Drag & Speak: A Multimodal Audio-haptic Approach for Arithmetic Operations for Screenreader Users

Author 1, Author 2, Author 3
Institution

1. Introduction

Many STEM (Science, Technology, Engineering and Mathematics) disciplines are dependent on non-linear notations, as characters and symbols are combined with graphics to write formulae and develop diagrams. Already basic arithmetic operations can be challenging to learn for students who are blind or have low vision. Addition, subtraction, multiplication, and division require placing digits and numbers both horizontally and vertically aligned to each other when reading and writing calculations.

Although screenreaders can indicate digits verbally, it is cumbersome to explore numbers spatially by text-to-speech output, in particular as empty space is neglected by screenreaders. As an alternative, mathematical braille and braille displays are suitable to write and read the terms needed including spaces. Still, braille displays are not presenting a two-dimensional layout. If students fail to learn such basic concepts in mathematics, their success in mastering STEM subjects can be limited.

Other non-visual approaches are needed to combine both visual and non-visual design of interactive stems for learning how sighted people solve mathematical tasks. We present a novel multimodal approach to basic arithmetic operations in a tabular layout by integrating voice input with placement of tangibles and provide feedback to the student with text-to-speech output as well as sonification. Such an approach may be extended to more complex notations and diagramming techniques.

2. Related Work

Both tactile as well as acoustic modalities can contain information about two-dimensional layout. Spatial sound can present spoken text in a horizontal layout if Header Related Transfer Functions (HRTFs) are implemented [4] but vertical resolution is rather limited and depends on individual preferences. Large tactile displays [8] can hold only a few lines with only some characters. The Hyperbraille display shows 12 lines with 40 characters each but is not easily affordable. [7] have developed one of the first eLearning systems to teach geometrical concepts on a tactile display with 7200 pins.

Low-cost physical bricks can carry braille labels and allow for tactile interaction, if accidental movements can be avoided. The Brannan Cubarithm Kit Slate 100 Cubes [2] allows placement of the braille labelled cubes in a tabular rubber-based frame.

Breiter et. al. suggest also groves in a plate to stabilize the position of wooden blocks [3]. In this work the concept of tangible objects is applied to allow digitalisation of the editing process. The Tangible Grid [6] consists of a tactile grid layout to design web pages by placing physical objects.

In a digital system individual feedback can be generated to support learners. Ali et. al. have developed a system guiding students who are blind or have low vision through structural information in a graphical user interface and a screenreader for solving mathematical tasks. They show an increase of efficiency but do not focus on effectiveness and learning.

The Tactonom[1] is built around affordable components: tactile feedback from microcapsule paper, a camera mounted above and a casing with a small computer generating text-to-speech output for regions identified through a finger. Similar audiohaptic devices are Talking Tactile Tablet and IVEO, but these rely on more expensive touch input sensing devices.

In this work, we extend the concept of a Tactonom by introducing tangible objects and real-time hand tracking as described below and evaluate the system with students who are blind or have low vision.

[1] https://www.tactonom.com/

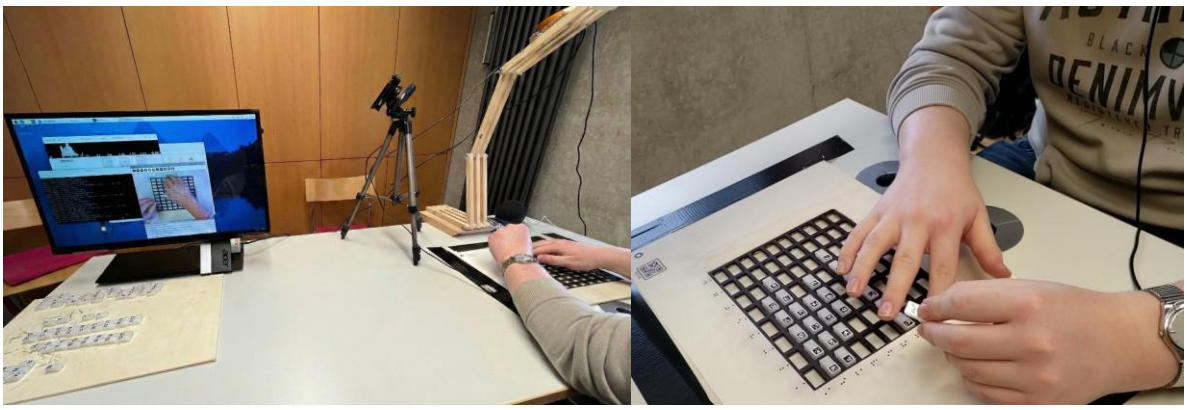

Fig. 1. System Overview. (a) The setup showing the custom wooden camera stand (right) and the experimenter's monitor (left) displaying real-time object detection and hand tracking, with a set of sorted tangibles arranged on a board in the left foreground. (b) Close-up view of a user exploring the tactile grid and placing 3D-printed tangibles.

3. Methodology

3.1 System Implementation

To achieve a low-cost, replicable, and portable solution, our prototype is built around a Raspberry Pi 4, which serves as a standalone computational unit. As shown in Fig. 1a, the hardware setup includes a custom, height-adjustable wooden stand equipped with a top-down Arducam camera. This design ensures that the tactile grid remains fully within the camera's Field of View for stable recognition.

The software architecture is implemented in Python. The computer vision module uses OpenCV for the parallel, real-time detection of multiple ArUco markers (position and ID). Concurrently, the MediaPipe framework is integrated for Hand Landmark Detection. This enables the system to differentiate between "palm occlusion" and deliberate "object removal," while also supporting deictic interaction via fingertip tracking. The recognition feedback is visualized in the monitor view shown in Fig. 1a.

For voice interaction, the system balances accuracy with processing speed by using low-latency online Automatic Speech Recognition (ASR) for user commands and a lightweight, local offline Text-to-Speech (TTS) engine for system feedback. This configuration ensures synchronization between user actions and system responses.

3.2 Tangible and Grid Design

The physical layer underwent several design iterations to optimize the haptic experience and recognition efficiency.

Tangibles: We replaced off-the-shelf solutions (e.g., Lego or wooden blocks) with custom 3D-printed tangibles. This approach reduced weight and allowed for precise sizing to match the ArUco markers. Compared to standard QR or Micro QR codes, ArUco markers demonstrated greater robustness and speed in multi-object tracking scenarios. To support tactile identification, the top surface of each tangible features a split layout: one section displays the ArUco marker for the camera, while the adjacent section features 8-dot Computer Braille printed on transparent foil. Unlike traditional 6-dot braille, the 8-dot format enables a more compact encoding of single digits and operators.

Grid: Early prototypes using Swell Paper and magnetic tangibles caused issues with unintended magnetic attraction and repulsion, as well as accidental displacement. Consequently, the final design uses a laser-cut thin wooden board. Each cell is a cutout with a diameter slightly larger than the base of the tangible, creating passive haptic constraints. This physical slotting mechanism stabilizes the tangibles without magnets while facilitating easy removal. Users can perceive the grid structure by touching the cutouts (see Fig. 1b) and identify the row and column coordinates via Braille labels on the edges of the board.

3.3 Interaction and Voice Assistance

The core interaction follows a "Prompt-Action-Feedback" loop. The system guides the user by providing a task prompt to set the context (e.g., "Set up the operands"), followed by specific action instructions (e.g., "Place number 2 at position B2"). The user then places a tangible, and the system verifies the position (Row, Col) and ID via computer vision. This allows for real-time error correction or confirmation.

To reduce cognitive load, a voice assistant is integrated, activated by the wake word "Hello Assistant". It supports four intent categories:

- **Task Prompting:** "What is the next step?"
- **Action Instruction:** "What is the next action"
- **Contextual Deictic Query:** Users can point to a tangible and ask, "What is this?". The system, using MediaPipe fingertip tracking, identifies not only the digit but also its place value context (e.g., "This is digit 2, representing the hundreds place in 256").
- **Calculation Support:** Queries such as "What is 5 times 9?" serve to reduce the mental arithmetic load.

The system is designed for robust interruptibility. User queries can interrupt TTS output at any time. Once the Q&A is completed, the system automatically resumes the previous state to ensure continuity. Similarly, if a navigation task is interrupted, it is automatically restarted.

4.User Study

We conducted a user study at the Educational Centre for the Blind and Visually Impaired (BBS Nürnberg) with eight participants who are blind or have low vision. Participant demographics, including their visual acuity and preferred calculation methods, are detailed in Table 1. The experimental task consisted of standard written arithmetic operations (e.g., 23×45), performed using the tangible grid layout. We used a within-subjects design with three experimental conditions:

- **V1 (Baseline):** The system provided step-by-step task prompts, action instructions, and verification feedback. Users relied solely on tactile exploration for spatial orientation on the grid.
- **V2 (Voice Assistant)**: This condition built upon V1 by adding the voice assistant, allowing users to ask questions (Q&A), but without active finger navigation.
- **V3 (Full Support):** This condition built upon V2 by adding finger navigation. Before each action, the system guided the user's finger to the target cell using clock-face directional cues (e.g., "9 o'clock") and sonification, where the pitch frequency increased as the finger approached the target.

We collected both quantitative and qualitative data, including demographic questionnaires and post-task evaluations. Subjective workload was measured using the NASA-TLX, and usability was assessed via the System Usability Scale (SUS). Additionally, custom 5-point Likert scales were used to evaluate the perceived effectiveness of the guidance and feedback mechanisms.

Table 1. Participant demographics. G: Gender (M=male, F=female), Age: in years, Hand: Handedness (R=Right, L=Left, B=Both), VA: Visual Acuity (TB = totally blind, LB = legally blind, VI = visually impaired), BRP: Braille Reading Proficiency, Calculation Method: Daily calculation habits

P	Age	G	Hand	VA	BRP	Calculation Method
P1	2004	M	B	TB	Advanced	Mental/Braille/App
P2	1995	F	R	VI	None	Mental
P3	2000	M	B	VI	Basic	Mental/Braille
P4	2008	F	R	TB	Advanced	Calculator/App
P5	2007	F	B	TB	Advanced	Calculator/App
P6	1998	F	R	TB	Advanced	App
P7	2005	F	L	TB	Advanced	Calculator/App
P8	2007	M	R	TB	Advanced	Audio/App

5. First Results

The overall usability of the system was evaluated using the System Usability Scale (SUS). As shown in Fig. 2, individual scores ranged from 57.5 to 95.0 with a mean score of 75.9 (SD=13.8). According to Bangor et al. [8], this falls into the “Good” range (Grade B). In our sample, this supports the feasibility of combining tangible interaction with audio-tactile feedback for blind and visually impaired learners.

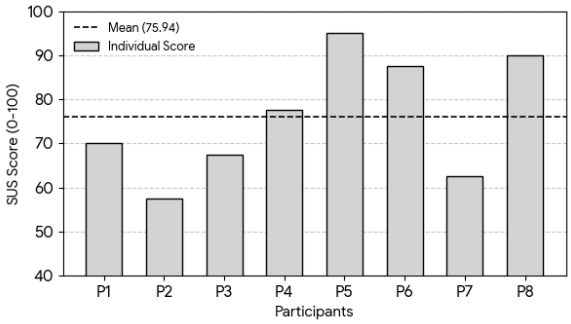

Fig. 2. Individual System Usability Scale (SUS) scores for the eight participants. The dashed line represents the mean score.

Subjective workload was analyzed using the NASA-TLX (see Fig. 3). The three conditions revealed distinct workload profiles. V1 (Baseline) and V2 (Voice Assistant) showed similar workload levels across most dimensions, with V2 yielding slightly more favorable scores in Mental Demand (M=2.38) and Performance (inverted score M=1.25).

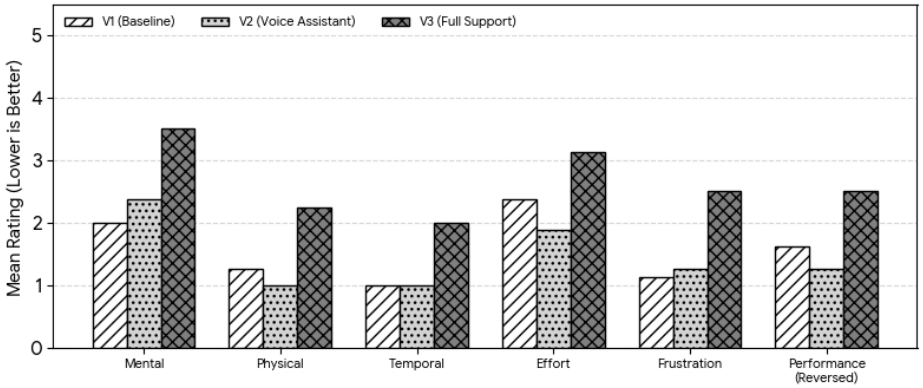

Fig. 3. Mean NASA-TLX scores across six dimensions for the three conditions: V1 (Baseline), V2 (Voice Assistant), and V3 (Full Support). Note: The "Performance" scale is inverted, so that for all dimensions, a lower bar indicates a better user experience.

Conversely, V3 (Full Support) exhibited higher workload ratings across all dimensions, receiving the highest Frustration rating (M=2.50). Qualitative feedback identified system latency as the primary cause for this increased stress. P1 explained that despite the guidance cues, it was difficult to confirm the correct position, often resulting in "circling" the target (e.g., receiving cues for 9 o'clock, then 6, then 3) because the audio feedback lagged behind the hand movement. Similarly, P3 found the continuous sonification "annoying" due to delays. However, several participants noted that they adapted to the system's latency after a few attempts. By deliberately slowing down their finger movements, the navigation function became usable. This suggests that the issues with V3 stem primarily from hardware performance limitations (latency) rather than the interaction concept itself. Despite these technical limitations, the physical and interactive design received positive validation. P7 explicitly commended the stability of the grid and noted that laser-cut holes held the tangibles more securely than magnets.

6. Discussion and Outlook

The study validates that the physical grid design and tangible interaction effectively support blind and visually impaired learners in understanding spatial arithmetic structures. Future work could focus on technical optimizations to reduce system latency, thereby significantly improving the overall user experience.

References

1. Ali, A., Khusro, S., Algamdi, S. A.. Accessible interactive learning of missing-digit arithmetic problems for students with visual disabilities. Scientific Reports 15.1, pp. 17804 (2025)
2. Brannan Cubarithm Kit Slate 100 Cubes. (2025) URL: https://www.maxiaids.com/product/brannan-cubarithm-kit-slate-100-cubes
3. Breiter, Y., Karshmer, A., & Karshmer, J.. Automathic blocks usability testing phase one. In International Conference on Computers for Handicapped Persons (pp. 191-195). Berlin, Heidelberg: Springer Berlin Heidelberg (2012)
4. [anonymized]
5. [anonymized]
6. Li, J., Yan, Z., Jarjue, E.H., Shetty, A. Peng, H. Tangiblegrid: Tangible web layout design for blind users. In: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology. pp. 1–12 (2022)
7. Schweikhardt W, Fehrle T, 1986. A computer-based drawing station for the blind (Ein Rechner unterstützter Zeichenplatz für Blinde, in German). In Proceedings of 5th International Workshop Computerized Braille Production, 30 October-1 November 1985, Winterthur, Switzerland, 251-261
8. [anonymized]
9. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 24(6), pp. 574–594 (2008)