

Gestural Input with Tangible Cubes on a Surface for Musical Expression: a Family of Experiments

NACERA LATRECHE, Université catholique de Louvain, Belgium

BERT SCHIETTECATTE, Noisetron LLC, USA

NUWAN T. ATTYGALLE, Université catholique de Louvain, Belgium

JEAN VANDERDONCKT, Université catholique de Louvain, Belgium

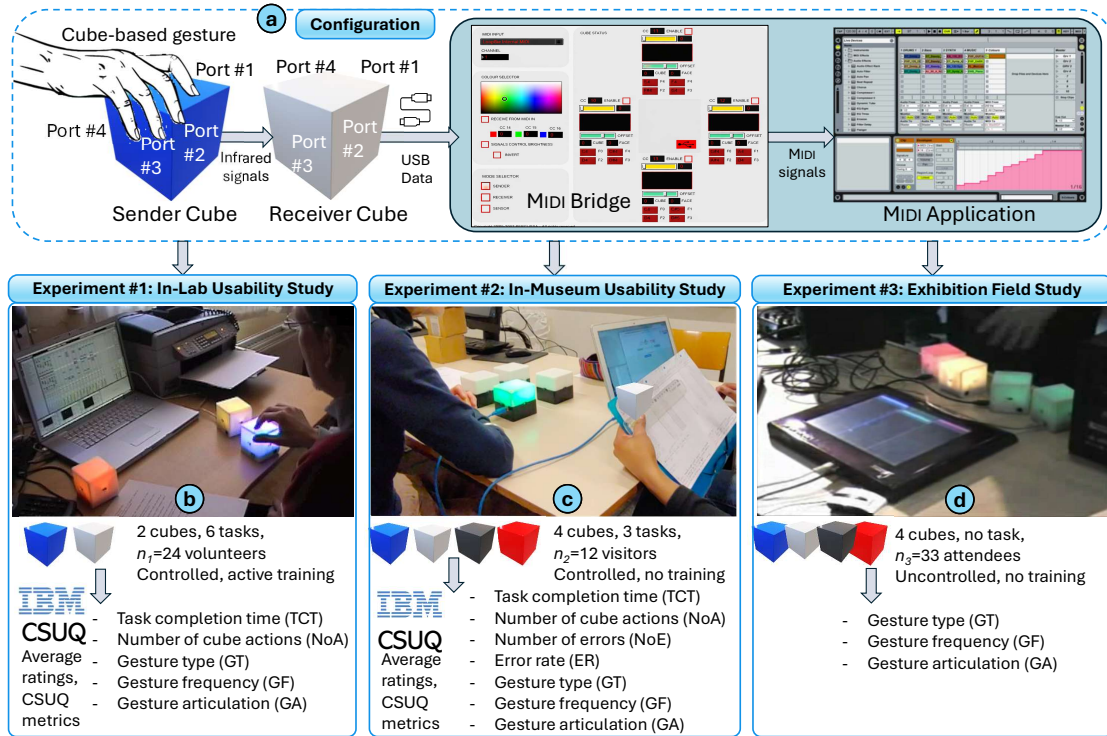


Fig. 1. Overview of our experiments for gestural input with tangible cubes on a surface– Configuration: a participant manipulates tangible cubes, which create infrared signals sent from source ports of a sender cube to the target ports of a receiver cube. Data are then sent via USB to MIDI BRIDGE connected to any MIDI-compliant application (a); Experiment #1: usability study controlled in a laboratory (b); Experiment #2: usability study controlled in a museum (c); Experiment #3: field study in an open exhibition (d).

Authors' Contact Information: Nacera Latreche, Université catholique de Louvain, Louvain Research Institute in Management and Organizations, Louvain-la-Neuve, Belgium, nacera.latreche@uclouvain.be; Bert Schiettecatte, Noisetron LLC, Percussa, Burlingame, California, USA, bert@noisetron.com; Nuwan T. Attygalle, Université catholique de Louvain, Louvain Research Institute in Management and Organizations, Louvain-la-Neuve, Belgium, nuwan.attygalle@uclouvain.be; Jean Vanderdonckt, Université catholique de Louvain, Louvain Research Institute in Management and Organizations, Louvain-la-Neuve, Belgium, jean.vanderdonckt@uclouvain.be.



This work is licensed under a Creative Commons Attribution 4.0 International License.

Cubes are increasingly used in tangible interaction, particularly when arranged in a topology on a surface such as a tabletop. To better understand how end users manipulate cubes as gestural input for tangible interaction on a surface, we report empirical findings from a family of three experiments conducted with AUDIOCUBES, a set of tangible cubes containing wireless computers with built-in sound and signal generating and processing capabilities that sense their topology for four action types on a surface: positioning, moving, rotating, and arranging. The two first experiments investigate the usability of AUDIOCUBES through a controlled study with and without an active training, inside and outside a laboratory. The third experiment determines which gestures are spontaneously produced by end users in an uncontrolled study and how they articulate these gestures by fingers. Tangible cubes offer a usable and intuitive medium for gestural interaction. Participants consistently favored face-based interactions over edges and corners, simpler one-to-one mappings over complex multi-object procedures, and physically stable arrangements over precision-dependent gestures. Usable cube interfaces should prioritize grasping, pushing, rotating, and aligning through cube faces, while minimizing reliance on corner-based or highly constrained actions. The actual composition of cubes covers only 32% of the theoretical space of 169 spatial compositions. Some practical implications for designing gestural input with tangible cubes on a surface, such as a tabletop, are offered.

CCS Concepts: • **Hardware** → **Sensors and actuators; Tactile and hand-based interfaces**; • **Human-centered computing** → *Interaction devices*; Laboratory experiments; *Graphical user interfaces*; **Gestural input**.

Additional Key Words and Phrases: Cubic object, Gesture-based user interfaces, Surface computing, Tangible user interaction.

ACM Reference Format:

Nacera Latreche, Bert Schietecatte, Nuwan T. Attygalle, and Jean Vanderdonckt. 2026. Gestural Input with Tangible Cubes on a Surface for Musical Expression: a Family of Experiments. *Proc. ACM Hum.-Comput. Interact.* 10, 8, Article ISS004 (November 2026), 36 pages. <https://doi.org/10.1145/3859316>

1 Introduction

Among *Tangible User Interfaces* (TUIs) [16, 47, 54], the cube is one of the most widely used physical forms [21, 27, 42, 45, 52, 62]. Cubic TUIs are physical objects with a cube-shaped form factor that serve as a device for various interaction modalities [29], such as multimodal input and gestural input. A syntactical query with the keywords “cube” AND “tangible user interface” performed on ACM Digital Library retrieves 264 references¹ belonging to different domains. Geometrically, the cube is one of the simplest symmetric polyhedra [49] known since Antiquity as one of the five Platonic solids, the five regular convex uniform polyhedra. Physically, the cube is one of the easiest objects for humans to grasp [54] and manipulate [10], because of all its symmetries that suggest many affordances [49].

From a cognitive psychology perspective, a cube evokes a natural manipulation that has been part of human habits and skills since childhood intuitively [23], cognitively [48], ecologically [46], and physically [42]. For example, GESTURE CUBE changes the current song by requiring minimal visual attention and cognitive load on the input device. From a Distributed Cognition [15] perspective, cubes can be handled by both hands of the same person (two-handed input [54]) or by the hands of different persons (multi-user input [51]) in a distributed fashion [35]. Different physical setups are possible, such as constructive assemblies [30].

From an interaction perspective, cubes support many types of interactive tasks in various domains of human activity, such as music creation [2], sound design [46], navigational tasks [57] in large datasets [23], a shopping menu [45], a 3D virtual model [19], exercises in eLearning [27], and storytelling with graphics and audio [67]. A cube is used for command input [24], data output [41], or data input/output [27]. The implementation of cubic TUIs has been technologically mastered for a number of years in a wide range of applications. Basic cube actions for gestural input, such as positioning, moving, rotating, and switching, have been known for years [47].

¹Query performed on February 6th, 2026, on <http://www.acm.org/dl>.

Table 1. Formulation of our research questions on gestural input with tangible cubes based on the Goal-Question-Metrics [4].

Goal	Purpose Issue Object Viewpoint	Investigate and explore gestural input through tangible cubes on a surface from the end user's viewpoint
Research question	RQ_1	How usable are tangible cubes for gestural input on a surface?
Metrics	M_1	Average rating of IBM CSUQ statements (Exp. #1 and #2), System Usefulness, Information Quality, Interaction quality, Overall usability (Exp. #1 and #2) to measure subjective satisfaction
	M_2	Task completion time and Number of cube actions (Exp. #1 and #2) to measure efficiency
	M_3	Number of errors and Error rate (Exp. #2) to measure effectiveness
Research question	RQ_2	What are the gestures performed with tangible cubes on a surface?
Metrics	M_4	Gesture type (Exp. #1, #2, and #3)
	M_5	Gesture frequency (Exp. #1, #2, and #3)
	M_6	Gesture articulation (Exp. #1, #2, and #3)

Still, we do not have a thorough knowledge about how people issue gestures corresponding to these actions. Only a relatively limited number of gestures have already been identified [49] and recognized as an important input source for interaction [64]. Moreover, there is no in-depth structural and usability analysis of the gestures produced by end users to manipulate cubes on a surface or in space. There is therefore a need to list and classify the gestures actually produced by users when they interact with cubes to ensure effective gesture recognition.

Following the Goal-Questions-Metrics (GQM) [4] framework, a hierarchical model that follows a top-down approach where a goal is first specified and associated with research questions collected and measured through metrics (Table 1), this paper makes the following contributions:

- Tangible cubes offer a usable and intuitive medium for gestural interaction based on the results of three experiments in descending order of control (Fig. 1).
- Face-based interactions are preferred over edges and corners, simpler one-to-one mappings over complex multi-object procedures, and physically stable arrangements over precision-dependent gestures. We should prioritize grasping, pushing, rotating, and aligning through cube faces, while minimizing reliance on corner-based or highly constrained actions.
- A heat map of how participants actually arrange cubes together that covers only 32% of the theoretical space of 169 spatial compositions.
- Some practical implications for designing gestural input with tangible cubes on a surface, such as a tabletop.

To this end, the remainder of this paper is structured as follows: Section 2 reviews some work pertaining to tangible interaction with cubes for gestural input. Section 3 presents methodological aspects that are common to the three experiments in our family, including a description of AUDIOCUBES, a set of distributed cubes equipped with sound and signal processing capabilities used in these experiments. Sections 4, 5, and 6 present the methodological aspects that are specific to each experiment, respectively, report their results, discuss them, and suggest some implications for designing gestural input with tangible cubes on a surface. Section 7 discusses the limitations of our work in terms of threats to validity. Section 8 concludes the paper by summarizing our findings and discussing some future avenues of this work.

Table 2. Sheridan et al. [49] non-verbal dynamics of a cube.

Action	Description	Event
Rotate	To turn the cube about an axis or center in a continuous, fluid motion, exposing three sides of the cube at one time	Turn + (exposed sides) + (speed)
Twist	To rotate the cube while taking a curving path or direction using the wrist	Turn + curve path
Turn	To cause the cube to move around an axis or a center, exposing one side at a time	Turn (right, left, up, down)
Flip	One fluid movement to cause the cube to turn over to expose the opposite side of the cube	Flip (top vs. bottom), Flip (front vs. back), Flip (side vs. side)
Hold	To have or maintain cube in the grasp	Hold (no movement)
Place	To put the cube in or as if in a particular place or position	Place hold, place
Pick up	To take hold of and lift up	Press (up)

2 Related Work

This research touches on three research topics, *i.e.*, tangible user interfaces in general and in particular with cubes, and gestural input with cubes, which are discussed in the next three subsections.

2.1 Tangible User Interfaces

Tangible User Interfaces [47] were first distinguished from other interfaces by Fitzmaurice [17] and called *Graspable User Interfaces*, which are defined as a physical handle to a virtual function where the physical handle serves as a dedicated functional manipulator [17]. TUIs provide a physical form to digital information and computation to facilitate the direct manipulation of data. Physical devices can be walls, building blocks, instruments, water flow, and light. The devices involved in TUIs should support and enhance people's natural abilities. For example, a device in the form of a phone could materialize the phone call service on a computer so that the end user does not feel unfamiliar. Learning to use the TUI should be easier because a TUI is a natural interface that supports novice users, works effectively for people doing things better with their hands and objects, and supports multi-user gestural input.

2.2 Tangible Interaction with Cubes

Lefeuvre et al. [29] classified cubic TUIs by analyzing 47 examples to explain the widespread use of cubes in interaction design, highlighting their design properties that resonate for our experiments: manipulation as input (*e.g.*, our cubes can be manipulated individually and independently), spatial placement (*e.g.*, our cubes can be placed on a surface), arrangement (*e.g.*, cubes can be arranged in a meaningful topology), multifunctionality (*e.g.*, multiple functions or options can be associated to a cube and its faces), randomness (*e.g.*, our cubes can be randomly manipulated), physical qualities (*e.g.*, our cubes are made of acrylic plastic), and as containers or pedestals for output. Since these design properties are applicable to any cubic interface in general and therefore to our cubes, we wish to investigate them further in our case and refine them.

Cubic TUIs are particularly appropriate for interaction on a surface as they combine maneuverability (they are easy to manipulate and move around), stability (they remain in place during interactions, such as on a tabletop), and scalability (they scale up/down to represent different data dimensions) [21, 47]. For example, sound and music design are creative tasks in which the musician or artist directly manipulates a musical instrument. Even



Fig. 2. A disk jockey executing a musical performance in front of an audience by manipulating cubes (Source: Mark Mosher.)

when electronic musical instruments became more popular and more available to musicians [3, 5], artists still rely on a direct interaction with their synthesizers and keyboards through the use of tangible sliders, knobs, buttons, and displays. When personal computers became affordable and widespread, they were introduced in most recording studios and used by musicians and artists for composition. Since then, desktop computers have become powerful enough to run real-time sound processing software, and companies have started to design and sell professional audio software that would run on any standard computer with an audio interface. More recently, replacing hardware equipment with physical buttons for sound control with software has become essential for mastering their complexity [46].

Most musicians still use electronic musical instruments; remembering the function of each button or the settings they control, one by one, remains a real challenge. AUDIOPAD [38, 39] and REACTABLE [24] were developed to support the expressiveness, legibility, and visibility of musicians by allowing them to create the sound by tracking the positions of TUIs on a tabletop or in space. The objects play the roles of generating and modifying sound or controlling sound parameters. The functions of the cubic TUIs and the connections between them and the real waveforms of the sound are made visible, thereby providing musicians with expressiveness and ease of use. The principal difference between the REACTABLE and the AUDIOPAD is that the REACTABLE uses modular synthesis while the AUDIOPAD uses loop-based synthesis. In AUDIOPAD, electromagnetic pucks are associated with different functions: once the musician has turned these objects on a surface, this action is transferred to the musical instrument that replicates the functions of the objects. Fig. 2 features a disk jockey executing a musical performance by manipulating cubes in the air that are connected to musical parameters, such as sound effects.

For these reasons, cubic TUIs have been extensively investigated in music and sound design [14], but not exclusively. MUSICCUBE [2] makes digital music tangible on a tabletop by controlling its access via cubic TUIs, like MUSICTABLE [6], BLOCKJAM [36], and REACTABLE [24]. AUDIOCUBES [46] alleviate the problems posed by GUIs for sound manipulation by replicating their physical counterpart devices [22]. AUDIOCUBES consist of a set of cubes based on their interaction range to users creating sounds. Augmented Groove [43] combines cubic TUIs

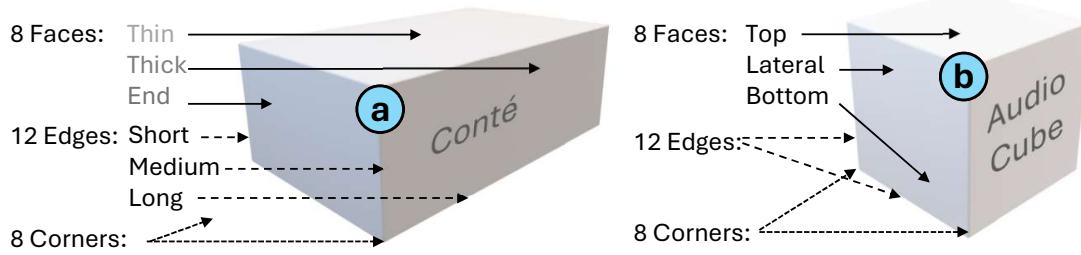


Fig. 3. Comparison of Conté [61] (a) against AUDIOCUBE (b) in terms of types of contact points and number of contact points.

with augmented reality and a 3D user interface to let users play music with or without any musical instrument. Tangible AudioGame Development Kit [55] allows audio game prototyping by placing and combining cubes.

Cubic TUIs are exploited for navigation in ubiquitous 2D/3D spaces, such as for visualizing 3D spatio-temporal data in mixed reality [21]. Since 1980, Bolt attached SPACESENSINGCUBES to his wrist to track the 3D position of his arm in the “Put-That-There” system while navigating through space in a multimodal way [7]. The CUBICMOUSE [19] is an input device for specifying any position in a 3D space by controlling its position and orientation in a virtual model. With INTERCUBE [45], the end users navigate a 3D shopping menu by rotating the cubes in any direction. Beyond navigation, Navigational Blocks [9] are arranged based on their position, orientation, and movement to produce narrative storytelling in a museum. Kranz et al. [27] augmented a cubic TUI with input capabilities, visual, and auditive feedback in the form of displays and speakers integrated in the DISPLAYCUBE. This platform deploys learning applications based on a cubic TUI, which is reported as intuitive and playful for quizzes, multiple choices, and tests for children.

When many cubes are involved in a setup, cultural and social aspects arise [27]: Computer-Supported Collaborative Work (CSCW) and Learning (CSCL) are promising areas of application by enabling end users to collaboratively manipulate several cubes for a shared goal, like in ACTIVECUBE [62], for writing an algorithm in ALGOBLOCKS [51], for learning in CUBETOLEARN [52] or to evaluate cognitive abilities for different tasks, such as spatial organization [37], thanks to a correlation with the cube [48].

Other areas of application also include, but are not limited to, monitoring large-scale systems [23], producing a 3D cube structure with a bijection between the TUI world and its representation in a GUI 2D world [26], architectural education with CUBEEXPLORER [50], and dual-arm manipulation for remote operation of robots [25]. This last study developed a cube as an instance of a Target-Object-Oriented User Interface (TOOUI) that enables the operator to express the orientation and contact state of the target object rather than the robot effectors. Like this cube, our study involve cubes enabling the selection of parameters and objects independently, but, unlike this study, our work targets gestures performed while manipulating cubes and arranging them on a surface.

2.3 Gestural Input with Cubes

Many existing cubic TUIs enable end users to issue gestures, but not always: manipulation as input [29] is neglected in favor of maximization of spatial placement or arrangement. For example, end users move around an ACTIVECUBE [26] in a 3D virtual world by assembling cubes into an arrangement to be recognized, but no gestures are captured. Only the arrangement is used as input [30]. Sheridan et al. [49] identified seventeen cube actions referred to as *non-verbal dynamics*. Participants proposed intuitive gestures for meaningful actions using a fictive wooden cube (Table 2): turning (86%), rolling (58%), shaking (49%), throwing (35%), and shifting (30%). Participants rated by importance these actions for controlling home appliances, computer applications, or games: turning (87%), shifting (54%), positioning (36%), orientation (32%), and shaking (28%). This study produced a first classification of cube actions according to (Table 2): action (manipulation), description of the action, and

events (*i.e.*, the type of action particular to that manipulation). For example, the “rotate” action is a turn event that depends on the speed of turning and the number of exposed sides. Like this study, our work is aimed at exploring a larger range of gestures on a cube, but, unlike this study, our work is expected to consider a larger scope of such gestures and analyze their usability in various contexts of use.

Yin and Liu [66] investigated how a cube can be flipped toward reaching a particular face: different behaviors for flipping cube faces emerged depending on the feedback, there was no significant difference between the dominant and the non-dominant hands; flipping times ranged between 1.2 sec and 6.8 sec., no single strategy to perform actions emerged. Cohé and Hachet [12]’ study is the closest to this paper in terms of gesture types: 16 users produced a dataset of 432 gestures for manipulating a cube in virtual reality using touch input. The cube is represented through three points of view for rotation, scaling, and translation (RST) [34]. Like this study, our work characterizes gestures by their initial point location (*i.e.*, face, edge, corner, or external reference), form (*i.e.*, one static finger, one static finger combined with one path finger, one path finger, two path fingers, four path fingers), and trajectory (*i.e.*, collinear to the transformation direction or not, with an intersection or not). Unlike this study conducted with a 2D touchscreen displaying a virtual cube, our study is conducted with a genuine cubic TUI as we believe that a 3D manipulation of a real cubic TUI is different than a 2D manipulation of a 3D cube represented in virtual reality.

Sheridan et al. [49] considered various cube materials and shape-changing ones, which were not be considered in the remainder of this paper: our cubic TUIs were not be subject to any physical deformation and influenced by their material. Consequently, actions assuming a shape-changing behavior [11] were considered out of scope. Instead, any gesture performed with a tangible cube on a surface was considered relevant: for example, roll (*i.e.*, impelling a cube forward by causing it to turn over and over on a surface), throw (*i.e.*, to propel a cube in space by a forward motion of the hand), spin (*i.e.*, to revolve a cube on a surface in a fast, fluid movement where all sides are exposed quickly), shake (*i.e.*, sharp, fluid movements up and down or side to side), squeeze [60] (*i.e.*, exert strong pressure on a cube with hands or fingers), press, tap, rub, and fiddle.

A closely related study is Conté [61], a multimodal rectangular prism used for gestural input inspired by an artist’s crayon (Fig. 3-a), while our work considers cubes like play bricks [3]. As a result, Conté’s gestures are subtle and complex, as they are applied according to the corners, edges, and faces of a cube in contact with a surface, whereas AUDIOCUBES’s gestures involve manipulating cubes as physical objects, with the resulting gestures being more coarse, remaining in contact with the surface (Fig. 3-b). Conté shares several characteristics with AUDIOCUBES (Table 3), such as the same number of contact points (26) and the ability to detect input from various gestures. However, these gestures are exploited for different purposes. Where Conté mimics an artist’s crayon for surface positioning, drawing, annotating, and selecting modes (*e.g.*, tilting the cube allows drawing with a fine point, while an edge allows drawing with broader strokes), AUDIOCUBE senses gestures to map them to basic action types only: positioning, moving, rotating, and arranging them in a topology. The granularity of Conté’s gestures is therefore fine-grained as opposed to gross-grained for AUDIOCUBES. To map gestural input to commands, Conté offers (Fig. 3) more types of contact types (7 instead of 4) to support fine-grained commands such as drawing, annotating, and pen-like input.

Let us now contextualize Conté and AUDIOCUBES according to Lefeuvre et al. [29]’s design properties for cubic TUIs (see Harvey Balls indicating the level of support in Table 3): input manipulation is Conté’s strongest point, whereas AUDIOCUBES only support gross-grained gestures; spatial placement is very precise in Conté and more moderate in AUDIOCUBE; the arrangement of several cubes and their togetherness are not the goal of Conté, whereas they are in AUDIOCUBE; the range of functions associated with a Conté cube is broader than with AUDIOCUBES; both allow the cube to be moved completely freely; the physical qualities of Conté are superior to those of AUDIOCUBES, as they perform all the functions of an artist’s stylus as opposed to bricks. By computer vision, Conté interprets gestures by their contact points on the surface, while this work senses directly applied cube-based gestures that are then transformed into infrared signals (Fig. 1-a).

Table 3. Comparison between Conté [61] and AUDIOCUBES used in our studies.

Characteristic	Conté	AUDIOCUBE
Metaphor	Artist's crayon	Play cube
Geometry	Rectangular prism	Cuboid
Number of contact point types	7	4
Number of contact points	26	26
Interaction	Multimodal	Tangible
Number of users	1	$1 \dots n$
Number of objects per user	1	≥ 1
Surface contact	Permanent	Intermittent
0D gestures	X-Y Positioning	X-Y Positioning
1D gestures	Attributes commands	Linear gestures
2D gestures	Contextual commands, pen-like input	Surface gestures
Gesture recognition	Image processing	X, Y, Z, α , t sensing
Gesture granularity	Fine-grained	Gross-grained
Gesture complexity	High	Low
Manipulation as input	●	○
Spatial placement	●	●
Arrangement	○	●
Multifunctionality	●	○
Randomness	○	○
Togetherness	○	●
Physical qualities	●	○

2.4 Summary

The increasing number of works devoted to the cube as a tangible object in general, and as a source of gestural input in particular, demonstrates the growing interest in this kind of interaction modality due to the cube's engaging properties. These works cover a wide spectrum of application domains and a highly variable construction space, ranging from non-interactive cubes captured by sensors to cubes equipped with screens on each face to display feedback. Most of these works concern a specific application of the cubes, optimally designed for a particular use, such as the visualization of 3D data or the manipulation of musical parameters. While the gestures implemented in these works are suited to their domain, they often remain specific and do not consider the whole space of gestures that could be produced with one or more cubes. What is needed is a broader study of the range of gestures associated with cubes, and a more in-depth investigation of how end-users can articulate these gestures on cubes, particularly when they are arranged on a surface. We did not identify any study that explores gestural input with tangible cubes on a surface in a wider way, apart from elicitation studies [56], which pursue the goal of creating a consensus set. In the realm of cubic TUIs, there is a need to explore this gestural input when action types, *i.e.*, positioning a cube, moving a cube, rotating a cube, and arranging them, are the main manipulation for input.

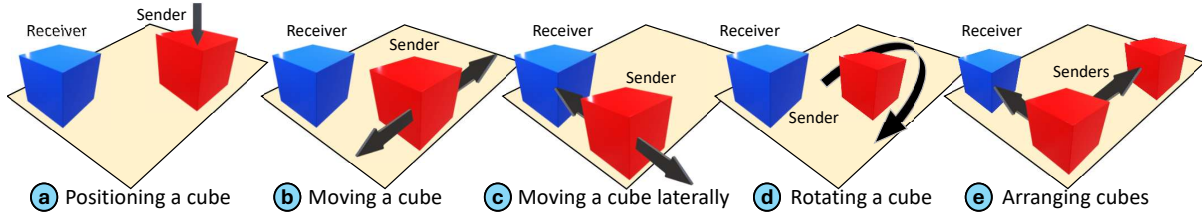


Fig. 4. The action types with AUDIOCUBES: positioning a Sender cube with respect to a Receiver cube (a), moving a Sender cube around a Receiver cube (b), moving a Sender cube laterally aligned with the Receiver (c), rotating a Sender cube on itself (d), and arranging Sender cubes together in a topology sensed by a Receiver (e).

3 Common Methodological Aspects

This section presents methodological aspects common to the three experiments conducted in this work. Specific details about each experiment and the corresponding results are given and discussed in Sections 4, 5, and 6, respectively. To collectively address the research questions (Table 1), the experiments were conducted in descending order of control to cover multiple use cases of musical expression (Fig. 1-b, c, d): the experiment #1 is a controlled usability study performed in a usability laboratory asking participants to perform 6 tasks with 2 cubes after an active training² (Section 4, Fig. 1-a); the experiment #2 invites museum visitors to perform 3 tasks with four cubes without any training (Section 5, Fig. 1-b), and the experiment #3 involves visitors of an on-site exhibition freely operating 4 cubes freely (Section 6, Fig. 1-c).

3.1 Apparatus

We introduce AUDIOCUBES, a cubic TUI that supports four types of action (Fig. 4): positioning, moving, rotating, and arranging them in topology at run-time.

3.1.1 Hardware and Software. An AUDIOCUBE consists of a wireless, battery-powered computer embedded in a plastic cube with built-in sound and general-purpose signal-generating and processing capabilities. There is only one RGB LED per cube to emit continuous visual feedback. An AUDIOCUBE is a small cube that contains a digital signal processor to receive and send signals from and to other AUDIOCUBES. All AUDIOCUBE have been invented and created by Bert Schiettecatte and are the intellectual property of Percussa, a professional audio consumer electronics brand of Noisetron LLC. The Receiver AUDIOCUBE communicates with a computer via a USB cable to send the raw data received from another AUDIOCUBE, called Sender (Fig. 1-a). Each AUDIOCUBE senses the

²An active training means a continuous and consistently applied program of tasks to enable a participant to become efficient, to function as independently as possible from the setup, and to maintain the best abilities possible [13].

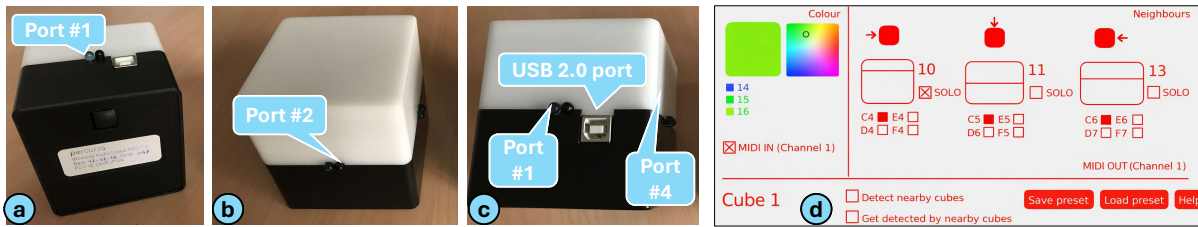


Fig. 5. Ports of a AUDIOCUBE: communication ports (a–c), USB 2.0 port (c), and cube selection (d).

location and orientation of each other and the distance to hands, fingers, and other objects. Each side, except the bottom and top sides, has communication ports (Fig. 5-a–c) capable of transmitting and receiving digital or analog signals, using infrared signals, to and from other AUDIOCUBES nearby. They can be arranged to form a topology that can be detected and recognized. AUDIOCUBES communicate through these ports and exchange sound or data signals depending on the application running.

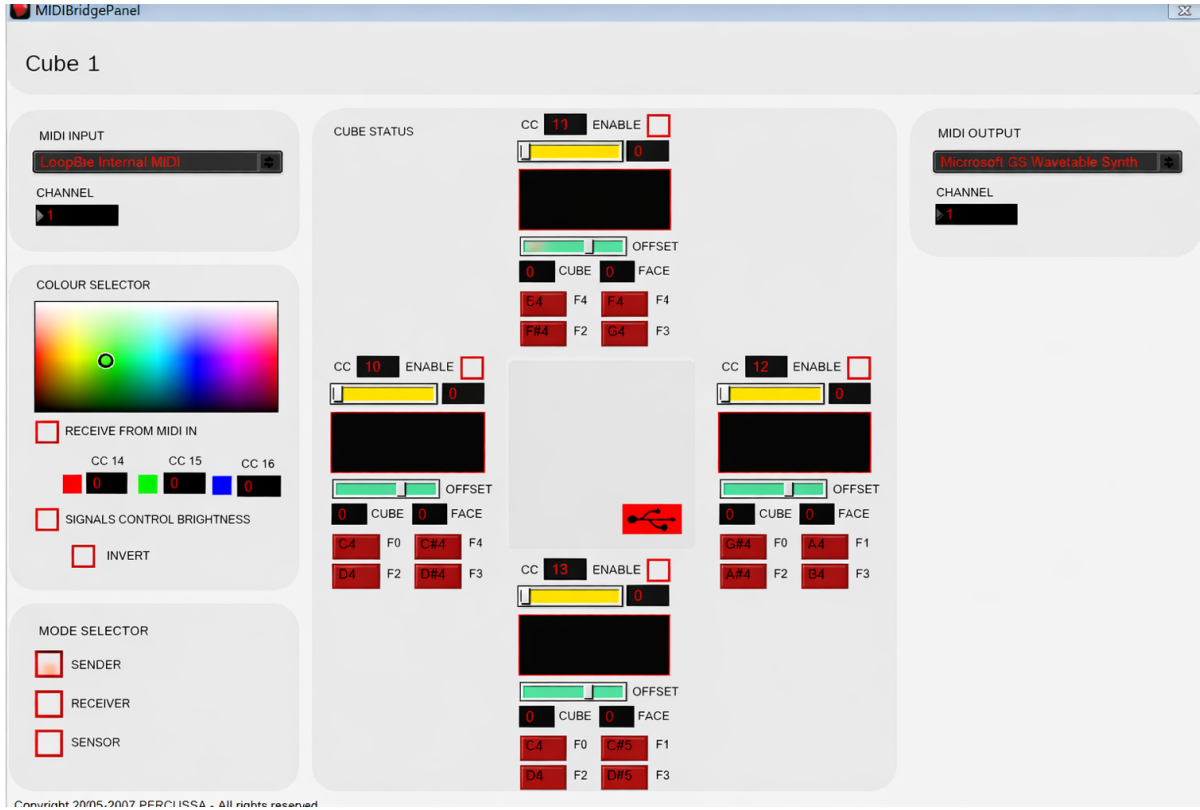
Each AUDIOCUBE is equipped with a $\frac{1}{4}$ inch input and output jack for processing or generating sounds and a USB 2.0 port (Fig. 5-c) to connect to a computer for the following functions: charge the battery, download and run a software “plugin”, update the firmware, change software parameters, change the configuration, change its color, and load a topology. When these functions are completed, the AUDIOCUBE can be disconnected and run.

To control an application running on a computer using gestural input, the MIDI BRIDGE (Fig. 6) acquires data from a Receiver AUDIOCUBE and transforms them into MIDI signals that can be understood by any MIDI-compatible device or application (Fig. 1-a). The configuration with any third-party software is performed by connecting one AUDIOCUBE to a computer via USB or wirelessly, and using a second unconnected one to send messages to the connected one. The connected cube detects the unconnected cube and reads its identification number and the number of the face that is pointed at the connected cube. This information then controls any MIDI device, for example, to play back a sound. Different modes are used in the MIDI BRIDGE to generate two types of MIDI information: (i) MIDI note-on messages (to start sounds in other software or MIDI devices), and (ii) MIDI continuous controllers (to change parameters in the sound). AUDIOCUBES operate in one of three modes:

- (1) *Sender-Receiver mode*. A cube plays the Sender and the other plays the Receiver, which transmits to the software the number on the face of the Sender and the number of the face perceiving the Sender.
- (2) *Sensor mode*. The cube transmits the level of obstruction of each face. The Sender and receiver modes are used together (one wireless cube as Sender, one wired cube as Receiver) to generate the data, while the Sensor mode is used to generate MIDI continuous controllers (one wired cube).
- (3) *Topology mode*. The cubes communicate with each other depending on the arrangement.

3.1.2 Positioning and Arranging AUDIOCUBES in a Topology. AUDIOCUBES [46] detect their relative position in terms of angles based on interaction range. Instead, AUDIOCUBES characterize a topology based on how the cubes are arranged together. Allen [1] identified thirteen interval relationships of two types: sequential (before) and simultaneous (equal, meets, overlaps, during, starts, or finishes). Each basic relationship has an inverse, except for the equal relationship, which is symmetric. Although Allen initially introduced the relationships to characterize uni-dimensional temporal intervals, they can be generalized to n dimensions to express similar constraints [58]. The 2D generalization [53] enables us to define the arrangement of two cubes in a topology by *Arrangement* ($C1, C2$) = (R_i, R_j), where $i, j \in \{1, \dots, 13\}$, R_i is the identifier of the relationship between $C1$ and $C2$ according to the X axis and R_j is the identifier of the relationship according to the Y axis. Fig. 7 shows the $13^2=169$ arrangements of two cubes where the two first lines (i.e., $R_{1,1}-R_{13,1}$ and $R_{1,2}-R_{13,2}$) and columns (i.e., $R_{1,1}-R_{1,13}$ and $R_{2,1}-R_{2,13}$), and the two last lines express arrangements without overlapping. They can be separated from each other (first and last lines and columns, represented in Fig. 7). Otherwise, the cubes touch each other or share one part of the convex envelope. The abscissa of Fig. 7-b characterizes the horizontal arrangement, while its ordinate represents the vertical arrangement (Fig. 7-c). A topology is then defined as a series of spatial arrangements between cubes.

3.1.3 Rotating a AUDIOCUBE. A cube has 13 symmetry axes: four diagonals, three axes that join the centers of opposite faces, and six axes that join the midpoints of two opposite edges of the cube. There are two rotations for each diagonal (by an angle of 120° or 240°), different from a non-rotation, bringing the cube into coincidence with itself. There are eight rotations around diagonals. For each axis joining the centers of two opposite faces, there are three different rotations from the non-rotation (by an angle of 90° , 180° , or 270°). Thus, three axes give nine rotations. A cube can be rotated 180° around any axis joining the midpoints of two opposite edges, leading



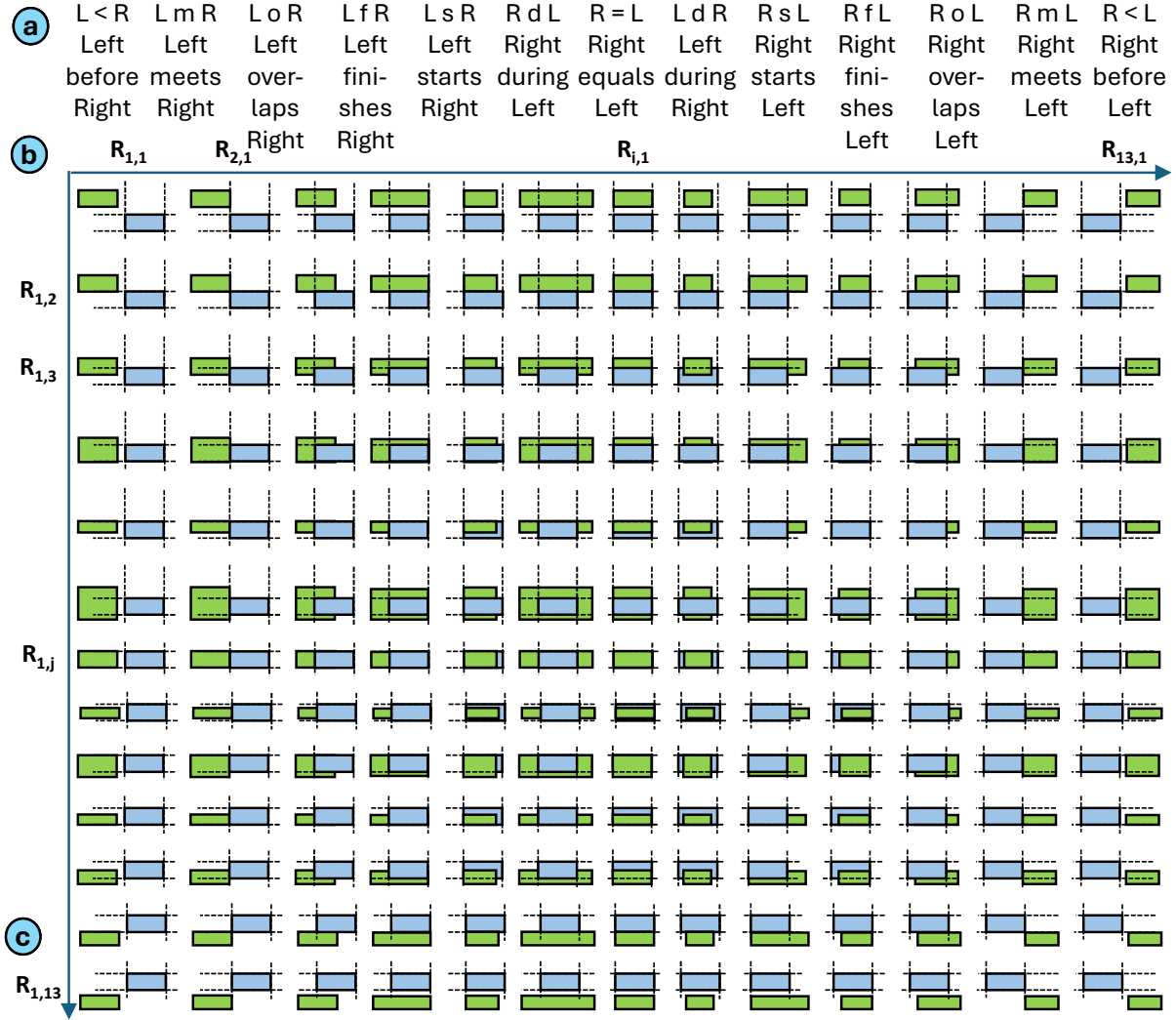


Fig. 7. The theoretical space for arranging two cubes based on Allen temporal relationships [1] and an associated notation [20] (a): the abscissa represents the horizontal arrangement (b) and the ordinate represents the vertical arrangement (c).

3.3 Metrics

Table 4 defines the metrics used in the three experiments. To evaluate the three usability factors defined by ISO 9241–11 [18] for RQ_1 (Table 4), two metrics measured *efficiency* (M_2 : the task completion time and the number of cube actions per task) and two metrics measured *effectiveness* (M_3 : the number of errors and the error rate). To evaluate the *subjective satisfaction*, we used the IBM Computer Satisfaction Usability Questionnaire (CSUQ)

Table 4. Definition of metrics used in the experiments.

ID	Metric (Abbreviation)	Definition
M_1	Average rating of each IBM CSUQ item $Q_1 - Q_{19}$	Mean of item responses, where higher scores indicate better perceived usability (1=strongly disagree to 7= strongly agree): $AR(Q_j) = \frac{\sum_{i=1}^n R(Q_{ij})}{n} \forall i = 1, \dots, n, \forall j = 1, \dots, 19$ where n = number of participants, $R(Q_{ij})$ is the response of the i^{th} participant to the j^{th} question, $\in \{1, \dots, 7\}$
	System Usefulness (SYSUSE)	$\sum_{j=1}^8 AR(Q_j)$
	Information Quality (INFOQUAL)	$\sum_{j=9}^{15} AR(Q_j)$
	Interaction Quality (INTERQUAL)	$\sum_{j=16}^{18} AR(Q_j)$
	Overall usability (OVERALL)	$\sum_{j=1}^{19} AR(Q_j)$
M_2	Task completion time (TCT)	Time spent executing a devised task in sec.
	Number of cube actions per task (NoA)	Total number of cube actions required for users to complete a task
M_3	Number of errors per task (NoE)	Total number of errors performed by users to complete a task
	Error rate	Ratio of the number of errors to the total number of actions = NoE / NoA
M_4	Gesture type (GT)	Type of cube gesture performed by the participant (e.g., grasp, touch, rotate, lift) [49], along with the number of hands involved (i.e., 1 or 2) [40], the number of cubes involved (i.e., 1 or 2), and their direction (i.e., linear up/down, left/right, rotational) [64]. This metric was captured in terms of occurrence.
M_5	Gesture frequency (GF)	Percentage of cube gestures in terms of four action types (i.e., positioning, moving, rotating, and arranging)
M_6	Gesture articulation (GA)	Percentage of manipulating cube gesture in terms of type (e.g., grasp, push), location (i.e., corner, edge, face) [61], number of cubes (i.e., 1 or 2), combination (i.e., stacked vs. unstacked), and arrangement relations (Fig. 7)

[31]³ and its related metrics (M_1 : the average rating of each item, the system usefulness, the information quality, the interaction quality, and the overall usability. To address RQ_2 , all actions were characterized according to the gesture type (M_4), the gesture frequency (M_5), and the gesture articulation (M_6).

4 Experiment #1: In-lab Usability Study

This experiment involves six predefined cube tasks with specific conditions to be performed as a controlled study in a usability laboratory. Prior to carrying out the experiment, participants enrolled in a 10-min active training program, where they were instructed to perform all action types repeatedly to develop their efficiency.

³This 19-statement questionnaire is selected because it has been empirically validated with a large number of participants on a significant set of applications [32], it is applicable to any interactive system, and it benefits from a proven $\alpha=0.89$ reliability coefficient between its results and the perceived usability [32]. Each CSUQ statement [31] is a closed question measured using a 7-point Likert [33] scale (1=strongly disagree, 2=largely disagree, 3=disagree, 4=neutral, 5=agree, 6=largely agree, 7=strongly agree)

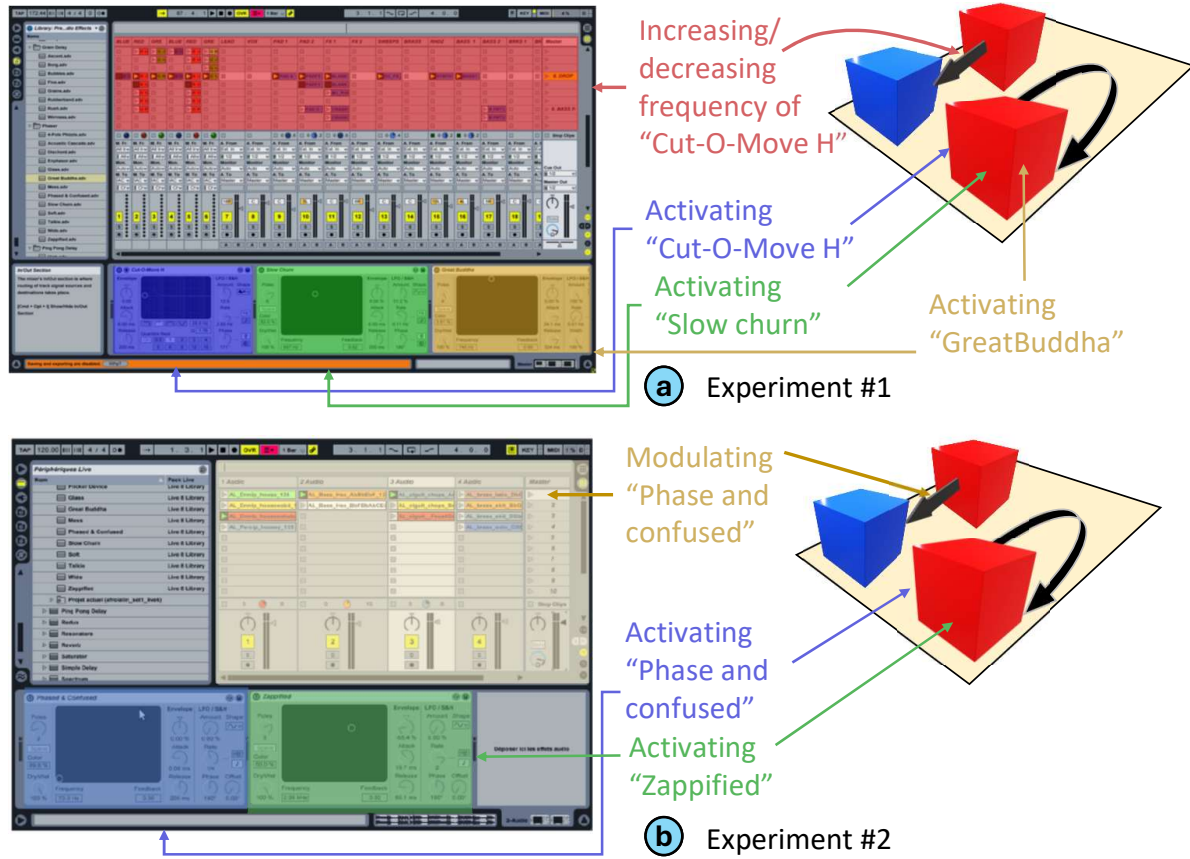


Fig. 8. Configuration of AudioCubes in Ableton Live for the experiments #1 (a) and #2 (b).

4.1 Method Specific to Experiment #1

4.1.1 Apparatus. We configured Ableton™ Live V9.6.2⁴, a Digital Audio Workstation (DAW) [5, 44] with the 14 tracks included in the software installation program and three commonly used effects (Fig. 8-a): "Great Buddha", "Slow Churn", and "Cut-o-move h". Ableton Live was selected over other DAWs for several reasons: its user interface consists of a single session view, making it appropriate for live performance and non-linear music production, such as for pop production (as in [44]), for ambient, electronic, and experimental music. This is very convenient for all types of our participants, who are not necessarily music or sound experts. Its flexible workflow, powerful MIDI sequencing capabilities, and built-in instruments and effects contribute to its wide use, including its direct interface with AUDIOCUBES (Fig. 1-a).

One AUDIOCUBE was configured to change the track ID that is played according to the Sender-Receiver mode and a second manages the three effects. The track was played according to the relative position of these two cubes. Both cubes form a coplanar arrangement on a tabletop. We configured three effects on the last cube using four lateral sensors of the cube: three to operate the effects by rotation, and one to change the parameters of an

⁴See <https://www.ableton.com/en/live/>

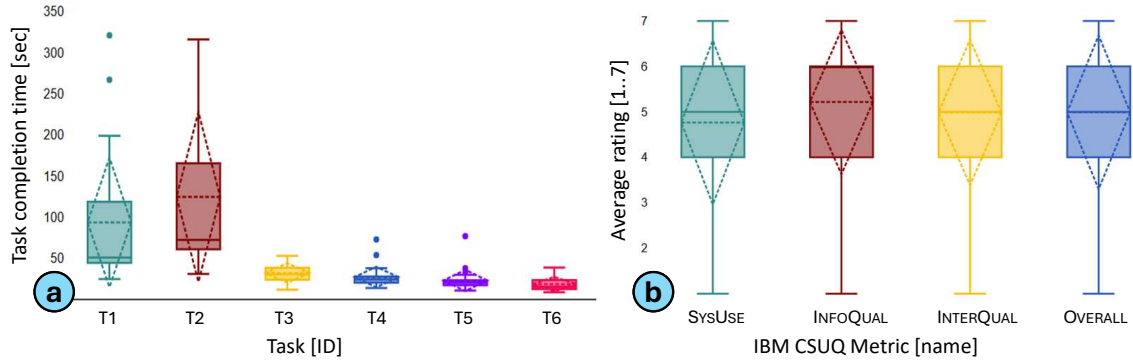


Fig. 9. Results of the first experiment: Task completion time (a), CSUQ metrics (b). Error bars show a 95% confidence interval.

effect by moving and positioning. The effect “Cut-o-move h” is selected for its control capability: a sensor is used to vary its frequency, changing the sound distortion.

4.1.2 Tasks. We devised six tasks: two tasks involve the transition from one track to another in a coordinated manner to check whether participants experienced any trouble in changing tracks when arriving at the number of the track they wanted; four tasks activate effects and the modification of a parameter to analyze changing tracks or modifying effects of, or both. We instructed participants to browse four tracks in a precise order without any mistake within a time limit. If a participant made more than 5 errors, we told them to continue the task anyway, thus allowing a complete analysis:

- T₁. Put the tracks in ascending order from 1 to 14. The participant should realize that track 7 was not working and switch from track 6 to track 8.
- T₂. Select the following tracks in the correct order without using additional tracks: 2, 8, 11, 14. If you make a mistake, start again from track 2.
- T₃. Select track 10 and then the effect “Great Buddha”.
- T₄. Select track 8 and select the effects “Great Buddha” and “Slow Churn”. Feel free to alternate them.
- T₅. Select Track 2 and select the effect “Cut-O-Move H” and vary the frequency.
- T₆. Choose a track by activating an effect while changing the track.

4.1.3 Procedure. Participants were given a free period of 5 minutes to become familiar with the setup before performing the six tasks with a time limit of 3 minutes and 1 minute for T_{1,2} and T_{3–6}, respectively. At the end of the experiment, the user was invited to complete the questionnaire plus two optional open questions for constructive feedback (*i.e.*, describe three positive aspects and three shortcomings of AUDIOCUBE).

4.1.4 Participants. 24 subjects ($n_1=24$, 12 women, 12 men) aged between 12 and 57 years ($M=28.08$ y, $SD=13.57$), were recruited outside our organization through a mailing list of volunteers with non-technical backgrounds. All were regular computer users ($M=5.91$, $SD=1.71$ on a 1-7 scale – “I use a computer frequently”: 1=strongly disagree to 7=strongly agree), with easiness ($M=5.20$, $SD=1.53$ – “I can use a computer with ease”: 1=strongly disagree to 7=strongly agree) without any prior exposure to TUI. We did not impose any previous knowledge or experience in musical expression.

4.2 Results and Discussion

Fig. 9-a shows the task completion time for the six tasks. T_6 , the most elementary task for selecting a track and an effect, required the shortest time ($M=18.75$ sec., $SD=8.69$ sec.), is significantly faster than T_3 (a pairwise Mann-Whitney exact test gives $U=102$, $p \leq .001^{***}$) and T_4 ($U=141$, $p=.002^{**}$). Indeed, selecting a predefined track and an effect takes longer than the same operations if not predefined. However, T_1 ($M=93.71$ sec., $SD=78.69$ sec.) involved 13 consecutive selections, which gives an average time of 7.21 sec per selection. Repeating the same selection consecutively therefore reduces the completion time by a factor of three, on average. T_5 ($M=23.17$ sec., $SD=12.95$ sec.) is also significantly faster than T_3 ($U=136.5$, $p=.0013^{**}$). The slowest tasks, T_1 and T_2 ($M=124.71$ sec., $SD=103.23$ sec.), are those related to the transition of tracks in a precise order without making any mistake. While the logical order of tracks varies from one candidate to another, finding a particular track or activating an effect does not induce any particular difficulty.

Fig. 10 shows the distribution of participants' answers to the usability questionnaire. Its internal consistency benefits from an excellent reliability (Cronbach's $\alpha=0.94$) and a slight level of agreement among participants (Kendall's coefficient of concordance $W=0.08$, $df=18$, $p=0.015^*$). In this figure, each divergent bar is scaled to five levels to facilitate the interpretation: a score between 6 and 7 is considered as "I strongly agree", a score of 5 as "Agree", a score of 4 as "I neither agree or disagree", a score of 3 is considered as "I disagree", and a score between 1 and 2 is considered as "I strongly Disagree". In general, a score between the median ($Mdn=4$) and the "Strongly agree" should not raise any particular concern regarding the statement, whereas a score between "Disagree" and "Strongly Disagree" should raise some discussion. All statements are averaged above the median, which is positive. In addition, we performed a two-tailed t test for each statement to check whether its distribution is significantly higher than the median. Eleven statements out of nineteen are significantly higher than the median; for example, Q_1 ($M=5.08$, $SD=1.29$, $p < .001^{***}$) is significantly higher while Q_4 is not ($M=4.21$, $SD=1.94$, $p=.61$, $n.s.$). In general, any statement having an average score of 5 with a standard deviation less than or equal to 1 is considered positive, if not optimal. Q_1 , Q_2 , Q_3 , Q_7 , Q_9 , Q_{11} to Q_{17} , and Q_{19} are all averaged above 5 but their deviation is always higher than 1.

Fig. 10 shows that the participants reported a generally high level of subjective satisfaction, reflected in an overall usability mean score of $M=4.99$ ($SD=1.69$). Users found the AUDIOCUBES easy to use ($M=5.08$) and easy to learn ($M=5.00$), suggesting a relatively low learning curve. The ability to effectively complete tasks received a strong rating ($M=5.42$), indicating that AUDIOCUBES supports users in accomplishing their intended work. Information-related aspects of the AUDIOCUBES were particularly well rated. Items related to clarity, understandability, and accessibility of information achieved some of the highest mean scores, including ease of understanding ($M=5.67$) and ease of finding needed information ($M=5.54$). TUI-related items were also positively perceived, with users expressing satisfaction with the pleasantness ($M=5.17$) and liking its overall design ($M=5.04$). However, comparatively lower scores were observed for efficiency- and speed-related items, such as completing work quickly ($M=4.21$) and becoming productive quickly ($M=4.25$), suggesting potential areas for improvement.

Fig. 9-b shows the CSUQ metrics. The *System Usefulness* (SysUSE, $M=4.77$, $SD=1.80$, $Mdn=5$) is the least appreciated metric due to efficiency problems: participants reported that they spent too much time before finding the mapping between a cube rotation and the corresponding effect. Q_4 ($M=4.21$, $SD=1.94$) on efficiency is the least appreciated statement of the whole questionnaire with the largest standard deviation. When participants acquired this mapping, the usage was estimated straightforwardly thanks to Q_3 ($M=5.42$, $SD=1.78$). Users admit having difficulty performing the tasks, but do not think that cubes are difficult to handle, considering that with time, they start performing the tasks easily. The issues concerning the interaction or the execution of basic actions (e.g., change the track and activate an effect) had the highest means. From the informal feedback collected in the open questions, we noticed that the participants were divided into two groups: a positive group that thought that

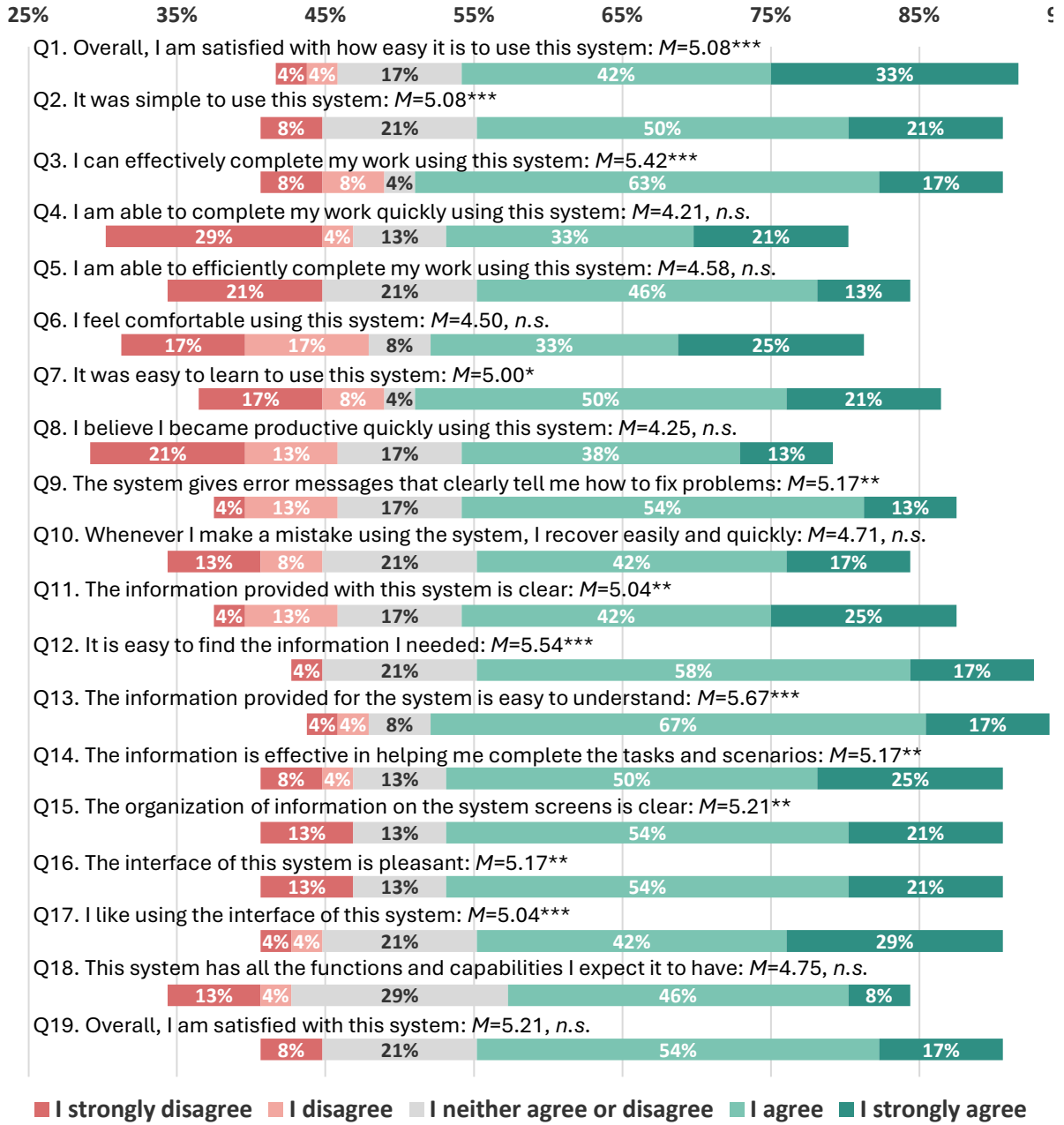


Fig. 10. Participants' answers to the first experiment questionnaire: statement, average rating, level of significance resulting from the calculation of the significant difference between the mean and the median.



Fig. 11. Stickers glued to cubes to remind them of their properties and behaviors.

feedback on Ableton was immediate enough for direct manipulation, and a negative group that estimated that the lack of feedback on the cubes impeded completing the task.

The *Interface quality* (INTERQUAL, $M=4.99$) further supports the observation that users found the AUDIOCUBES visually pleasant and acceptable. While not the highest-rated scale, these results indicate that this design does not present major barriers and contributes positively to the overall usability. Error handling and recovery were rated positively, with users indicating that error messages were clear ($M=5.17$) and that recovery from mistakes was relatively easy ($M=4.71$). This suggests that AUDIOCUBES provide adequate feedback and support during interaction, which is an important aspect of usability, especially for creative TUIs.

The *Interaction quality* (INTERQUAL, $M=4.99$, $SD=1.60$, $Mdn=5$) was appreciated more positively than *information quality*, mainly for its pleasant aspects and easy manipulation of cubes, even if their mapping is not obvious. All cube actions observed were equally appreciated, with less real 3D cube rotations, perhaps because the cubes were lying down on a tabletop.

The *Overall satisfaction* (OVERALL, $M=4.99$, $SD=1.69$) received a score almost equal to 5: although cubic actions were understood as natural actions whose mapping is not always straightforward, participants reported a positive attitude towards their usage for simple to moderately complex tasks. The participants' scores did not necessarily depend on the results they had in the time taken to perform the tasks. Knowing that the sample involves TUI novice users, but with a large distribution in their reported experience with computers (none of them were computer scientists), the results suggest that participants generally share the same degree of subjective satisfaction. However, improving some cube aspects was felt by all types of users. They seem to like the system, regardless of their age or sex. The use of effects does not pose any problem while the tracks present a mapping problem, which can confuse users. In sum, overall satisfaction was high, indicating that despite some efficiency-related limitations, users generally viewed the system favorably. We should focus on enhancing task efficiency and reducing the time required to achieve productivity.

4.3 Implications for Gestural Input with Tangible Cubes Resulting from Experiment #1

Based on our findings, we now propose some practical implications for interaction design addressing gestural input with tangible cubes, as they result from experiment #1, with potential concrete applications.

4.3.1 Optimize for Single-Object Direct Manipulation. The cube interactions worked best in a one-to-one mapping established between a cube action and a system command. Users performed efficiently when interaction mappings are direct and physically intuitive. Indeed, T_3 – T_6 (single track operations and switching) were faster (they belong to the 18–32 sec. range) with low standard deviations (8–13 sec.), suggesting a consistent performance. Applications: one-to-one mappings are recommended, such as one cube is associated to one element (e.g., a track), one rotation increases/decreases the intensity or the value of a parameter, flipping a cube enables mode changing, push a cube confirms an action, shake a cube to undo the last action, apply a clockwise circle to the cube to (re)do and a counterclockwise to reverse.

4.3.2 Minimize Physical Complexity for Multi-Cube Tasks. The tasks requiring multiple cube interactions (e.g., ordering, grouping, selecting several items) should use simple, low-effort gestures rather than requiring repeated physical rearrangements. Indeed, T_1 (ordering) and T_2 (multi-selection) have the highest mean times ($M=93.71s$, $M=124.71s$) and largest standard deviations ($SD=78.69$, $SD=103.23$), which led to inconsistent performance. This variability indicates that spatial ordering creates cognitive and motor workload. Applications: grouping operations could be enabled via stacking or clustering cubes. For such tasks, the cube could privilege proximity-based grouping instead of precise alignment. For ordering tasks, there is no need to require exact spatial sequences of cubes, only sequence.

4.3.3 Provide Immediate Feedback for Cube State and Selection. The cube interactions should generate clear feedback that is either monomodal (e.g., visual, haptic, or audio) or multimodal, indicating for example what is selected, what is grouped, what effect is active, what mode the cube is in. Indeed, users reported high satisfaction with information quality ($M=5.21$), but speed ratings are lower ($M=4.21$ for completing work). Applications: while the cubes can display any color from a palette (only color was used to differentiate the properties and the behavior of cubes, but a cube can change its color saturation depending on variation), there could be a color coding scheme associated to each cube state. Glowing or pulsing effects could be added to reflect their activation, deactivation when (un)selected. When cubes are successfully grouped or arranged in a topology, sound cues could express this state. The feedback could also be reflected in the target application, not just in the locus of control, e.g., by highlighting the affected GUI widgets. Stickers can be glued on cube faces (Fig. 11), but they remain static until they are removed.

4.3.4 Reduce Work Load in Sequential Cube Arrangements. The cube interactions should minimize work load and potential errors when arrangements are sequential. Complex multi-step tasks show slowdown and cumulation of errors decrease the quality of the whole sequence. Indeed, large gap between satisfaction (~ 5.0) and perceived speed/productivity (~ 4.2 – 4.5) were observed. Tangible cubes should in principle offload cognition into a physical arrangement. Applications: after a cube has been assigned to a command, there is a need to display how new cubes could be placed in the arrangement (e.g., by projecting a ghost of the cube on the surface).

4.3.5 Support Fast Recovery and Cube Reconfiguration. The cube interactions should enable effortless correction and encourage exploration without penalizing end users, therefore asking for quickly recovering a previous state. High variability in complex tasks suggests frequent errors. Indeed, users rated errors as important ($M=5.17$) and recovery score ($M=4.71$) was good but improvable. Applications: A dedicated “reset face” could be assigned to a cube or a cube can be entirely dedicated to undo/redo/reset. The interaction could support easy de-selection (e.g., “lift a cube away”).

4.3.6 Maintain a Feedback that is Consistent with the Application. The cube interactions should be a consistent proxy of the visual and auditory feedback produced by the application, which is challenging to design due to the variety of values to be controlled. Indeed, in the experiment, the visual feedback was executed in Ableton Live, whereas the locus of control was on the cubes. Applications: consistent feedback can be provided by the cube itself: visually if an icon is displayed on each face (like in [30]), but preferably via other channels not to interfere with the visual, auditory outputs. Force feedback expressing relief could repulse the hand for a forbidden action, while a groove could attract the hand for an allowed action. Tactile feedback, e.g., via tactile actuators, can give information on the status, texture, and temperature of the cube. Vibrotactile vibrations can be transmitted through the skin. In this way, discrete and continuous behaviors can be differentiated: a stepper for a discrete behavior (e.g., a selection of a value) and a progressive vibration for a continuous behavior (e.g., navigating in a range of values).

5 Experiment #2: In-Museum Usability Study

5.1 Method Specific to Experiment #2

This second experiment was conducted in a museum with visitors passing by who volunteered to do the study (Fig. 1-c), which may induce self-selection bias. They did not have any physical/motor impairment of the hand, wrist, or forearm (no biometric test was performed to assess this claim). All participants were new to avoid any fatigue or learning effect. We wanted to assess the usability in a more ecological context of use, thus opening the scope of all previous metrics.

5.1.1 Apparatus. To test a more complex and flexible setup than in experiment #1, this experiment involved four AUDIOCUBES: the first cube is a SENDER, the second is a RECEIVER, and the last two are Senders that need to be coordinated in topology to produce some effect. We configured Ableton™ Live with the set “Afrolat-in-1” provided in the software installation program (Fig. 8-b). This set is composed of 13 tracks in 4 audio groups. An audio track in an audio group can be played once, simultaneously with the other groups. We used two effects in group 3: “Phased and Confused” and “Zappified”. Each face of the Receiver was associated with an audio group. The Senders are used to select the group of tracks to be played. For example, placing face 3 of a Sender in front of face 1 of a Receiver activates track 3 in group 1. The Sender activates the effects and modulates one of them. The obstruction of face 1 by the hand activates the effect “Zappified”, which does not require any modulation. Obstruction of face 2 activates “Phased and confused”, which can be modulated by the obstruction of face 3.

5.1.2 Tasks. We devised three tasks with an increasing level of manipulation complexity.

The *first task* T_1 was jointly operating a Sender with a Receiver to select a track in an audio group, and from one audio group to another. After activating the first audio track of each group, the participant is invited to proceed as follows: (1) Play track 3 in the audio group 1, (2) Play track 2 in the audio group 2, (3) Play track 2 in the audio group 3, and (4) Play track 3 in the audio group 4. The track can be changed by passing through intermediate tracks, except for the last change. For example, when the participant played track 3 in audio group 1 and passed through track 2 before playing track 3, we did not count it as an error. When the participant played track 3 in group 4 (last change), s/he was unable to pass through an intermediate track; otherwise, we counted it as an error. If the user changed the track in the wrong audio group, we recorded it as an error.

The *second task* T_2 handled two Senders concurrently with a Receiver to select two tracks in two audio groups. Simultaneous changes are always done in audio groups 1 and 3. Participants are invited to proceed as follows: (1) Play track 1 of audio group 1 and simultaneously play track 1 of group 3, (2) Play track 2 of audio group 1 and simultaneously play track 1 of group 3, (3) Play track 3 of group 1 and simultaneously play track 2 of group 3, and (4) Play track 4 of group 1 and simultaneously play track 1 of group 3. After completion, participants reinitialized track 1 in all groups.

The *third task* T_3 handled a Receiver cube and a Sender cube simultaneously to activate an effect in a track using the Receiver, to modulate this effect, and simultaneously change a track using the Sender cube. Changing a track must be done directly without going through an intermediate track, otherwise, we recorded it as an error. We asked users to perform the following procedure: (1) Enable the effect "Zamplified" with the Receiver cube, (2) Enable the effect "Face and confused" with the Receiver and modulate this effect, and (3) Enable the effect "Face and confused" and modulate this effect and simultaneously play track 2 of the audio group.

5.1.3 Procedure. The procedure was similar to experiment #1, with a questionnaire composed of open questions, whose purpose is to qualitatively assess the AUDIOCUBE usability.

5.1.4 Participants. Twelve new subjects ($n_2=12$, 6 women, 6 men), aged between 18 and 56 years ($M=32.33y$, $SD=11.59$), participated in this experiment, none of them participated in the first one. The participants had various backgrounds, ranging from student and employee to CEO in administration, communication, marketing, arts, and restaurants. They reported themselves as regular computer users ($M=6.00$, $SD=1.65$ on a 1-7 scale – "I use a computer frequently": 1=strongly disagree to 7=strongly agree) with easiness ($M=6.42$, $SD=1.38$ – "I can use a computer with ease": 1=strongly disagree to 7=strongly agree), with excellent manipulation of a mouse ($M=7.00$, $SD=0.0$ – "I can use a mouse with ease": 1=strongly disagree to 7=strongly agree), low skills with a stylus ($M=2.25$, $SD=1.96$ – "I can use a stylus with ease": 1=strongly disagree to 7=strongly agree), and moderate experience with sound ($M=3.75$, $SD=1.66$ – "I am familiar with sound expression": 1=strongly disagree to 7=strongly agree), all without any prior exposure to TUIs.

5.2 Results and Discussion

Fig. 12 shows the distribution of participants' answers to the usability questionnaire that are aggregated in the CSUQ metrics in Fig. 13-b. The internal consistency for this questionnaire is assessed as quite reliable (Guttman's split-half $G=.88$) and well agreed among participants (Kendall's $W=0.29$, $df=19$, $p\leq.001^{***}$). We performed a two-tailed t test for each statement to check whether its distribution was significantly higher than the median ($Mdn=4$), which was the case for all statements.

Overall, the results indicate very high perceived usability, with most statements receiving high mean scores (well above 5.5), high medians (usually 6) and relatively low standard deviations (less than 1.19), suggesting strong agreement among participants. Users reported high subjective satisfaction with the ease of use ($M=6.25$, $SD=0.83$) and rated the AUDIOCUBES as simple to use ($M=6.50$, $SD=0.65$) and easy to learn ($M=6.58$, $SD=0.49$). Task effectiveness and efficiency were also rated positively, with participants indicating they could complete their work effectively ($M=6.50$) and quickly ($M=6.50$). Comfort and rapid productivity were rated slightly lower but still favorable ($M=5.50$).

Information-related aspects of the AUDIOCUBES received the strongest evaluations. Participants found it easy to locate the information they needed ($M=6.83$, $SD=0.37$), and rated guidance as clear and helpful ($M=6.25$). This is reflected in the high *Information Quality* (INFOQUAL, $M=6.18$, $SD=0.97$, Fig. 13-a). These aspects were positively assessed due to the easiness of the cubes, the surprise effect, and the creative setup. Only one participant, who was not considered an outlier, was negative in these aspects, mainly because the playability was not counterbalanced with any form of guidance, other than the visual feedback in Ableton Live and the auditory feedback produced by AUDIOCUBES.

Interface-related statements were generally positive at the individual level. But the *Interface Quality* (INTERQUAL) metric was notably lower ($M=4.19$, $SD=2.22$, Fig. 13-a), indicating variability in users' perceptions. While most participants appreciated the cubes and found them pleasant, others may have experienced visual, layout, or interaction issues. One notable outlier is the very low score for the item "This system has all the functions and capabilities I expect it to have" ($M=1.25$). This sharply contrasts with the otherwise positive findings and may

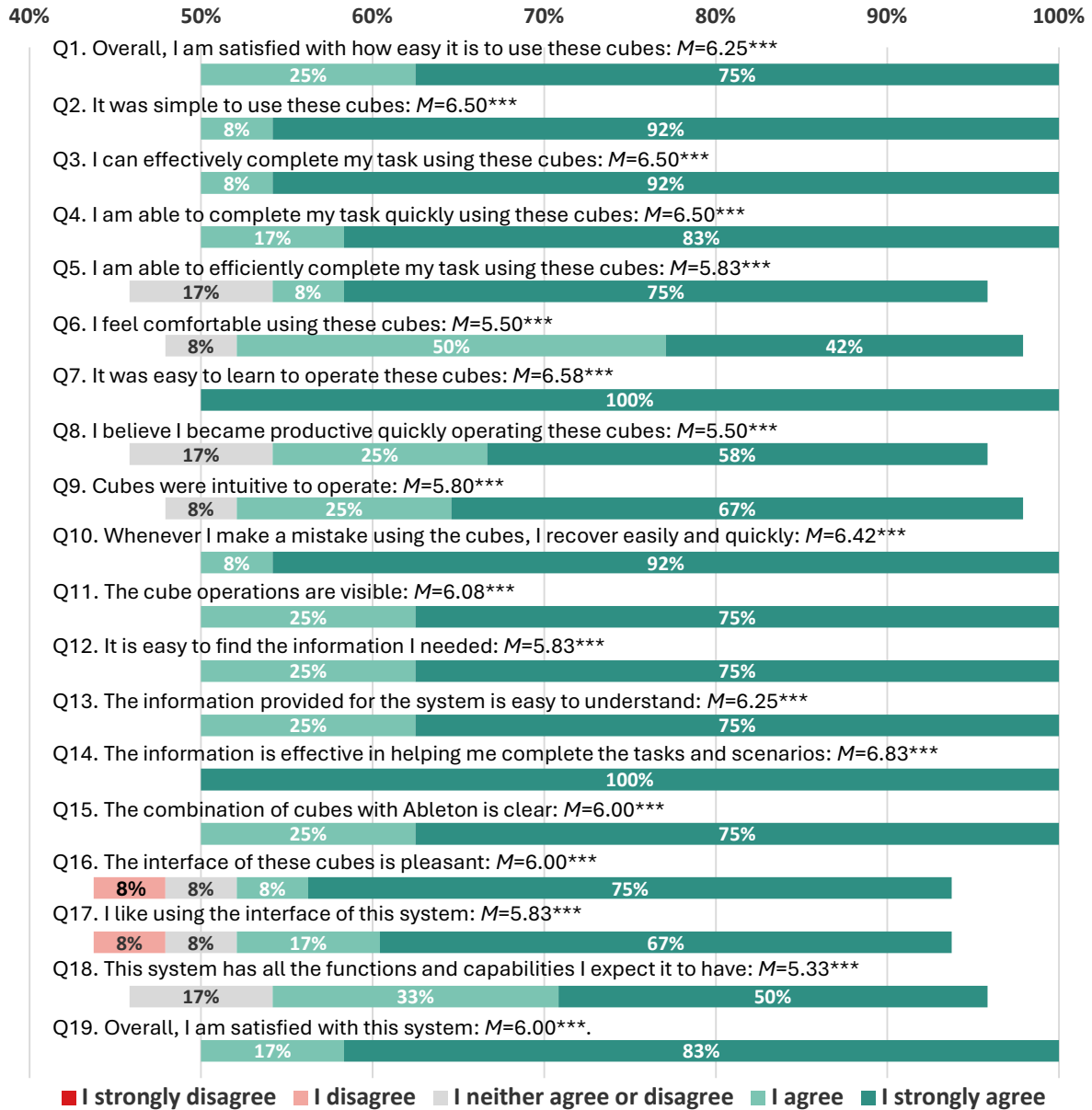


Fig. 12. Participants' answers to the second experiment questionnaire: statement, average rating, Average answers to the first experiment questionnaire: statement, average rating, level of significance resulting from the calculation of the significant difference between the mean and the median.

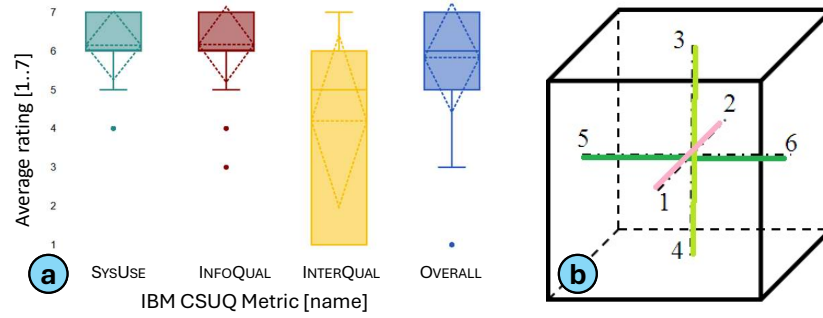


Fig. 13. Results of the second experiment: CSUQ-Metrics (a) and frequencies of rotations (b). Error bars show a 95% confid. interval.

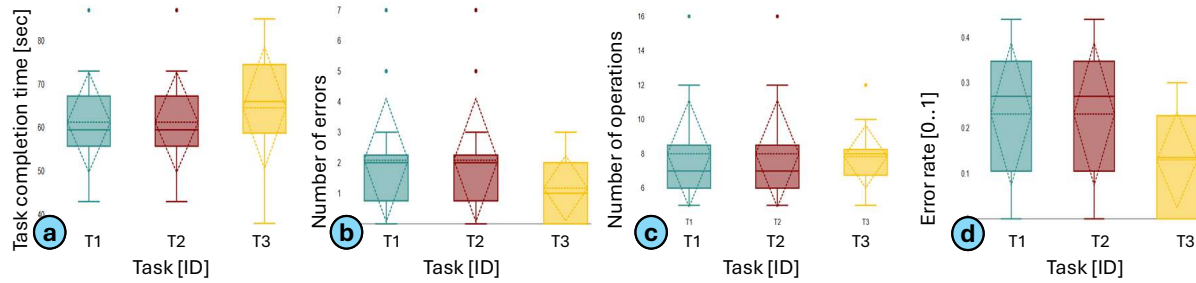


Fig. 14. Results of the second experiment: Task completion time (a), Number of Errors per task (b), Number of cube actions per task (c), and Error rate per task (right). The error bars show a confidence interval of 95% ($\alpha=0.05$).

indicate unmet expectations regarding functionality, feature completeness, or system scope. This variability suggests that the interface may work well for some users or contexts but could benefit from refinement to improve consistency and appeal across a broader user base. Overall satisfaction with the system remained high ($M=5.83$, $SD=0.55$).

The *System Usefulness* (SysUSE, $M=6.15$) confirms that AUDIOCUBES effectively supported participants in completing their tasks. High ratings for speed and effectiveness imply that once users engage with the cubes, it enables effective interaction without unnecessary friction. Slightly lower scores for efficiency may reflect differences in experience levels or task complexity, but these values still fall within a positive range.

Fig. 14 shows the other metrics measured in the experiment #2 (Table 5): Task completion time, Number of Errors per task, Number of cube actions per task, and Error rate per task. All tasks were completed to 100% for all users, suggesting that effectiveness was maximum. T_2 took the longest time ($M=64.58$ sec, $SD=14.04$ sec) because participants performed two actions simultaneously among the eight, thus producing a few errors. The participants needed some thinking time before moving cubes for every task: ± 10 sec. T_1 required less time because some participants were used to performing simultaneous actions, while others were not so used and had to discover the concurrent behavior. T_3 took the shortest time ($M=34.92$ sec, $SD=6.18$ sec): initially, the participants must manipulate only one cube and did not have to think about which face of which cube must be set in front of the other. Activating an effect and blocking one side of a cube is very similar to pressing a button, a familiar action. The last T_3 actions involved two cubes, but these actions have been done before, thus suggesting a learning effect. T_3 also benefits from almost no errors and an even a more limited number of actions, both being consistent

with the completion time, thus suggesting that participants converged to actions they learned. T_1 users were more likely to try a direct cube action, even if they were wrong. This training was reused for T_2 and T_3 . The operation mode is different for T_2 where the simultaneity of actions encouraged them to think more before any action. These observations are consistent with the time taken by the participants, who made fewer errors with subsequent tasks. They become accustomed to cubes and their usage as they learn from their errors. With thinking time decreasing, they finally succeeded in performing progressively more complex cube actions, even simultaneous ones.

5.3 Implications for Gestural Input with Tangible Cubes Resulting from Experiment #2

Based on our new findings cumulated with the findings of the experiment #1, we now propose some implications for interaction design addressing gestural input with tangible cubes. Before doing so, we discuss the overall evaluation.

First, the consistently high scores for ease of use, learnability, and task effectiveness indicate that the fundamental interaction paradigm of cube-based interaction (at least with the AUDIOCUBES) should be preserved. Users were able to understand how to use the cubes quickly and complete tasks effectively, suggesting that the mapping between user actions and system responses is intuitive. Future design iterations should therefore avoid introducing unnecessary complexity that could compromise this learning curve.

Second, the very good ratings for information quality highlight the importance of maintaining clear and well-organized feedback. Design decisions related to guidance and error management appear to support user confidence and recovery from mistakes.

Table 5. Metrics measured by task: Task completion time (TCT), Number of errors per task (NoE), and Number of cube actions per task (NoA).

Part.	Task 1			Task 2			Task 3		
	TCT	NoE	NoA	TCT	NoE	NoA	TCT	NoE	NoA
1	87	7	16	82	3	12	49	0	4
2	66	2	8	67	0	8	32	0	3
3	43	0	6	38	0	6	27	1	4
4	46	0	6	42	0	5	29	0	5
5	62	1	7	63	1	8	32	0	8
6	62	5	12	61	3	10	36	1	6
7	73	0	8	65	2	8	36	1	4
8	71	3	10	85	1	8	44	0	5
9	57	2	7	52	0	6	28	1	2
10	55	2	6	70	2	9	33	0	4
11	56	1	5	74	1	7	38	1	5
12	57	2	5	76	1	7	35	0	5
Avg.	61.25	2.08	8.00	64.58	1.17	7.83	34.92	0.42	4.58
Min.	43.00	0.00	5.00	38.00	0.00	5.00	27.00	0.00	2.00
Max.	87.00	7.00	16.00	85.00	3.00	12.00	49.00	1.00	8.00
Med.	59.50	2.00	7.00	66.00	1.00	8.00	34.00	0.00	4.50
St.dev.	12.02	2.11	3.25	14.66	1.11	1.90	6.46	0.51	1.51

Third, the relatively lower and more variable interface quality scores suggest a need to improve visual and interaction consistency. While many users found the interface pleasant and enjoyable, others may have encountered issues related to layout, visual hierarchy, or aesthetic coherence. Design improvements could focus on refining visual alignment, standardizing interaction elements, and improving affordances to ensure that the cubes communicate their function more uniformly across different user contexts of use.

A particularly important design implication arises from the very low rating for the item concerning expected system functions and capabilities. This result indicates a potential mismatch between user expectations and the system's current feature set. From a design perspective, this suggests two possible directions: either expanding functionality to better align with user expectations, or more clearly communicating the intended scope and purpose of AUDIOCUBES.

Based on the experimental results, the following implications are derived, now extended with concrete design implications and examples for tangible cube interaction systems.

5.3.1 Prioritize Error Prevention in Cross-Group Selection Tasks. The cube interactions should reduce ambiguity and prevent invalid actions when cubes are used across spatial or logical groups. The goal is to reduce accidental cross-group operations and lower cognitive demand. Indeed, T_1 showed the highest error rate ($M=0.23$) and the highest mean number of errors ($M=2.08$), indicating difficulty when selecting and moving tracks across audio groups. Applications: The interaction surface on which cubes are placed could be segmented into clearly marked zones (e.g., by projected boundaries or LED outlines) to prevent errors. When a cube moves close to a correct target zone, cubes could be “snapped”. When a cube is lifted over the surface, potential target groups glow to indicate valid drop locations.

5.3.2 Support Coordinated Multi-Object Interaction with Clear State Awareness. The cube interactions should make multi-cube states and relationships perceptually explicit. The goal is to support distributed cognition by externalizing relationships between cubes. Indeed, T_2 (simultaneous selection in two groups) has high completion time (64.58 s) and variability, indicating coordination challenges. Applications: When two cubes are selected simultaneously, they could share the same color or become connected by a projected line. Two successfully grouped cubes could pulse together in the same frequency. The users should be able to physically stack or align cubes to create temporary links.

5.3.3 Leverage Direct Manipulation for Integrated Parameter Control. The cube interactions should exploit continuous, embodied mappings between cube movement and digital parameters. Continuous mappings should be immediate and proportional (e.g., a small rotation should indicate a small change), reinforcing the sense of control and reducing cognitive translation effort. Indeed, T_3 is significantly faster (34.92 s) with fewer errors (0.42), suggesting efficiency in continuous, effect-based manipulation. Applications: Rotate a cube to modulate the depth, Tilt a cube to change the frequency or a filter, Press and hold a cube to activate its corresponding effect, Slide a cube across the surface to track its changing position.

5.3.4 Reinforce Feedback and Lightweight Recovery. The cube interactions should maintain strong feedback and make error recovery effortless. Low-cost recovery encourages exploration and reduces risk of errors. Indeed, high scores for information clarity ($M=6.17$), error management ($M=6.42$), and recovery ($M=5.83$) indicate that feedback is a strength to reinforce. Applications: To ensure an immediate confirmation, a cube glows green when action is successful or a mode is activated, and red when action is unsuccessful or a mode is deactivated. Subtle confirmation tones could reinforce successful operations.

5.3.5 Improve Cube Coherence and Functional Completeness. The cube interactions should strive for consistency in interaction mappings and align physical affordances with user expectations. The goal is to preserve consistency and completeness to improve perceived interface quality and trust in the system. Indeed, despite high usefulness

($M=6.15$) and information quality ($M=6.17$), interface quality was lower ($M=4.19$, $SD=2.22$), and the capability expectation item scored very low ($M=1.25$). This suggests perceived inconsistency or missing features. Applications: To preserve consistency, any cube action should always be mapped to the same command. For example, if rotation controls modulation in one context of use, it should not control volume in another context. A small display or LED icon should indicate the current function of the cube in the same way. Basic cube interactions could be assigned to default cube faces while advanced controls could be revealed by flipping the cube.

6 Experiment #3: Exhibition Field Study

This experiment allowed participants to operate the cubes without any guided task during a technology exhibition on site. People passing by stopped and explored the effects of performing any gestures on the AUDIOCUBE (Fig. 1-c).

6.1 Method Specific to Experiment #3

6.1.1 Procedure. No particular task was given to participants. No explanation was given to them, thus enabling them to freely test the same apparatus as in the experiment #2 with the common protocol (Section 3.2).

6.1.2 Participants. 33 subjects ($n_3=33$, 10 women, 23 men) participated in this experiment. They were not asked demographic or experience questionnaire to let them feel free throughout the experiment. None of them had previous exposure to TUI. Although no physical or experience tests were performed, no particular disability was observed.

6.2 Results and Discussion

The results of gestures captured in this experiment #3 are aggregated with the corresponding ones from the two experiments #1 and #2.

6.2.1 Gesture Types and Frequencies. Table 6 shows the distribution of positioning gestures performed by participants in terms of percentage, individually by experiment, and overall for all experiments. Fig. 16, c-p, visually reproduces the gestures corresponding to these results in decreasing order of their overall frequency. Based on the gesture frequencies and the gesture grasping types, we classified positioning gestures in the following categories.

Uni-manual gesture with one cube. The most frequently observed positioning gestures were (Fig. 16): *Grab a cube* (34%) and (2) *push a cube* (30%). Grabbing is most frequent on two opposite faces (22% of all gestures), on opposite edges (8%), on corners (4%) to perform one action at a time (planar translation, then curve, then rotation – non-planar curve, then rotation) or several operations combined (any combination of the previous ones). To carry out the tasks, most participants tend to start by grabbing a cube, then continue with the simplest gestures (e.g., linear translation of a cube while grabbing) to end up with rotation (like a Rubik's cube), first with rectilinear path, then non-rectilinear (e.g., lift and drop on the surface or for (un)stacking). Pushing is the second most frequent gesture on a face (14%), an edge (12%). The corner (4%) was avoided because of its unpredictability. Some participants coincidentally tested it.

Uni-manual gesture with two cubes. Nobody moved more than two cubes at once because the human hand cannot hold more than two cubes in grabbing mode. The 1-5 *hand span*, i.e., the distance from the tip of the thumb to the tip of the forefinger, is just higher than two tiled cubes: the mean hand span is about 22.6 cm (8.9 inches) for male and about 20.1 cm (7.9 inches) for female [8] as opposed to $2 \times 7.7\text{cm}/\text{cube} = 15.4\text{cm} \approx 6.6$ inches. Grabbing and pushing two cubes cover 26%, 10% of gestures, respectively. Almost nobody used 'by corner' because of the lack of precision: pushing on a corner does not allow any prediction of the resulting movement of cubes. For both faces and edges, the participants preferred grabbing stacked cubes at their bottom by faces (13%) and edges (3%),

Table 6. Distribution of gesture articulations performed by participants over all experiments in percentage.

Cubes	Gesture	Location	Actions	Exp. #1	Exp. #2	Exp. #3	All
1	Grab a cube	opp. faces	1	19	13	18	50
			many	6	3	4	13
			total	25	16	22	63
		opp. edges	1	9	7	7	23
			many	2	2	1	5
			total	11	9	8	28
		opp. corners	1	2	3	4	9
			many	0	0	0	0
			total	2	3	4	9
		grand total		38	28	34	100
	Push a cube	on face	1	24	12	9	45
			many	7	5	5	17
			total	31	17	14	62
		on edge	1	11	6	8	25
			many	2	2	4	8
			total	13	8	12	33
		on corner	1	0	1	4	5
			many	0	0	0	0
			total	0	1	4	5
		grand total		44	26	30	100
2	Grab 2 cubes	by faces	unstacked	10	10	8	28
			stacked-bottom	12	13	9	34
			stacked-top	8	5	4	17
			total	30	28	21	79
		by edges	unstacked	4	2	2	8
			stacked-bottom	4	2	2	8
			stacked-top	2	2	1	5
			total	10	6	5	21
		by corners	stacked	0	0	0	0
			unstacked	0	0	0	0
			total	0	0	0	0
		grand total		40	34	26	100
	Push two cubes	on face	top	24	12	9	45
			bottom	7	5	5	17
			total	31	17	14	62
		on edge	top	11	6	8	25
			bottom	2	2	4	8
			total	13	8	12	33
		on corner	top	0	1	4	5
			bottom	0	0	0	0
		total		0	1	4	5
		grand total		44	26	30	100

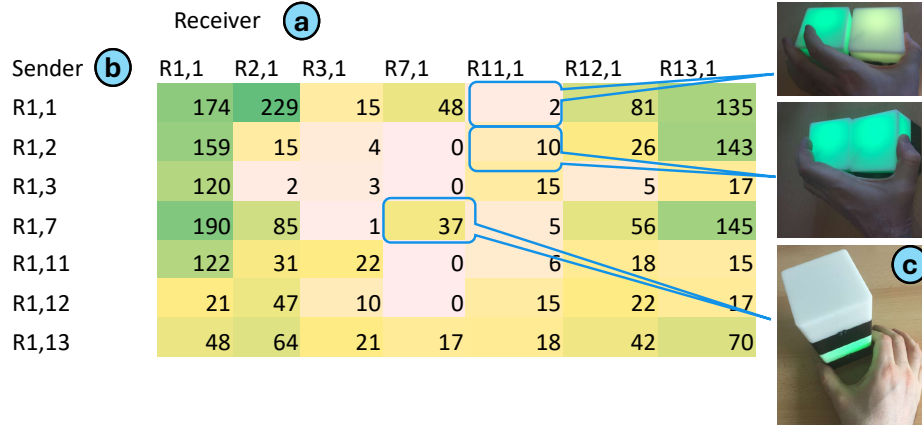


Fig. 15. The actual space for arranging two cubes found in the experiments, as a heat map of the theoretical space of Fig. 7. Every arrangement of two cubes is represented on the abscissa by the Receiver cube and on the ordinate by the Sender cube

then unstacked by faces (8%) and edges (5%). Pushing was considered second when cubes were aligned and tiled towards a particular direction, but only on a face. Pushing two cubes was preferred on their top (4%) or bottom (1%) face, then on their edges (3%), then on their bottom edge (1%). Other gestures were revealed to be imprecise.

Bi-manual gesture with one cube. This case is analog to the case of using one hand, except that users preferred to rely on two hands to ensure precision, fine-tuning, and slight movement.

Bi-manual gesture with two cubes. Bi-manual gestures were repeated for each cube, such as for actions such as swapping (e.g., moving cubes from one hand to another), switching (i.e., exchanging two cubes across hands), permutation (i.e., exchanging more than two cubes across hands), or parallel lift-and-drop of cubes, which are symmetric [40].

6.2.2 Arrangement Frequencies. Fig. 15 shows the heat map expressing the arrangement frequencies between two cubes cumulated throughout the experiments. Each arrangement is represented on the abscissa by the Receiver cube (Fig. 15-a) and on the ordinate by the Sender cube (Fig. 15-b). The darker the cell color, the more frequent the arrangement was observed. For example, $R_{1,1}-R_{1,1}$ represents an arrangement corresponding to cell (1, 1) in Fig. 7: none of the cubes are aligned ($\frac{174}{2348}=7\%$ of the arrangements). The second cell represents an arrangement in which the Sender is aligned to the left of the Sensor ($\frac{229}{2348}=9.7\%$). The left column contains the most frequent arrangements, representing the typical positions of the Receiver to the Sender. The subsequent rows express similar arrangements with different alignments, such as the left vs. right alignment to the right/left of the Receiver. $R_{7,1}-R_{1,7}$ represents the case where the two cubes were perfectly stacked one top of the other ($\frac{37}{2348}=1.6\%$). Rows 4-6 and 8-10 were never formed because they represent arrangements in which objects are of different sizes. Only $169 - [(6 \text{ rows} \times 13 \text{ positions}) + (4 \text{ blocks} \times 9 \text{ positions})] = 169 - 114 = 55$ positions were finally used which cover 32% of the theoretical space (Fig. 7).

6.2.3 Rotation Frequencies. Fig. 13-b shows the heat map for rotation gestures observed throughout the experiment: 5-6-axis rotations were the most frequently observed, followed by 3-4 and 1-2.

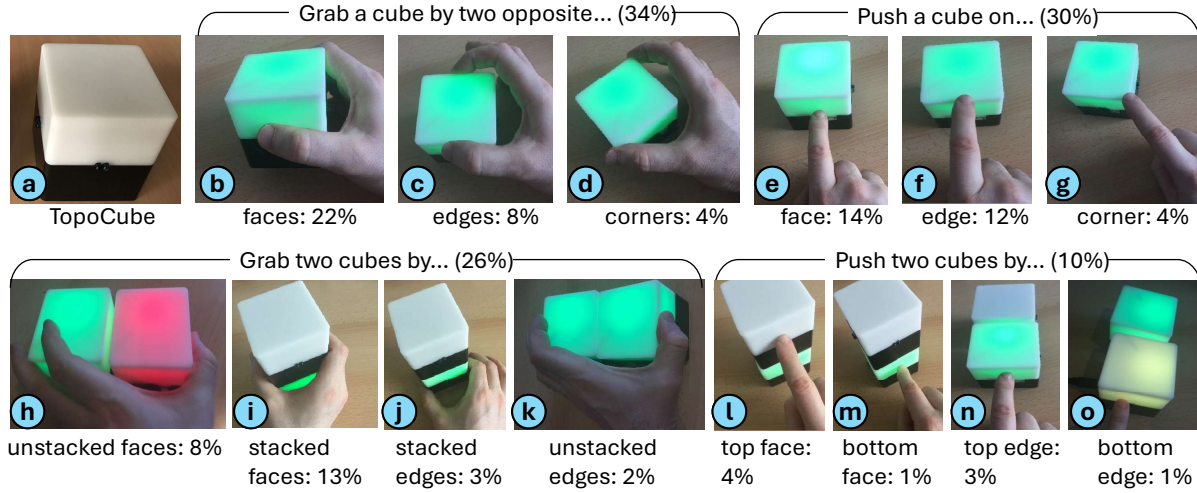


Fig. 16. Gesture articulation with a AUDIOCUBE (a): Grab a cube (b–d), Push a cube (e–g), Grab two cubes (h–k), and Push two cubes (l–o), sorted in decreasing order of their frequency.

6.3 Implications for Gestural Input with Tangible Cubes Resulting from Experiment #3

Based on our new findings cumulated with the findings of the previous experiments, we now propose implications for designing cube interaction by gestural input, with a focus on gesture type and articulation.

6.3.1 Prioritize Face-Based Interaction as the Primary Input Surface. The cube interactions should be designed primarily around faces, not edges or corners. Users naturally grasp and push cubes via faces. Designing around this increases intuitiveness and reduces motor strain. Indeed, single-cube grab by two opposite faces were preferred in 63% of cases, "Push on face" in 62%, and two-cube grab by faces in 79%, whereas corner articulations were consistently near 0–5%. Applications: The touch sensors, displays, or LEDs should focus primarily on detecting cube faces, especially algorithms for gesture recognition. The most important functions, *e.g.*, select, activate, confirm, should be mapped to face contact. In contrast, face-to-face contact (*e.g.*, placing cubes flush together) should be preferred to trigger cube linking or grouping. In all cases, cube-based gesture recognizers should avoid requiring precise corner-based gestures and accept some tolerance. The next implication is a consequence of this prioritization.

6.3.2 Avoid Corner-Based Interaction for Important Functions. The cube interactions should not assign essential functions to corner-based manipulation. The goal is to avoid any corner interaction, which demand precision and are not part of users' natural articulation patterns. Indeed, Grabbing two cubes by corners were observed in 0% of the cases, as well as pushing on corners (5% or less). No multi-corner gestures were captured. Applications: Corner taps should not be used, such as to activate modes. Corner-based stacking detection should not be used as a primary gestural input. The cube corners should be reserved for passive structural cube arrangement rather than active interaction.

6.3.3 Design for One-Handed, Single-Cube Efficiency. The cube interactions should ensure all important functions can be preferably executed with one hand on a single cube. The goal is to support accessibility, fluid workflow, and reduced physical fatigue. Indeed, The dataset shows strong dominance of one-hand, one-cube gestures for both grabbing and pushing. Applications: For example, rotating a single cube adjusts parameters, pressing a cube

face toggles activation and pressing it again toggles deactivation, sliding a cube across the surface changes track position.

6.3.4 Support Cube Stacking and Top-Focused Interaction. The cube interactions should design stacking as a meaningful and stable interaction state, especially emphasizing the top cube as the active control element. Users naturally interact more with the top cube, so stacking can encode hierarchical relationships. The interaction prioritizes input from the top cube when stacked because visual feedback appears primarily on the top cube since the bottom cube is less visible and manipulable. Indeed, two-cube gestures were preeminent for grabbing by faces when stacked (bottom: 34%, top: 17%). Pushing two cubes on face—top interactions dominate (45%) vs. bottom (17%). Applications: When cubes are stacked, the top cube controls modulation, the bottom cube defines context of use (e.g., a track). More generally, an action+object paradigm could be established by having the bottom cube for designating the object and the top cube applying the action on this object. Stacking cubes could trigger mode switching (e.g., a bottom cube selects an instrument and the top cube selects the effect applied to this instrument).

6.3.5 Enable Multi-Cube Interaction through Face Alignment Rather than Edge Precision. The cube interactions should design multi-cube operations by relying on face alignment or surface contact rather than edge matching. Face-based alignment is both mechanically stable and cognitively intuitive. Indeed, two-cube grabs by faces were preferred (79%), far exceeding edge-based (21%) and corner-based (never) interactions. Applications: A shared color or LED animation appears when multiple cubes are face-aligned. Placing cubes side-by-side on their flat faces on the surface could trigger a group selection. Again, edge-to-edge docking as a primary gesture should be discarded.

7 Limitations of this Work

There are several threats [59, 65] that have influenced the results of our three experiments, which we discuss in the next subsections.

7.1 External Validity

Our experiments were designed to investigate and explore (Table 1) the proposed cube-based interaction approach progressively across complementary contexts. The first experiments provided a controlled laboratory setting to assess baseline usability, task performance, and initial gesture preferences. The second experiment extended the evaluation to a more realistic environment, allowing to examine how participants interacted with the AUDIOCUBES under contextual constraints and less controlled circumstances. The third experiment focused on ecological validity by capturing spontaneous, naturalistic use from a broader and more diverse participant group. Together, these experiments enabled a staged validation strategy that balanced external validity, contextual realism, and real-world applicability.

However, despite the coverage of these experiments, there are significant differences between them. Thus, our empirical results should be interpreted with the understanding that they reflect a partial representation of real settings. The representativeness of the results, as well as the number and complexity of tasks, can affect the generalizability of the results. The results' representatives could have been affected by these parameters. The gestures analyzed in our experiments can be considered a baseline. They cannot represent the whole design space of gestural input with all possible tangible cubes since, here, AUDIOCUBES have been used for musical expression. This domain was chosen for its simplicity, its universality, and no need for participants to have domain expertise, as perhaps required in other domains. With regard to the size and complexity of the tasks, we decided to use relatively small tasks to enable the participants to complete the experimental session within 10 minutes. Replications with different and more complex tasks with cubes are necessary.

The three experiments adopted an applied and practice-oriented research method by systematically investigating how tangible cubes function for gestural input across progressively realistic contexts. Across these contexts, participants used AUDIOCUBES to perform authentic manipulation tasks involving positioning, moving, rotating, and arranging cubes on a surface. This staged methodology enabled us to assess usability, capture naturally occurring gestures, and validate interaction patterns under both guided and unguided conditions. As a result, the paper contributes empirical findings, actionable implications for designing cube-based tangible interfaces, but does not establish any theory for cube-based TUI.

7.2 Internal Validity

The learning effect, the understandability of the tasks, and the validity of the instrumentation must be considered. The learning effect was mitigated by randomizing the order in which the tasks were instructed. The understandability of the tasks was assessed during the experiments, after some familiarization. To avoid possible sources of bias, the experimental materials were evaluated by an experienced researcher. We mitigated the instrumentation validity by using AUDIOCUBES, which are specific tangible cubes that only represent a fraction of the potential tangible cubes in their implementation. We only used the infrared ports of the AUDIOCUBES to determine the gestures and the built-in capabilities of the MIDI BRIDGE, which restrict the scope of potential gestures that can be captured. Since the top face of AUDIOCUBES does not have any sensor, stacking cannot be recognized in this way. The C++ toolkit for capturing raw signals should be used instead.

7.3 Construct Validity

Potential learning effects in consecutive tasks, such as those involved in experiment #1, can be mitigated by counterbalancing task order. Latin square designs can also distribute order effects evenly across participants. The metrics measured to evaluate the AUDIOCUBES and the participants' apprehension about being evaluated must be considered. We mitigated the threat regarding the metrics used by employing varying metrics, keeping them consistent across experiments. The participants' possible apprehension about being evaluated was mitigated as they were instructed that the AUDIOCUBES were the subject of evaluation. The construct validity remains constrained because the experiments relied on AUDIOCUBES as a single implementation of tangible cubes, meaning that the measured constructs may capture interaction with this specific device more than the broader concept of cube-based tangible interaction. No direct comparison with alternative cubic TUIs was conducted because the primary objective of the research was exploratory and formative rather than comparative. Moreover, suitable comparison baselines are difficult to define, as cube-based tangible interfaces differ substantially depending on the software and hardware used.

7.4 Conclusion Validity

The data collection and the validity of the statistical tests applied must be considered. To mitigate the threat related to data collection, we systematically applied the same measurement and data extraction procedures in each session with each participant in each experiment. The most appropriate analyses were performed as they were selected by considering the design of the experiments and the nature of the variables and their assumptions.

8 Conclusion and Future Work

We reported in this work empirical findings from a family of three experiments conducted with AUDIOCUBES, a set of wireless battery-powered cubic TUIs with built-in multimedia and signal generating and processing capabilities that can detect and recognize their arrangement on a surface, such as a tabletop. We used the AUDIOCUBES as a representative example of cubic TUIs to conduct these experiments to better understand how end users interact with them by gesture: RQ_1 =*how usable are tangible cubes for gestural input on a surface?* was addressed in

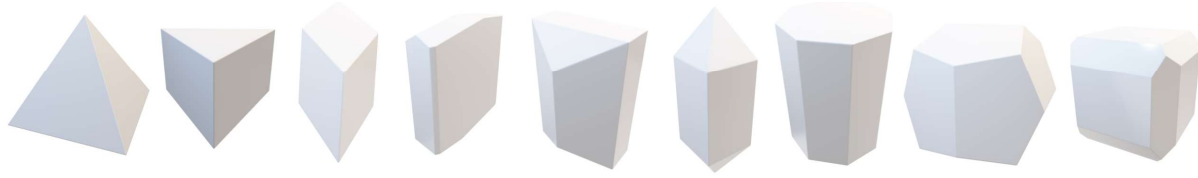


Fig. 17. Other polyhedra to study as tangible objects for gestural input.

experiments #1 and #2, while $RQ_2 = \text{What are the gestures performed with tangible cubes on a surface?}$ was addressed throughout the three experiments with a decreasing level of control to maximize the coverage of gestural input. Observations led to user-performed gestures with tangible cubes based on parameters coming from both the user and the cube itself, such as "Grab a cube" and "Push a cube" for positioning cubes. Although cubes offer 24 rotations and 169 spatial arrangements, the gestural input vocabulary used by the participants was much more limited, especially with one cube. Only 32% of the theoretical space of all potential arrangements of cubes was used as depicted in a heatmap and the frequency analysis of grasping techniques highlighted how end users preferred to position cubes by gesture. A first major conclusion is that tangible cubes offer a usable and intuitive medium for gestural interaction, but effective design must align with users' naturally preferred manipulation strategies. Participants consistently favored face-based interactions over edges and corners, simpler one-to-one mappings over complex multi-object procedures, and physically stable arrangements over precision-dependent gestures. Usable cube interfaces should prioritize grasping, pushing, rotating, and aligning through cube faces, while minimizing reliance on corner-based or highly constrained actions.

A second major conclusion is that this paper provides a scalable empirical foundation for future TUIs: by identifying gesture frequencies, articulation patterns, and actual arrangement spaces, the cube is no longer a generic physical object but now a measurable interaction vocabulary that can inform recognition algorithms and mappings to commands. More specifically, cube-based interaction can serve as an effective bridge between physical manipulation and abstract digital information (here, sound design only). These findings represent a baseline rather than an exhaustive design space.

Based on these findings, we suggested a set of practical implications for designing gestural input with tangible cubes. We finally discussed the threats to validity to which our family of three experiments were exposed.

Our future work would explore avenues towards various directions as follows:

- Conduct an experiment involving actions requiring uncovered gestures such as spatial motion gestures, shaking, accelerating, and decelerating, producing a 3D assembly, drawing symbols based on a topology, tapping signals, and authenticating an end user [63].
- Evaluate which colors of LED lights are actually assigned to actions and how different colors affect the heat map of spatial compositions. We observed that the colors most frequently selected are green, blue, yellow, light red, light green, pink, and purple, but we do not have any clue about how combinations of colors could be produced for related tasks.
- Conduct a gesture elicitation study to explore which gestures are preferred by end users depending on their task and to compare them with those discussed in this paper. A first study already identified three gestures preferred by end users for performing 6 tasks [56]. We should expand the set of referents to more tasks, such as Internet-of-Things (IoT) tasks and to the most frequent configurations allowed by AUDIOCUBE.
- Repurpose some of these experiments with other polyhedra (see Fig. 17 for some samples) than the cube to identify their effect on gestural input and to compare the gestures observed with those reported in this work. For example, a rectangular prism could be considered as in Conté [61] and a cube with truncated corners or edges (Fig. 17-rightmost polyhedra) to examine the effects of these cuts on gestures.

Acknowledgments

Nacera Latreche, Nuwan Attygalle, and Jean Vanderdonckt are supported by the EU EIC Pathfinder-Awareness Inside challenge "Symbiotik" project (1 October 2022-30 September 2026) under Grant no. 101071147.

References

- [1] James F. Allen. 1983. Maintaining Knowledge About Temporal Intervals. *Commun. ACM* 26, 11 (November 1983), 832–843. doi:10.1145/182.358434
- [2] Miguel Bruns Alonso and David V. Keyson. 2005. MusicCube: making digital music tangible. In *Proceedings of ACM International Conference on Human Factors in Computing Systems, Extended Abstracts* (Portland, OR, USA) (*CHI EA '05*). Association for Computing Machinery, New York, NY, USA, 1176–1179. doi:10.1145/1056808.1056870
- [3] Ana M. Barbancho and Isabel Barbancho. 2024. Building music with Lego bricks and Raspberry Pi. *Multim. Tools Appl.* 83, 4 (2024), 10503–10523. doi:10.1007/S11042-023-15902-Z
- [4] Victor R. Basili and David M. Weiss. 1984. A Methodology for Collecting Valid Software Engineering Data. *IEEE Transactions on Software Engineering* SE-10, 6 (1984), 728–738. doi:10.1109/TSE.1984.5010301
- [5] Adam Bell. 2015. Can we afford these affordances? GarageBand and the double-edged sword of the digital audio workstation. *Action, Criticism, and Theory for Music Education* 14, 1 (04 2015), 44–65. https://act.maydaygroup.org/articles/Bell14_1.pdf
- [6] R. Berry, M. Makino, N. Hikawa, and M. Suzuki. 2003. The augmented composer project: the music table. In *Proceedings of The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003*. IEEE Computer Society Press, Piscataway, NJ, USA, 338–339. doi:10.1109/ISMAR.2003.1240749
- [7] Richard A. Bolt. 1980. "Put-that-there": Voice and Gesture at the Graphics Interface. *SIGGRAPH Comput. Graph.* 14, 3 (July 1980), 262–270. doi:10.1145/965105.807503
- [8] Rhonda Boyle, Robin Boyle, and Erica Booker. 2015. Pianist Hand Spans: Gender and Ethnic Differences and Implications for Piano Playing. In *Proceedings of the 12th Australasian Piano Pedagogy Conference* (Victorian College of the Arts Secondary School, Melbourne, Australia) (*APCC '15*). The University of Adelaide, Australia, Adelaide, Australia, 1–79. <https://www.appca.com.au/past-proceedings/#toggle-id-3>
- [9] Ken Camarata, Ellen Yi-Luen Do, Brian R. Johnson, and Mark D. Gross. 2002. Navigational Blocks: Navigating Information Space with Tangible Media. In *Proceedings of the 7th International Conference on Intelligent User Interfaces* (San Francisco, California, USA) (*IUI '02*). ACM, New York, NY, USA, 31–38. doi:10.1145/502716.502725
- [10] Salvador Cobos, Manuel Ferre, M. Ángel Sánchez-Urán, Javier Ortego, and Rafael Aracil. 2010. Human hand descriptions and gesture recognition for object manipulation. *Computer Methods in Biomechanics and Biomedical Engineering* 13, 3 (2010), 305–317. PMID: 20146129. arXiv:<https://doi.org/10.1080/10255840903208171> doi:10.1080/10255840903208171
- [11] Marcelo Coelho and Jamie Zigelbaum. 2011. Shape-changing Interfaces. *Personal Ubiquitous Comput.* 15, 2 (February 2011), 161–173. doi:10.1007/s00779-010-0311-y
- [12] Aurélie Cohé and Martin Hachet. 2012. Understanding User Gestures for Manipulating 3D Objects from Touchscreen Inputs. In *Proceedings of Graphics Interface 2012* (Toronto, Ontario, Canada) (*GI '12*). Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 157–164. <http://dl.acm.org/citation.cfm?id=2305276.2305303>
- [13] David Cosio and Erica Lin. 2018. Role of Active Versus Passive Complementary and Integrative Health Approaches in Pain Management. *Global advances in health and medicine* 7 (2018), 1–7. doi:10.1177/2164956118768492
- [14] Arfib Daniel and Dudon Jacques. 2002. A Digital Emulator of the Photosonic Instrument. In *Proceedings of the 2002 Conference on New Interfaces for Musical Expression* (Dublin, Ireland) (*NIME '02*). National University of Singapore, Singapore, Singapore, 1–4. <http://dl.acm.org/citation.cfm?id=1085171.1085202>
- [15] Hutchins Edwin. February 1995. *Cognition in the Wild*. MIT Press, New York, NY, USA.
- [16] Kenneth P. Fishkin. 2004. A taxonomy for and analysis of tangible interfaces. *Personal and Ubiquitous Computing* 8, 5 (01 Sep 2004), 347–358. doi:10.1007/s00779-004-0297-4
- [17] George W. Fitzmaurice. 1996. *Graspable user interfaces*. Ph.D. Dissertation. School of Engineering, Canada. Advisor(s) Buxton, William. AAINN18871.
- [18] International Organization for Standardization. 1998. *ISO 9241-11: Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs), Part 11: Guidance on Usability*. ISO, Geneva, Switzerland.
- [19] Bernd Fröhlich and John Plate. 2000. The Cubic Mouse: A New Device for Three-dimensional Input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) (*CHI '00*). ACM, New York, NY, USA, 526–531. doi:10.1145/332040.332491
- [20] Bogdan-Florin Gheran, Radu-Daniel Vatavu, and Jean Vanderdonckt. 2018. Ring x2: Designing Gestures for Smart Rings using Temporal Calculus. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems* (Hong Kong, China) (*DIS '18 Companion*). Association for Computing Machinery, New York, NY, USA, 117–122. doi:10.1145/3197391.3205422

- [21] Shuqi He, Haonan Yao, Luyan Jiang, Kaiwen Li, Nan Xiang, Yue Li, Hai-Ning Liang, and Lingyun Yu. 2024. Data Cubes in Hand: A Design Space of Tangible Cubes for Visualizing 3D Spatio-Temporal Data in Mixed Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Honolulu</city>, <state>HI</state>, <country>USA</country>, </conf-loc>) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 209, 21 pages. doi:10.1145/3613904.3642740
- [22] Andy Hunt, Marcelo M. Wanderley, and Matthew Paradis. 2003. The Importance of Parameter Mapping in Electronic Instrument Design. *Journal of New Music Research* 32, 4 (2003), 429–440. arXiv:https://www.tandfonline.com/doi/pdf/10.1076/jnmr.32.4.429.18853 doi:10.1076/jnmr.32.4.429.18853
- [23] Elliot Jaffe, Aviva Dayan, and Amnon Dekel. 2007. Cube Management System: A Tangible Interface for Monitoring Large Scale Systems. In *Proceedings of the 2007 Symposium on Computer Human Interaction for the Management of Information Technology* (Cambridge, Massachusetts) (CHIMIT '07). ACM, New York, NY, USA, Article 4, 4 pages. doi:10.1145/1234772.1234778
- [24] Sergi Jordà, Günter Geiger, Marcos Alonso, and Martin Kaltenbrunner. 2007. The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction* (Baton Rouge, Louisiana) (TEI '07). Association for Computing Machinery, New York, NY, USA, 139–146. doi:10.1145/1226969.1226998
- [25] Shingo Kitagawa, Shun Hasegawa, Naoya Yamaguchi, Kei Okada, and Masayuki Inaba. 2021. Miniature Tangible Cube: Concept and Design of Target-Object-Oriented User Interface for Dual-Arm Telemanipulation. *IEEE Robotics and Automation Letters* 6, 4 (2021), 6977–6984. doi:10.1109/LRA.2021.3096475
- [26] Yoshifumi Kitamura, Yuichi Itoh, and Fumio Kishino. 2001. Real-time 3D Interaction with ActiveCube. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems* (Seattle, Washington) (CHI EA '01). ACM, New York, NY, USA, 355–356. doi:10.1145/634067.634277
- [27] Matthias Kranz, Dominik Schmidt, Paul Holleis, and Albrecht Schmidt. 2005. A Display Cube as a Tangible User Interface. Adj. In *Adjunct Proceedings of 7th International Conference on Ubiquitous Computing (UbiComp'2005)*. Springer, Berlin, Germany, 2 pages. http://eprints.lancs.ac.uk/41523/
- [28] Nacera Latreche, Bert Schiettecatte, Paolo Roselli, Nuwan T Attygalle, and Jean Vanderdonckt. 2026. Designing User-Defined Gestures for Tangible Interaction with Cubes for Smart Home Control. In *Proceedings of the 2026 Designing Interactive Systems Conference* (Singapore, Singapore) (DIS '26). Association for Computing Machinery, New York, NY, USA, 93–122. doi:10.1145/3800645.3813062
- [29] Kevin Lefevre, Soeren Totzauer, Michael Storz, Albrecht Kurze, Andreas Bischof, and Arne Berger. 2018. Bricks, Blocks, Boxes, Cubes, and Dice: On the Role of Cubic Shapes for the Design of Tangible Interactive Devices. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) (DIS '18). Association for Computing Machinery, New York, NY, USA, 485–496. doi:10.1145/3196709.3196768
- [30] Joanne Leong, Florian Perteneder, Hans-Christian Jetter, and Michael Haller. 2017. What a Life!: Building a Framework for Constructive Assemblies. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction* (Yokohama, Japan) (TEI '17). ACM, New York, NY, USA, 57–66. doi:10.1145/3024969.3024985
- [31] James R. Lewis. 1995. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction* 7, 1 (1995), 57–78. arXiv:https://doi.org/10.1080/10447319509526110 doi:10.1080/10447319509526110
- [32] James R. Lewis. 2006. Sample Sizes for Usability Tests: Mostly Math, Not Magic. *interactions* 13, 6 (November 2006), 29–33. doi:10.1145/1167948.1167973
- [33] Rensis Likert. 1932. A technique for the measurement of attitudes. *Archives of Psychology* 22, 140 (1932), 55–. http://psycnet.apa.org/record/1933-01885-001
- [34] Nathan Magrofuoco, Paolo Roselli, and Jean Vanderdonckt. 2022. μV : An Articulation, Rotation, Scaling, and Translation Invariant (ARST) Multi-stroke Gesture Recognizer. *Proc. ACM Hum. Comput. Interact.* 6, EICS (2022), 150:1–150:25. doi:10.1145/3532200
- [35] Nicolai Marquardt and Saul Greenberg. 2007. Distributed Physical Interfaces with Shared Phidgets. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction* (Baton Rouge, Louisiana) (TEI '07). ACM, New York, NY, USA, 13–20. doi:10.1145/1226969.1226973
- [36] Henry Newton-Dunn, Hiroaki Nakano, and James Gibson. 2003. Block Jam: A Tangible Interface for Interactive Music. *Journal of New Music Research* 32, 4 (2003), 383–393. arXiv:https://www.tandfonline.com/doi/pdf/10.1076/jnmr.32.4.383.18852 doi:10.1076/jnmr.32.4.383.18852
- [37] James Patten and Hiroshi Ishii. 2000. A Comparison of Spatial Organization Strategies in Graphical and Tangible User Interfaces. In *Proceedings of DARE 2000 on Designing Augmented Reality Environments* (Elsinore, Denmark) (DARE '00). ACM, New York, NY, USA, 41–50. doi:10.1145/354666.354671
- [38] James Patten, Ben Recht, and Hiroshi Ishii. 2002. Audiopad: a tag-based interface for musical performance. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (Dublin, Ireland) (NIME '02). National University of Singapore, SGP, 1–6. doi:10.5555/1085171.1085175
- [39] James Patten, Ben Recht, and Hiroshi Ishii. 2006. Interaction techniques for musical performance with tabletop tangible interfaces. In *Proceedings of the ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (Hollywood, California, USA) (ACE '06). Association for Computing Machinery, New York, NY, USA, 27–es. doi:10.1145/1178823.1178856

- [40] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-Defined Gestures for Augmented Reality. In *Human-Computer Interaction – INTERACT 2013*, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 282–299.
- [41] Pol Pla and Pattie Maes. 2013. Display Blocks: A Set of Cubic Displays for Tangible, Multi-perspective Data Exploration. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction* (Barcelona, Spain) (TEI '13). ACM, New York, NY, USA, 307–314. doi:10.1145/2460625.2460677
- [42] G. Michael Poor, Samuel D. Jaffee, Laura Marie Leventhal, Jordan Ringenberg, Dale S. Klopfer, Guy Zimmerman, and Brandi A. Klein. 2016. Applying the Norman 1986 User-Centered Model to Post-WIMP UIs: Theoretical Predictions and Empirical Outcomes. *ACM Trans. Comput.-Hum. Interact.* 23, 5, Article 30 (October 2016), 33 pages. doi:10.1145/2983531
- [43] Ivan Poupyrev. 2000. Augmented Groove : Collaborative Jamming in Augmented Reality. *Proc. of Int. ACM Conf. SIGGRAPH'2000, Abstracts and Applications* (2000). https://modin.yuri.at/tangibles/data/ATR_poupyrev_augmentedgroove.pdf
- [44] Anders Reuter. 2022. Who let the DAWs Out? The Digital in a New Generation of the Digital Audio Workstation. *Popular Music and Society* 45, 2 (2022), 113–128. arXiv:<https://doi.org/10.1080/03007766.2021.1972701> doi:10.1080/03007766.2021.1972701
- [45] Benjamin Salem and Harold Peeters. 2007. InterCUBE: A Study into Merging Action and Interaction Spaces. In *Human-Computer Interaction – INTERACT 2007*, Cécilia Baranauskas, Philippe Palanque, Julio Abascal, and Simone Diniz Junqueira Barbosa (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 57–70.
- [46] Bert Schiettecatte and Jean Vanderdonck. 2008. AudioCubes: A Distributed Cube Tangible Interface Based on Interaction Range for Sound Design. In *Proceedings of the 2Nd International Conference on Tangible and Embedded Interaction* (Bonn, Germany) (TEI '08). ACM, New York, NY, USA, 3–10. doi:10.1145/1347390.1347394
- [47] Orit Shaer and Eva Hornecker. 2010. Tangible User Interfaces: Past, Present, and Future Directions. *Foundations and Trends in Human-Computer Interaction* 3, 1–2 (2010), 4–137. doi:10.1561/1100000026
- [48] Ehud Sharlin, Yuichi Itoh, Benjamin Watson, Yoshifumi Kitamura, Steve Sutphen, and Lili Liu. 2002. Cognitive Cubes: A Tangible User Interface for Cognitive Assessment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) (CHI '02). ACM, New York, NY, USA, 347–354. doi:10.1145/503376.503438
- [49] J. G. Sheridan, B. W. Short, K. Van Laerhoven, N. Villar, and G. Kortuem. 2003. Exploring cube affordance: towards a classification of non-verbal dynamics of physical interfaces for wearable computing. In *Proceedings of IEE Eurowearable* (Birmingham, UK) (Eurowearable 2003). IEEE Press, Piscataway, NJ, USA, 113–118. doi:10.1049/ic:20030156
- [50] Hyunyoung Song, François Guimbretière, Michael A. Ambrose, and Carl Lostritto. 2007. CubeExplorer: An Evaluation of Interaction Techniques in Architectural Education. In *Human-Computer Interaction – INTERACT 2007*, Cécilia Baranauskas, Philippe Palanque, Julio Abascal, and Simone Diniz Junqueira Barbosa (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 43–56.
- [51] Hideyuki Suzuki and Hiroshi Kato. 1995. Interaction-level Support for Collaborative Learning: AlgoBlock&Mdash;an Open Programming Language. In *The First International Conference on Computer Support for Collaborative Learning* (Indiana Univ., Bloomington, Indiana, USA) (CSCL '95). L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 349–355. doi:10.3115/222020.222828
- [52] Lucia Terrenghi, Matthias Kranz, Paul Holleis, and Albrecht Schmidt. 2006. A Cube to Learn: A Tangible User Interface for the Design of a Learning Appliance. *Personal Ubiquitous Comput.* 10, 2-3 (January 2006), 153–158. doi:10.1007/s00779-005-0025-8
- [53] Daniela Gorski Trevisan, Jean Vanderdonck, and Benoît Macq. 2004. Conceptualising mixed spaces of interaction for designing continuous interaction. *Virtual Reality* 8, 2 (01 Jun 2004), 83–95. doi:10.1007/s10055-004-0140-2
- [54] Brygg Ullmer, Hiroshi Ishii, and Robert J. K. Jacob. 2005. Token+Constraint Systems for Tangible Interaction with Digital Information. *ACM Trans. Comput.-Hum. Interact.* 12, 1 (March 2005), 81–118. doi:10.1145/1057237.1057242
- [55] Michael Urbanek and Florian Güldenpfennig. 2017. Tangible Audio Game Development Kit: Prototyping Audio Games with a Tangible Editor. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction* (Yokohama, Japan) (TEI '17). ACM, New York, NY, USA, 473–479. doi:10.1145/3024969.3025077
- [56] Consuelo Valdes, Diana Eastman, Casey Grote, Shantanu Thatte, Orit Shaer, Ali Mazalek, Brygg Ullmer, and Miriam K. Konkel. 2014. Exploring the Design Space of Gestural Interaction with Active Tokens Through User-defined Gestures. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (CHI '14). Association for Computing Machinery, New York, NY, USA, 4107–4116. doi:10.1145/2556288.2557373
- [57] Kristof Van Laerhoven, Nicolas Villar, Albrecht Schmidt, Gerd Kortuem, and Hans Gellersen. 2003. Using an Autonomous Cube for Basic Navigation and Input. In *Proceedings of the 5th International Conference on Multimodal Interfaces* (Vancouver, British Columbia, Canada) (ICMI '03). ACM, New York, NY, USA, 203–210. doi:10.1145/958432.958472
- [58] Michael Vazirgiannis, Yannis Theodoridis, and Timos Sellis. 1998. Spatio-temporal composition and indexing for large multimedia applications. *Multimedia Systems* 6, 4 (01 Jul 1998), 284–298. doi:10.1007/s005300050094
- [59] Roberto Verdecchia, Emelie Engström, Patricia Lago, Per Runeson, and Quynh Song. 2023. Threats to validity in software engineering research: A critical reflection. *Information and Software Technology* 164 (2023), 107329. doi:10.1016/j.infsof.2023.107329
- [60] Santiago Villarreal-Narvaez, Arthur Sluÿters, Jean Vanderdonck, and Efrem Mbaki Luzayisu. 2022. Theoretically-Defined vs. User-Defined Squeeze Gestures. *Proc. ACM Hum. Comput. Interact.* 6, ISS (2022), 73–102. doi:10.1145/3567805

- [61] Daniel Vogel and Géry Casiez. 2011. Conté: multimodal input inspired by an artist's crayon. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) (*UIST '11*). Association for Computing Machinery, New York, NY, USA, 357–366. doi:10.1145/2047196.2047242
- [62] Ryoichi Watanabe, Yuichi Itoh, Masatsugu Asai, Yoshifumi Kitamura, Fumio Kishino, and Hideo Kikuchi. 2004. The Soul of ActiveCube: Implementing a Flexible, Multimodal, Three-dimensional Spatial Tangible Interface. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (Singapore) (*ACE '04*). ACM, New York, NY, USA, 173–180. doi:10.1145/1067343.1067364
- [63] Alexander Wiethoff, Robert Kowalski, and Andreas Butz. 2011. inTUIT: Simple Identification on Tangible User Interfaces. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction* (Funchal, Portugal) (*TEI '11*). ACM, New York, NY, USA, 201–204. doi:10.1145/1935701.1935741
- [64] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) (*CHI '09*). ACM, New York, NY, USA, 1083–1092. doi:10.1145/1518701.1518866
- [65] Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell, and Anders Wesslén. 2012. *Empirical Strategies*. Springer Berlin Heidelberg, Berlin, Heidelberg, 9–36. doi:10.1007/978-3-642-29044-2_2
- [66] Jibin Yin and Hua Liu. 2011. The Empirical Study of Input Modalities Sensed from Flipping a Cube. In *Proceedings of the 4th International Conference on Intelligent Networks and Intelligent Systems* (Kuming, China). IEEE Press, Piscataway, NJ, USA, 5–8. doi:10.1109/ICINIS.2011.61
- [67] ZhiYing Zhou, Adrian David Cheok, and JiunHorng Pan. 2004. 3D Story Cube: An Interactive Tangible User Interface for Storytelling with 3D Graphics and Audio. *Personal Ubiquitous Comput.* 8, 5 (September 2004), 374–376. doi:10.1007/s00779-004-0300-0