EmbodiComp: Embodied Interaction for Mixing and Composition

Dalia El-Shimy
Centre for Interdisciplinary Research
in Music, Media and Technology
McGill University
dalia@cim.mcgill.ca

Steve Cowan
Professional Guitarist
and Composer
scowan14@hotmail.com

Jeremy R. Cooperstock
Centre for Interdisciplinary Research
in Music, Media and Technology
McGill University
jer@cim.mcgill.ca

ABSTRACT

We introduce EmbodiComp, a novel system that leverages simple and common gestures to allow for simultaneous mixing and composition. Through the use of a "band performance" metaphor that offers users the illusion of being part of an ensemble, musicians are able to play and mix their instruments with pre-recorded tracks in real-time through embodied interactions. Using five unique features, our system allows musicians to experiment seamlessly with volume and reverb levels, as well as the degree to which instruments are mixed, as they simply move about their space. As such, users can easily explore various settings and arrangements during composition, and determine how an instrument might best fit with others in the final piece. The system evolved, in part, as a result of a collaboration between an engineer and a composer that is also described in this paper. The outcomes of this participatory design cycle indicate that EmbodiComp could prove beneficial for musicians seeking to facilitate the process of composition through alternatives to traditional mixing tools.

1. INTRODUCTION

Musical performance and mixing have traditionally been treated as separate processes, which is natural since musicians can hardly be expected to step over repeatedly to a mixing console or computer in order to adjust settings mid-performance. The exception, perhaps, is the case where the computer is also the instrument. We use the term "mixing" to denote "the adjustment of relative volumes, panning and other parameters corresponding to different sound sources, in order to create a technically and aesthetically adequate sound sum" [1]. Digital audio workstations (DAWs) continue to be the gold standard for audio recording, editing and mixing, with possibilities that range from simple two-channel editors to complete recording suites, and include both hardware and software components. However, the vast majority of stations continue to operate according to the same "multitrack tape recorder" metaphor, utilizing mixing consoles that allow musicians to control multiple channels—each carrying an audio track—through pan pots, faders

and sliders, or software solutions that simply simulate such mixing consoles.

The drawbacks to such traditional mixing technology are that it significantly constrains composition activities that wish to mix musical input as it is being generated, and its requirement of hands-on interaction is ill-suited to supporting musicians who wish to exercise independent control over their mix during performance. As a solution to these problems for the musician-composer, we propose EmbodiComp, an alternative to the DAW interface that leverages simple gestures as a means of controlling and mixing various audio channels. This approach employs the idea of embodied interactions to allow for hands-free, seamless, dynamic control of musical parameters during performance. By allowing musicians to play *and* mix their instruments with pre-recorded tracks in real-time—thereby effectively bridging the gap between mixing and performance—such embodied interactions can help enhance creativity during composition.

We note that EmbodiComp is not necessarily meant for producing polished, final works. Rather, it aims to help single musicians experiment seamlessly with various mix possibilities during the process of composition, in order to determine how an instrument might best fit among others in a final recording.

2. BACKGROUND AND RELATED WORKS

In spite of the tremendous potential afforded by the advent of digital audio, mixing interfaces have changed very little in the decades following their introduction [1, 2]. As exemplified through such systems as Avid Technology's Pro Tools, Apple's Logic Pro, Ableton Live and Steinberg's CueBase, the software systems most commonly used by professionals and amateurs alike take their inspiration from the mixing console: faders, knobs and sliders are considered standard tools for mix control [3]. However, although a number of systems have sought to facilitate or improve the mixing process through novel solutions, most continue to reflect the console analogy. For instance, while the Lemur2 and Dexter interfaces, both developed by JazzMutant, offer multi-touch to allow users to take advantage of common pinching and expansion gestures for added precision, their layout still emulates that of the mixing console [1, 4]. As another example, the Cuebert system, which also utilizes a multi-touch interface to allow for flexible display of dynamic and context-sensitive content in the "high-pressure" environment of musical theatre, relies

on a traditional mixing board paradigm as well [2].

Nonetheless, a few alternatives have been proposed. For instance, Pachet et al. introduced the concept of "dynamic audio mixing", which offers listeners direct control over the spatialization of musical pieces [5]. To facilitate this process, while allowing users to move more than one sound source at a time, the authors employ a constraint paradigm that aims to preserve the properties of the configuration of sound sources that need to be satisfied in order to maintain "coherent, nice-sounding mixings". Such ideas were implemented through MusicSpace, a system whereby speaker icons representing sound sources, and an avatar representing the listener, can be moved graphically to induce real-time changes in the spatial arrangement of an overall piece [6]. This work can also be seen as an example of the emerging active music listening paradigm, which gives listeners the ability to mix and manipulate the different constituent sources, or "stems", of a musical piece on their own [7]. Similarly, Carrascal et al. developed an interface that allows its users to manipulate spatially arranged sound sources, in an attempt to take into account modern mixing technologies such as surround and 3D audio [1]. Another example is the waveTable, a tabletop audio waveform editor that combines multi-touch and tangible interaction techniques, allowing users to manipulate sound samples directly [8]. Furthermore, the Chopping Board allows users to "chop" and re-sequence tracks through interaction with a physical "editing pad" that can detect their gestures through a combination of infrared and touch sensors [9]. Our final example is Noisescape, a 3D first-person computer game where users can collaboratively compose complex musical structures, by creating and combining elements with varying physical attributes [10]. However, much like those inspired by mixing consoles, the systems described here do not support simultaneous performance with an instrument and mixing by the same user. Therefore, we turn instead to the concept of embodied interactions as a solution that allows for hands-free, seamless, dynamic control of musical parameters mid-performance.

The idea of embodiment is deeply rooted within the musical context, with Godøy et al. describing the well-established links between musical sounds and sound-producing movement as an "embodied understanding of music perception and cognition" [11]. Embodied music cognition views the relationship between sound and movement as having its roots in the broader paradigm of embodied cognition, which stipulates that people relate perception to mental stimulations of associated actions. For our purposes, however, we use the related notion of embodied interaction commonly found in human-computer interaction research, and described by Antle et al. as "leveraging users' natural body movement in direct interaction with spaces and everyday objects to control computational systems" [12]. Examples of this notion within the context of music include the Sound Maker system, which was designed to map a user's location and movement to changes in the pitch, tempo and volume of an electronically-generated percussive stream, and can also be seen as providing an alternative to traditional mixing techniques. Furthermore,

the Ariel system, designed by Corness and Schirphorst, system responds to gestures utilized by musicians during improvisation with simulated breathing sounds. Ariel was specifically designed to capitalize on the ability of skilled musicians to exchange, detect and tacitly respond to cues for interpersonal interactions [13]. Finally, Bakker et al. advocate the use of embodied interaction within the context of musical learning for children. As an example, the authors developed the Moving Sounds Tangibles, a system that allows children to learn abstract sound concepts such as pitch, volume and tempo by manipulating a set of interactive tangibles designed in accordance with various schemata, or higher-order cognitive structures that emerge from recurring patterns of bodily or sensori-motor experience [14].

3. SYSTEM DESCRIPTION

EmbodiComp allows for simultaneous performance and mixing according to a "band performance" metaphor: a musician using the system is given the illusion of performing alongside two virtual "band members", each of whom is assigned a pre-recorded track. A graphical user interface (GUI), seen in Figure 1, offers a top down view of all participants, including the user, as avatars. The musician can then play their instrument and interact with the other band members' tracks according to the system features described next.

3.1 Features

EmbodiComp currently offers musicians the following five features:

- **Dynamic volume:** As a user moves *towards* the avatar of another band member, the pre-recorded track associated with that band member is experienced as gradually increasing in volume. The converse holds true as the user moves away from that band member's.

- **Dynamic reverb:** As a user moves *away* from the avatar of another band member, the pre-recorded track associated with that band member is experienced as gradually increasing in reverberation. The converse holds true as they move towards that band member's avatar.

- **Mix control:** This feature allows the user to change the mix of their instrument with the pre-recorded tracks by tilting their head. Tilting to the left will move the sound of their instrument, along with that of the band member whose avatar is to their left, entirely to the left headphone. The track of the band member whose avatar is to their right will be heard unaccompanied through the right headphone. The converse holds true when the user tilts their heads to the right.

- **Track panning:** A user can isolate each of the pre-recorded tracks by changing their body's orientation. Turning their body to the left will allow them to hear only the track of the band member whose avatar is
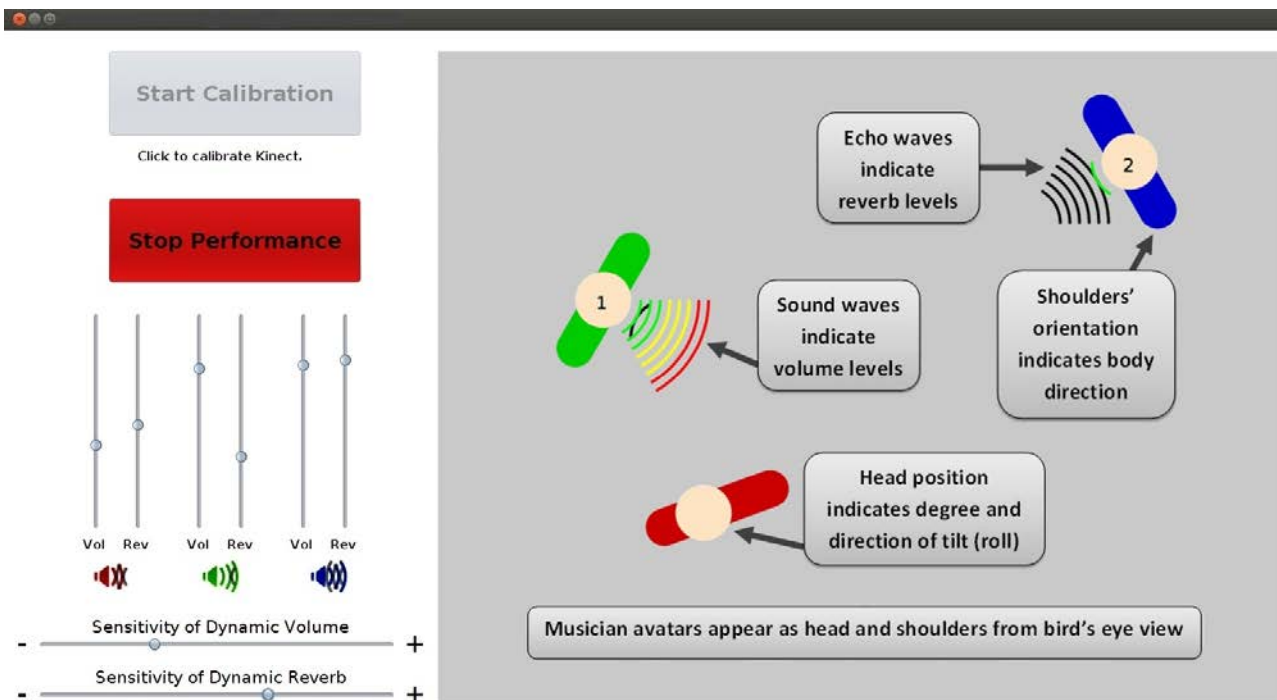
Figure 1. Main graphical user interface, which includes a control panel and animated graphics. The user's avatar is in red.

to their left, entirely through the left headphone. The track of the band member whose avatar is to their right will become silent. The user's own instrument will continue to sound the same, coming through both headphones. The converse holds true when the user turns their body to the right.

- **Musician spatialization:** This features allows a user to experience the pre-recorded tracks as spatialized sound sources within their own space. The spatialization effect is determined by the user's body orientation, and changes accordingly.

3.2 Graphical User Interface

As mentioned above, EmbodiComp offers musicians access to a main graphical user interface, seen in Figure 1, that serves a number of functions. First, the avatars representing the user among the band members are dynamically animated to graphically reflect the changes in sound effected by the system features. In addition, the panel on the left side of the main GUI, allows users to set the base volume and reverb levels for themselves and the pre-recorded tracks at the very start of a session. It is those base values that are subsequently affected by the system features. The panel also allows users to start and stop the system, calibrate the tracking device, and select the sensitivities of the dynamic volume and dynamic reverb features.

Users also have access to the secondary GUI seen in Figure 2, which allows them to select the system features they would like to use, and move the avatars of the virtual band members, independently of their actual physical positions. Moving the avatars allows users to experiment with the subset of the overall dynamic volume and dynamic reverb ranges they experience. Specifically, the range for both

features is determined as a function of the minimum and maximum possible distances between any two avatars. If a user moves one band member's avatar significantly closer, this in turn reduces the maximum distance that can be achieved relative to that avatar as the user moves about in their physical space. As a result, they will experience a subset of volume changes closer to the higher end of the possible dynamic volume range, and a subset of reverb changes closer to the lower end of the possible dynamic reverb range for the track associated with that particular avatar.

3.3 Configuration

Our system configuration can be seen in Figure 3. The musician's instrument is captured by an audio interface, such as the Roland Edirol FA-101. It is then routed, along with two pre-recorded tracks loaded in a sequencer such as Ardour, to our SuperCollider (SC) software via the JACK Audio Connection Kit. The musician's position and orientation information is tracked by a Microsoft Kinect, and also sent to our SuperCollider software via Open Sound Control messages. Such information is then used to process the audio streams according to the user's choice of system features described above. Subsequently, the resulting mix is sent back to JACK, where it can be routed to the audio interface for playback, and to the sequencer for recording. We note that, as an alternative to loading pre-recorded tracks in a sequencer, a musician can also choose to mix his instrument with tracks recorded on-the-fly and played back through a Loop Station connected to the audio interface. In either case, the tracks can be routed to SuperCollider as separate channels.
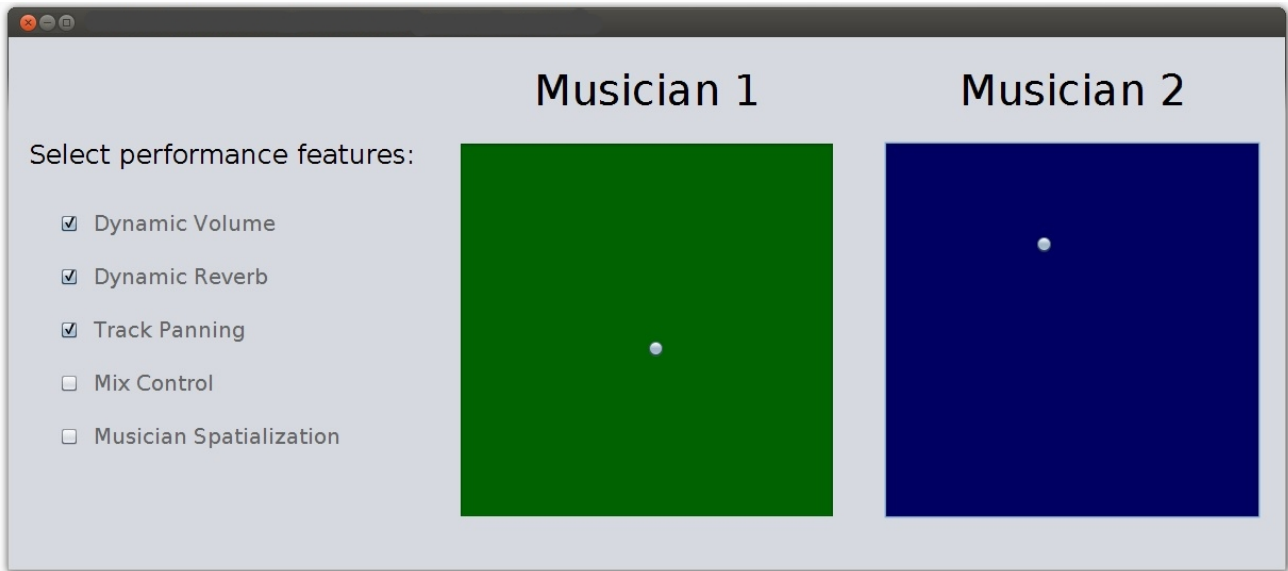
Figure 2. Secondary graphical user interface for feature selection and avatar control.

Figure 3. Configuration of the system's hardware and software components.

4. PARTICIPATORY DESIGN CYCLE

Inspired by a previous project on augmented distributed performance described in reference [15], we had developed a prototype for EmbodiComp that encompassed three of the features described in Section 3.1: dynamic volume, track panning and musician spatialization. In a bid to further refine the system's existing features and explore new ones, while simultaneously gauging the extent to which it could support the creative process, we invited a composer to take part in a participatory design cycle. We opted for the "cooperative prototyping" participatory design technique, which entails delivering a system to its end-users as a series of iterative prototypes, each of which gradually adds functionality. Cooperative prototyping offers several advantages, including enhanced communication by grounding discussions in concrete artefacts, and improved working relations through a sense of shared ownership of the resulting system. The success of this technique hinges on presenting each prototype as a "crucial artifact in the end user's work", which allows them to form ecologically valid impressions of the system [16]. As a result, the composer was simply asked to write a few musical pieces using EmbodiComp, and informed that his criticisms and suggestions, no matter how extensive, would play a crucial part in shaping any further iterations of the system.

4.1 Methodology

Our collaboration with the composer lasted 14 weeks, with sessions being held on a regular basis every 1-2 weeks. The composer spent the first few sessions familiarizing himself with the system, and determining how to best approach his given task. After this introductory phase, he began shifting his focus towards experimentation. Each session would begin with a discussion of any changes made to the system as a result of previous suggestions. Subsequently, he would spend a few hours playing music and interacting with the

system. During this exploratory stage of the session, the composer would typically record his impressions in point-form notes, while we provided our assistance on demand, and only in a technical capacity to resolve any glitches with the system, or make clarifications. Afterwards, a discussion would be held, allowing the composer to share the notes he had made, and describe how our prototype could be improved for the following week's session. The composer would then take a few days to expand on the ideas contained in his notes, before sending us a full report that typically included additional details and explanations for his recommendations, and comments on the progress of the pieces thus far. In the final weeks, as the composer determined the system to have reached a satisfactory state and, with fewer recommendations to make, he began to immerse himself fully in the process of composition.

4.2 Outcomes

In addition to making recommendations for improving existing features, the composer was the source behind new additions to EmbodiComp. For instance, he introduced the idea behind the mix control feature, and was in large part responsible in shaping the dynamic reverb feature. He also made extensive recommendations to help improve the system's overall sound quality, the design of the graphical user interface, and the animated avatars.

In a final report summarizing his experience with our system, the composer found that embodied interactions lent themselves particularly well to seamless experimentation with various mix settings, which, in turn, helped facilitate the process of composition. He explained that he previously had a tendency to avoid the post-composition mixing process:

"Almost every musician I know these days has some sort of recording software on their computer, and thus has the ability to record and produce multi-track recordings at home. Personally, I find all the clicking and computer-based activity in this to drain my creative energy and make the process frustrating."

In contrast, however, he found the ability to compose and mix simultaneously to be particularly beneficial:

"Using the performance system here, I was able to get some great solutions for these issues without having to do anything other than play my music in real time, and move my body a bit. I was easily able to see which tracks sounded best panned left, or right, or in the center; I was able to hear which textures were better off in the foreground, and which sounded better off more "distant", perhaps with a hint of reverb; I was able to iron out how two musical ideas interacted one on one, and then with a slight 90 degree turn, could hear how it then sounded with a third musical idea in the mix. "

The composer further detailed how certain features proved to be particularly well-matched to specific stages of the compositional process:

"Other than dynamic manipulations to volume and reverb, the three features I worked with also provided a logical succession for the creative process. Track panning allows the ability to work on ideas one on one, by cutting out

one of the 3 musicians with a simple torso pivot. The mix control brings all 3 players into the mix, but with the ability to pan your own part around to see how everything is blending/working together. Then the spatialization is a good final step, fleshing out the music ideas into their own space within the panning, and hearing how it works in a situation that will sound closer to the eventual desired final product (be it a live performance or a recording)."

In summary, the composer had a positive impression of the overall system:

"In conclusion, the features that this system offered were fun, useful, and helped me come up with new musical and production ideas."

However, he also offered important criticisms, explaining, for instance, that the system's current motion tracking technique may prove inadequate for instruments that require musicians to be seated, such as the keyboard. Furthermore, he anticipated that the lack of precise, numerical representation of the various levels effected by the system features might make it more difficult to correctly re-create the mix when working on the final, polished product.

5. FUTURE WORK

The participatory design cycle we held with the composer was beneficial in helping improve our system, and shedding some light on its potential for facilitating mixing and composition. However, we would like to further validate the generalizabilty of this collaboration's outcome, and determine whether the idea of embodied interaction for mixing and composition is one that a broader set of users would also find advantageous. As such, we hope to conduct formal user experiments in order to investigate further improvements, and explore the possibility of supporting new features.

Furthermore, our current prototype only supports two pre-recorded tracks in addition to the instrument being played by the user. As elaborate compositions can involve a far greater number of instruments, we would like to expand our system to allow for more complex pieces. This would require updating our current features to support various spatial arrangements of the user in relation to an increasing number of virtual band members, each associated with a pre-recorded track.

Finally, as per the composer's criticism, we would like our system to better accommodate seated musicians. The current implementations of the dynamic volume and dynamic reverb, which respond to motion, and even features such as track panning or musician spatialization, which rely on body orientation, cannot be used to their full potential by such musicians. Therefore, we wish to investigate alternative embodied gestures as input for these features, while still maintaining a reasonably clear mapping to the resulting auditory output.

6. CONCLUSION

A system that leverages embodied interactions for simultaneous mixing and composition was developed. EmbodiComp differs from the ubiquitous digital audio workstation paradigm in its reliance on a "band performance" metaphor,

whereby users are given the illusion of playing as part of an ensemble whose instruments can be mixed with their own in real-time. Through the use of several gesture-based features, musicians are able to adjust their mix mid-performance seamlessly, simply by moving around their space. The current system was designed alongside a composer who provided recommendations for new features and overall improvements to sound quality. The composer found that bridging the gap between mixing and performance helped improve his creative process, allowing him to experiment with various settings in real-time and, in turn, determine how an instrument could best fit within a piece. As such, we believe that the system described here could prove beneficial for other musicians seeking alternatives to traditional mixing solutions that may enhance their creativity during composition.

7. REFERENCES

[1] J. P. Carrascal and S. Jordà, "Multitouch Interface for Audio Mixing," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, ser. NIME '11, 2011, pp. 100–103.

[2] N. Liebman, M. Nagara, J. Spiewla, and E. Zolkosky, "Cuebert: A New Mixing Board Concept for Musical Theatre," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, ser. NIME '10, 2010, pp. 51–56.

[3] J. Forsyth, A. Glennon, and J. P. Bello, "Random Access Remixing on the iPad," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, ser. NIME '11, 2011, pp. 487–490.

[4] C. Roberts, "Multi-touch, Consumers and Developers," Media Arts and Technology Program, University of California, Tech. Rep., 2008. [Online]. Available: http://www.mat.ucsb.edu/200C/2008_Students/MAT-200C_2008_Files/charlie_roberts/charlesRoberts_200C_finalPaper.pdf

[5] F. Pachet, O. Delerue, and P. Hanappe, "Dynamic Audio Mixing," in *Proceedings of ICMC*. Berlin: ICMA, 2000.

[6] F. Pachet and O. Delerue, "MusicSpace: a Constraint-based Control System for Music Spatialization," in *Proceedings of ICMC 1999*. Beijing: ICMA, 1999, pp. 272–275.

[7] N. Sturmel, A. Liutkus, J. Pinel, L. Girin, S. Marchand, G. Richard, R. Badeau, and L. Daudet, "Linear Mixing Models for Active Listening of Music Productions in Realistic Studio Conditions," in *Audio Engineering Society Convention 132*, Apr 2012. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16232

[8] G. Roma and A. Xambó, "A Tabletop Waveform Editor for Live Performance," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, ser. NIME '08, 2008, pp. 249–252.

[9] J. Lee, "The Chopping Board: Real-time Sample Editor," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, ser. NIME '06, 2006, pp. 77–78.

[10] M. Grierson, "Noisescape: An Interactive 3D Audio-visual Multi-User Composition Environment," in *Proceedings of International Computer Music Conference*, ser. ICMC '07, 2007.

[11] R. I. Godøy, A. R. Jensenius, and K. Nymoen, "Chunking in Music by Coarticulation," *Acta Acustica united with Acustica*, vol. 96, no. 4, pp. 690–700, 2010.

[12] A. N. Antle, G. Corness, and M. Droumeva, "Human-Computer-Intuition? Exploring the Cognitive Basis for Intuition in Embodied Interaction," *International Journal of Arts and Technology*, vol. 2, no. 3, pp. 235–254, 2009.

[13] G. Corness and T. Schiphorst, "Performing with a system's intention: Embodied cues in performer-system interaction," in *Proceedings of the 9th ACM Conference on Creativity & Cognition*, ser. C&C '13. New York, NY, USA: ACM, 2013, pp. 156–164.

[14] S. Bakker, A. N. Antle, and E. Van Den Hoven, "Embodied Metaphors in Tangible Interaction Design," *Personal Ubiquitous Computing*, vol. 16, no. 4, pp. 433–449, Apr. 2012.

[15] D. El-Shimy and J. R. Cooperstock, "Reactive Environment for Network Music Performance," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, ser. NIME '13, May 2013.

[16] M. J. Muller, "Participatory Design: The Third Space in HCI," in *The Human-Computer Interaction Handbook*, J. A. Jacko and A. Sears, Eds. Hillsdale, NJ, USA: L. Erlbaum Associates Inc., 2003, pp. 1051–1068.