Chair for Computer Science 10 (Media Computing and Human-Computer Interaction)

RWTH AACHEN UNIVERSITY

Setup and Evaluation of a Robotic Lamp Mediating Musical Co-Creation

Master Thesis at the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Thomas Oliver Pollert

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Ulrik Schroeder

Registration date: 02.09.2025
Submission date: 13.04.2026

Zentrales Prüfungsamt/Central Examination Office

RWTHAACHEN UNIVERSITY

Eidesstattliche Versicherung
Declaration of Academic Integrity

Pollert, Thomas Oliver

Name, Vorname/Last Name, First Name

406215

Matrikelnummer (freiwillige Angabe)
Student ID Number (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/Masterarbeit* mit dem Titel

I hereby declare under penalty of perjury that I have completed the present paper/bachelor's thesis/master's thesis* entitled

Setup and Evaluation of a Robotic Lamp Mediating Musical Co-Creation

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt; dies umfasst insbesondere auch Software und Dienste zur Sprach-, Text- und Medienproduktion. Ich erkläre, dass für den Fall, dass die Arbeit in unterschiedlichen Formen eingereicht wird (z.B. elektronisch, gedruckt, geplottet, auf einem Datenträger) alle eingereichten Versionen vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

independently and without unauthorized assistance from third parties (in particular academic ghostwriting. I have not used any other sources or aids than those indicated; this includes in particular software and services for language, text, and media production. In the event that the work is submitted in different formats (e.g. electronically, printed, plotted, on a data carrier), I declare that all the submitted versions are fully identical. I have not previously submitted this work, either in the same or a similar form to an examination body.

Aachen,13.04.2026

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen/Please delete as appropriate

Belehrung:
Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt
Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.

§ 156 StGB (German Criminal Code): False Unsworn Declarations
Whosoever before a public authority competent to administer unsworn declarations (including Declarations of Academic Integrity) falsely submits such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment for a term not exceeding three years or to a fine.

§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt
(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.
(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.

§ 161 StGB (German Criminal Code): False Unsworn Declarations Due to Negligence
(1) If an individual commits one of the offenses listed in §§ 154 to 156 due to negligence, they are liable to imprisonment for a term not exceeding one year or to a fine.
(2) The offender shall be exempt from liability if they correct their false testimony in time. The provisions of § 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:
I have read and understood the above official notification:

Aachen,13.04.2026

Ort, Datum/City, Date

Unterschrift/Signature

Contents

List of Figures and Tables

Abstract

Musical group improvisation is a cognitively demanding activity that relies heavily on non-verbal communication between performers. For novice musicians in particular, the lack of an established movement vocabulary makes coordination challenging and can reduce the enjoyment and accessibility of the experience. This work investigates the potential of an embodied robotic mediator to support collaborative musical improvisation by introducing gaze and rhythmic movement as communication channels within the group.

We first define a design space for musical mediators, mapping possible input modalities, analytical processes and output channels relevant to such a system. Drawing on this design space, we implement the Dancing Desk Lamp (DDL), a robotic prototype based on a 6DOF robotic arm augmented with a 3D-printed lampshade, whose appearance and movement vocabulary are inspired by the animated character Luxo Jr. The system uses pre-choreographed animations synchronized to a musical beat, with gaze direction handled via Wizard of Oz control.

We evaluate the DDL in a mixed-methods user study with 24 participants across 8 sessions, comparing four conditions that vary in beat synchronization and gaze behavior. Quantitative results suggest that directed gaze influences turn-taking dynamics, with the robot being identified as the leader of the improvisation more frequently than any individual participant in all robot conditions. Participants in conditions with directed gaze reported feeling that attention within the group was more focused on individual participants, and the robot's influence on who participants paid attention to was significantly higher in the beat-matched direct gaze condition compared to the avoidant gaze condition. Flow and enjoyment were consistently high across all conditions with no significant differences, though trends suggest a slight advantage of directed gaze conditions. Qualitative findings reveal that participants most commonly perceived the robot as a motivator, entertainer and leader, and that directed gaze was widely interpreted as a solo cue. A secondary gaze parameter study provides design insights into how activity level, gaze height, orientation order and duration each shape the social and expressive perception of a robotic gaze.

Überblick

Musikalische Gruppenimprovisation ist eine kognitiv anspruchsvolle Aktivität, die stark auf nonverbaler Kommunikation zwischen den Musikern beruht. Gerade für unerfahrene Musikerinnen und Musiker ist die Koordination ohne ein etabliertes Bewegungsvokabular schwierig, was den Einstieg in die Improvisation erschweren kann. Diese Arbeit untersucht, ob ein verkörperter robotischer Mediator diesen Prozess unterstützen kann, indem er Blickverhalten und rhythmische Bewegung als Kommunikationskanäle in die Gruppe einbringt.

Zunächst wird ein Designspace für musikalische Mediatoren entwickelt, der mögliche Eingabemodalitäten, Analyseverfahren und Ausgabekanäle systematisch erfasst. Darauf aufbauend wird die Dancing Desk Lamp (DDL) implementiert — ein robotischer Prototyp auf Basis eines 6DOF-Roboterarms mit einem 3D-gedruckten Lampenschirm, dessen Erscheinungsbild und Bewegungssprache von der Animationsfigur Luxo Jr. inspiriert sind. Das System setzt vorchoreografierte, beatgenaue Animationen ein, während die Blickrichtung per Wizard-of-Oz-Steuerung gesteuert wird.

Die DDL wird in einer mixed-method Nutzerstudie mit 24 Teilnehmenden in 8 Sitzungen evaluiert. Dabei werden vier Bedingungen verglichen, die sich in Beatsynchronisation und Blickverhalten unterscheiden. Die quantitativen Ergebnisse deuten darauf hin, dass ein gerichteter Blick die Gruppendynamik beeinflusst: In allen Roboterbedingungen wurde der Roboter häufiger als Leiter wahrgenommen als einzelne Teilnehmende. Bedingungen mit gerichtetem Blick gingen zudem mit dem Gefühl einher, dass die Aufmerksamkeit der Gruppe stärker auf einzelne Personen fokussiert war. Erlebnis- und Beteiligungswerte lagen durchgehend hoch, mit nicht signifikanten Trends zugunsten der Direktblick-Bedingungen. Flow-Erleben zeigte keine messbaren Unterschiede zwischen den Bedingungen. Qualitativ wurde der Roboter am häufigsten als Motivator, Unterhalter und Leiter beschrieben, wobei direkter Blick von vielen als Solosignal interpretiert wurde. Eine ergänzende Studie zu Blickparametern liefert Hinweise darauf, wie Aktiv-

itätsniveau, Blickhöhe, Orientierungsreihenfolge und Blickdauer die soziale und expressive Wirkung eines robotischen Blicks beeinflussen können.

Acknowledgments

At first, I want to thank Prof. Dr. Jan Borchers and Prof. Dr.-Ing. Ulrik Schroeder for examining this thesis.

Secondly I want to express my appreciation for the help of my advisor Paul Preuschoff who offered great guidance throughout the whole process of my thesis.

I also want to thank Doğa Kıralioğlu for her help in designing the head of the Dancing Desk Lamp, Frederik Hilpert for assisting the printing process, as well as Lennart Becker for helping with any hardware problems concerning the robotic arm.

Also, I want to thank the whole Media Computing Group for the great atmosphere at the chair, all the helpful tips I received and the open ears whenever I was struggling in the process.

Special thanks also to all my reviewers who caught quite a few things that could be improved.

A very important contribution also came from everyone taking their time to take part in my studies. Thanks a lot, this work wouldn't have been possible without you.

I want to thank my friends for motivating me to stay on track as well as listening to me talking about my thesis quite frequently, I hope you found what I was talking about as interesting as me.

And finally, I want to deeply thank my family for supporting me throughout the whole of my academic career. Not only financially, which got me the privilege of fully focussing on my studies, but also emotionally whenever it was necessary!

(Disclaimer: during this thesis, I used AI tools for language editing namely Claude Sonnet 4.6 (Anthropic) for grammar and spell checking. Furthermore, ChatGPT (OpenAI) was used as a coding assistant in the prototyping phase. All content was reviewed and verified, and full responsibility for the accuracy and integrity of the work remains with me.)

Conventions

Throughout this thesis we use the following conventions:

- The thesis is written in American English.
- The first person is written in plural form.
- Unidentified third persons are described in neutral form.
- bar plots always include the 95% CI as a a line on the plot

Chapter 1

Introduction

Musical improvisation in groups is a widely practiced and socially enriching activity. Beyond its creative value, group music-making has been shown to foster a sense of belonging [Verneert et al., 2021], support well-being [MacDonald and Wilson, 2014], and build shared group identity [Lage-Gómez and Cremades-Andreu, 2021]. At the same time, improvisation is cognitively demanding, requiring performers to anticipate musical events, express themselves emotionally, and continuously interpret cues from their co-musicians [Biasutti and Frezza, 2009]. For novices in particular, these demands can make the experience feel overwhelming rather than rewarding.

Musical improvisation is an endeavour worth supporting

When multiple novices improvise together, the process lacks the established communication channels that experienced musicians rely on. Gaze, body sway, head nods and gesture serve as the primary means through which musicians coordinate entries, signal solos, and support synchronization [Bishop et al., 2019; Davidson, 2012]. Novices, still focused on the instrument itself, are ill-equipped to give these cues in the musical process, leaving the group without a shared structure to hold onto. A mediator that could introduce these cues artificially, in a way that fits naturally into the physical and social space of the performance, has the potential to lower this barrier and make the experience more accessible and enjoyable.

Novices might benefit from an artificial leader in improvisation processes

The Dancing Desk Lamp - an embodied musical mediator

This work proposes the Dancing Desk Lamp (DDL), a robotic mediator designed to participate in musical group improvisation through rhythmic movement and gaze. Inspired by Hu et al. [2025] and resembling the animated character Luxo Jr., the DDL uses a 6DOF robotic arm augmented with a 3D-printed lampshade to communicate through the same physical channels musicians use: body sway, rhythmic movement and directed gaze. To investigate the design possibilities for such a system, we first define a design space for embodied musical mediators, mapping possible inputs, analytical processes and outputs. We then implement a prototype and evaluate it in a user study with 24 participants in groups of 3 across 8 sessions, examining how the robot's gaze behavior influences the social dynamics of a collaborative improvisation process.

Key findings - The DDL improves enjoyment - is frequently identified as a leader - its gaze is perceived as a solo cue

Our results show that the robot's gaze direction meaningfully influences turn-taking dynamics. Participants in conditions with directed gaze more frequently attributed leadership to the robot, and reported feeling that attention within the group was more focused on individual participants at a time. Across all sessions including the robot, it was most commonly perceived as a motivator, entertainer and leader, and was rated as likeable without being perceived as strongly human-like. Directed gaze was widely interpreted as a solo cue, prompting increased musical activity in the gazed-at participant, though some experienced this as pressure rather than encouragement. The gaze parameter study revealed that gaze height, activity level, order of orientation and duration each carry distinct social meaning, with long gaze reading as confident but potentially threatening, and gaze from below reading as friendly and curious.

Contributions - a design space for musical mediators - the DDL - exploratory evidence for the benefits of the DDL in musical improvisation

The main contributions of this work are: (1) a design space for embodied mediators in collaborative musical improvisation, mapping inputs, analytical processes and outputs relevant to a musical mediating system; (2) the Dancing Desk Lamp, a robotic prototype implementing physical and gaze-based communication as a mediator in a musical improvisation setting and (3) exploratory evidence suggesting that an embodied mediator's gaze behavior influences turn-taking dynamics, leadership attribution and social engagement in a group improvisation process, supported by both quantitative ratings and qualitative accounts from participants describing the robot's role as motivator, leader and solo cue, alongside design insights into how specific gaze parameters shape impression of the robot on observers.

1.1 Outline

Chapter 2 reviews related work on collaborative musical improvisation, the role of embodied interaction in supporting creative collaboration, and existing robotic systems in musical and social contexts. Chapter 3 defines a design space for embodied musical mediators and describes the Dancing Desk Lamp as one instantiation of that space, covering the prototype hardware, animation pipeline, control architecture and movement design. Chapter 4 presents the study design and evaluation methods, followed by the quantitative and qualitative results and a discussion of the findings, before closing with the study limitations. Chapter 5 summarizes the contributions of this work and outlines directions for future research.

Chapter 2

Related Work

2.1 Collaborative musical improvisation

Musical collaboration can foster well-being, group formation and stress relief

Understanding the value of musical group improvisation motivates the design of systems that support and enhance it. Musical improvisation is described as real-time creative performance of novel music. Participants sing or play an instrument to take part in the activity. In group scenarios, performers need to collaborate to create a coherent piece of music, thus developing complex non-verbal communication skills for working together [Biasutti, 2017]. This highly interactive and collaborative nature of improvisation contributes to many observed benefits beyond mere enjoyment [MacDonald and Wilson, 2014]. In educational contexts, they can foster active participation and encourage a focus on a common goal, thereby supporting the formation of a shared group identity [Lage-Gómez and Cremades-Andreu, 2021]. This group identity and coherence through music making can contribute to a general sense of well-being [Verneert et al., 2021]. Furthermore, benefits of group music making on individuals with chronic illnesses have been observed [Mukherjee and Mutnury, 2021]. It is also associated with stress relief and the mitigation of burnout effects [Bittman et al.].

These positive effects provide strong motivation to explore musical group improvisation, with the aim of enhancing engagement, enjoyment, and the quality of the improvisational process

2.1.1 Communication during musical collaboration

Movement serves an expressive function in musical collaboration

Body movement or gestures are a crucial part of musical performances. It is shown that the physical aspect of a performance can convey the same structural information to an audience as the auditory counterpart [Camurri et al., 2004]. A similar communicative function of movement can be seen in clarinet duos, who use body sways, bobs and nods to enhance their coordination [Davidson, 2012]. They achieve this by indicating the pulse of the music and developing a movement vocabulary to signal expressive and timing cues. Davidson [2012] also observed that the communication between musicians relied less on facial expressions. This opens the way for a mediator that relies on a movement vocabulary.

Gaze as another important communication channel

Another important means of communication in musical performance is gaze. The direction of the gaze of one person is able to bring attention of another person to where they are gazing to [Frischen et al., 2007], enabling coordination of attention in collaborative settings. Ensemble musicians use gaze to stabilize temporal coordination [Bishop et al., 2019]. This makes gaze a useful means of communication for a system mediating musical collaboration processes, by shifting attention between performers, as well as stabilizing a shared meter.

Musical conduction, a more explicit form of physical musical communication

A more mediating form of bodily communication in music is the art of conducting. Predefined movements communicate the pulse of the music, demonstrate timing, dynamics, phrasing, and expressive intent, as well as signaling entry points, solos or changes in the musical character [Green and Malko, 1987]. Conductors also often use gaze in their conducting to indicate entry points by their gaze direction, as well as more sophisticated signals, often relying on facial expressions. These can communicate calls for attention, as well as expressive cues such as crescendo and accelerando [Poggi et al., 2020]. Gaze cues also do not rely on mutual gaze, meaning they can be perceived from the periphery [Yokoyama and Takeda, 2019]. These insights into conduction further back the claim that rhythmic movements and gaze can be a meaningful channel of communication for an artificial mediator for music-making.

2.2 Supporting collaboration and creativity with technology

A mediator should fit into the social context

Drawing on the description by Dourish [1999], social computing is a disciplinary partnership between HCI and sociology. Instead of mainly focussing on cognition and the individual user, social computing acknowledges interaction as a practice embedded in relationships and activities. Good interaction design should therefore study the relationships surrounding the interaction, as well as an understanding of how and why users perform certain activities [Suchman, 1987]. In the context of musical improvisation, this means a mediating system should not be designed around individual user goals but around the relational dynamics of the group performance.

This work is colse to the field of Computer supported collaborative work

The importance of social components varies across application domains, becoming critical when interaction occurs in a collaborative context. A popular example of a discipline incorporating social principles is Computer Supported Cooperative Work (CSCW). Systems focus on real time coordination, gestures and timing in collaborative processes and how to integrate systems into them [Schmidt and Bannon, 1992]. Musical improvisation is a prime example of such a collaborative process, where real-time coordination between performers is critical, and any supporting system must integrate into the social flow of the activity rather than interrupting it.

A mediator for musical collaboration is a Creativity Support Tool

While CSCW systems generally support coordination in a variety of collaborative tasks, Creativity Support Tools (CSTs) are a class of systems that focus specifically on enabling creative work, both for individuals and for groups.[Fischer, 2004; Shneiderman, 2000]. Unlike general CSCW systems, which may be concerned primarily with task completion and process efficiency, CSTs are designed to foster exploration, idea generation, and creative problem-solving, while still supporting the social processes that underlie successful collaboration.

This work on collaborative musical improvisation falls into the intersection of CSTs and CSCW: it aims to support a creative domain through a system that mediates social coordination and respects the norms and emergent practices of its participants. By situating the Dancing Desk Lamp within this framework, the design can address not only task-oriented coordination but also the subtler dynamics of attention, timing, and social signaling that are critical to group creativity.

The goal for the prototype therefore becomes to design a CST that communicates in a similar fashion as musicians in collaborative musical tasks, as worked out in Section 2.1.1.

2.3 Interaction model for a musical collaboration system

Tangible interaction fits real world intuitions

Tangible interaction describes interfaces that employ some form of physical interaction. These interfaces built upon real world experiences of users, which lowers the threshold for interactivity [Hornecker and Buur, 2006]. By offering input modalities that are haptically manipulatable as well presenting output in a more peripheral way, interaction with digital media becomes more similar to how we experience our ordinary physical environment [Ishii and Ullmer, 1997]. For a mediator in a musical improvisation context, this peripheral quality is particularly valuable since performers cannot shift their center of attention away from their instruments and co-musicians.

Embodiment - the intersection of social computing and tangible interaction as a paradigm for this work

Embodied interaction was proposed as a unification of social computing and tangible interaction in Dourish [1999]. This mainly analytical work frames interaction as engagement with ones physical and social environment. This challenges the old view of systems merely representing information and shift interaction into the world of practical participation. Later refinements in Dourish [2004] claimed that systems should be context aware, meaning they should be evaluated in how they participate in action.

Following this line of thought, a mediator for musical improvisation should be physically present in the process, as well as actively participate in it. Its cues overall behaviour should fit the social context, meaning they should in some form resemble the behaviour of a human mediator in musical tasks.

As seen before in 2.1.1, communicative social norms during musical performance revolve around rhythmic body movements, gestures and gaze direction. This is the reasoning behind the setup of the Dancing Desk Lamp, which communicates via its movements and posture with special regard to its head position.

2.4 Robotics

Robots are frequently used as CSTs

Robots have the ability to elicit creativity through interaction with them during the creative process of different creative disciplines [Ali et al., 2021; Kahn et al., 2016]. This has also been shown in several cases of music creation, in solo and collaborative approaches. Rythmic body movements, gestures and gaze as expressive means in robots have been studied in several examples already. The following section presents research about different designs and results of robots in musical context, that help with designing the Dancing Desk Lamp.

Robotic gaze as a means to shift attention

A robot's gaze direction influences nearby humans attention and gaze, enabling controlled attention shifts in group interactions. [Yonezawa et al., 2007]. The use of this effect for mediation has been studied by, Pham et al. [2024], who applied gaze-based mediation in a non-humanoid robot to enhance equal participation during group ideation. Their goal was to use the robot, visible on the left in Figure 2.1, to give attention to the least involved group member in the

Figure 2.1: Left: Group ideation mediation robot,
Right: Shimon, used for robot-human musical interaction

collaboration. They identified two main capabilities of their system as supporting equal participation and "breaking the ice". These results suggest that head-orienting robots are promising candidates for mediating turn-taking in musical collaboration.

Shimon - A robotic system supporting musical creation

A very similar robot to the Dancing Desk Lamp is Shimon, a well known robotic marimba player [Hoffman and Weinberg, 2010]. Its architecture has been used for several music related projects. In Gao et al. [2024], Shimon was used to explore the use of physical bodies for robots to communicate in musical context, you can see the robot and its moveset on the right side of Figure 2.1. It uses head and body movements to communicate tempo changes, which a piano player should adjust to. Their key findings: the use of expressive gestures for human robot co-musicianship can enhance synchronization just as in interhuman setups, as well as creating a more engaging and enjoyable experience. Similar expressive principles are applied in the Dancing Desk Lamp, though it differs in head shape, facial features, and its use in a collaborative context compared to Shimon.

Musical conduction robots

Several research projects studied the use of robots as conductors in orchestra setups. The varieties of robots reach from simpler time keeping assistant which visualize the meter, to more expressive variants that incorporate gestures in their conducting, which either communicate patterns in the music [Geuther et al., 2010] or more recently express emotionality of musical passages [Chawla et al., 2025].

Mood contagion through bodily expressions

The body of a robot is an important channel of communication for robots. Humanoid robots, like Nao, are shown to be capable of displaying identifiable emotions [Beck et al., 2012]. Robotic lamps specifically have been shown to express identifiable emotions through movement, as demonstrated by Gertie [Metzler et al., 2012], a lamp-shaped robot whose expressive motion design serves as a direct precedent for the Dancing Desk Lamp. Valence refers to the degree to which an emotion is positive or negative. Prior work has demonstrated that positive valence of bodily expressions in robots evokes positive feelings in humans [Xu et al., 2014]. Mood contagion enables enhancement of collaborative activities, as positive mood has been associated with smoother interaction, higher satisfaction, and improved task performance [de Rooij et al., 2024].

Conveying a specific emotion or valence requires carefully designed movements that are interpretable by observers. If they cannot recognise the intention, the effect becomes arbitrary. There is a variety of frameworks for human emotional expressions [Chang et al., 2019; Dael et al., 2012], as well as robotic ones [Chang et al., 2025; Rogel et al., 2022; Savery et al., 2020]. However, expressive movement vocabularies are specific to the morphology of the robot they were designed for and cannot be directly transferred to a differently shaped platform [Peng et al., 2015]. Since the Dancing Desk Lamp has a unique lamp-shaped, non-humanoid structure, a dedicated movement set needs to be designed, drawing on classical animation principles [Thomas and Johnston, 1981], conducting technique [Green and Malko, 1987] and the frameworks named above. This design is described in Section 3.2.3.

Research gap - Influencing turn taking in musical improvisation with a robotic mediator via gaze direction

While several robotic systems have been studied in musical and social contexts, they are rarely examined as mediators of a collaborative musical process between multiple human performers. Similarly, while gaze has been studied both as a communication channel in music-making and as a tool for robotic turn-taking in group settings, its use by a robot to influence turn-taking dynamics within a musical improvisation remains largely unexplored. This work aims to close this gap by defining a design space for embodied musical mediators and examining the influence of a robot's gaze behavior on the turn-taking mechanisms and social dynamics of a collaborative improvisation process.

Chapter 3

Design Space for a Musical Mediator

3.1 Design Space

To explore what a mediator in a musical improvisation process could be and do, we defined a design space structured around the inputs it might draw upon, the outputs it could offer and the analytical processes necessary to bridge between them, all grounded in the musical context and the social dynamics of collaborative music-making.

3.1.1 Possible Inputs

Possible input methods for the system should consider that musical improvisation is already a very cognitively challenging task that demands continuous cognitive and physical engagement. Furthermore, the hands of performers are typically occupied by their instruments. For this reason, we exclude possible inputs that are conscious or manually executed by hand and rather focus on the possibilities to capture information of the ongoing improvisation process passively, without the need for performers to actively control the system.

The mediating system should be informed from the ongoing musical process - Users are already occupied with music making

Audio Input

The first carrier of information that could be used is the music being played. A suitable system could record the auditory information in the room and analyze it to gain information.

Midi Input

A simpler approach would be to use the mediator in a setup with digital musical controllers. These instruments typically use the MIDI protocol to pass their inputs to some music software. This protocol transfers the notes being played, their length and velocity. This way the system receives already structured data that is easier to be analyzed.

Video Input

Another form of gathering information about the musical process is the visual domain. As mentioned in Davidson [2012], musicians use rhythmic movements and gaze to communicate during musical collaboration. These movements and gestures can be captured and analyzed to inform the system about the improvisation process.

3.1.2 Possible Outputs

As established in the related work, communication in collaborative musical tasks is often experienced in the periphery. We propose that an artificial mediator should align with this with the use of output methods that suit peripheral perception.

Screen based visual output

Graphical user interfaces have historically been the dominant output modality in human–computer interaction [Hornbæk et al., 2025]. Screens have the advantage of being capable to display a wide range of visual information. System designers could animate a human conductor and show them to the participants, display a color fitting the mood of the current music or incorporate abstract animations. The screen approach can be taken a step further via 3D-tabletop displays, which would make it possible to see the same object or entity from multiple different sides [Saad et al., 2022].

Auditory output

Another approach for a mediator to be received during musical collaboration is the auditory channel. There are two distinct approaches for this. The mediator could become a co-musician adding to what's already played [Hoffman and Weinberg, 2011], or interfere by giving verbal commands, like an initial counting or reinforcing signals at fitting times of the musical piece.

Physical output

Physical output in the form of movement is another option. Signals can be designed resembling inter-human communication during musical collaboration as done by [Gao et al., 2024]. This approach involves the choice or design of a robot, creation of a useful move set and some degree of choreography.

3.1.3 Analysis Tools

Audio Analysis Tools for recognizing meter, mood or energy level

Since the input channels are all capturing data implicit in the process rather than direct commands that can be mapped to the output, some sort of analysis and classification needs to be integrated into the system. Audio analysis tools like ESSENTIA [Bogdanov et al., 2013] can capture the tempo, mood[Chawla et al., 2025] and energy level of a piece. Classification on how much each participant is contributing to the performance can also be valuable information for a mediating system aiming to balance participation.

Video Analysis Tools - reading the body language of musicians

Video analysis tools [Camurri et al., 2004] and gaze tracking [Bishop et al., 2019] can extract information inherent to the performers' movement. This way the system could identify the pulse of the music through body sway and head nods, detect the leader of an improvisation through gesture and gaze direction, or infer the emotional engagement of individual performers.

MIDI Analysis - making use of a structured format

Even if already in a structured format, midi data needs some sort of analysis as well. Moments of silence would be easily detectable due to the lack of incoming notes, mood could be classified by identifying the played scales, harmonies, or melodies and comparing them with music theory, as already done by Savery et al. [2020]. Velocity and the number of notes played during a time frame can be used to identify the energy level of a performance.

Bridging in and output

The relationship between inputs and outputs is central to the success of a musical mediating system. Collecting and analyzing input is not sufficient on its own. The mapping between an analyzed dimension and a chosen output modality is itself a meaningful design decision that shapes how the mediator is perceived and whether it is effective. Table 3.1 provides an overview of possible mappings across the design space. For instance, the pulse of a musical piece could be extracted from audio, MIDI, or video input and expressed through the movement tempo of a physical mediator. The mood of the ongoing improvisation could be analyzed and reflected through visual animations, color, or a musical response from an artificial collaborator. The energy level of the performance could modulate the intensity of physical movement or visual output. More socially oriented mappings are also possible. The least contributing musician could be identified and engaged through directed gaze or designated gestures, with the aim of encouraging more balanced participation. As the table illustrates, which dimensions are taken into consideration and which output modalities are used to express them is ultimately a creative endeavor. The design space presented here is intentionally broad. The prototype developed in the following chapter focuses on a subset of these possibilities, specifically physical output of an embodied mediator and its capabilities to support and shape the collaboration.

Table 3.1

Input Signal	Analyzed Dimension	Output
Audio / MIDI notes	Mood	Visual animation / Musical response
Audio dynamics	Tension	Movement urgency / Visual intensity
MIDI scale / harmony	Tonality	Visual color palette
MIDI note density	Complexity	Visual complexity / Movement complexity
Audio / MIDI velocity	Energy level	Movement intensity / Visual intensity
Audio / MIDI rhythm	Pulse	Dance rhythm / Movement tempo
Audio / MIDI notes	Play quality	Feedback signal
Facial expression	Emotional state	Empathetic musical response
Video / Gaze	Participation balance	Directed gaze / Designated gesture
Video / Body movement	Leadership	Responsive gesture
Audio + Video	Player synchronization	Coordination cue

Table 3.1: Possible mappings of input signals, analyzed dimensions, and outputs in a musical mediating system.

3.2 The DDL-Prototype

To investigate one promising aspect of the design space, we developed a robotic prototype called the Dancing Desk Lamp (DDL) which can be seen in Figure 3.2. We focus on the physical, embodied output modality, while using a combination of a backing track as a timing anchor, pre-choreography of moves and wizard of oz techniques to simplify the implementation of the system, while designing it in a way that allows for later introduction of input reading and analysis. We decided to use an embodied approach since rhythmic movements have the potential to reinforce synchronization among players, while gestures can communicate entry points, solos, or provide feedback to musicians, as described Section 2. Furthermore, an embodied mediator can make use of the same channels of communication that musicians already employ during collaboration. Gaze direction, rhythmic movement and body sway allow it to fit naturally into the social dynamics of the improvisation process rather than introducing a foreign interaction paradigm.

We made the design choice of an embodied mediator

The system is based on a robotic arm, augmented with a custom 3D-printed lampshade and bulb. The design is inspired by Hu et al. [2025] and resembles the animated character Luxo Jr., as a design assumption to support intuitive interpretation of its movements due to the widespread familiarity with the character. The choice of a non-anthropomorphic appearance is deliberate, as it avoids the uncanny valley effect that arises when a robot closely but imperfectly resembles a human [Geller, 2008], allowing the focus to remain on the expressiveness of the movement rather than the plausibility of the form.

We deliberately chose a non-antropomorphic character - avoiding the uncanny valley

3.2.1 Hardware and Animation pipeline

The Widow-X 250 6DOF[1] offers 6 degrees of freedom as well as a beginner-friendly interface for easy manipulation

We control a WidowX250s robotic arm via direct joint manipulation

[1] https://www.trossenrobotics.com/widowx-250, *last accessed 10.04.2026*

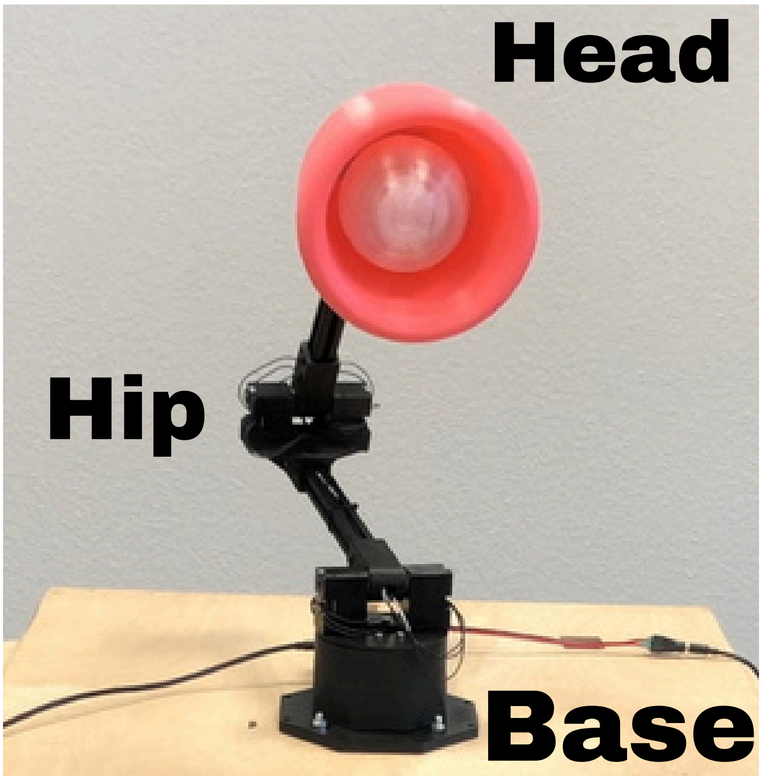

Figure 3.2: The Dancing Desk Lamp - Appearence and naming conventions

through its various high level libraries. We used the python ROS interface that offers the set_joint_position function. This interface allows replaying recorded motion sequences by sending joint configurations with associated timing parameters. This way lists of intermediate joint angles and time deltas can be iteratively fed to the robot. We refer to such sequences as animations, which are stored in CSV files. Using this approach allows for several ways to create animations.

Animations can be designed in Blender

First, we explored creating animation files in Blender by importing the robots model and attaching an armature to it, which could then be manipulated over time and exported to CSV via a simple Python script. While this gave us the opportunity to design movements on a visual clone of the robot in simulation and use Blender's animation tools, the resulting movements were rigid and the approach required substantial time and expertise to produce lively results, so we switched to a different approach.

Record and Replay for rapid lively animation generation

The interbotix library offers a record and replay function. First the robot is set into recording mode. The servos are then non-blocking and the robot's position can be manipulated by hand. During this manipulation, the script saves all joint angles and timestamps in a database file that later can be replayed or converted to a CSV file. This allows for intuitive creation of fluent movements on the physical robot in a very short time frame. It is also helpful that a beat can be played while manipulating, allowing the movement to already follow a consistent rhythm and meter. We chose this approach since it allowed for rapid design of fluent natural movement.

We stretched animations to a BPM parameter by defining keyposes - poses supposed to be on the beat

Creating an animation to a beat with a fixed bpm, while helpful, has the problem of reusability for different BPMs. It would be too tedious to create a new movement for every round number bpm, and impossible for a continuous range. We therefore created a time stretching technique to fit a recorded animation to a bpm. To do this, we identified keyposes, which are the positions the robot should be in when the beat hits. In dancing and conducting, the moment the beat hits is often one of a physical direction change. We walked through the animations frame by frame, identified poses corresponding to directional changes, and indicated them in an additional CSV column. Instead of using the recorded timestamps when replaying the animation, we compute the duration of a single beat at the current BPM and linearly distribute it across the frames between consecutive keyposes. When doing this iteratively for each keypose, it is ensured that the robot hits a desired pose every beat.

Gaze direction is handled with pre defined positions of performers

To enable gaze direction toward individual musicians, movements were designed around a fixed point in space where a musician would later be positioned. Since the system does not incorporate facial or position tracking, musicians need to be placed at defined positions for the gaze direction to be meaningful. Directing attention toward other musicians is achieved by offsetting the base joint, introduc-

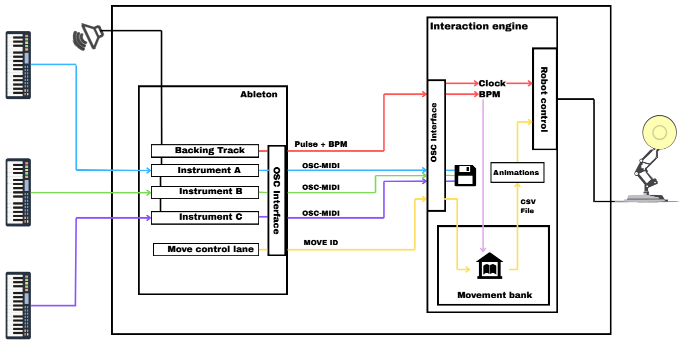

Figure 3.3: High-level Architecture of the mediating system

ing a turn of the robot that simulates a body sway with the head following. This turning behavior was the component handled via wizard of oz.

3.2.2 Control Architecture

Our interaction engine dynamically introduces animation snippets

To control the robot we built an interaction engine which can replay pre-choreographed animations on the robot. The interaction engine runs a base animation on loop, while allowing new animation snippets to be dynamically introduced. Transitions between animations are achieved by interpolating joint positions between the outgoing and incoming animation over a fixed number of frames, weighted by a sigmoid function to produce smooth acceleration and deceleration. A new animation is gradually faded in, run, and faded out shortly before it ends or after a new one replaces it.

High level control of the robot happens from within Ableton

Figure 3.3 illustrates the high level communication architecture of the system, showing the signal flow between Ableton Live and the interaction engine with color-coded communication channels.The system uses Ableton Live (a digital audio workstation) as a global timing anchor, avoiding the need for live tempo analysis by introducing a backing track that fixes the tempo on Ableton's beat grid. A custom Max4Live (or M4L, a visual programming environ-

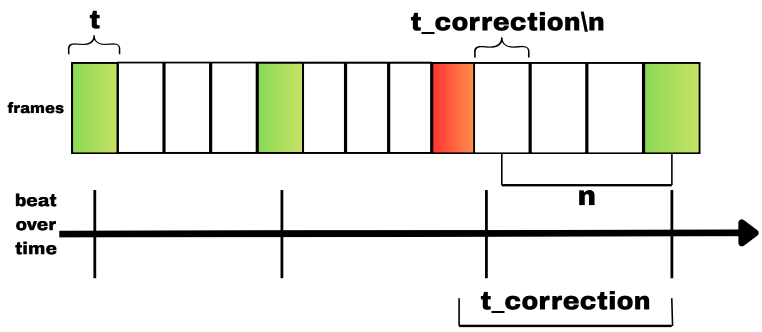

Figure 3.4: Visualization of the time timestretching technique to match keyframes to the beat

ment integrated into Ableton Live) plugin was developed to send OSC messages (Open Sound Control) over Ethernet to the interaction engine on every beat, carrying the current BPM value. An incoming OSC message therefore serves as a beat pulse, while the BPM value is updated in the interaction engine whenever it changes. A second instance of the plugin was placed on a dedicated control lane and modified to broadcast the start and end of MIDI notes. This allows choreographies to be composed directly in Ableton's MIDI editor by mapping notes to animation primitives in the interaction engine. The start of the note brings the new animation in, the end starts fading it out. The control lane was also used to trigger directional turns toward specific participants or to signal the robot to move freely without beat synchronization.

Animations are synched to a fixed grid

Although the animations are time-stretched so that keypose frames align with beats, temporal drift can still occur. To compensate, the system evaluates each incoming pulse message and determines whether the current frame corresponds to a keypose. If it does not, the remaining time until the next beat is estimated based on the current BPM and linearly distributed across the frames leading to the next keypose. This approach ensures that keyposes remain synchronized with the beat over the course of the animation. A visualization of this technique can be seen in Figure 3.4.

This setup allows the robot to synchronize to the pulse of an Ableton project with dynamically changing BPM while also supporting primitive-based control of its movements directly from Ableton. The following section describes the movement primitives designed for this system and the considerations behind them.

One shortcoming of the current implementation concerns the fallback behavior when an animation fades out. The system keeps the most recently introduced animation aligned to the beat grid, but does not maintain alignment for animations lower in the stack. When the top animation fades out and the system falls back to the one below, that animation may be at an arbitrary point in its cycle, far from a keypose. This causes the robot to either rush through frames or pause briefly to realign, which was observed to be perceived as unnatural in preliminary testing. In retrospect, a cleaner solution would be to retrigger the fallback animation rather than resuming it mid-cycle, which would ensure keypose alignment and produce smoother transitions.

3.2.3 Movement design

We recorded small dances that can be triggered dynamically

We tested two approaches of picking recorded movements to create choreographies. The first one was using very short primitives like "moving up" or "turn to the left" and then trigger and blend them on to a baseline movement. This allows for very fine-grained control of the choreography or even the creation of heuristics choosing a primitive move to create unique dances. The problem with this was one of blending, since a short movement needs a rather short blending phase. These short blendings lead to sudden jumps, which slightly shocked or distracted people around the lamp. We therefore took a different route and created longer recordings of movements, small dance artifacts. We picked one of them as a baseline dance and blended in new dances to have some variety in the choreography.

The primary degrees of freedom contributing to the lamp's expressiveness are the head joint and the elbow joint of the

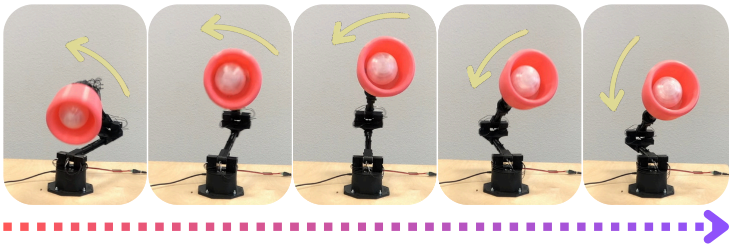

Figure 3.5: The cyclic hip bop - Serving as a base fallback movement

robotic arm. In the lamp configuration, the elbow joint visually resembles a hip due to the orientation of the structure. For clarity, we therefore refer to this joint as the hip in the following descriptions.

All movements were designed with a positive valence in mind. Therefore, we created them very dynamically. We therefore kept movements dynamic and oriented upward more than downward. We did this to introduce a positive mood into the process with the help of the mood contagion effect described in the Related Work Section 2.

Positive valence - introducing positive mood into the collaboration

The cyclic hip bop (see Figure 3.5), our baseline movement, is centered around periodic motion of the hip joint. The movement begins in a lowered position and follows an upward arc to a mirrored configuration. Rather than strictly reversing this trajectory, a slightly varied motion is executed in the opposite direction to avoid perfect symmetry, which was perceived as unnatural in preliminary tests. After several such variations, the sequence repeats, forming a cyclic pattern.

A baseline movement continually runs even if no new animation is tiggered to achieve a non stopping behaviour

The first, middle, and final frames of the sequence correspond to key poses. These were selected as alignment points for the Interaction Engine, which synchronizes movement with musical beats. The selection is motivated by observations from human dance, where changes in movement direction often coincide with beat accents.

Keyposes at moments of direction change

The wiggle movement (Figure 3.6) compresses the robot's

The wiggle movement

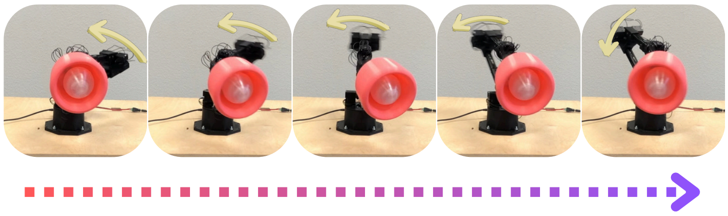

Figure 3.6: The wiggle - Horizontal hip movement with a stationary neck, the head looks around

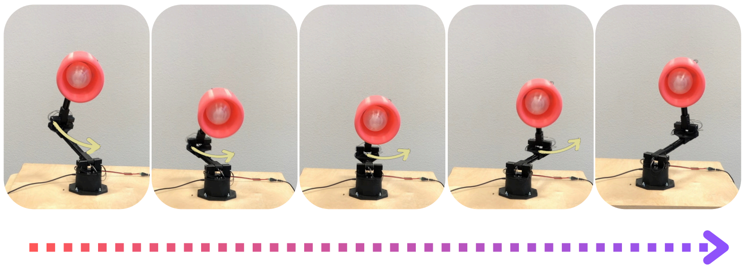

Figure 3.7: The conducting hip - Fastest at the bottom direction change, like the stick of a conductor

structure to a minimal height and introduces lateral oscillations of the hip joint. Simultaneously, the head joint moves in the opposite direction which creates the effect reminding a dog lowering its head and wiggling with its tail.

The conducting hip movement

The conducting hip movement (Figure 3.7) is inspired by the motion trajectory of a conductor's baton. The defining characteristic is the acceleration peak at points of directional change, particularly at the lowest point of the movement. This produces a pendulum-like effect, visualizing the beats with short stops at the top and a swing at the bottom.

The hip hump

The hip hump movement (Figure 3.8) consists of forward and backward displacements of the hip joint without lateral movement. This creates a squash and stretch pattern, since moving the hip forward, forces the robot to move up and therefore gaining height.

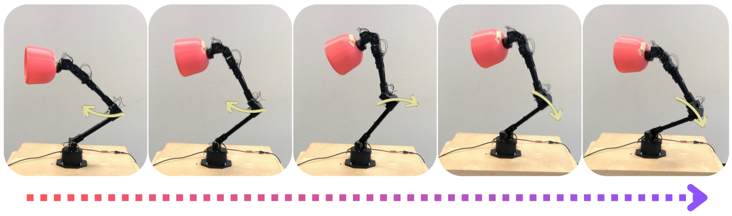

Figure 3.8: The hip hump - Forward and backward movement of the hip

3.2.4 Integration of classic animation principles

The movement design of the DDL was informed by the classical animation principles outlined in Thomas and Johnston [1981]. Several of these principles were incorporated either deliberately or as a natural consequence of the record-and-replay approach, even though they are only partially implementable due to the physical and stationary nature of the robot.

Arcs

Arcs were deliberately incorporated into the movement design, as natural motion rarely follows straight lines but instead traces curved paths. This is particularly visible in Figure 3.7, where the pendulum-like trajectory shows a pronounced arc.

Squash and stretch

Even tho parts of the robot cant be physically squashed or stretched, Squash and stretch was applied by arranging the joints to compress or extend the robot's overall appearance throughout animations, making it appear smaller or larger at different moments. Staging was addressed by placing the robot in the center of the participant circle, ensuring it was visible and readable from all positions at all times.

Staging

Pose-to-pose and Straight-ahead

The two animation creation approaches reflect a natural parallel to the principles of pose-to-pose and straight-ahead animation. The Blender-based approach, where keyframes were defined explicitly, corresponds to pose-to-pose animation, while the record-and-replay method, where movements were created in a single continuous pass by physically manipulating the robot, mirrors the straight-ahead approach. The straight-ahead method ultimately produced more natural and fluid results, consistent with its reputation for generating livelier motion.

Follow-through and overlapping action

Follow-through and overlapping action emerged organically from the record-and-replay process. When manipulating the robot by hand, body parts naturally overshot their target position and pulled back slightly before settling, producing a subtle follow-through effect without explicit design intent.

Slow in-Slow out

Similarly, the physical manipulation process naturally produced slow-in and slow-out timing, as movements tended to accelerate out of and decelerate into held positions when moving them by hand.

Appeal

Appeal was considered in the visual design of the prototype. The lampshade was printed in a warm coral color to give the robot a friendly and characterful appearance, complementing the expressive movement vocabulary and supporting the assumption that participants would engage with it as a social presence rather than a mechanical device.

Chapter 4

Evaluation

4.1 Study design

This section contains a description of our explorative study setup up to gain insight into the question:

RQ1: How do gaze, rhythmic body movements and motivational gestures of an embodied mediator influence the creative process, flow and turn taking mechanisms in an improvisational, musical co-creation process?

Furthermore, we wanted to examine the influence of different parameters of gazing movements and therefore designed a second part of the study which aims to explore the question:

RQ2: How do different parameters of gaze behavior of an embodied agent shape the attribution of social, expressive, and cognitive qualities?

4.1.1 Study procedure

Study Part 1: Collaborative Musical Improvisation

Our study is a mixed methods approach aiming to get a broad spectrum of subjective opinions about the usefulness of the Dancing Desk Lamp in the process of musical improvisation. Additionally, we investigate different parameters of the lamps gaze, to gain insight into expressive movement design for our prototype.

The setup consisted of 3 digital MIDI Instruments placed on a circle in equal distances from each other. Instrument A was a keyboard, B a finger drum pad with a synth pad assigned, and C a drum pad with different drum sounds and some bass notes. The DDL was placed in the middle of the circle. We conducted 8 rounds of the study, each in groups of 3 participants. Once they arrived into the room they were instructed to freely chose one of the three stations where the different digital instruments were placed. The principal investigator explained the information on the informed consent and once participants agreed to it, the study started.

The first part of the study was split into four conditions. In each condition, the participants were advised to perform an improvisational piece of music to a drum loop that played for about two minutes. To give some variation, the drum loop played at a different BPM each round. The conditions were:

Fiona Improvisation without the robot

Sally Improvisation with a beat matched robot that shifts its gaze direction from participant to participant

Dori Improvisation with a robot moving slower than the beat, that shifts its gaze direction from participant to participant

Row	Letter Sequence	Number Sequence	Combined
1	F S E D	80 65 70 75	F80 S65 E70 D75
2	F E D S	65 75 80 70	F65 E75 D80 S70
3	S E F D	70 80 65 75	S70 E80 F65 D75
4	S D F E	75 70 80 65	S75 D70 F80 E65
5	E D S F	80 65 75 70	E80 D65 S75 F70
6	E F S D	65 75 70 80	E65 F75 S70 D80
7	D F E S	70 80 65 75	D70 F80 E65 S75
8	D S E F	75 70 80 65	D75 S70 E80 F65

Table 4.1: Sequence of conditions and BPMs for the experiment

Eve Improvisation with a beat-matched robot that avoids directed gaze by turning back and forth around its base

We selected these conditions to examine differences in beat-synchronized and unsynchronized behavior, the effect of direct versus absent gaze, and the differences between improvisation with and without the robot.

In the two directed gaze conditions, the investigator triggered turns toward different participants at contextually appropriate moments (roughly at the end of 8 or 16 bars of the backing track) using a wizard of oz technique, with the aim of distributing direct gaze time equally across all three participants.

Although the system supports dynamic choreography, we used a fixed movement set across all robot conditions. This ensured that any differences between conditions could be attributed to the gaze behavior and beat synchronization rather than to variations in the movements themselves.

To counterbalance order effects, we shuffled the conditions such that every condition appears twice in each position, and twice after every condition. We then shuffled the BPM measures and assigned them to our conditions. This results in the combined condition Table 4.1.

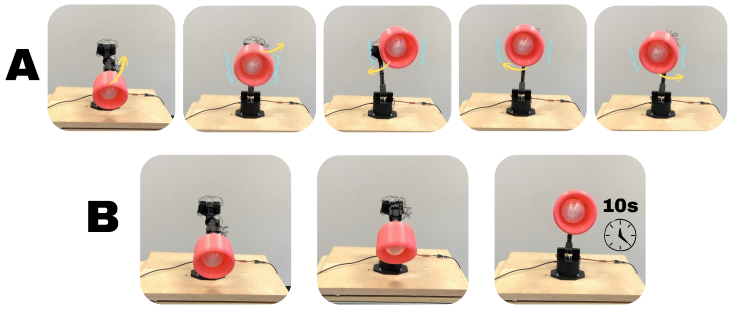

Figure 4.2: Energy level - A: very active shaky gazing, B: totally still gazing

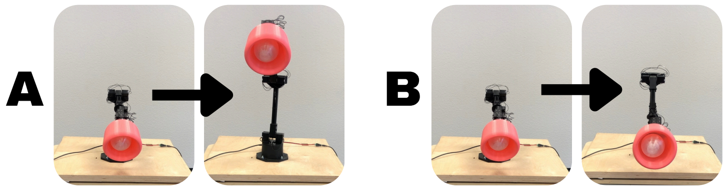

Figure 4.3: Height of the head - A: gazing from above, B: gazing from below

After each round, participants were asked to fill out a questionnaire, with two additional questionnaires after all 4 rounds.

Study Part 2: Gaze Parameter Study The second part of the study examined different ways the lamp could shift its gaze towards a subject. We decided to pick four parameters that are prevalent in gazing behavior[Cañigueral and Hamilton, 2019; Vázquez et al., 2017], namely the **activity level**, the **height of the head**, the **order of body and head orientation** and the **duration of the gaze**. For the activity level, we recorded a shaky movement that looks at the participant, while actively moving the head from side to side minimally, compared to a gaze where the lamp orientates there gaze to the subject and rests in that pose for a similar amount of time (Figure 4.2). To examine the height parameter, we fully extended the arm and oriented the head to be looking down at the subject, which we compared to po-

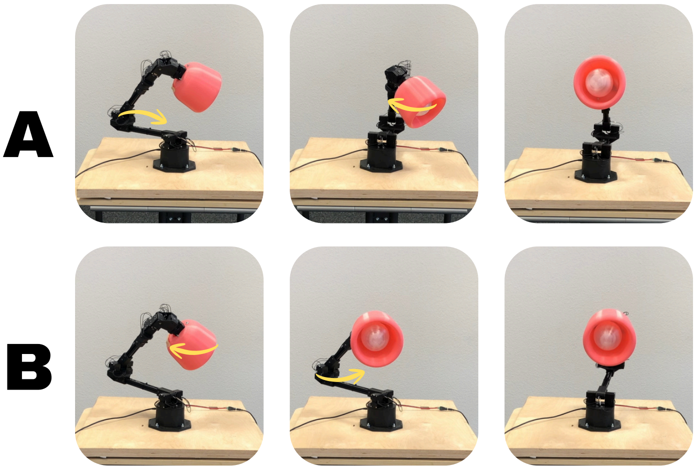

Figure 4.4: Order of body and head orientation: A: body first, B: head first

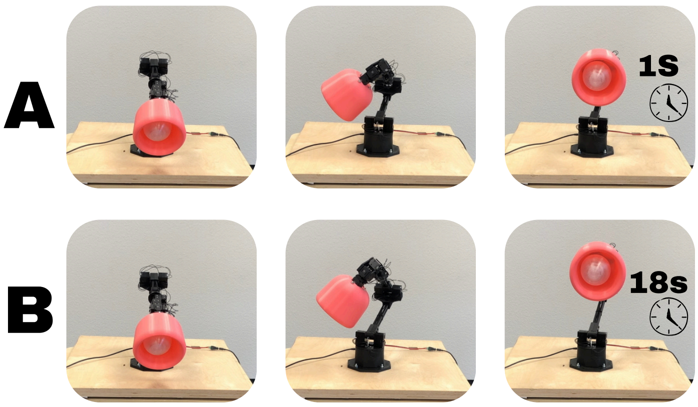

Figure 4.5: Duration of gaze: A: short gaze, B: long gaze

sitioning the head at the level of the table and looking up (Figure 4.3). The third pair of movements consisted of a body sway towards the subject, followed by directing the gaze towards them and a second movement that switches this order (Figure 4.4). The final pair had a large difference in the duration of the gazing, short gaze(1 second) and long gaze (18 seconds) (Figure 4.5).

The order of movement pairs was counterbalanced using a 4x4 Latin square design, with each row appearing twice and the order of movements within each pair reversed between repetitions.

4.1.2 Measurements

Our mixed-method approach consisted of a questionnaire administered after every round, two additional questionnaires after all rounds, a separate questionnaire for the second part of the study, and a semi-structured interview at the end.

The assessment questionnaire for every round of improvisation

The questionnaire that we handed out after every round mainly consisted of 23 7-point Likert scale items. They were either anchored as "1=very poor, 7= Very good" or "1 =Strongly disagree, 7=Strongly agree". The first two questions assessed the general impression of the robot condition and the quality of improvisation in that round. The third asked to identify who led the improvisation, followed by questions on enjoyment, involvement, and motivation. Except for condition Fiona, where we removed these questions since the robot was inactive, we asked about the robot's helpfulness, influence on the improvisation with respect to turn-taking and whether it was a distraction in the process. At the end, participants completed the short-flow questionnaire [Martin and Jackson, 2008] to assess whether the robot had any influence on flow.

Assessment of the robot after the improvisations

After all rounds were completed, we handed out the Godspeed questionnaire [Bartneck, 2023] which is a standard to assess the performance of a robot. It consists of 5 categories relating anthropomorphism, animacy, likeability, perceived intelligence and perceived safety, from which we excluded safetey from our assessment, since the participants werent in direct proximity to the robot. Additionally, an open-ended questionnaire was given, to gain qualitative insight into the participants experiences with the robots. They were asked to describe the role of the robot in the improvisation, explain whether it was helpful, describe its influence on their performance referring to certain movements, report what happened when the robot looked at them and suggest changes for the whole system.

In the second part of the study, we asked participants to evaluate pairs of gazing movements differing in the before described dimensions. We selected a set of adjectives to capture different dimensions of social or interactive behavior. Inspired by Bartneck [2023] and Carpinella et al. [2017], we picked *Friendly, Engaging, Lively* and *Curious* for a social component, *Confident* and *Intelligent* as indicators of competence or agency as well as *Threatening* to reveal potential sources of discomfort. Instead of Likert scale rating we went for a comparative approach ("Which move fits best to the Attribute? Movement A, Movement B, Both or None"), to focus participants' attention on the contrast between the two movements rather than individual scale judgments.

We used a comparative approach for the second part of the study

In the end we conducted a video-recorded short semi-structured interview (2-10 minutes) with all three participants together, to catch things that did not fit any question, as well as to reveal emotional nuance connected to statements that could be lost in the text based answers. We did not split the participants in the interview to encourage discussion between participants. It consisted of three questions we asked in every study run, with the intention to follow up on things that seemed important to the participants. Which impression did the robot make on you? Do you associate the robot with something lively or is it more mechanical to you? How do you think the robot influenced your improvisation?

Semi-structured group interview to reveal emotional nuance missing in the text based answers

Additionally, the entire study was video-recorded to capture notable events that might not be reflected in the questionnaires or interview data.

4.1.3 Methods of evaluation

The qualitative data was organized using Thematic Analysis. Thematic Analysis aims to identify recurring patterns or themes within a dataset. As no predefined coding frame-

Thematic analysis for qualitative data

work aligned with the research questions, an inductive approach was used.

The analysis followed the six phases proposed by Braun and Clarke [2006]:

Familiarisation with the data Reading through answers to open-ended questions and interview transcripts.

Generating initial codes Producing initial codes.

Searching for themes Organization into potential themes.

Reviewing themes Refinement through merging, splitting, or removal where appropriate.

Defining and naming themes Committing to a set of themes and describing what they capture.

Producing the report Presenting the analysis of valuable themes for the research question with examples of codes in the results Section 4.2.2 of this work.

We used MAXQDA[1] for coding and theme development.

Statistical methods- Friedmann test, Wilcoxon signed-rank test, Bonferroni correction

To test for significant differences across conditions on Likert scale items, we applied the Friedman test [Friedman, 1937], a non-parametric test suited for ordinal data with small sample sizes, which allowed us to assess whether participants rated conditions differently overall. Since we were also interested in which specific conditions differed from each other, we additionally performed pairwise comparisons using the Wilcoxon signed-rank test [Wilcoxon, 1992]. Since the study was of an explorative kind, we did not exactly define how to group questions and what to look for, but rather ask a broad set of questions with the aim of identifying patterns in the data unknown to us before. That's why we looked for significances for single Likert items between pairs of conditions. Running multiple pairwise comparisons increases the risk of incorrectly identifying a difference as significant by chance, so p-values were compared to their Bonferroni correction [Dunn, 1964].

[1] https://www.maxqda.com, *last accessed 09.04.2026*

For the comparative gaze parameter study, participants selected which of two movements best fit each attribute. To assess whether a clear preference existed for one movement over the other, we used the chi-square test [Pearson, 1900], which tests whether the observed distribution of choices differs significantly from what would be expected if participants had no preference.

Chi-square test

4.2 Results

Participants (aged 20-33) rated their experience across relevant domains on a 7-point scale. Music experience was highest with a mean of 4.79, followed by dance experience (M=3.75) and animation experience (M=3.00), while acting and robotics experience were lowest (M=2.63 each). Participants who rated music experience highly most commonly referenced playing instruments, DJing or music production. High dance experience was mainly associated with training in ballet, hip-hop or salsa, while robotics experience was mostly gained through their academic studies.

4.2.1 Quantitive findings

Study Part 1: Collaborative Musical Improvisation

Participants were asked to identify who led the improvisation after each round, with options being one of the three participants (A, B, C), the robot (R), or no clear leader (NC). No clear leader was the most common response across all conditions. As visible in Figure 4.6, the robot was nevertheless the most frequently identified individual leader in all three robot conditions, with 10 out of 24 participants identifying it as leader in condition S and 8 in condition D. In condition E, where the robot avoided directed gaze, leadership attribution dropped to 5. Across all three robot conditions, the robot was identified as leader more often than the sum of all individual participant attributions combined. Among individual participants, A was consistently identi-

Subjective leader identification - Mostly no clear leader or robot perceived as leader

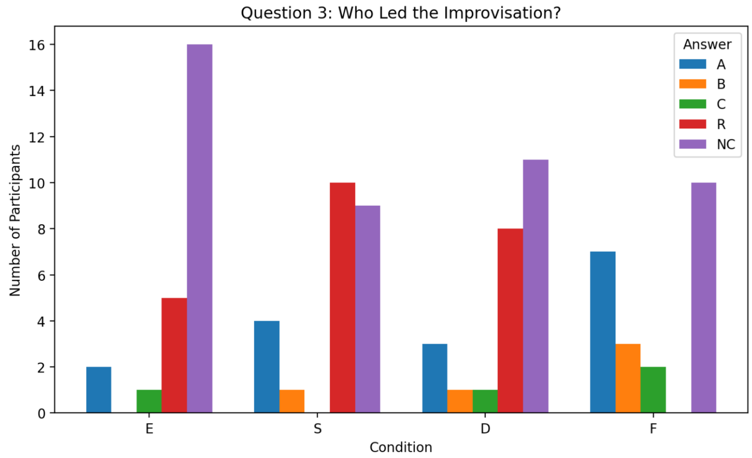

Figure 4.6: Answer count regarding the perceived leader of the improvisation

fied as the most dominant leader across all four conditions, and was the most frequently named leader in condition F where no robot was present.

Gaze manipulation check - successful

As a manipulation check, participants rated whether the robot shifted its focus between participants over time. Results confirmed the intended distinction between conditions ($\chi^2 = 13.861$, $p = 0.001$), with S and D scoring significantly higher than E (E vs S: $p_{uncorr} = 0.003$; E vs D: $p_{uncorr} < 0.001$), validating that the gaze manipulation was perceived as intended.

Beat synchronization manipulation check - unsuccessfull

Participants also rated whether the robot was consistently hitting the beat. No significant difference was found across conditions ($\chi^2 = 0.692$, $p = 0.707$), suggesting that participants did not reliably perceive the difference between the beat-matched conditions S and D and the slower unsynced condition Dori. This is consistent with qualitative reports that participants generally did not notice a difference between conditions with respect to rhythm.

Attention direction to singular participants was highest in condition Dori

Participants rated the degree to which attention within the group felt directed toward one participant at a time (Figure 4.7). Although the Friedman test did not reach significance ($\chi^2 = 7.055$, $p = 0.070$), exploratory pairwise comparisons

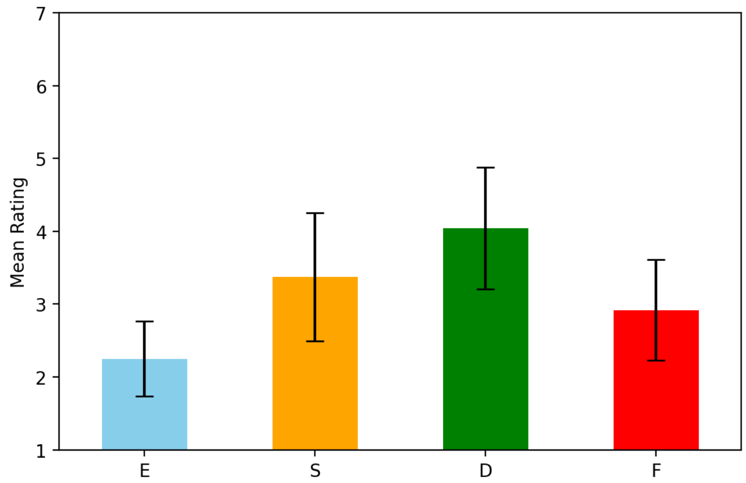

Figure 4.7: I felt that attention within the group was directed toward one participant at a time

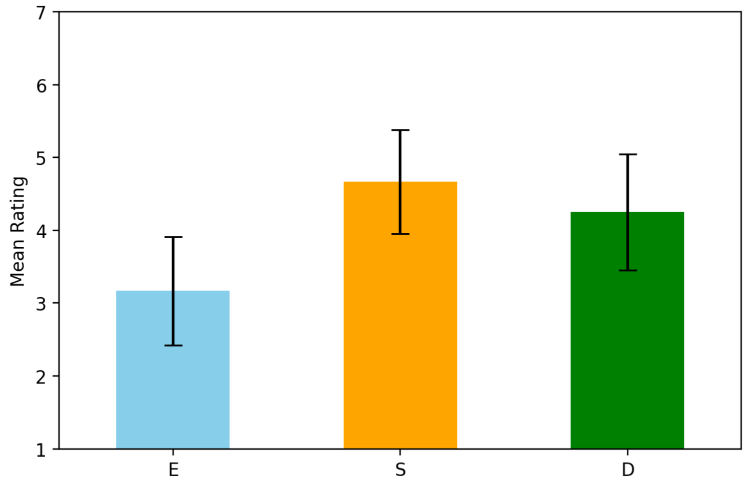

Figure 4.8: The robot influenced who I paid attention to during the improvisation

revealed significant differences between conditions E and D ($p_{uncorr} = 0.005$) and E and F ($p_{uncorr} = 0.048$), with D scoring higher on average ($M = 4.04$, $SD = 1.99$) than E ($M = 2.25$, $SD = 1.22$). Conditions S ($M = 3.38$, $SD = 2.08$) and F ($M = 2.92$, $SD = 1.64$) fell in between with no further significant pairwise differences.

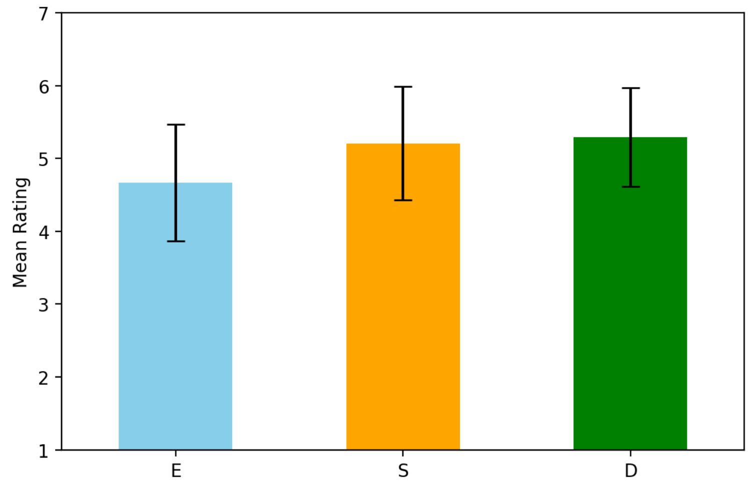

Figure 4.9: When the robot focused on me, I felt more engaged

Directed gaze seems to influence participants attention

The robot's influence on who participants paid attention to (Figure 4.8) showed a significant difference across conditions ($\chi^2 = 10.167$, $p = 0.006$). Pairwise tests revealed significant differences between E and S, even after correction ($p_{corr} = 0.012$; $p_{uncorr} = 0.004$) and between E and D ($p_{uncorr} = 0.038$), with S ($M = 4.67$, $SD = 1.69$) and D ($M = 4.25$, $SD = 1.89$) both scoring higher than E ($M = 3.17$, $SD = 1.76$). Conditions S and D did not differ significantly from each other.

Slight trend suggesting directed gaze is engaging

Participants reported feeling more engaged when the robot focused on them (Figure 4.9) with consistently high means across all three robot conditions (E: $M = 4.67$, $SD = 1.90$; S: $M = 5.21$, $SD = 1.84$; D: $M = 5.29$, $SD = 1.65$), with S and D being slightly higher than E. The Friedman test was not significant ($\chi^2 = 1.500$, $p = 0.472$) and no significant pairwise differences were found.

The data suggests trends of a gaze directed robot positively affecting enjoyment

To assess whether the conditions influenced the enjoyment of the improvisation process, participants rated the overall impression of the condition (Figure 4.11), their enjoyment of the round (Figure 4.12), and how encouraged they felt to contribute musically (Figure 4.10). None of the three items showed significant differences across condi-

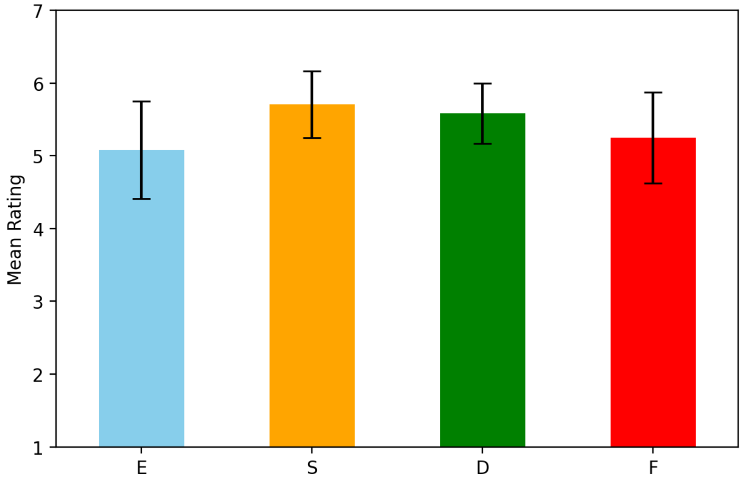

Figure 4.10: I felt encouraged to contribute musically

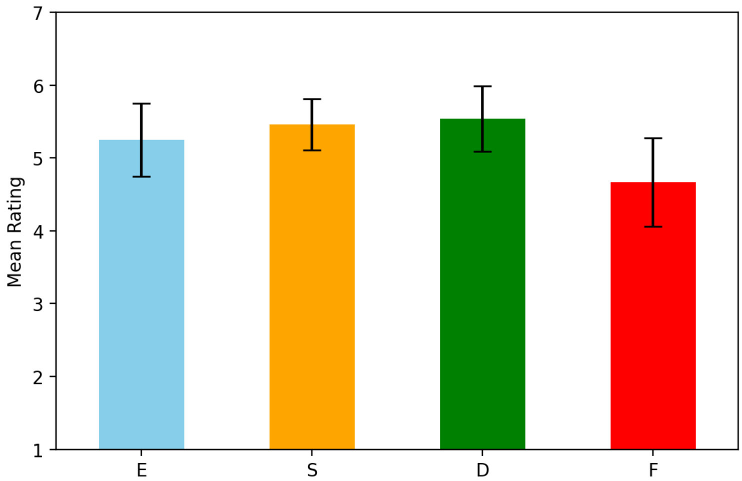

Figure 4.11: How would you rate the condition you've just seen?

tions ($\chi^2 = 6.553$, $p = 0.088$; $\chi^2 = 2.529$, $p = 0.470$; $\chi^2 = 1.175$, $p = 0.759$ respectively). Mean ratings were consistently high across all conditions and all three items. For the overall impression of the condition, means ranged from $M = 4.67$ ($SD = 1.43$) in F to $M = 5.54$ ($SD = 1.06$) in D. For enjoyment, means ranged from $M = 5.13$ ($SD = 1.54$) in E to $M = 5.71$ ($SD = 0.95$) in S. For musical encouragement,

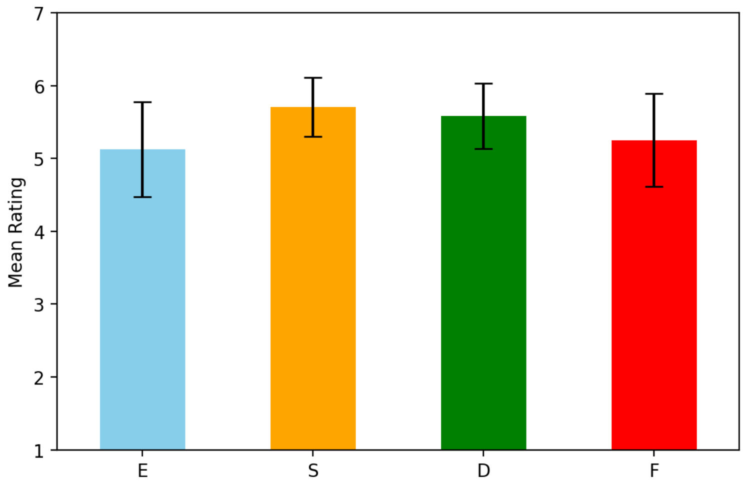

Figure 4.12: I enjoyed this round of improvisation

means ranged from $M = 5.08$ ($SD = 1.59$) in E to $M = 5.71$ ($SD = 1.08$) in S. Conditions S and D consistently received the highest mean ratings, while condition F tended toward the lower end. Exploratory pairwise comparisons revealed that S and D both rated the overall condition significantly higher than F ($p_{uncorr} = 0.027$ and $p_{uncorr} = 0.010$ respectively), and E also rated it higher than F ($p_{uncorr} = 0.044$). For enjoyment specifically, S was rated significantly higher than E ($p_{uncorr} = 0.047$), suggesting a possible advantage of directed gaze conditions over the avoidant gaze condition in the enjoyment measurements.

Feeling of involvement was low for undirected condition Eve

Participants also rated how actively involved they felt in the improvisation (Figure 4.13). The Friedman test was not significant ($\chi^2 = 4.042$, $p = 0.257$). Mean ratings were high across all conditions (E: $M = 5.04$, $SD = 1.55$; S: $M = 5.71$, $SD = 1.27$; D: $M = 5.71$, $SD = 0.81$; F: $M = 5.58$, $SD = 1.41$), with condition E again receiving the lowest mean rating and conditions S and D the highest. An exploratory pairwise comparison revealed a significant difference between E and D ($p_{uncorr} = 0.041$).

No significant effect on flow

The Short Flow Scale [Martin and Jackson, 2008] demonstrated acceptable internal consistency across all conditions (Cronbach's α ranging from 0.75 to 0.78). It was used to

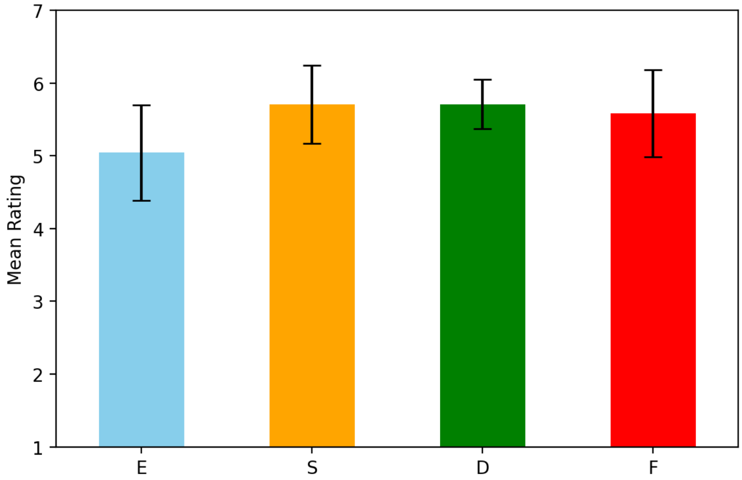

Figure 4.13: I felt actively involved in the improvisation during this round

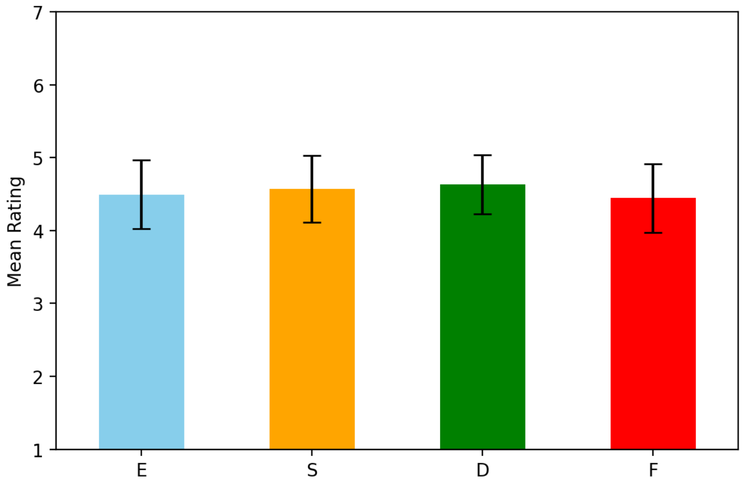

Figure 4.14: Results of the short flow scale

assess whether the conditions influenced the experience of flow during the improvisation. Mean flow scores (Figure 4.14) were consistently high and nearly identical across all conditions (E: $M = 4.48$, $SD = 1.08$; S: $M = 4.58$, $SD = 1.02$; D: $M = 4.61$, $SD = 0.92$; F: $M = 4.39$, $SD = 1.03$). The Friedman test was not significant ($\chi^2 = 0.822$, $p = 0.844$) and no significant pairwise differences were found, sug-

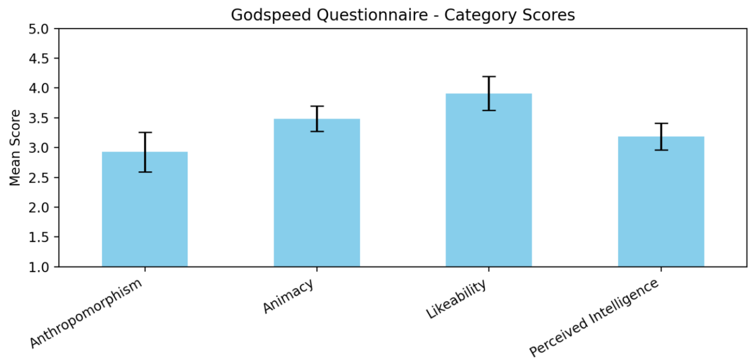

Figure 4.15: Mean answers on the Godspeed questionnaire

gesting that the presence and behavior of the robot had no measurable effect on the overall flow experience.

Godspeed revealed decently high likeability ratings

The Godspeed questionnaire was administered after all rounds to assess participants' overall perception of the robot. As visible in Figure 4.15, likeability received the highest mean rating ($M = 3.91$, 95% CI $[3.62, 4.19]$), followed by animacy ($M = 3.49$, 95% CI $[3.27, 3.70]$), perceived intelligence ($M = 3.18$, 95% CI $[2.96, 3.41]$) and anthropomorphism ($M = 2.93$, 95% CI $[2.60, 3.26]$). The robot was thus perceived as likeable and moderately animate, while scoring lowest on anthropomorphism, suggesting that participants responded positively to the robot without attributing strongly human-like qualities to it.

Tempo had no significant effect on ratings

An exploratory analysis of the effect of tempo on participant ratings across all Likert items revealed no significant differences between BPM conditions, suggesting that the specific tempo used in each round did not systematically influence the results.

Study Part 2: Gaze Parameter Study

The second part of the study examined how different parameters of the robot's gaze behavior shape the attribution of social, expressive, and cognitive qualities. Participants evaluated four pairs of movements, each differing in one

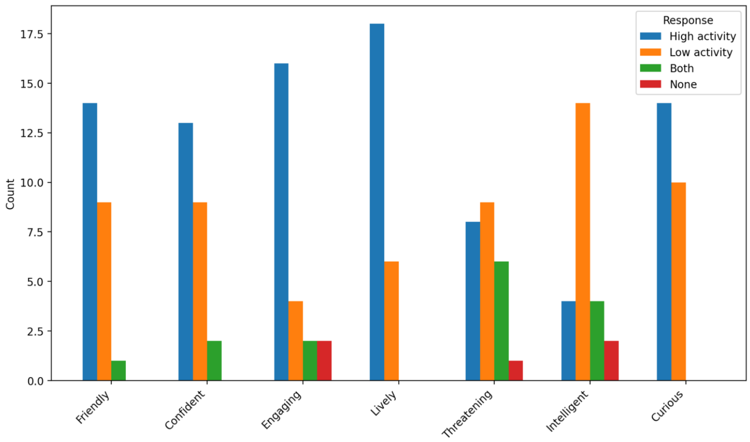

Figure 4.16: Comparative results for difference in gaze activity level

gaze parameter, by selecting which movement best fit a set of adjectives. The results are reported as preference counts and chi-square goodness-of-fit tests comparing the two options directly.

Participants compared a high activity gaze, featuring continuous wiggling movement, to a low activity gaze that rested in place after orienting toward the subject (Figure 4.16). High activity was significantly preferred for *Engaging* (A=16, B=4, $\chi^2 = 7.200$, $p = 0.007$) and *Lively* (A=18, B=6, $\chi^2 = 6.000$, $p = 0.014$), suggesting that continuous movement is perceived as more dynamic and energetic. Low activity was significantly preferred for *Intelligent* (A=4, B=14, $\chi^2 = 5.556$, $p = 0.018$), indicating that a calm, deliberate gaze reads as more considered and competent. No significant preferences were found for *Friendly*, *Confident*, *Threatening* or *Curious*.

Activity level - High: Engaging and Lively - Low: Intelligent

Participants compared a gaze directed downward from above to one directed upward from below (Figure 4.17). Gaze from above was significantly preferred for *Confident* (A=17, B=5, $\chi^2 = 6.545$, $p = 0.011$), consistent with the association between a dominant posture and self-assurance. Gaze from below was significantly preferred for *Friendly* (A=4, B=14, $\chi^2 = 5.556$, $p = 0.018$) and *Curious* (A=5,

Height of the head - Above: Confident - Below: Friendly, Curious

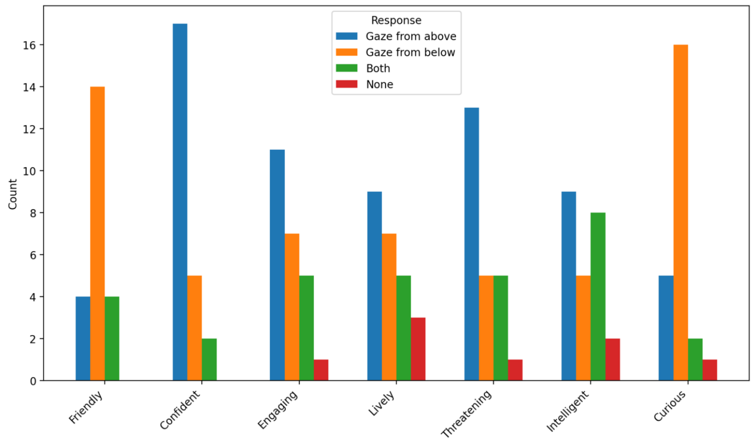

Figure 4.17: Comparative results for difference in height of the head position

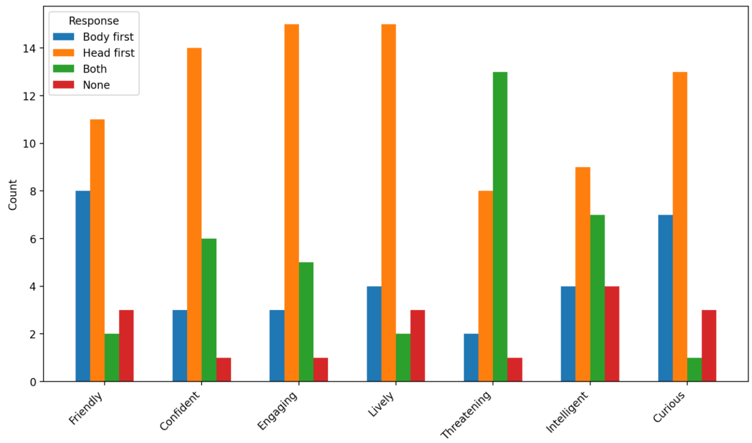

Figure 4.18: Comparative results for difference in order of body and head orientation

B=16, $\chi^2 = 5.762$, $p = 0.016$), suggesting that looking up at a subject reads as more approachable and as a sign of interest. *Threatening* showed a non-significant trend toward gaze from above ($p = 0.059$). No significant preferences were found for *Engaging*, *Lively* or *Intelligent*.

Order of body and head orientation - Head first: Confident, Engaging, Lively - Body First: No significant preferences

Participants compared a movement where the body orients toward the subject before the head follows, to one where the head leads and the body follows (Figure 4.18). Head

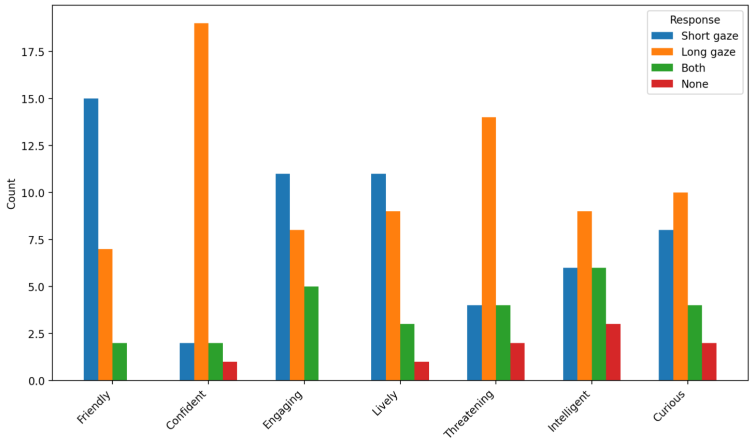

Figure 4.19: Comparative results for difference in gaze duration

first was significantly preferred for *Confident* (A=3, B=14, $\chi^2 = 7.118$, $p = 0.008$), *Engaging* (A=3, B=15, $\chi^2 = 8.000$, $p = 0.005$) and *Lively* (A=4, B=15, $\chi^2 = 6.368$, $p = 0.012$), suggesting that leading with the gaze reads as more intentional and socially directed. *Threatening* showed a non-significant trend toward head first ($p = 0.058$). No significant preferences were found for *Friendly*, *Intelligent* or *Curious*.

Participants compared a short gaze of approximately one second to a long gaze of eighteen seconds (Figure 4.19). Long gaze was overwhelmingly preferred for *Confident* (A=2, B=19, $\chi^2 = 13.762$, $p < 0.001$) and also significantly preferred for *Threatening* (A=4, B=14, $\chi^2 = 5.556$, $p = 0.018$), suggesting that a sustained gaze feels intrusive, consistent with qualitative reports of participants feeling stared at. *Friendly* showed a non-significant trend toward short gaze ($p = 0.088$), suggesting a brief glance reads as warmer and less intrusive. No significant preferences were found for *Engaging*, *Lively*, *Intelligent* or *Curious*.

Duration of gaze - Long gaze: Confident, Threatening - Short gaze: Friendly

4.2.2 Qualitative findings

Thematic analysis of the role of the robot in the improvisa-

Role of the robot - Motivator Entertainer, Listener and Leader tion resulted in the following categories:

- Motivator (4A, 4C, 5B, 7A, 7B, 7C, 8A, 8C)
- Entertainment (1B, 2B, 3A, 4B, 4C, 6B, 6C, 7A, 8B, 8C)
- Listener/Audience (2A, 3A, 4C, 5A, 6A, 6C, 7A, 8B)
- Dancer (3A, 4C, 7A, 8A)
- Leader (1B, 2A, 3B, 3C, 4A, 5A, 5B, 5C)
- Inspiration (1A)
- Distraction (2C, 3B, 7B, 7C, 8A, 8C)

Influence of gaze - Gaze is a perceived as solo cue

Most participants interpreted the robot's direct gaze as a signal to take the lead (1B, 1C, 2A, 2B, 2C, 3B, 3C, 4B, 4C, 5A, 5B, 5C, 7A, 7B, 7C). This was accompanied by reports of increased musical activity, with participants indicating that they played faster (2B, 7A, 7B), more actively (1B, 1C, 2B, 3C, 5A, 5B, 7A, 7B), louder (3B, 5A, 7C), with greater intensity (2B), or more confident (1C, 5A, 7C). Some explicitly described gazing as giving them a solo cue(5B, 7C). Additionally, some participants felt like the direct gazing improved their focus on the task (1A, 1B, 4B, 6B). At the same time, others perceived the gaze of the robot as pressuring (3B, 4B, B5, 7A, 7B, 8C). They reported they felt like they "have to deliver" (4B) or "perform" (7A). In some cases, the gaze was characterized as uncomfortable or intrusive, for example as "staring" (7B) or "staring into their soul" (3A).

Helpfulness - Mixed opinions

Relating the helpfulness of the robot during the improvisation tasks, 7 participants found the system to not be helpful due to the following reasons:

- They were more concentrated on the music (making) itself (2B, 2C, 3A, 3B, 3C, 8B)
- they missed an understanding of the moves (2C, 3A, 8B)
- they rather wanted to do "their own thing" (4B)

Some participants, reported both helpful and unhelpful aspects at the same time (1A, 1C, 2B, 7B, 7C). They felt the system was either only helpful in certain moments, like when it was looking at them (B2, 1C), making certain movements that served as an inspiration (1A) or at late stages where they felt like they could focus more on the robot (1C). 13 participants found the robot to be helpful in the improvisation process, giving the following reasons:

- The robot helped by visualizing the beat (1B, 1C, 4C, 5C, 6B)
- The robot was motivating (4A, 4C, 8A)
- The experience was more fun due to the robot (6C, 7A, 8C)
- The robot removed their perceived pressure to perform well (2A, 4A, 5A)
- The robot helped to decide who plays a solo (3B, 5B)
- They could focus better (6B)

Liveliness - An attribute depending on the moveset

Mainly stemming from the semi-structured interview, participants gave answers to whether they experience the robot as something lively or machine like. Those who described it as machine-like with an explicit reason did this due to:

- repetitiveness of movements (2A, 7A)
- their conscious awareness of it being a machine (5A, 5B)
- the missing intentionality in condition Eve (2A)
- some certain movements (4C)
- the noise of the motors (5B, 7A)

Participants describing the lamp as lively almost always did this with the same limitation.

"A few movements looked very lively." (8C)

With similar responses from 1A, 2A, 4C and 7A. It was also described as cute (5A, 8A), friendly (5A), threatening (4A), creepy (7B), clumsy (8B) and scared (4A).

Cognitive overload - Not everyonecould pay attention to the robot

Several participants reported that they could not pay attention to the robot as much due to them focussing on what other participants were playing and making music themselves (1C, 2A, 2B, 2C, 3C, 4A, 4C, 8C). Two of them mentioned this got better in later rounds (1C, 3C), since they had to learn to use the instrument first. Two others also related their lack of attention to the lamp to not being able to effortlessly play their instruments (4A, 4C, 8C).

Suggested changes

The questionnaires revealed a couple of suggested changes for the system, namely:

- Change of movements (3A, 4C, 5A, 6B, 7A, 7B, 7C, 8A, 8C)
- Introduce reactivity (1A, 3B, 4B, 5B, 6C, 8B)
- Better beat synchronization
- An illuminated bulb (3A, 2A, 3C, 5C)

Those participants who suggested a change of the movements often did this because certain movements were too sudden (4C, 5A, 6B, 7C, 8A, 8C). Participant 8A mentioned that it moved too abruptly, which "disturbed the flow". 7B described some of the movements as silly, 6B suggested "cooler" movements and 3A was missing an understanding of what the moves aimed to communicate. The non reactivity of the robot was another point participants wanted to see a change in. Some of them wanted feedback for their quality of play (4B, 8B), while others actively tried to influence the robot with their playing to find out if and how the robot reacts (3A, 6A, 6B 6C)

4.3 Discussion

Majority of participants found the robot to be helpful in the improvisation process. The robot was described as a fun, motivating part of the experience and helped by removing pressure to perform. This indicates a partial achievement of our goal of introducing a generally more positive mood into the improvisation to support creativity. Quantitative trends point in the same direction, with conditions S and D consistently receiving higher enjoyment and involvement ratings than E and F, though these differences did not reach significance, potentially due to the small sample size. Another helpful aspect seems to have been the visualization of the beat during the process. What is interesting here is that none of the participants mentioned a difference between conditions with respect to beat synchronization, despite Dori being unsynced to the beat.

The DDL can support musical improvisation in several ways

The leadership attribution data suggests that when a clear leader was perceived, it was most often the robot rather than any individual participant. Across all three robot conditions, the robot was identified as leader more often than the sum of all individual participant attributions combined, pointing toward a degree of acceptance of the robot as an active participant and mediator in the improvisation process. This is further supported by the significant effect of the robot's gaze direction on who participants paid attention to, as well as the trend toward higher scores in the directed gaze conditions on the item assessing whether attention within the group felt directed toward one participant at a time. Together these results suggest that the robot's gaze was not merely noticed but actively shaped the social dynamics of the improvisation.

The robot was accepted as a mediator in the process

While some participants generally disliked the robot, most of those who found the system to be unhelpful reported a general cognitive overload. Improvising with two new musical partners on a never before played instrument in a study setting is already challenging without focusing on a robotic mediator at the same time. Even though there was a phase to familiarize with the musical setup in the beginning, it was still relatively short. Two participants noted

Cognitive overload and limited ability to react that their ability to engage with the robot improved in later rounds once they had grown more comfortable with their instrument, suggesting that the effect of the mediator may be more pronounced with experienced musicians or with longer familiarization phases.

Ambiguous or distracting movements can hinder the support of the process

Several participants suggested changes to the movement set, most commonly requesting smoother and less abrupt transitions. Sudden movements were described as distracting or as disturbing the flow of the improvisation. This suggests that the design of expressive movements for a robotic mediator requires careful attention not only to what is communicated but also to how transitions between movements are handled, as abrupt changes can pull attention away from the music rather than supporting it.

Liveliness strongly depends on moveset

The qualitative answers suggest that the liveliness level of the robot is heavily dependent on move set. While only two participants described it as lively without restrictions, 5 answers related their assessment of the lamp appearing as lively to certain moves. Participants viewing the system as rather machine-like also did this due to the repetitiveness of the choreography, which was the same for every condition, as well as their conscious awareness of the system being a robotic arm. This last point could indicate that changing the visual appearance by concealing the technical parts of the robot might improve the system to be recognized as something lively.

Godspeed results suggest positive reception

The Godspeed results indicate that the robot was perceived as likeable and moderately animate, but scored lowest on anthropomorphism. The moderate animacy score aligns with the qualitative finding that liveliness was perceived selectively, tied to specific movements rather than the robot as a whole. While there is room for improvement in the animacy dimension specifically, the relatively high likeability score is encouraging and suggests that participants responded positively to the robot as a presence in the improvisation. The low anthropomorphism score is unsurprising given the lamp-shaped appearance of the DDL, and is arguably not a limitation in this context, since the mediating role the robot fulfills does not require it to appear humanlike. What matters is that it communicates meaningfully through movement and gaze, which the results suggest it was able to do.

Flow results are inconclusive given study design

The Short Flow Scale showed no significant differences across conditions, with consistently high scores throughout. Given that each improvisation round lasted only approximately two minutes, it is unclear whether this duration was sufficient for participants to enter a flow state at all, let alone detect differences between conditions. This is a limitation of the current study design rather than a definitive finding about the robot's influence on flow, and future work with longer improvisation sessions would be needed to investigate this properly.

Gaze parameter study informs future movement design

The results of the gaze parameter study offer preliminary design considerations for future implementations of the DDL and similar robotic mediators. The preference for head-first orientation across several attributes suggests that initiating gaze shifts with a head movement rather than a body turn reads as more intentional and socially directed. The duration results suggest that a mediator should avoid dwelling on a single participant for too long without accompanying movement, as a static long gaze tended to be associated with threat as well as confidence, which aligns with qualitative reports of participants feeling stared at. If the goal of the mediator is to signal interest in what a musician is playing, gazing from below may be a suitable choice, given its association with friendliness and curiosity in this study. Similarly, if the mediator is intended to convey competence, for example for an implementation focussing more on feedback mechanisms, a low activity gaze with minimal head movement could be worth exploring, as it tended to be associated with perceived intelligence. These findings are exploratory and based on a small sample, but they provide a starting point for more systematic investigation of socially meaningful gaze behaviors in robotic mediators.

4.4 Limitations

The study has several limitations that should be considered when interpreting the results.

24 participants limit the expressive power of quantative resluts

The sample size of 24 participants limits the statistical power of the quantitative analysis, which likely contributed to the large number of non-significant results despite visible trends in the data. All participants were novices improvising on unfamiliar instruments, introducing a level of cognitive load that reduced their ability to engage with the robot, as reflected in participant reports. Whether the observed effects generalize to experienced musicians remains an open question.

Participants mostly with technical backround

Additionally, the participant pool was relatively homogeneous, with ages ranging from 20 to 33 and the majority coming from technical academic backgrounds. A considerable number of participants were personally acquainted with the researcher, which may have introduced a bias, with participants potentially inclined to evaluate the system more favorably than they otherwise would have.

Short rounds hinder reaching a flow state

The improvisation rounds lasted approximately two minutes each, which may have been insufficient for participants to enter a flow state, making the flow scale results difficult to interpret meaningfully.

High cognitive load due to novelty of experience

The cognitive load introduced by improvising on unfamiliar instruments in a group setting also limited participants' ability to engage with the robot, as reflected in participant reports.

Beat synchronization was not perceived as intended

The beat synchronization manipulation was not perceived as intended by participants, weakening the intended distinction between conditions S and D and limiting the conclusions that can be drawn about the role of rhythmic synchronization specifically.

Stopping and quick movements of robot were perceived as a distraction

Another practical limitation of the current implementation concerns the execution of movement transitions. Although animations were blended using a sigmoid-weighted interpolation, the individual animations were not designed with sequential composition in mind, meaning that transitions between them could result in movements that lacked logical continuity. Additionally, the robot occasionally came to a brief stop when an animation faded out, which was perceived as unnatural and distracting. These issues point to the need for further movement design and a fix in the implementation concerning fading into the base animation in the beat synched conditions.

The gaze direction of the robot was controlled manually by the researcher during the study. This means that the decisions about when to shift the robot's attention and toward whom were made in the moment by a single person, which may have introduced inconsistencies across sessions and influenced the results in ways that are difficult to account for.

Wizard-of-oz decisions in gaze direction might have intoduced a bias

In the gaze parameter study, the comparative forced-choice format only captures associations to the attributes included in the questionnaire and does not reflect the magnitude of preference between the two options. The small sample additionally limits the reliability of the chi-square results. Furthermore, the seating position was visualized but not formally analyzed, leaving open whether being the direct target of the robot's gaze produced systematically different attributions compared to observing it from the side.

The second part of the study dictated possibly attributes, reducing the effectiveness of explorative nature of the study

Chapter 5

Summary and Future Work

5.1 Summary and Contributions

This work investigated the potential of an artificial mediator to support collaborative musical improvisation. Starting from a broad design space mapping possible inputs, analytical processes and outputs for such a system, we decided to further investigate the potential of embodied interaction as one instantiation of the design space. We therefore implemented the Dancing Desk Lamp, a non-anthropomorphic robotic prototype that communicates through rhythmic movement and directed gaze. We evaluated it in a mixed-methods user study examining how the presence of an actively participating robot and different gaze behaviors influence the social dynamics of a group improvisation process, complemented by a secondary study on the expressive qualities of specific gaze parameters.

The results suggest that the robot's gaze direction influences turn-taking dynamics and attention within the group, and that the robot was broadly accepted as an active participant in the improvisation. Participants most commonly described it as a motivator, entertainer and leader, and rated the experience positively across conditions. The gaze parameter study provides preliminary design insights into how activity level, head height, orientation order and gaze duration each shape the social perception of a robotic gaze.

The main contributions of this work are:

1. A design space for mediators in collaborative musical improvisation, exploring inputs, analytical processes and outputs relevant to such a system.
2. The Dancing Desk Lamp, a robotic prototype implementing rhythmic movement and gaze-based communication as a mediator in a musical improvisation setting.
3. Exploratory evidence suggesting that an embodied mediator's gaze behavior influences turn-taking dynamics and social engagement in a group improvisation process, supported by quantitative ratings and qualitative participant accounts.
4. Preliminary design guidelines for expressive gaze behavior in robotic mediators, derived from a comparative gaze parameter study.

5.2 Future Work

The results of this work suggest that an embodied mediator can meaningfully influence the dynamics of musical improvisation, opening up a range of directions for future work.

Further exploration of the design space

The prototype developed here represents only a brief instantiation of the design space defined in Chapter 3, focusing exclusively on physical output driven by a fixed backing track. The most impactful direction for future work is therefore the systematic exploration of further dimensions of the design space. Incorporating audio analysis would allow the mediator to respond directly to the musical content of the performance, removing the dependency on a backing track and enabling the use of non-digital instruments. This opens the possibility of fitting the robot's movements to the emotional character or energy level of the ongoing performance, making the mediator feel more like a genuine participant rather than a pre-programmed choreography. Identifying the current leader of the improvisation through movement or audio analysis and reinforcing ongoing dynamics rather than imposing predetermined ones would be another valuable extension. Multi-modal input, combining audio, MIDI and video, could enable richer analysis than any single channel alone, for instance detecting synchronization between players or identifying the least contributing musician and engaging them through directed attention.

Introducing reactivity and feedback

Reactivity and feedback were among the most frequently requested changes by participants in this study. Future systems could analyze the ongoing performance and respond to it in real time, providing musicians with a signal of how they are doing. One particularly accessible way to introduce feedback would be through an illuminated bulb, which was suggested by multiple participants and would add a lightweight expressive output channel without requiring additional hardware complexity. Combined with audio analysis, the light could reflect the mood or energy of the performance, or signal a solo cue to a specific participant.

Further analyzing movements to express intentions

The gaze parameter study conducted in this work provides a concrete starting point for designing more nuanced and socially readable gaze behaviors in future implementations. Future work could explore combinations of gaze parameters rather than testing them in isolation, to understand how they interact when composed into a full gaze behavior.

Using the system with expert musicians

The study in this work was conducted with novice musicians on unfamiliar instruments, which introduced considerable cognitive load. Testing the system with experienced instrumentalists would allow participants to engage more freely with the mediator, potentially revealing effects that were obscured by the demands of the task. Similarly, testing with different group sizes could reveal how the mediator scales to larger ensembles and whether its turn-taking and attention-directing functions remain effective. Finally, testing robots with different visual appearances, ranging from more abstract to more humanoid, could shed light on how appearance shapes the perception of the mediator and its integration into the social dynamics of the improvisation.

Appendix A

Code System

Code	Description	N
Associations	*Mentions of associations participants had with the robot*	**8**
Animated Character	Robot reminded participant of some animated character	2
Human	Robot reminded the participant of something human	4
Luxo	Robot reminded participant of Luxo Jr., the Pixar lamp	2
Attributes	*Attributes associated with the robot*	**13**
Scared	Robot was described as scared	1
Friendly	Robot was described as friendly	1
Creepy	Robot was described as creepy	1
Clumsy	Robot was described as clumsy	1
Threatening	Robot was described as threatening	2
Cute	Robot was described as cute	4
Animal	Robot reminded participants of an animal	3
Liveliness	*Comments on whether the robot seemed lively or machine-like*	**16**
Machine-like	Robot was described as a machine, something mechanical	8
Living	Robot was described as living	8
Mediating Goals	*Answers directly addressing the goals set for the mediating system*	**24**
Influencing turn-taking	Mentions of the robot's influence on turn-taking	20
Improve enjoyment	Mentions that the robot was enjoyable or made the process more enjoyable	4
Helpfulness in Improvisation	*Comments addressing helpfulness in improvisation*	**37**
Not helpful	Participant considered the robot to be unhelpful	8
Remove Pressure	Participants felt the robot removed performance pressure	4
Visualizing beat	Participants felt the robot's movements visualized the beat	6
Helpful	Participant considered the robot to be helpful	13
Neutral/Unsure	Participant considered the robot to be neither helpful nor unhelpful	6
Role of Robot	*Explicit mentions of the role the robot took in the improvisation*	**73**
Entertainment	The robot was seen as a form of entertainment	13
Encouragement/Motivation	The robot motivated or encouraged participants	20
Listener	Participant felt the robot was paying attention or listening	9
Critic	Robot was described as a critic of the performance	1
Dancer	Robot was described as a dancer	5
Leader	Participant felt the robot was leading the improvisation	8
Conductor	Robot was described as a conductor	5
Distraction	Participant felt distracted in their musical process	9

Code	**Description**	**N**
Inspiration	Participant felt inspired or gathered new ideas because of the robot	3
Gaze	*Effects of the robot's gaze on participants*	**49**
Group Inclusion	Participant felt more included in the collaboration when gazed at	1
Being Heard	Participant felt the robot was giving them attention	3
Solo Cue	Participant felt encouraged, motivated or pressured to play a solo	32
More Focused	Participant felt more focused on the task when gazed at	4
Pressuring	Participant felt pressured to play more when gazed at	7
Negatively Feeling Watched	Participants disliked being gazed at by the robot	2
Notable	*Notable observations that did not fit any other category*	**32**
Expressing Emotions	Reference to the robot expressing emotions	1
Sound	References to the motor sounds	2
Too Focused on Instrument	Participants could not pay attention to the robot due to instrument focus	13
Focus on Others	Participants focused on other participants rather than the robot	4
Musically Restricted	Musical ability restricted participants from reacting to the robot	6
Dislike	Participant disliked the robot or the experience	3
Mapping	Participants created a direct mapping of moves to their playing	3
Suggested Changes	*Explicit or implicit mentions of changes to improve the robot*	**35**
Add Light	Participants wanted the bulb to be illuminated	4
Increased Reactivity	Participants wished for increased reactivity or interactivity	10
Smoothness	Participants wanted smoother, less abrupt movements	7
Less Direct Look	Participants wanted to be gazed at less intensely	2
Changed Moveset	Participants wished for a better moveset	7
On Beat Improvement	Participants felt the robot was not rhythmically consistent	5
Interactivity	*Observations regarding reactivity and interactivity*	**16**
Seemed Interactive	Participants thought the robot was interactive	4
Tried Influencing Robot	Participants actively tried to find out whether the robot was interactive	12

Table 1: Full code system from thematic analysis of qualitative study data.

Appendix B

Study Material

- 1: Questionnaire for rounds Eve, Dori and Sally (Q18-26 Short flow scale)
- 2: Questionnaire for round Fiona (Q12-20 Short flow scale)
- 3: Open ended questionnaire after last improvisation
- 4: Comparative answer boxes Part 2 of the study
- 5: Godspeed questionnaire without safety items

Post-Round Questionnaire

Participant: ____________________

Instructions

Please answer the following questions based on your experience in this round of improvisation. There are no right or wrong answers. Please respond based on your immediate impression.

1. How would you rate the condition you've just seen?
(1 = Very poor, 7 = Very good)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

2. How would you rate the quality of the improvisation in this round?
(1 = Very poor, 7 = Very good)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

3. Who do you feel led the improvisation in this round? (circle one)

- Participant A
- Participant B
- Participant C
- The robot
- No clear leader

4. I enjoyed this round of improvisation.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

5. I felt that attention within the group was directed toward one participant at a time.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

6. The robot influenced who I paid attention to during the improvisation.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

Participant: ____________________

7. I felt that the robot shifted its focus between participants over time.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

8. I felt actively involved in the improvisation during this round.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

9. I felt encouraged to contribute musically.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

10. When the robot focused on me, I felt more engaged.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

11. The improvisation felt smooth and continuous.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

12. I felt rhythmically synchronized with the group.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

13. I feel like the robot was consistently hitting the beat.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

14. The robot's movement helped me stay on the beat.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

15. The robot's movement distracted me from the music.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

Participant: ____________________

16. At any point, I felt interrupted or disrupted.
☐ Yes ☐ No

17. If yes, please briefly describe what caused the interruption:

____

____

18. I feel I am competent enough to meet the high demands of the situation.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

19. I do things spontaneously and automatically without having to think about them.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

20. I have a strong sense of what I want to do.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

21. I have a good idea while I am performing about how well I am doing.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

22. I am completely focused on the task at hand.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

23. I have a feeling of total control.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

24. I am not worried about what others may be thinking of me.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

25. The way time passes seems to be different from normal.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

26. The experience is extremely rewarding.

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

Post-Round Questionnaire

Participant: ____________________

Instructions

Please answer the following questions based on your experience in this round of improvisation. There are no right or wrong answers. Please respond based on your immediate impression.

1. How would you rate the condition you've just seen?
(1 = Very poor, 7 = Very good)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

2. How would you rate the quality of the improvisation in this round?
(1 = Very poor, 7 = Very good)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

3. Who do you feel led the improvisation in this round? (circle one)

- Participant A
- Participant B
- Participant C
- No clear leader

4. I enjoyed this round of improvisation.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

5. I felt that attention within the group was directed toward one participant at a time.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

Participant: ____________________

6. I felt actively involved in the improvisation during this round.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

7. I felt encouraged to contribute musically.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

8. The improvisation felt smooth and continuous.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

9. I felt rhythmically synchronized with the group.
(1 = Strongly disagree, 7 = Strongly agree)

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7

10. At any point, I felt interrupted or disrupted.
□ Yes □ No

11. If yes, please briefly describe what caused the interruption:

____

____

Participant: ____________________

12. I feel I am competent enough to meet the high demands of the situation.

○ ○ ○ ○ ○ ○ ○
1 2 3 4 5 6 7

13. I do things spontaneously and automatically without having to think about them.

○ ○ ○ ○ ○ ○ ○
1 2 3 4 5 6 7

14. I have a strong sense of what I want to do.

○ ○ ○ ○ ○ ○ ○
1 2 3 4 5 6 7

15. I have a good idea while I am performing about how well I am doing.

○ ○ ○ ○ ○ ○ ○
1 2 3 4 5 6 7

16. I am completely focused on the task at hand.

○ ○ ○ ○ ○ ○ ○
1 2 3 4 5 6 7

17. I have a feeling of total control.

○ ○ ○ ○ ○ ○ ○
1 2 3 4 5 6 7

18. I am not worried about what others may be thinking of me.

○ ○ ○ ○ ○ ○ ○
1 2 3 4 5 6 7

19. The way time passes seems to be different from normal.

○ ○ ○ ○ ○ ○ ○
1 2 3 4 5 6 7

20. The experience is extremely rewarding.

○ ○ ○ ○ ○ ○ ○
1 2 3 4 5 6 7

Dancing Robot Study – Participant Questionnaire

Participant: ____________

Participant Questionnaire

Please answer the following questions based on your experience during the improvisation tasks with the robot. Use the space provided or use the back of the page with indication which question you are answering.

Question 1: How would you describe the role of the robot in the improvisation tasks you just participated in?

Question 2: Do you think the robot was helpful in the improvisation process? Please describe why or why not.

Question 3: Did the robot influence what you were playing on the instrument? If yes, describe how? If no, describe why you think it didn't.

1

Figure B.1: Open ended questionnaire

Movement Comparison Questionnaire

Instructions

For each pair of moves, indicate which element (A or B) feels more like the attribute. Put A or B if one fits the attribute more than the other. Put None if both don't fit at all and Both if both fit equally.

Pair 1: Move 1

Which is more? (A / B / Both / None):

Attribute	Consideration
Friendly	
Confident	
Engaging	
Lively	
Threatening	
Intelligent	
Curious	

Pair 2: Move 2

Which is more? (A / B / Both / None):

Attribute	Consideration
Friendly	
Confident	
Engaging	
Lively	
Threatening	
Intelligent	
Curious	

Pair 3: Move 3

Which is more? (A / B / Both / None):

Attribute	Consideration
Friendly	
Confident	
Engaging	
Lively	
Threatening	
Intelligent	
Curious	

Pair 4: Move 4

Which is more? (A / B / Both / None):

Attribute	Consideration
Friendly	
Confident	
Engaging	
Lively	
Threatening	
Intelligent	
Curious	

1

Figure B.2: Comparative questionnaire second part of the study

Questionnaire

Please rate your overall impression of the robot on these scales:

	1	2	3	4	5	
Machinelike	○	○	○	○	○	Humanlike
Ignorant	○	○	○	○	○	Knowledgeable
Awful	○	○	○	○	○	Nice
Unkind	○	○	○	○	○	Kind
Unintelligent	○	○	○	○	○	Intelligent
Dislike	○	○	○	○	○	Like
Dead	○	○	○	○	○	Alive
Foolish	○	○	○	○	○	Sensible
Unconscious	○	○	○	○	○	Conscious
Artificial	○	○	○	○	○	Lifelike
Fake	○	○	○	○	○	Natural
Movingrigidly	○	○	○	○	○	Movingelegantly
Incompetent	○	○	○	○	○	Competent
Mechanical	○	○	○	○	○	Organic
Stagnant	○	○	○	○	○	Lively
Unpleasant	○	○	○	○	○	Pleasant
Apathetic	○	○	○	○	○	Responsive
Unfriendly	○	○	○	○	○	Friendly
Irresponsible	○	○	○	○	○	Responsible
Inert	○	○	○	○	○	Interactive

Bibliography

[1] Safinah Ali, Nisha Devasia, Hae Won Park, and Cynthia Breazeal. Social Robots as Creativity Eliciting Agents. *Frontiers in Robotics and AI*, 8, September 2021. doi.org/10.3389/frobt.2021.673730.

[2] Christoph Bartneck. Godspeed Questionnaire Series: Translations and Usage. In Christian U. Krägeloh, Mohsen Alyami, and Oleg N. Medvedev, editors, *International Handbook of Behavioral Health Assessment*, pages 1–35. Springer International Publishing, Cham, 2023. doi.org/10.1007/978-3-030-89738-3_24-1.

[3] Aryel Beck, Brett Stevens, Kim A. Bard, and Lola Cañamero. Emotional body language displayed by artificial agents. *ACM Transactions on Interactive Intelligent Systems*, 2(1):1–29, March 2012. doi.org/10.1145/2133366.2133368.

[4] Michele Biasutti. Teaching Improvisation through Processes. Applications in Music Education and Implications for General Education. *Frontiers in Psychology*, 8, June 2017. doi.org/10.3389/fpsyg.2017.00911.

[5] Michele Biasutti and Luigi Frezza. Dimensions of Music Improvisation. *Creativity Research Journal*, 21(2-3):232–242, May 2009. doi.org/10.1080/10400410902861240. _eprint: https://doi.org/10.1080/10400410902861240.

[6] Laura Bishop, Carlos Cancino-Chacón, and Werner Goebl. Eye gaze as a means of giving and seeking information during musical interaction. *Consciousness and Cognition*, 68:73–96, February 2019. doi.org/10.1016/j.concog.2019.01.002.

[7] Barry Bittman, Karl T Bruhn, Christine Stevens, James Westengard, and Paul O Umbach. Recreational Music-Making: A Cost-Effective Group Interdisciplinary Strategy for Reducing Burnout and Improving Mood States in Long-Term Care Workers.

[8] Dmitry Bogdanov, Nicolas Wack, Emilia Gómez, Sankalp Gulati, Perfecto Herrera, Oscar Mayor, Gerard Roma, Justin Salamon, José Zapata, and Xavier

Serra. ESSENTIA: an open-source library for sound and music analysis. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 855–858, Barcelona Spain, October 2013. ACM. doi.org/10.1145/2502081.2502229.

[9] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, January 2006. doi.org/10.1191/1478088706qp063oa. _eprint: https://doi.org/10.1191/1478088706qp063oa.

[10] Antonio Camurri, Barbara Mazzarino, Matteo Ricchetti, Renee Timmers, and Gualtiero Volpe. Multimodal Analysis of Expressive Gesture in Music and Dance Performances. In Antonio Camurri and Gualtiero Volpe, editors, *Gesture-Based Communication in Human-Computer Interaction*, pages 20–39, Berlin, Heidelberg, 2004. Springer. doi.org/10.1007/978-3-540-24598-8_3.

[11] Colleen Carpinella, Alisa Wyman, Michael Perez, and Steven Stroessner. *The Robotic Social Attributes Scale (RoSAS): Development and Validation*. March 2017. doi.org/10.1145/2909824.3020208. Pages: 262.

[12] Roser Cañigueral and Antonia F. de C. Hamilton. The Role of Eye Gaze During Natural Social Interactions in Typical and Autistic People. *Frontiers in Psychology*, 10, March 2019. doi.org/10.3389/fpsyg.2019.00560.

[13] Andrew Chang, Haley E. Kragness, Steven R. Livingstone, Dan J. Bosnyak, and Laurel J. Trainor. Body sway reflects joint emotional expression in music ensemble performance. *Scientific Reports*, 9(1):205, January 2019. doi.org/10.1038/s41598-018-36358-4.

[14] Xiaoyu Chang, Fan Zhang, Kexue Fu, Carla Diana, Wendy Ju, and Ray Lc. A Constructed Response: Designing and Choreographing Robot Arm Movements in Collaborative Dance Improvisation. *Proceedings of the ACM on Human-Computer Interaction*, 9(7):1–31, October 2025. doi.org/10.1145/3757452.

[15] Rashmi Chawla, Raffaele Longo, Francesco Pupo, and Giancarlo Fortino. EAC-HRC: Emotion-Aware Communication for Human-Robot Collaboration in a Music Orchestra System. In *2025 IEEE International Conference on Cyber Humanities (IEEE-CH)*, pages 1–6, September 2025. doi.org/10.1109/IEEE-CH65308.2025.11279369.

[16] Nele Dael, Marcello Mortillaro, and Klaus R. Scherer. Emotion expression in body action and posture. *Emotion*, 12(5):1085–1101, 2012. doi.org/10.1037/a0025737.

[17] Jane W. Davidson. Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case stud-

ies. *Psychology of Music*, 40(5):595–633, September 2012. doi.org/10.1177/0305735612449896.

[18] Alwin de Rooij, Simone van den Broek, Michelle Bouw, and Jan de Wit. Co-Creating with a Robot Facilitator: Robot Expressions Cause Mood Contagion Enhancing Collaboration, Satisfaction, and Performance. *International Journal of Social Robotics*, 16(11):2133–2152, December 2024. doi.org/10.1007/s12369-024-01177-3.

[19] Paul Dourish. Embodied interaction: Exploring the foundations of a new approach to HCI. January 1999.

[20] Paul Dourish. What we talk about when we talk about context. *Personal and Ubiquitous Computing*, 8(1):19–30, February 2004. doi.org/10.1007/s00779-003-0253-8.

[21] Olive Jean Dunn. Multiple Comparisons Using Rank Sums. *Technometrics*, 6(3): 241–252, August 1964. doi.org/10.1080/00401706.1964.10490181. _eprint: https://www.tandfonline.com/doi/pdf/10.1080/00401706.1964.10490181.

[22] Gerhard Fischer. Social creativity: turning barriers into opportunities for collaborative design. In *Proceedings of the eighth conference on Participatory design Artful integration: interweaving media, materials and practices - PDC 04*, volume 1, page 152, Toronto, Ontario, Canada, 2004. ACM Press. doi.org/10.1145/1011870.1011889.

[23] Milton Friedman. The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, 32(200):675–701, 1937. doi.org/10.1080/01621459.1937.10503522.

[24] Alexandra Frischen, Andrew P. Bayliss, and Steven P. Tipper. Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, 133(4):694–724, 2007. doi.org/10.1037/0033-2909.133.4.694.

[25] Xuedan Gao, Amit Rogel, Raghavasimhan Sankaranarayanan, Brody Dowling, and Gil Weinberg. Music, body, and machine: gesture-based synchronization in human-robot musical interaction. *Frontiers in Robotics and AI*, 11, December 2024. doi.org/10.3389/frobt.2024.1461615.

[26] Tom Geller. Overcoming the Uncanny Valley. *IEEE Computer Graphics and Applications*, 28(4):11–17, July 2008. doi.org/10.1109/MCG.2008.79.

[27] Brain Geuther, Autumn Breese, and Yunfeng Wang. A study on musical conducting robots and their users. In *2010 10th IEEE-RAS International Conference on Humanoid Robots*, pages 124–129, December 2010. doi.org/10.1109/ICHR.2010.5686302. ISSN: 2164-0580.

[28] Elizabeth A. H. Green and Nicolai Malko. *The modern conductor : a college text on conducting based on the technical principles of Nicolai Malko as set forth in his The conductor and his baton*. Prentice-Hall, 4th ed edition, 1987.

[29] Guy Hoffman and Gil Weinberg. Shimon: an interactive improvisational robotic marimba player. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 3097–3102, New York, NY, USA, April 2010. Association for Computing Machinery. `doi.org/10.1145/1753846.1753925`.

[30] Guy Hoffman and Gil Weinberg. Interactive improvisation with a robotic marimba player. *Autonomous Robots*, 31(2):133–153, October 2011. `doi.org/10.1007/s10514-011-9237-0`.

[31] Kasper Hornbæk, Per Ola Kristensson, and Antti Oulasvirta. Introduction to user interfaces. In Kasper Hornbæk, Per Ola Kristensson, and Antti Oulasvirta, editors, *Introduction to Human-Computer Interaction*, page 0. Oxford University Press, June 2025. `doi.org/10.1093/oso/9780192864543.003.0023`.

[32] Eva Hornecker and Jacob Buur. Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 437–446, New York, NY, USA, April 2006. Association for Computing Machinery. `doi.org/10.1145/1124772.1124838`.

[33] Yuhan Hu, Peide Huang, Mouli Sivapurapu, and Jian Zhang. ELEGNT: Expressive and Functional Movement Design for Non-anthropomorphic Robot, January 2025. URL `http://arxiv.org/abs/2501.12493`. arXiv:2501.12493 [cs].

[34] Hiroshi Ishii and Brygg Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, pages 234–241, Atlanta Georgia USA, March 1997. ACM. `doi.org/10.1145/258549.258715`.

[35] Peter H. Kahn, Takayuki Kanda, Hiroshi Ishiguro, Brian T. Gill, Solace Shen, Jolina H. Ruckert, and Heather E. Gary. Human creativity can be facilitated through interacting with a social robot. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 173–180, March 2016. `doi.org/10.1109/HRI.2016.7451749`. ISSN: 2167-2148.

[36] Carlos Lage-Gómez and Roberto Cremades-Andreu. Group Identity in a Secondary School Classroom Constructed through Musical Creation / Oblikovanje grupnoga identiteta glazbenim stvaranjem u srednjoj školi. *Croatian Journal of Education - Hrvatski časopis za odgoj i obrazovanje*, 23(1), April 2021. `doi.org/10.15516/cje.v23i1.3824`.

[37] Raymond AR MacDonald and Graeme B. Wilson. Musical improvisation and health: a review. *Psychology of Well-Being*, 4(1):20, December 2014. doi.org/10.1186/s13612-014-0020-9.

[38] Andrew J. Martin and Susan A. Jackson. Brief approaches to assessing task absorption and enhanced subjective experience: Examining 'short' and 'core' flow in diverse performance domains. *Motivation and Emotion*, 32(3):141–157, September 2008. doi.org/10.1007/s11031-008-9094-0.

[39] Torsten Metzler, Fabian Gerlinghaus, Brennand Pierce, Iestyn Jowers, K. Shea, and Gordon Cheng. Design and Emotional Expressiveness of Gertie (An Open Hardware Robotic Desk Lamp). page 1, 2012.

[40] Oindrila Mukherjee and Suma Lavanya Mutnury. Management of Chronic Illness through Music Therapy: A Review. *Indian Journal of Health Studies*, 03 (02):55–94, 2021. doi.org/10.56490/IJHS.2021.3204.

[41] Karl Pearson. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, July 1900. doi.org/10.1080/14786440009463897. _eprint: https://doi.org/10.1080/14786440009463897.

[42] Hua Peng, Changle Zhou, Huosheng Hu, Fei Chao, and Jing Li. Robotic Dance in Social Robotics—A Taxonomy. *IEEE Transactions on Human-Machine Systems*, 45(3):281–293, June 2015. doi.org/10.1109/THMS.2015.2393558.

[43] Tuan Vu Pham, Thomas H. Weisswange, and Marc Hassenzahl. Embodied Mediation in Group Ideation – A Gestural Robot Can Facilitate Consensus-Building. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, DIS '24, pages 2611–2632, New York, NY, USA, July 2024. Association for Computing Machinery. doi.org/10.1145/3643834.3660696.

[44] Isabella Poggi, Loredana Ranieri, Ylenia Leone, and Alessandro Ansani. The Power of Gaze in Music. Leonard Bernstein's Conducting Eyes. *Multimodal Technologies and Interaction*, 4(2):20, June 2020. doi.org/10.3390/mti4020020.

[45] Amit Rogel, Richard Savery, Ning Yang, and Gil Weinberg. RoboGroove: Creating Fluid Motion for Dancing Robotic Arms. In *Proceedings of the 8th International Conference on Movement and Computing*, MOCO '22, pages 1–9, New York, NY, USA, June 2022. Association for Computing Machinery. doi.org/10.1145/3537972.3537985.

[46] Muhammad Saad, Shahid Iqbal, and Shoaib R. Soomro. Design and Development of a Multi-Sided Tabletop Augmented Reality 3D Display Coupled

with Remote 3D Imaging Module. *Journal of Engineering and Technological Sciences*, 54(6):220606, December 2022. doi.org/10.5614/j.eng.technol.sci.2022.54.6.6.

[47] Richard Savery, Ryan Rose, and Gil Weinberg. Establishing Human-Robot Trust through Music-Driven Robotic Emotion Prosody and Gesture, January 2020. URL http://arxiv.org/abs/2001.05863. arXiv:2001.05863 [cs].

[48] Kjeld Schmidt and Liam Bannon. Taking CSCW seriously. *Computer Supported Cooperative Work (CSCW)*, 1(1):7–40, March 1992. doi.org/10.1007/BF00752449.

[49] Ben Shneiderman. Creating creativity: user interfaces for supporting innovation. *ACM Transactions on Computer-Human Interaction*, 7(1):114–138, March 2000. doi.org/10.1145/344949.345077.

[50] Lucille Alice Suchman. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, November 1987. ISBN 978-0-521-33739-7. Google-Books-ID: AJ_eBJtHxmsC.

[51] Frank Thomas and Ollie Johnston. *The Illusion of Life: Disney Animation*. Abbeville Press, New York, NY, USA, 1981. ISBN 978-0-89659-224-6.

[52] Filip Verneert, Luc Nijs, and Thomas De Baets. A Space for Collaborative Creativity. How Collective Improvising Shapes 'a Sense of Belonging'. *Frontiers in Psychology*, 12:648770, March 2021. doi.org/10.3389/fpsyg.2021.648770.

[53] Marynel Vázquez, Elizabeth J. Carter, Braden McDorman, Jodi Forlizzi, Aaron Steinfeld, and Scott E. Hudson. Towards Robot Autonomy in Group Conversations: Understanding the Effects of Body Orientation and Gaze. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 42–52, Vienna Austria, March 2017. ACM. doi.org/10.1145/2909824.3020207.

[54] Frank Wilcoxon. Individual Comparisons by Ranking Methods. In Samuel Kotz and Norman L. Johnson, editors, *Breakthroughs in Statistics: Methodology and Distribution*, pages 196–202. Springer, New York, NY, 1992. doi.org/10.1007/978-1-4612-4380-9_16.

[55] Junchao Xu, Joost Broekens, Koen Hindriks, and Mark A. Neerincx. Effects of bodily mood expression of a robotic teacher on students. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2614–2620, September 2014. doi.org/10.1109/IROS.2014.6942919. ISSN: 2153-0866.

[56] Takemasa Yokoyama and Yuji Takeda. Gaze Cuing Effects in Peripheral Vision. *Frontiers in Psychology*, 10:708, April 2019. doi.org/10.3389/fpsyg.2019.00708.

[57] Tomoko Yonezawa, Hirotake Yamazoe, Akira Utsumi, and Shinji Abe. Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze-tracking. In *Proceedings of the 9th international conference on Multimodal interfaces*, ICMI '07, pages 140–145, New York, NY, USA, November 2007. Association for Computing Machinery. `doi.org/10.1145/1322192.1322218`.

Index

Typeset April 13, 2026