Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 81–90
https://doi.org/10.54941/ahfe1008243

From Visual Observation to Auditory Interaction: A Human-Centered Tangible Music System Based on Physical Objects

Yishan Hu, Liwen Liu, Qihang Yu, Sishuo Liu, and Yiqi Wang

School of Architecture and Design, Harbin Institute of Technology, Harbin 150001, China

ABSTRACT

Visual information plays a central role in many contemporary interactive experiences, while sound is often treated as a supplementary feedback channel. This study presents Contour Sound Chamber, an object-based tangible interactive music system that explores how physical object properties and bodily actions can be transformed into auditory experiences. The system integrates AI-based object recognition, spatial sensing, sound mapping, and tangible interaction. Objects are classified according to material and color temperature, forming four predefined categories. Material determines the basic timbre, color temperature influences tonality and instrumental layers, and contour data acquired through ToF scanning are quantized into rhythmic patterns. Users further participate in the generated music through four tangible interaction modes: striking, plucking, pressing, and stroking, corresponding to different instrumental sounds. A user study with 34 participants evaluated object recognition, interaction task performance, and subjective perception. The system achieved a 97.1% first-attempt recognition accuracy across 136 trials, while the interaction tasks achieved an overall mean efficiency of 90.1% and a completion rate of 96.3%. The subjective evaluation showed an overall Likert score of 4.60/5. Participants were able to perceive differences between sounds generated by different objects, recognize the influence of their actions on musical output, and perceive certain object characteristics through sound. Together, these findings provide initial support for a perceptible relationship between object properties, sound representation, and user interaction, highlighting the potential of sound as an active component of multisensory interaction.

Keywords: Human-computer interaction, Human-centered design, Tangible interaction, Interactive music, Multimodal interaction, Object recognition, Spatial sensing

INTRODUCTION

Visual information plays a central role in many contemporary interactive experiences, while sound is often used primarily for feedback, communication, or entertainment. Beyond these functions, sound can also represent information and support the perception of physical characteristics through auditory patterns (Gaver, 1986). This provides an opportunity to explore sound as an active component of interaction rather than only as supplementary feedback (Franinović and Serafin, 2013).

Tangible interaction allows users to engage with digital information through physical objects and bodily actions. In interactive music systems,

Received September 22, 2026; Revised October 23, 2026; Accepted October 28, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

physical actions can also influence sound generation. However, the intrinsic characteristics of objects, such as material, color, and spatial structure, are less frequently incorporated into the sound-generation process. This study therefore focuses on the relationship between object properties, sound representation, and user interaction, and investigates the following question: How can physical object characteristics and user actions be transformed into auditory experiences that users can actively perceive and explore?

To investigate this question, this study presents Contour Sound Chamber, an object-based tangible interactive music system that integrates object recognition, spatial sensing, sound mapping, and tangible interaction. Material, color temperature, and contour information are mapped to different musical elements, while users interact with the generated music through physical actions. A user study with 34 university participants evaluates object recognition performance, interaction task performance, and users' perception of the relationship between objects, actions, and sound.

FEASIBILITY ANALYSIS

Tangible interaction provides a theoretical basis for connecting physical objects, bodily actions, and digital information. Previous studies have demonstrated that physical objects can function as carriers of digital information and support embodied interaction with computational systems (Ishii and Ullmer, 1997; Hornecker and Buur, 2006). However, objects in tangible systems are often used primarily as input or control tools, while their intrinsic characteristics are less frequently incorporated into the interaction content.

Sound also provides a means of representing non-linguistic information. Sonification research has shown that data can be transformed into auditory signals, allowing patterns and changes to be perceived through listening (Kramer et al., 1999; Hermann, Hunt, and Neuhoff, 2011). Meanwhile, embodied interaction and tangible music systems demonstrate that bodily actions can become meaningful ways of generating and controlling sound (Dourish, 2001; Jordà, 2008). Based on these foundations, this study connects these three elements—object properties, auditory representation, and bodily interaction—to explore how the physical characteristics of objects can become part of an interactive auditory experience.

RESEARCH METHOD AND SYSTEM DESIGN

Research Method

This study adopts a Research through Design (RtD) approach. Zimmerman, Forlizzi, and Evenson (2007) proposed RtD as a method for interaction design research in HCI, emphasizing the role of design practice and design artifacts in generating research contributions.

Based on this approach, the study explores how physical objects can be represented and perceived through sound. The research process consists of four main stages: problem identification, concept development, prototype design, and user evaluation. Through this process, the prototype serves not

only as the final design outcome but also as a research medium for examining the proposed concept and exploring the relationship between physical objects and sound.

System Overview

Based on the Research through Design approach, this study focuses on three objectives: exploring object-based auditory perception, developing an interactive object-to-sound system, and evaluating its feasibility and user experience.

To achieve these objectives, Contour Sound Chamber establishes the following process: *Object → Recognition & Scanning → Sound Mapping → Interactive Sound → Auditory Perception → User Evaluation.*

The system uses HUSKYLENS to identify object attributes and ToF10120 to capture spatial information. These data are transformed into musical elements, while users' physical interactions further influence sound generation. The final user evaluation examines the feasibility and experience of this object-to-sound interaction.

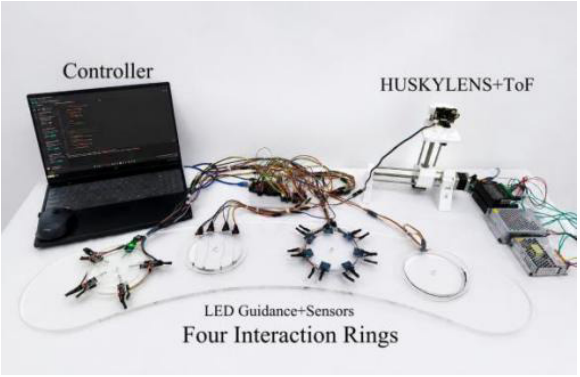


Figure 1: Prototype setup of the contour sound chamber.

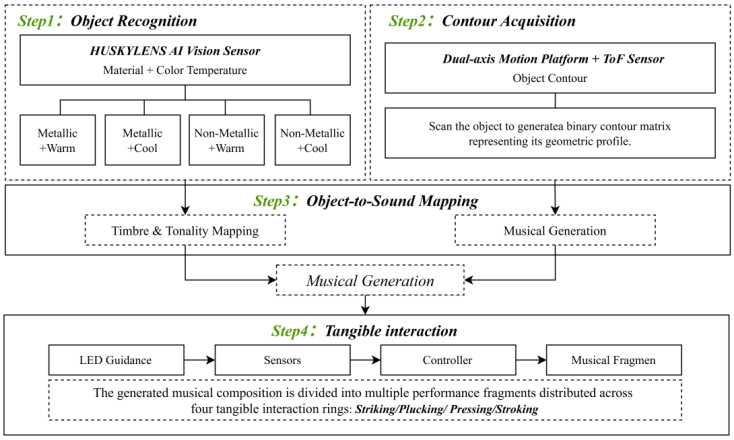


Figure 2: System flowchart of contour sound chamber.

Data Processing and Sound Mapping

Before system operation, a musical sample library was established by collecting instrument sounds corresponding to the four interaction rings from Splice. Each instrument was assigned two different tonal variations. The four interaction rings correspond to drums, guitar, piano, and harp, which are triggered through striking, plucking, pressing, and stroking gestures, respectively. The collected samples were standardized into a consistent audio format and organized into a sound library according to predefined mapping rules between object properties, timbre, tonality, and scanning data.

For object recognition, a HUSKYLENS AI vision sensor connected to an Arduino controller is used to identify the object's material and color temperature. Training samples were used to establish classifications of metallic/non-metallic materials and warm/cool colors. The resulting categories were defined as ID1 (metallic + warm), ID2 (metallic + cool), ID3 (non-metallic + cool), and ID4 (non-metallic + warm). The recognition result is transmitted to the Arduino as a classification signal and used to activate the corresponding musical mapping rules.

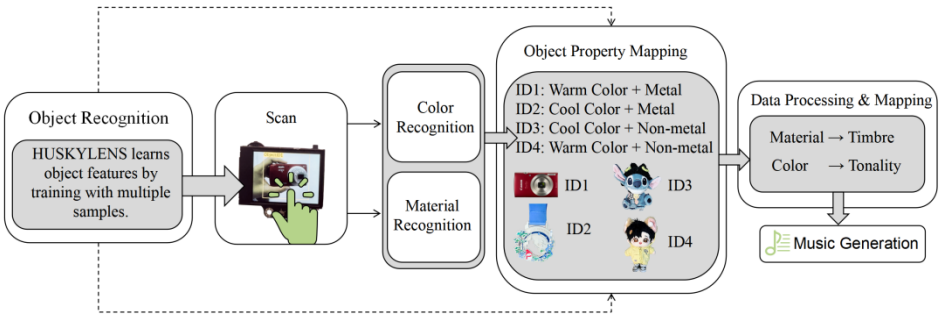


Figure 3: HUSKYLENS-based object recognition and property classification.

Meanwhile, a dual-axis stepper motor drives a ToF10120 sensor to scan the object and obtain two-dimensional contour data. The sensor performs one distance measurement every three seconds. A distance below 200 mm is recorded as "1", while a distance of 200 mm or above is recorded as "0". Every 20 measurements form one data group, and 20 groups are collected to construct the contour data.

In the musical mapping process, material primarily determines the basic timbre, while color temperature determines the tonality and corresponding instrumental layers. In the proposed mapping scheme, ID1 and ID2 use guitar as the primary melodic instrument for metallic objects, while ID3 and ID4 use piano for non-metallic objects. Warm-color objects are paired with drum patterns to create a relatively lively musical character, whereas cool-color objects are paired with harp sounds to create a calmer auditory impression.

The contour data obtained through ToF scanning are quantized into musical rhythm and aligned to a unified rhythmic grid. These rhythmic

patterns are then combined with the corresponding instrumental layers, transforming the object's material, color, and spatial contour into structured musical output with distinct rhythmic and emotional characteristics.

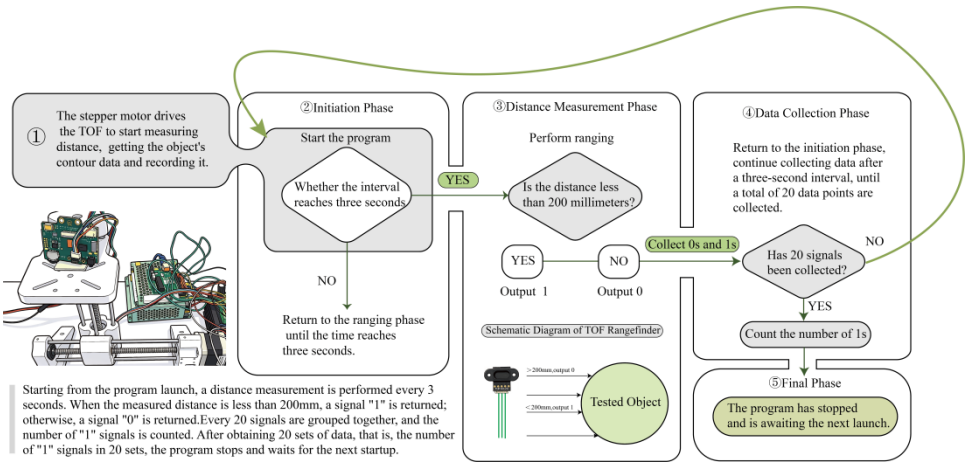


Figure 4: ToF-based object contour acquisition and data collection process.

Tangible Interaction and Musical Performance

The object-to-sound mapping establishes the auditory representation of each object, while tangible interaction determines how users actively experience this generated musical content. In this way, object properties provide the basis for sound generation, and user actions provide the means through which the resulting sound is explored and performed. The system therefore connects object perception and user participation within a continuous interaction process.

Following the sound mapping process, users interact with the system through four tangible interaction rings, each corresponding to a specific musical instrument. Striking, plucking, pressing, and stroking gestures activate the drum, guitar, piano, and harp sounds, respectively. After the object is processed, the generated composition is divided into corresponding musical fragments, while sequential LED illumination guides users to the appropriate interaction ring. As shown in Figure 5, when users perform the required gestures on the interaction rings, the corresponding sensors detect their actions and trigger the assigned musical fragments.

If no interaction is detected within one second, the system automatically advances to the next fragment to maintain temporal continuity. After all four interaction stages are completed, the generated fragments are replayed as a complete musical composition. This interaction design allows users to actively participate in the music-generation process and explore the relationship between object properties, bodily actions, and sound.

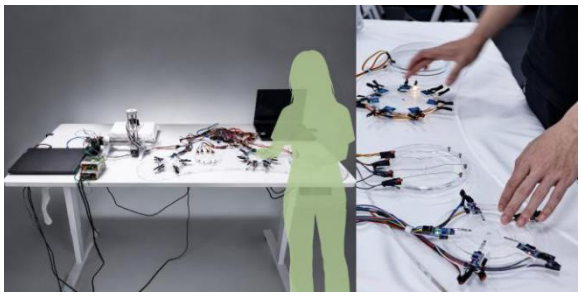


Figure 5: User interaction with the tangible music interface.

USER STUDY AND EVALUATION

A user study involving 34 participants was conducted to evaluate the system from three aspects: object recognition accuracy, interaction task performance, and subjective perception. Each participant tested and interacted with four predefined objects. The evaluation consisted of three consecutive stages: object recognition, interaction task performance, and subjective evaluation.

Object Recognition Accuracy Evaluation

During the object recognition stage, each participant tested the four objects once, resulting in a total of 136 recognition trials. Object placement, sensing distance, camera position, environmental lighting, and system parameters were kept relatively consistent. A trial was considered successful when the classification ID detected by the system matched the predefined category of the object. If the first recognition attempt failed, a second recognition attempt was conducted.

Of the 136 first-attempt recognition trials, 132 were successful, resulting in an overall first-attempt recognition accuracy of 97.1%. Object 1 had two failed first attempts, corresponding to an accuracy of 94.1%, while Objects 2 and 4 each had one failed attempt, corresponding to an accuracy of 97.1%. All 34 trials for Object 3 were successfully recognized, achieving an accuracy of 100%. At the participant level, 30 of the 34 participants successfully completed the first-attempt recognition of all four objects without error, accounting for 88.2% of the participants. The four failed first attempts occurred for Object 4 with Participant 6, Object 1 with Participants 11 and 16, and Object 2 with Participant 28. All four cases were successfully recognized on the second attempt.

Table 1: Object recognition performance.

Object	First-Attempt Success	First-Attempt Failure	First-Attempt Accuracy	Second-Attempt Success
Object 1	32	2	94.1%	2/2
Object 2	33	1	97.1%	1/1
Object 3	34	0	100%	-
Object 4	33	1	97.1%	1/1
Overall	**132**	**4**	**97.1%**	**4/4**

These results indicate that the system was able to recognize the predefined object-property categories with relatively stable performance, with only minor differences among the four test objects. The unsuccessful first attempts were corrected during the second recognition attempt, supporting the feasibility of the current recognition method. However, recognition stability may still be affected by factors such as object placement, environmental lighting, and temporary variations in the recognition process.

Interaction Task Performance Evaluation

Following the recognition stage, participants completed a task-based interaction experiment to evaluate the usability and interaction efficiency of the system. For each object, the required number of interactions was predefined according to its corresponding musical sequence. During the experiment, the actual number of interactions performed by each participant, task completion, and assistance requirements were recorded. An interaction was considered successful when the required action was correctly detected and completed within one second. Interaction efficiency was defined as the proportion of predefined interactions successfully completed by participants (Actual Interactions / Required Interactions × 100%).

The results showed an overall mean interaction efficiency of 90.1% (SD = 7.3%), with an overall task completion rate of 96.3% and an assistance rate of 10.3%. Descriptively, Object 3 showed the highest mean efficiency (92.6%), while Object 1 showed a lower mean efficiency (86.4%). Objects 2 and 4 showed similar mean efficiencies of 90.8% and 90.6%, respectively. In terms of task completion, Objects 2, 3, and 4 achieved a 100% completion rate, while Object 1 achieved 85.3%. Assistance rates varied across the four objects, with Object 4 showing the highest rate at 23.5%.

However, a Friedman test showed no statistically significant difference in interaction efficiency across the four objects ($\chi^2 = 5.654$, $p = .130$). Overall, participants were generally able to complete the object-based musical interaction tasks with relatively high efficiency and completion rates. Differences in assistance requirements and task completion suggest that some interaction sequences may still require further refinement.

Table 2: Task performance across four objects.

Object	Required Interactions	Actual Interactions (Mean)	Efficiency (Mean ± SD)	Completion Rate	Assistance Rate
Object 1	27	23.32	86.4% ± 8.3%	85.3%	14.7%
Object 2	26	23.62	90.8% ± 7.1%	100%	2.9%
Object 3	23	21.29	92.6% ± 5.2%	100%	0%
Object 4	32	29.00	90.6% ± 7.1%	100%	23.5%
Overall	-	-	**90.1% ± 7.3%**	**96.3%**	**10.3%**

Note. Interaction efficiency represents the proportion of predefined interactions successfully completed by participants (Actual Interactions / Required Interactions × 100%). A Friedman test showed no significant difference in interaction efficiency among the four objects ($\chi^2 = 5.654$, $p = 0.130$).

Subjective User Evaluation

After completing the interaction tasks, participants evaluated their experience using a six-item 5-point Likert-scale questionnaire. The questionnaire was designed to examine the system from three main aspects: interaction usability, object-to-sound perception, and overall user experience. Q1 and Q2 focused on whether participants could understand the interaction process and perceive the influence of their physical actions on the generated music, reflecting the usability and responsiveness of the tangible interaction. Q3 and Q4 examined whether differences between objects could be perceived through sound and whether auditory output could convey certain physical characteristics of the objects, directly addressing the feasibility of the proposed object-to-sound mapping. Q5 and Q6 focused on participants' interest, willingness to continue exploring, and their overall perception of the system's interactivity, innovation, and experiential value.

Table 3: Results of the 5-point likert-scale questionnaire.

Item	Questionnaire Statement	Mean ± SD
Q1	I could easily understand and complete the interaction with the system.	4.38 ± 0.82
Q2	My actions directly affected the generation or change of the music.	4.79 ± 0.48
Q3	I could perceive clear differences between the sounds generated by different objects.	4.85 ± 0.36
Q4	I could perceive some physical properties or characteristics of the objects through sound.	4.53 ± 0.75
Q5	I found this object-based musical interaction interesting and wanted to continue exploring it.	4.47 ± 0.79
Q6	I consider the system to have good interactivity, innovation, and experiential value.	4.59 ± 0.66
Overall	-	**4.60**

The results showed generally positive evaluations across all items. Participants were able to perceive differences between the sounds generated by different objects and recognized a clear relationship between their physical actions and the resulting musical output. They also reported that sound helped them perceive certain characteristics of the objects, supporting the feasibility of representing physical object properties through auditory information. In addition, participants showed positive engagement and willingness to further explore the interaction, suggesting that the combination of tangible interaction and musical feedback encouraged active participation rather than passive listening.

DISCUSSION

The results provide initial support for the proposed object-to-sound interaction framework. The recognition experiment confirmed that predefined object properties could be reliably identified and used as the basis for sound

mapping, while the interaction experiment showed that users were generally able to complete the corresponding tangible interaction tasks. Subjective evaluation further indicated that participants could distinguish sounds generated by different objects and perceive certain object characteristics through sound. Together, these findings suggest that the system established a perceptible connection between object properties, sound representation, and user interaction. However, the current study mainly evaluated users' subjective perception of this relationship and did not directly test whether specific object properties could be identified through sound alone. Future studies could therefore introduce sound-only identification tasks and a wider range of objects to further evaluate the effectiveness of the proposed mapping approach.

CONCLUSION

This study investigated how the physical characteristics of objects and user actions can be transformed into auditory experiences that users can actively perceive and explore. To address this question, Contour Sound Chamber, an object-based tangible interactive music system, was developed. The system captures material, color temperature, and contour information through object recognition and spatial scanning, maps these properties to different musical elements, and integrates bodily actions into the interaction process. In this way, the system establishes a continuous process from object perception and sound generation to user interaction and auditory experience.

The user study evaluated the system in terms of object recognition, interaction task performance, and subjective perception. The results showed that the system could reliably recognize the predefined object categories, that participants were generally able to complete the designed interaction tasks, and that they could perceive differences between sounds generated by different objects as well as some of their physical characteristics. Together, these findings provide initial support for establishing a perceptible relationship between object properties, sound representation, and user interaction.

In a broader interaction design context, this study also draws attention to perceptual channels beyond vision. Contemporary interactive experiences are often dominated by visual information, while sound is frequently used as a supplementary channel for notification and feedback. This study explores a different role for sound by allowing it to participate in the representation of object information, enabling users to engage with physical objects not only through visual observation, but also through auditory perception and bodily interaction. Rather than replacing vision with sound, the proposed approach explores a shift from vision-centered interaction toward greater multisensory participation, while highlighting the potential of sound to connect users, physical objects, and digital information.

Future work will expand the range of object categories and sound mappings and introduce sound-only object-property identification tasks to more directly examine whether users can infer material, color temperature, and contour characteristics through auditory information. The approach may also be further explored in multisensory experiences, non-visual information representation, and other interactive contexts.

REFERENCES

Dourish, Paul. (2001). Where the Action Is: The Foundations of Embodied Interaction. Cambridge, MA: MIT Press.

Franinović, Karmen and Serafin, Stefania. (eds.) (2013). Sonic Interaction Design. Cambridge, MA: MIT Press.

Gaver, William W. (1986). "Auditory Icons: Using Sound in Computer Interfaces." Human–Computer Interaction, 2(2), pp. 167–177.

Hermann, Thomas, Hunt, Andy and Neuhoff, John G. (eds.) (2011). The Sonification Handbook. Berlin: Logos Verlag.

Hornecker, Eva and Buur, Jacob. (2006). "Getting a Grip on Tangible Interaction: A Framework on Physical Space and Social Interaction." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 437–446.

Ishii, Hiroshi and Ullmer, Brygg. (1997). "Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms." Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 234–241.

Jordà, Sergi. (2008). "On Stage: The Reactable and Other Musical Tangibles Go Real." International Journal of Arts and Technology, 1(3/4), pp. 268–287.

Kramer, Gregory, Walker, Bruce, Bonebright, Terri, Cook, Perry, Flowers, James, Miner, Nadine and Neuhoff, John. (1999). The Sonification Report: Status of the Field and Research Agenda. International Community for Auditory Display.

Zimmerman, John, Forlizzi, Jodi and Evenson, Shelley. (2007). "Research Through Design as a Method for Interaction Design Research in HCI." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 493–502.