Field report for Platform mBox
Designing an Open MMLA Platform
Li, Zaibei; Jensen, Martin Thoft; Nolte, Alexander; Spikol, Daniel

Published in:
LAK24 Conference Proceedings

DOI:
10.1145/3636555.3636872

Publication date:
2024

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Field report for mBox: Designing an Open MMLA Platform

Zaibei Li
zali@di.ku.dk
University of Copenhagen
Copenhagen, Denmark

Alexander Nolte
a.u.nolte@tue.nl
Eindhoven University of Technology
Eindhoven, Netherlands
Carnegie Mellon University
Pittsburgh, PA, USA

Martin Thoft Jensen
mtj@di.ku.dk
University of Copenhagen
Copenhagen, Denmark

Daniel Spikol
ds@di.ku.dk
University of Copenhagen
Copenhagen, Denmark

ABSTRACT
Multimodal Learning Analytics (MMLA) is an evolving sector within learning analytics that has become increasingly useful for examining complex learning and collaboration dynamics for group work across all educational levels. The availability of low-cost sensors and affordable computational power allows researchers to investigate different modes of group work. However, the field faces challenges stemming from the complexity and specialization of the systems required for capturing diverse interaction modalities, with commercial systems often being expensive or narrow in scope and researcher-developed systems needing to be more specialized and difficult to deploy. Therefore, more user-friendly, adaptable, affordable, open-source, and easy-to-deploy systems are needed to advance research and application in the MMLA field. The paper presents a field report on the design of mBox that aims to support group work across different contexts. We share the progress of mBox, a low-cost, easy-to-use platform grounded on learning theories to investigate collaborative learning settings. Our approach has been guided by iterative design processes that let us rapidly prototype different solutions for these settings.

CCS CONCEPTS
- Applied computing → Collaborative learning; • Human-centered computing → Interaction design process and methods; • Multimodal Learning Analytics, Human-centered computing, Open Learning Analytics Tools, Scenario-based design;

KEYWORDS
Multimodal Learning Analytics, Sociometric Wearable Devices, Prototyping

1 INTRODUCTION
Multimodal Learning Analytics (MMLA) has shifted from an emerging field to a stable part of learning analytics over the last decade. MMLA has shown promise in capturing the rich multiple streams of interaction data that make up learning and collaboration [13]. With the advent of low-cost and high-performance sensors, MMLA has become more accessible for educational research [4]. This is especially important for research into group work, where gaining better insights into collaborative learning is a key goal for research and education. Yet, designing, enacting, and assessing group work is complex and challenging [21]. Using MMLA can provide new insights into how people cooperate in these collaborative educational activities and how to design better activities that engage learners [28].

However, the technical infrastructure and the system behind collecting multiple modalities of interactions like proximity, patterns of interaction including physical engagement, patterns of speech, and physiological data are complex and tend to be specialized. Commercial systems can be beyond the cost of researchers or limited to individual subjects [9]. In contrast, systems developed by researchers are often specialized and are hard for other teams to deploy and use. Such specialization poses challenges, particularly an absence of user-friendly and adaptable systems that can be easily integrated into educational settings for robust data collection and analysis. Over the last decade, researchers have explored different approaches using both systems developed in-house and commercial systems; however, these platforms have yet to be scalable, open, lower cost, and easy to maintain. Therefore, opportunities exist to create more affordable, open source and open hardware and easier-to-use systems [35]. Benefits from easier-to-use systems would provide shareable data sets for establishing shareability and replication for the field.
1.1 Research Aim
To address these gaps, we aim to create a platform that can be used in different configurations and contexts and is shareable and usable with other researchers. The main research question is as follows: How do we design and prototype an MMLA platform that supports group work across diverse learning contexts? The platform needs to provide meaningful data that aligns with and potentially helps develop educational theories. Additionally, the platform needs to be open, useable, and scalable.

This paper introduces mBox, an MMLA platform under development designed to make it easier to research group work. mBox enables various sensors to connect via local networks (Wi-Fi and LAN) with base stations and servers that process, analyze, store, and share diverse data types. The system utilizes badges equipped with sensors and tags. Participants wear these badges, which collect raw data while minimally interfering with participants’ activities, and the badges are traceable with unique badge IDs aligned with the wearers’ identities. These badges are integrated with Internet of Things (IoT) technology and offer near-real-time monitoring of participant activities across multiple modalities. Furthermore, the mBox system is designed for easy scalability; it can accommodate additional participants simply by adding more badges. Our approach investigates how using different types of smart badges can yield rich data while preserving an acceptable level of learner privacy.

1.2 Paper Organization
The paper is organized as follows: section 2 briefly presents a background of different MMLA platforms and systems, shows intelligent badges, and discusses theoretical concerns. Section 3 outlines our methodological approach of sketching with technology while being grounded with theory to provide a holistic approach. Section 4 presents the iterative development cycles of mBox. Section 5 details the proposed system requirements. Finally, section 6 presents the discussion and the next steps.

2 BACKGROUND
2.1 Different MMLA Platforms/Systems
This section provides an overview rather than a systematic mapping or scoping review of the different MMLA platforms based on experience. For more systematic reviews, see the following work [2, 4, 25, 26]. MMLA research has used different platforms created for research, commercial, and physiological systems to capture, process, and visualize data. The Social Signal Interpretation Framework (SSI) [32] and Platform Situated Intelligence (PSI) [5] are designed for multiple uses that range from robots to performance and include mixed reality systems. The continued development of these two platforms has needed to be more consistent, and getting them set up requires specific platforms and expertise. From the research side, Multimodal Learning Hub [29], Oral Presentation Automated Feedback (OPAF) [22], Sensor-based Regulation Profiler Web Services [34], and Kunitron [12] have developed specialized systems that have provided multi-modal data for investigating diverse learning situations. However, these systems are specialized to the specific learning contexts with specific technologies. Commercial systems that use multi-modal data streams for behavioral observation have also been used for learning analytics, such as Noldus and iMotions [10, 36]. However, these systems are more tuned for individual research focusing on observational tools. Commercial tools generally offer video analysis and data synchronization of physiological sensors from human and animal behaviors. These platforms are costly and not geared directly for learning.

2.2 Badges
Sociometric wearable devices (SWDs) for investigating patterns of human behavior and interaction have been investigated from the early 2000s forward by Lederman and colleagues [20] with the Rhythm and Open Badges systems that developed an open-source framework based on a modular system. The system allowed researchers to collect, monitor, and analyze interaction data from people in real-life settings. The Open Badges system relied on voice analysis, proximity, and biometric sensors (vibration) to analyze group interactions in business and social settings.

In more recent work, Ito-Masui and colleagues examined the utility of sociometric wearable devices (SWDs) in corporate and healthcare environments [16]. While their primary focus differed from studies on educational collaboration, the multimodal data sources used are comparable. A key comparison between commercial SWDs is their technical strengths and limitations. Notably, in corporate contexts, they enhance teamwork and resource strategies. However, the study emphasizes the need to be aware of the devices’ technological constraints. Yamaguchi and colleagues have introduced the Sensor-based Regulation Profiler, which offers a method to discern and depict collaboration phases in collaborative learning scenarios automatically [34]. These badges use a business card-sized sensor worn around a participant’s neck, employing infrared detection for device interaction, audio-sensing for determining speakers, and an RF module for time coordination. The benefits of SWDs in education present opportunities for collecting data with more control for privacy and limited facial recognition technologies.

2.3 Theory and Technology
When examining how theory is used to guide the design of Learning Analytics, we can see tensions between technology, research, and practice. Wise and Shaffer [33] have argued the need to integrate a stronger theoretical grounding into learning analytics. Ochoa and colleagues [23] have presented a strong case for learning analytics to engage more in the critical and social aspects of how we impact education [1, 13, 17].

The development and use of Multimodal Learning Analytics platforms and systems provide diverse research opportunities. However, the field still needs systematic mappings and comprehensive comparative studies of the different types of MMLA systems to gauge their efficacy and versatility across varied learning contexts to develop more adaptable and universally applicable platforms.

The specialized nature of existing systems necessitates exploration into creating systems with broader applicability across different learning environments. The integration and alignment of theory in designing learning analytics represent another key ground for
investigation, aiming to reconcile existing tensions between technology, research, and practice in MMLA. Moreover, developing easier-to-use, more affordable, and adaptable systems that can used by more researchers across diverse learning scenarios can enrich the MMLA field. Smart badges and sociometric wearable devices (SWDs) can provide more scalable ways to provide privacy without excessive facial recognition and other technologies.

3 METHODOLOGICAL APPROACH

This work iteratively designed an MMLA platform that provides an easy-to-use system to investigate group work. The design of mBox is grounded in learning theories (Embodied Cognition, Cognitive Load Theory, and Control-Value Theory of Achievement Emotions) [13], clarifying the data types essential for learning analytics and directly influencing the modular design of our platform, which encompasses distinct base stations and badges. We followed a broad design science approach [15, 31] during the development, using the pragmatic approaches of agile software development and sketching with technology [6]. Since we are in the initial phase of developing the platform, we focused on 1) awareness of problems, 2) Suggestions for solutions, 3) development of prototypes, and 4) evaluation of outcomes through a series of sequential cycles. These steps aim to start identifying design principles for the platform [14].

The platform’s initial design, development, and evaluation were broken into five main cycles (details in the next section). Cycle 1, Proto Vision, focused on exploring the smart badges and computer vision. Cycle 2 focused on the audio analysis tools Proto-Audio, which investigated conversational patterns with a focus on speaker diarization tasks. After these two cycles, we performed various evaluations (Cycle 3) and began to refine and combine the two strands in Cycle 4 for Platform-Alpha. The aim of Cycle 4 was to develop an integrated system that combines vision badges and audio analysis. To achieve this, we refined the existing vision badges and introduced a new voice badge for audio capture. Concurrently, we developed base stations to identify these badges within a given space and process the associated audio data. We also implemented a time-series database and visualization tools. For Cycle 5, we conducted different evaluations that formed the current results, Platform-Beta. These evaluations were inspired by scenario design and embodied interaction to align with the needs of the learners [3, 8].

4 CYCLES

4.1 Proto Vision

For the initial investigation, we built an SWD around the Arduino Nicla Vision board, a compact microcontroller capable of running different computer vision models for face and object detection. Technical limitations and privacy concerns led us to explore using fiducial markers that proved unique identification without the computational overhead. This approach promised improved performance, enhanced privacy, and ensured unbiased identification, opening avenues for future enhancements. We adopted the AprilTag fiducial marker due to its optimized detection algorithm and compatibility with Nicla Vision. Comparative tests showed superior distance and accuracy performance than the initial face detection model while consuming fewer resources, solidifying our decision to proceed with AprilTag. Figure 1 section A) shows the initial badge design featuring a 6x6cm AprilTag and the Nicla Vision board.

Then, we focused on designing a simple system that allowed us to test multiple badges with a camera-affixed base station. This concept allowed the base station to recognize the badges around a space, revealing who is around the table. At the same time, the badges identified which person was looking at each other. Bluetooth Low Energy (BLE) was the communication protocol between badges and the base station. This phase highlighted the feasibility of implementing real-time, robust data transmission and presentation strategies for future systems, with the prototype base station operating on a Raspberry Pi 4 model B, as depicted in Figure 1 section B.

Finally, we aimed to develop a refined prototype or “tech sketch” for the Proto-Vision module, allowing real testing and evaluation of using vision badges for collaboration analysis. The Nicla Vision board’s camera captured frames at a resolution of 160x120 pixels with a view angle of only 80°, requiring the precise positioning of badges at closer distances for optimal results. Consequently, the idea of a wide-angle-camera equipped base station was developed, primarily focusing on detecting and visualizing students/collaborators around a table. The Proto-Vision module comprised three vision badges and a base station connected to a monitor. The base station detected the existence of badges. It synchronized the onboard detection results from the badges, displaying real-time video and network graphs that depicted badge relationships on the monitor. This approach simplified the testing process and focused on evaluating badge interaction and detection capabilities.

4.2 Proto Audio

Audio modalities offer significant potential for analyzing collaborative dynamics in learning environments. In line with our design science methodology, the second cycle, “Proto Audio”, focused on audio analysis to study collaborative dynamics within learning environments. Traditional machine learning techniques have offered some insights into classroom interactions but lack the nuanced understanding required for deeper analysis.

We, therefore, turned to advanced deep learning techniques to explore audio analysis; emerging models like Titanet [18] offer promising avenues. Titanet is a novel model for speaker representation that utilizes global context and channel attention pooling to achieve state-of-the-art performance in speaker verification and diarization tasks. It leverages NeMo [19], an open-source conversational AI toolkit, allowing researchers to fine-tune AI models for specific audio tasks. Despite these advancements, deep learning techniques’ practical and scalable application in real-time or near-real-time educational settings remains under-explored.

Acknowledging these gaps in the literature, we introduced the Proto-Audio module for mBox as shown in figure 2. This lightweight voice analysis data pipeline leveraged deep learning and specialized audio processing techniques. In the audio pre-processing phase, the Proto-Audio module incorporated specialized audio processing techniques to clarify speech and differentiate voice from silence.

1 See https://store.arduino.cc/products/nicla-vision

2 See https://www.raspberrypi.com/products/raspberry-pi-4-model-b/
These included volume normalization, noise reduction, and voice activity detection (VAD) using existing models [11, 30].

After pre-processing, the TitaNet-L model served as the feature extractor. A 15-second audio clip from each participant was processed during the registration phase to generate a speaker embedding. These embeddings were securely stored in a local audio base station to maintain privacy. The system captured brief (1.5-second) audio segments for real-time speaker recognition, comparing them against stored embeddings using cosine similarity.

### 4.3 Evaluation of Proto-Vision and Audio

The third cycle focused on assessing the Proto system’s real-world efficacy in alignment with key objectives: cost-effectiveness, ease of deployment, and user-friendliness. This assessment was informed by user feedback and technical observations from activity conducted in Estonia.

Stakeholders participated in a collaborative activity using the Proto system. After the study ended, one key takeaway was that participants had minimal concerns regarding data privacy and security, as the system neither retained recordings nor tied data to individuals. The potential for research in collaboration analytics using unique badge detection and additional data modalities like hand movements and physiological data was recognized. Still, considerations regarding user comfort in wearing the badges throughout the day were discussed, leading to solutions like removing or turning off the badge when needed. Figure 1 section C shows the running test.

Technical limitations were also identified, requiring adjustments to both Proto-Vision and Proto-Audio modules. The Proto-Vision badge, which utilized Nicla Vision's onboard camera for AprilTag detection, suffered from the narrow field of view (80°) and low image resolution (160x120 pixels), limited by the device’s memory while running the AprilTag detection algorithm. This impacted the onboard detection’s accuracy and reliability. Additionally, communication between the vision badges and the base station via BLE proved unstable and limited in range. Furthermore, although the camera-affixed base station can detect the presence of badges, the wide-angle lens introduced distortion, compromising both position estimation and the accuracy of badge directional interaction. Similarly, onboard detection could not offer comprehensive details, such as tags’ precise location and orientation. The Proto-Audio module, which employed a Jabra device for recording, performed adequately in confined settings but faced limitations in larger spaces like classrooms due to many speakers. The Jabra’s effective sound pick-up range was limited to 2.3 meters, leading to sub-optimal voice capture for distant participants, wherein voices closer to the device were more distinctly recorded and thus more likely to be recognized. These findings informed subsequent development cycles as we continued to iterate on the system to better align with our initial research aims.

### 4.4 Platform Alpha

To address the abovementioned limitations, we developed the alpha version of mBox, incorporating key improvements. Both vision and audio modules are now synchronized via a local network, and we transitioned from BLE to Wi-Fi to ensure more reliable and expansive data transmission.

**Enhanced Vision Module.** To enhance AprilTag detection, we used an external webcam (Logitech C920) on a light stand connected to the vision base. This setup improved detection and provided additional data like translation and rotation vectors of tags. The outward normal vector $\mathbf{N}$ perpendicular to the tag’s surface, calculated from the tag’s rotation matrix, indicated the badge wearer’s orientation. We also computed vector $\mathbf{X}$, pointing from the center of one tag to another. Cosine similarity between $\mathbf{N}$ and $\mathbf{X}$ revealed if tags faced each other, indicated by a cosine value near -1.

**Optimized Audio Module.** To overcome the limitations of the Jabra device, we designed voice badges equipped with Nicla Vision, serving as individual recorders. This enhanced the system’s scalability and resolved the issue of distance limitations. We also incorporated real-time transcription using OpenAI Whisper [27].

**Communication Protocols.** Message Queuing Telemetry Transport (MQTT) was used for communication between the vision badges and the vision base station. At the same time, User Datagram Protocol (UDP) was employed for audio streaming between the voice badges and the audio base station.

**Data Synchronization and Storage.** We employed Redis as the broker for data synchronization across badges, facilitating data exchange between base station processes. We generated five key measurements from the vision and audio modules: participant orientations, participant locations, participant network, speaker recognition, and speech transcriptions. These measurements were stored in InfluxDB on the Uber server for real-time and post-hoc analyses.
4.5 Evaluation of Platform-Alpha

In a recent empirical evaluation of Platform-Alpha, conducted during a Hackathon event, participants engaged in a roundtable discussion to create a business plan. The study aimed to assess the system’s feasibility and stability. Preliminary findings showed that the system achieved an approximate 85% accuracy rate in speaker recognition under realistic conditions. Additionally, there was a 20-40 ms recording delay between consecutive 1.5s audio segments. The audio and vision modules were also effectively synchronized with real-time timestamps.

However, we identified limitations in the vision module, particularly concerning the single-camera setup. The camera’s limited 78° field of view and issues of light interference and human occlusion impaired the effective capture of all AprilTags. These findings indicated the necessity for future enhancements. Specifically, we proposed integrating an RFID proximity module and transitioning to a multi-camera system for synchronized AprilTag detection. Figure 3 illustrates the second-generation badges for vision and audio (Section A), the vision badges in action along with the visualization of group interaction (Section B), and the speaker diarization presented in pie charts, including a close-up segment displaying the speaker timeline (Section C).

5 RESULTS

Drawing from the previous cycles that started with Proto-Vision and Proto-Audio that ended with Platform-Alpha, we developed from the heuristic evaluations a set of emerging design guidelines that can act as starting principles to define Platform-Beta. Figure 4 details the four major modules of mBox Platform-Beta. The Data Acquisition module contains the external sensors and the different types of badges. The Data Processing module contains various base stations that handle (through the Data Storage and Exchange module) the processing and timestamping of the different data streams. The Data Storage and Exchange module manages data flow from acquisition sensors to processing bases, facilitates intercommunication among the bases, and provides a database for data storage. Finally, the Data Analysis and Visualization module offers visualization tools for analysis.

The architecture of Platform-Beta revolves around multi-faceted badges for on-person data collection and specialized base stations for data processing and synchronization. The badges come in four distinct types, each serving a specific purpose. Vision Badges are equipped with Nicla Vision and AprilTag, facilitating onboard AprilTag detection that aids in constructing a participant network graph. Voice Badges, also utilizing Nicla Vision, are designed to stream audio data to the audio base station for speaker and speech recognition. Regular Badges feature a basic AprilTag to provide essential data on location and orientation, while RFID Badges include an RFID tag to measure proximity.

Transferring from badges to base stations, these hubs are specialized units designed to handle specific data types. Vision Base Station, powered by either Raspberry Pi 4 or high-performance computers, processes visual data captured from multiple web cameras. The Vision Base Station detects AprilTags on badges to pinpoint their location and orientation. At the same time, Vision Badges contribute additional AprilTag data, enhancing the base station’s grasp of badge-to-badge spatial relationships. AprilTag detection results from base station webcams and badge onboard cameras are synchronized every second to form a network graph, elucidating participant spatial relationships. Similarly, the Audio Base Station, utilizing Raspberry Pi 4 or high-performance computers, handles audio data from either Jabra devices or Voice Badges. It synchronizes the recognition results from different Voice Badges to identify the most dominant speaker in each segment. Proximity Base Station uses an Arduino board or a computer to connect with a Simultaneous RFID Tag Reader to triangulate participants’ proximity.

The Uber Client is the system’s dashboard, providing two key visualizations. Real-Time Visualization is achieved by subscribing to MQTT/Redis topics or InfluxDB for instant insights, while Post-Time Visualization offers retrospective analysis based on archived InfluxDB measurements. Lastly, the Uber Server acts as the data backbone. It utilizes InfluxDB to retain time-series measurements and manages data traffic through MQTT and Redis brokers, ensuring a seamless flow of information across the system.

6 CONCLUSIONS & FUTURE WORK

This paper explores how to design and develop an MMLA platform to support group work across different contexts. We share the progress of mBox, which aims to create a low-cost, easy-to-use system grounded on learning theories to investigate collaborative learning. Driven by design science principles, mBox has demonstrated advancements in leveraging deep learning and IoT technologies to provide granular, moment-to-moment analytics for educational engagement. While the current version showcases
promising capabilities, continued development is essential to address the educational landscape’s demands fully.

Our commitment to advancing MMLA will guide mBox’s evolution into an even more versatile and effective tool for educational research. We envision its potential to unveil engagement patterns and learning efficacy and contribute to developing new, educationally grounded theories grounded in real-world data. mBox exemplifies cost-effective educational technology, with badge prices around 100 USD and base stations ranging from 35 to 1000 USD. Its design offers remarkable flexibility, accommodating stationary and dynamic learning environments through a scalable IoT network of badges and bases. This combination of affordability and adaptability makes mBox accessible to a broad spectrum of educational institutions and learning contexts. Our iterative prototyping process has yielded technological sketches enabling collaboration measurements from machine and human perspectives. Future work will focus on further development and thorough empirical validation of different modules, exploring incorporating natural language processing to enhance the analysis of transcribed text for deeper insights into participant interactions and behaviors.

ACKNOWLEDGMENTS

This project is partially funded by the Creative Impact Research Centre Europe (CIRCE). This work was (partially) conducted when the author, Nolte, was affiliated with the University of Tartu (Estonia). We also thank the participants of the Empowering Women Ukraine project.

REFERENCES
