jennings. 2010. Developing creativity: Artificial barriers in artificial intelligence. *Minds and Machines* 20, 4 (2010), 489–501.

[35] Sergi Jordà, Günter Geiger, Marcos Alonso, and Martin Kaltenbrunner. 2007. The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction*. 139–146.

[36] Purnima Kamath, Fabio Morreale, Priambudi Lintang Bagaskara, Yize Wei, and Suranga Nanayakkara. 2024. Sound Designer-Generative AI Interactions: Towards Designing Creative Support Tools for Professional Sound Designers. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.

[37] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences* 103 (2023), 102274.

[38] Chris Kiefer, Nick Collins, and Geraldine Fitzpatrick. 2008. HCI Methodology For Evaluating Musical Controllers: A Case Study.. In *NIME*. 87–90.

[39] W. Bradley Knox and Peter Stone. 2009. Interactively Shaping Agents via Human Reinforcement: The TAMER Framework. In *Proceedings of the Fifth International Conference on Knowledge Capture*. Association for Computing Machinery, 9–16.

[40] George E Lewis. 2000. Too many notes: Computers, complexity and culture in voyager. *Leonardo Music Journal* 10 (2000), 33–39.

[41] Steven Robert Livingstone and William Forde Thompson. 2009. The emergence of music from the Theory of Mind. *Musicae Scientiae* 13, 2_suppl (2009), 83–115.

[42] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J Cai. 2020. Novice-AI music co-creation via AI-steering tools for deep generative models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[43] Pedro Lucas and Stefano Fasciani. 2023. A Human-Agents Music Performance System in an Extended Reality Environment. In *New Interfaces for Musical Expression (NIME)*.

[44] Brady D Lund, Ting Wang, Nishith Reddy Mannuru, Bing Nie, Somipam Shimray, and Ziang Wang. 2023. ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology* 74, 5 (2023), 570–581.

[45] Charles P Martin, Kai Olav Ellefsen, and Jim Torresen. 2018. Deep predictive models in interactive music. *arXiv preprint arXiv:1801.10492* (2018).

[46] Stephanie M Merritt. 2011. Affective processes in human–automation interactions. *Human Factors* 53, 4 (2011), 356–370.

[47] Sarah Nabi, Philippe Esling, Geoffroy Peeters, and Frédéric Bevilacqua. 2024. Embodied exploration of deep latent spaces in interactive dance-music performance. In *Proceedings of the 9th International Conference on Movement and Computing*. 1–9.

[48] Mohammad YM Naser and Sylvia Bhattacharya. 2023. Empowering human-AI teams via Intentional Behavioral Synchrony. *Frontiers in Neuroergonomics* 4 (2023), 1181827.

[49] Olivier Perrotin and Christophe d'Alessandro. 2014. Visualizing gestures in the control of a digital musical instrument. In *14th international conference on new interfaces for musical expression (NIME 2014)*. 605–608.

[50] Jeff Pressing. 2007. Improvisation: Methods and models. In *Physical Theatres: A Critical Reader*. Routledge, 66–78.

[51] Harish Ravichandar, Athanasios S. Polydoros, Sonia Chernova, and Aude Billard. 2020. Recent Advances in Robot Learning from Demonstration. *Annual Review of Control, Robotics, and Autonomous Systems* 3, 1 (2020), 297–330.

[52] Jeba Rezwana and Mary Lou Maher. 2022. Understanding User Perceptions, Collaborative Experience and User Engagement in Different Human-AI Interaction Designs for Co-Creative Systems. *arXiv preprint arXiv:2204.13217* (2022).

[53] Curtis Roads. 1985. Research in music and artificial intelligence. *ACM Computing Surveys (CSUR)* 17, 2 (1985), 163–190.

[54] David Rokeby. 2019. Perspectives on Algorithmic Performance through the Lens of Interactive Art. *TDR/The Drama Review* 63, 4 (2019), 88–98.

[55] Robert Rowe. 1992. *Interactive music systems: machine listening and composing.* MIT press.
[56] Robert Rowe. 2001. *Machine musicianship.* MIT press.
[57] Patrick Saint-Dizier. 2020. Music and Artificial Intelligence. In *A Guided Tour of Artificial Intelligence Research.* Springer, 503–529.
[58] Richard Savery, Lisa Zahray, and Gil Weinberg. 2021. Shimon sings-robotic musicianship finds its voice. *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity* (2021), 823–847.
[59] Hugo Scurto, Bavo Van Kerrebroeck, Baptiste Caramiaux, and Frédéric Bevilacqua. 2021. Designing deep reinforcement learning for human parameter exploration. *ACM Transactions on Computer-Human Interaction (TOCHI)* 28, 1 (2021), 1–35.
[60] Jason Smith and Jason Freeman. 2021. Effects of Deep Neural Networks on the Perceived Creative Autonomy of a Generative Musical System. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 17. 91–98.
[61] Jason Brent Smith and Jason Freeman. 2023. Effects of Visual Explanation on Perceived Creative Autonomy in an AI-Based Generative Music System. In *Companion Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23 Companion).* Association for Computing Machinery, New York, NY, USA, 25–28.
[62] Paul Steinbeck. 2018. George lewis's voyager. *The Routledge companion to Jazz studies* (2018), 261–270.
[63] Minhyang Suh, Emily Youngblom, Michael Terry, and Carrie J Cai. 2021. AI as Social Glue: Uncovering the Roles of Deep Generative AI during Social Music Composition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–11.
[64] Shigeki Takahashi. 2020. https://github.com/Kazuhito00/hand-gesture-recognition-using-mediapipe.
[65] Kıvanç Tatar and Philippe Pasquier. 2019. Musical agents: A typology and state of the art towards musical metacreation. *Journal of New Music Research* 48, 1 (2019), 56–105.
[66] Porcia Vaughn and Cherie Turner. 2016. Decoding via coding: Analyzing qualitative text data through thematic coding and survey methodologies. *Journal of Library Administration* 56, 1 (2016), 41–51.
[67] Federico Ghelli Visi and Atau Tanaka. 2020. Towards assisted interactive machine learning: exploring gesture-sound mappings using reinforcement learning. In *ICLI 2020—the fifth international conference on live interfaces.* 9–11.
[68] Qiaosi Wang, Koustuv Saha, Eric Gregori, David Joyner, and Ashok Goel. 2021. Towards mutual theory of mind in human-ai interaction: How language reflects what students perceive about a virtual teaching assistant. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–14.
[69] Garrett Warnell, Nicholas Waytowich, Vernon Lawhern, and Peter Stone. 2018. Deep tamer: Interactive agent shaping in high-dimensional state spaces. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.
[70] Gil Weinberg. 2005. Interconnected musical networks: Toward a theoretical framework. *Computer Music Journal* 29, 2 (2005), 23–39.
[71] Scott S Wiltermuth and Chip Heath. 2009. Synchrony and cooperation. *Psychological science* 20, 1 (2009), 1–5.
[72] Asterios Zacharakis, Maximos Kaliakatsos-Papakostas, Stamatia Kalaitzidou, and Emilios Cambouropoulos. 2021. Evaluating Human-Computer Co-creative Processes in Music: A Case Study on the CHAMELEON Melodic Harmonizer. *Frontiers in Psychology* 12 (2021), 322.
[73] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214* (2020).

Question	Appendix	Scale
Music Experience	A.1	Likert scale (1-7)
ML Experience	A.2	
Autonomy Ratings	C	
Control Ratings		
Creativity Support Index	B.1	
Explanation Satisfaction Scale	B.2	
Trust	B.3	
Change in Perception	D	Open-ended
Surprises (pos/neg)		
Suggestions for Improvements		
Mutual Theory of Mind		
Synchronization		
Changes since prev. trial		

Table 4: Survey and interview questions asked during the study evaluating *GestAlt*.

A Pre-Questionnaire

A.1 Music Experience

- I have listened/been exposed to live electronic music (Mus1).
- I have performed live electronic music (Mus2).
- I have performed live electronic music collaboratively (Mus3).
- I have performed live electronic music with an artificial agent collaboratively (Mus4).
- I am interested in performing live electronic music with an artificial agent collaboratively (Mus5).
- I actively listen to music, and regularly try to find new music to listen to (Mus6).
- I am able to understand music well by listening to it, including mistakes in a performance (Mus7).
- I consider myself a trained musician (Mus8).
- Listening to music can evoke memories, emotions, or "shivers" in me (Mus9).

A.2 Machine Learning Experience

- I feel confident in my ability to remember machine learning concepts from my computer science education (ML1).
- I feel confident in my ability to describe machine learning concepts (ML2).
- I feel confident in my ability to compare machine learning models (ML3).
- I feel confident in my ability to evaluate a machine learning model (ML4).
- I feel confident in my ability to create a machine learning model (ML5).
- I am familiar with the use of machine learning in a live/interactive music context (ML6).

B Post-Questionnaire

B.1 Modified Creativity Support Index

- I was able to collaborate musically with the system (CS1).
- I enjoyed using this system (CS2).
- I was able to be expressive while using this system (CS3).
- I would be happy to use this system on a regular basis (CS4).
- The musical output of the system is of a high standard (CS5).
- The musical output of the system is like that of a human (CS6).

B.2 Modified Explainability Satisfaction Scale

- From the visualization, I understand how the system works (Ex1).
- This visualization of how the system works is satisfying (Ex2).
- This visualization has sufficient detail (Ex3).
- This visualization is complete (Ex4).
- This visualization tells me how to use the system (Ex5).
- This visualization is useful to my goals (Ex6).
- This visualization tells me how accurate the system is (Ex7).
- This visualization lets me judge when to trust the system (Ex8).

B.3 Affective Trust Survey

- I believe the system is a competent performer (T1).

- I trust the system (T2).
- I have confidence in the advice given by the system (T3).
- I can depend on the system (T4).
- I can rely on the system to behave in consistent ways (T5).
- I can rely on the system to do its best every time (T6).

C Creative Autonomy Ratings

- **Autonomy:** For each musical interaction in the recording, please indicate how much the musical output felt of an instrument that you were controlling or an autonomous collaborator (1: Instrument, 7: Autonomous agent).
- **Control:** For each musical interaction in the recording, please indicate how much you felt like you were influencing the system or the system was influencing you (1: I was in control, 7: The system was in control).

D Interview Questions

D.1 On all trials

- Did this trial change your perception about performing with an artificial agent collaboratively?
- Were there any behaviors or musical interactions you found unexpected or surprising, and were they positive or negative surprises?
- Do you have any suggestions for improvements to the system?
- How would you describe your process of coming to understand the system while performing with it?
- Were there any actions that you performed specifically due to the system's visualizations or musical output?
- How "synchronized" or "in sync" were you and the AI?
- What would your preferred levels of creative autonomy and control in an interactive music system be?

D.2 On repeat trials

- How have you adjusted to using the system?
- X rating has changed since your last trial. Why do you feel differently now?