Techniques for Augmented-Tangibles on Mobile Devices for Early Childhood Learning

Victor Cheung, Alissa N. Antle, Shubhra Sarker, Min Fan, Jianyu Fan, Philippe Pasquier
School of Interactive Arts and Technology, Simon Fraser University
Surrey, Canada
{vca45, aantle, ssarker, minf, jianyuf, pasquier}@sfu.ca

ABSTRACT
Integrating physical learning materials with mobile device applications may have benefits for early childhood learning. We present three techniques for creating a hybrid tangible-augmented reality (T-AR) enabling technology platform. This platform enables researchers to develop applications that use readily available physical learning materials, such as letters, numbers, symbols or shapes. The techniques are visual marker-based; computer-vision and machine-learning; and capacitive touches. We describe details of implementation and demonstrate these techniques through a use case of a reading tablet app that uses wooden/plastic letters for input and augmented output. Our comparative analysis revealed that the machine-learning technique most flexibly sensed different physical letter sets but had variable accuracy impacted by lighting and tracking lag at this time. Lastly, we demonstrate how this enabling technology can be generalized to a variety of early learning apps through a second use case with physical numbers.

Author Keyword
Early childhood learning; augmented reality; tangible interaction; tablets; education mobile apps.

CSS Concepts
•Human-centered computing~Interaction design~Systems and tools for interaction design

INTRODUCTION
Over the past decade, the education sector has been incorporating mobile devices (e.g., tablets, touch laptops) as part of its teaching tools, taking advantage of their interactivity and portability [30]. In particular, early education venues such as elementary classrooms are now using tablet-based applications for basic subjects including language learning, arithmetic, music, design, geometry and art. A similar trend can also be observed in informal environments including homes and museums, as evidenced by numerous mobile apps available for purchase and download [5, 45]. However, mobile device applications are largely limited to touch-based interactions with 2D digital (screen-based) content. This misses the opportunity to use techniques that would enable incorporation of physical objects (e.g., peg boards, letter blocks, geometric shapes), which are integral to many early childhood education approaches (e.g., Fröbel [9], Montessori [37]) and have shown to benefit learning in tangible learning applications (e.g., languages [3, 15, 19, 41], programming [20, 28, 33]). While there is a recent development in including physical objects in digital applications, this concept often requires specialty hardware (e.g., tangibles), which is typically inaccessible and unscalable due to high cost of replication and maintenance. Moreover, such hardware is often only useful for the specific application it was developed for; it is not customizable for other educational and research uses.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
IDC '20, June 21–24, 2020, London, United Kingdom
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-7981-6/20/06…$15.00
https://doi.org/10.1145/3392063.3394412

What is needed is an enabling technology platform that would support HCI researchers and educational developers to easily incorporate everyday physical learning objects (e.g., physical letters, numbers, arithmetic operators, geometric shapes) into smart device applications. Such an enabling system would bring together a set of capabilities around sensing, tracking and augmenting physical shape forms using mobile technology. Developing such a system requires technology development research [22] exploring different techniques that can be used to create this set of capabilities on mobile devices. This form of research does not have the direct goal of addressing or evaluating end-user interaction (e.g., through a usability study). Instead the research focuses on the development, implementation, evaluation of proof-of-concepts, and illustration of generalization through multiple cases. Once an enabling technology platform is shown to work well and generalizable to multiple cases, it can be released as open source so others can use it. This type of validated and open source enabling technology would make it easier, faster and more cost effective to develop and deploy hybrid physical-digital mobile applications for early childhood learning; and thus benefits both researchers, looking to better understand the benefits of hybrid systems; and developers wanting to use physical materials alongside digital applications.

There are several commercial products for mobile devices that combine physical objects with software applications. These products typically include a set of physical objects (e.g., letter tiles, shapes, toy pieces) with an app. For example, there are products for young children aimed at language learning, arithmetic principles and geometric shapes (e.g., Osmo [40], Tiggly [47], Marbotic [32]). However, our analysis of these products found that they do not fully utilize the potential benefits of physicality. For example, in the language learning Osmo app, the letters are represented as 2D symbols on tiles, rather than as 3D letter shapes, which enable tracing and can be distinguished tactilely as well as visually. Moreover, many of these commercial products are specialty products with single-use input objects, thus are not practical or cost-effective for classrooms and homes. Lastly, because they come "pre-packaged" they do not lend themselves well to research that evaluates their effectiveness and as a result most have not been validated under formal research studies.

In the academic domain, there has been considerable research evaluating the combination of physical objects and digital content. In most studies, the systems were designed specifically using custom hardware to explore particular research questions and/or to support specific research requirements (e.g., experimental control). For example, in the area of early reading systems, Tangible PhonoBlocks [3] used a touch tablet combined with a custom-made electronic letter-making platform, 3D plastic physical letters with embedded LEDs, and pogo pins. LinguaBytes [19] and Tiblo [41] also used custom hardware modules. These systems enabled researchers to conduct rigorous studies. Yet, the customized hardware in these prototypes made them difficult to scale beyond single-use research instruments or become available for use outside specific scenarios (e.g., labs, small field studies). Replicating such systems is costly, hindering uses by other researchers or widespread availability in schools and homes.

In this paper we report on our technology development research, motivated by the potential benefits of integrating physical objects with digital content for mobile device-based learning, and the challenges faced by this approach in terms of ease of development and broad deployment. We aim to introduce a platform that includes physical objects in mobile device-based educational applications, using materials that are readily available in classrooms or homes and augmented reality (AR) technology. We position this work as a technical development research in human computer interaction (HCI) that enables further research [22] in early childhood educational technology. We validate our research through four steps: (1) specifying a set of requirements for such an enabling platform; (2) developing three different techniques into proof-of-concept prototypes using an illustrative case of a mobile reading app with physical letters; (3) conducting a comparative analysis of the benefits and limitations of the techniques; and (4) selecting the "best" platform and showing generalizability to another application area (physical numbers).

The results from our investigation will enable HCI researchers and educational developers to create a variety of hybrid physical-digital learning applications on mobile devices, which can then be used to further explore the benefits of both hands-on interaction with physical materials and augmented reality (AR) in learning. We also contribute by suggesting future directions for research and development for this class of enabling technologies.

RELATED WORK

Early childhood education has been characterized by a long history of using concrete representations and physical objects in the learning process. This approach rests on the theoretical foundation that sensory-motor activity is critical to cognitive development [17]. That is, children develop cognitively from physical engagement in reasoning with materials in real world settings. Examples can be seen in the early pedagogical "hands-on" materials and curriculums of Dewey [12], Fröbel [9], and Montessori [37]. Based on predicted benefits of physicality and concreteness, researchers in HCI and educational technology began to explore if and how tangible user interfaces might provide benefits in early childhood learning applications. For example, in the late 1990's, the AlgoBlocks and Programming Bricks systems were developed to support the physical expression of programs through the constructive assembly of physical blocks [48].

Progression from Explorations to Application Validity

Since the early 1990's early stage research in tangible learning for children has moved from anecdotal reports, proof-of-concepts, and case studies to more rigorous empirical studies that examined both learning outcomes (e.g., viability, effectiveness) and learning processes (e.g., benefits of hands-on interaction). A mixed picture has emerged, in which some of the proposed benefits have been supported with evidence; largely in cases where hands-on physical manipulation was integral to learning and system were well-designed and theoretically grounded. For example, a tangible puzzle system that enabled hands-on interaction with materials was shown to scaffold spatial problem skills development through epistemic actions [2]. Results from several studies have shown that tangible systems for learning programming improve not only engagement and motivation but concepts related to learning outcomes (e.g., [28, 33]). Our study of a tangible system for early literacy acquisition showed viability and evidence that hands-on interaction with physical letter shapes augmented with dynamic colours led to gains in learning the alphabetic principle, through mechanisms including shape tracing, epistemic organization strategies, and attention to objects in hands [15]. Yet, there have continued to be calls to action to improve methodological rigor that links proposed benefits to empirical evidence [4, 55]. Much remains to be explored in hybrid physical-digital learning applications for children. An open-source platform that enables fast, simple prototyping with physical learning materials and digital applications would benefit this research agenda.

Making Early Childhood Learning More Accessible

In an effort to make systems more accessible in classrooms and homes, there has been research in tangibles that utilize mobile devices (e.g., tablets, smart phones) rather than desktops [30]. Some of these are research prototypes that employ commercially available systems such as Osmo (e.g., [21, 31, 44]), which sells kits including letter/number tiles, tangram pieces, and Fröbel's Sticks and Rings, along with accompanying apps and a custom-made mirror for the tablet to detect the objects [40]. Other systems include Tiggly [47] and Marbotic [32] which have toys shaped as letters and geometric shapes embedded with conductive materials, along with accompanying apps implemented to detect the objects and show associated digital content.

However, many of these systems use proprietary physical objects, and few of these systems have been rigorously deployed or evaluated in naturistic settings. What is needed is a platform for mobile devices that researchers can use to create prototypical learning applications and can also be scaled for broader deployments.

Mobile AR for Learning

Augmented Reality (AR) is a technology where the display of an otherwise real environment is augmented with virtual objects by means of computer graphics [34]. While AR has been an area of research interest for decades, it has only recently reached a level of maturity to be used in mobile devices [25]. For example, recent technological advancements and availability of development kits (e.g., Vuforia [51], ARToolkit [6]) have now made mobile AR accessible to HCI and educational technology researchers.

In preliminary studies, researchers have shown that mobile AR increases students' learning motivation in various subjects and across various ages, for example, visual art in middle school [13] and mathematics in elementary classes [10]. Of our particular interest is using mobile AR for early childhood learning. Recent research has demonstrated that mobile AR helps providing contextual and location-specific information to young learners [1] and promotes learning both inside [7, 8, 11] and outside [29, 43] of classrooms. More importantly, a special stream of mobile AR that combines physical objects and digital information has been proposed, and calls for more research in addressing its usefulness from a psychological perspective [10]. In a recent work by Yilmaz [54] that combined traditional toys and AR technology, the author reported that both the teachers and students enjoyed the combination, but also observed a lower cognitive attainment, which was linked to less cognitive effort exhibited by the students as they were mostly watching the multimedia content. This suggests that further work is needed to examine the use of physical objects with digital content, and the balance between them.

In light of the well-documented educational benefits of concrete representations and physical objects, and the diversity of learning areas and benefits of mobile AR, we are interested in the question: *What techniques can we use to create a robust, scalable hybrid tangible-AR tablet platform that can be used to develop and deploy a wide variety of early childhood education applications?*

REQUIREMENTS

We present four requirements for a hybrid tangible-AR enabling platform to develop early childhood learning apps that utilize readily available physical objects. Our scope is learning that utilizes symbols and/or shapes, which underlies a large variety of early childhood education application areas. To identify important aspects for this design space, our requirements were sourced from our own experience in tangible learning applications, prior work in early education, and literature in tangibles and augmented reality (e.g., [2, 10, 14, 15, 52]). Furthermore, to encompass a wider variety of applications (thus more generalizable), we focus on the system development rather than application-specific content. For example, we identify persistence in object tracking as a requirement rather than having requirements related explicitly to learning theory.

REQ1: Readily Available Physical Objects

One of the main issues with current commercial and research application systems for early childhood learning is the use of specialty hardware, thus the lack of ease of development, cost-effectiveness, scalability and broad availability. To address this issue, we require that the system should only consist of components that are readily available, either purchasable or makeable with few skills.

The application should be able to run in one or both of the most popular mobile operating systems: iOS and Android (installed in over 99% of tablets worldwide [46]) and should require no modification nor jailbreaking (i.e., unofficially escalating system privilege to use otherwise inaccessible features) to the device. This ensures that our enabling system will be widely accessible to researchers and developers; and the apps created will be accessible to teachers, parents and other caregivers to ensure accessibility in both research and real-world settings.

In addition, to keep the components minimal, the rest of the system should contain no more than the physical objects. These objects should be common in homes and classrooms and/or can be readily purchased/created by researchers, developers or even school teachers (e.g., cut out of thick cardboard, simple 3D printed projects).

REQ2: Hands-On Interaction with Physical Objects

To support hands-on interaction in learning [2, 4], the system should be able to accommodate sets of physical objects with the following properties. The physical objects should be easy to handle by children ranging in age from 4 to 8. They should also be easy to store and move around. For example, letters that are about 1-2.5 cm in both width and height and 0.5-0.8 cm in thickness would be suitable. Moreover, to facilitate tactile feedback [35] and tracing [15] in learning, the objects should be hi-fidelity; that is, they should be shaped in the same way as the shapes or symbols they represent (versus, for example, generic shapes with symbols printed on them). We envision the shape or symbol objects with the widest applicability in learning including common symbols like Latin lower case letters (e.g., a, b, c), Arabic numbers (e.g., 0, 1, 2), basic arithmetic operators (e.g., +, -, ×), musical notes (e.g., ♪, ♭, ♯) and shapes such as geometric shapes (e.g., □, △, ○). More complex ones like language specific accent marks (e.g., é, è, â) or a subset of Mandarin radical characters (e.g., 金, 木, 水) could also be supported. However, these more specialized objects are outside of the scope of our current study.

We also require that the physical objects used should withstand long-term use, be durable, and even washable [31]. Since they will be used by young children, they should be safe (e.g., made with materials that are certified as kids-safe). To maximize long term use and minimize the need for maintenance, they should also be passive objects rather than electronically augmented objects (e.g., Bluetooth).

REQ3: Sensing and Tracking of Physical Objects

Our third requirement relates to input. To interact, a child must be able to move the physical objects and the system must sense and track these objects in real time. Sensing involves accurately determining object attributes such as individual object size, shape, identity, position, and orientation. Some symbols have similar shapes but have completely different meanings (e.g., letters: i/l, d/p/q/b numbers: 2/5, or music notes: half/quarter). The system should also be able to accurately sense and distinguish these symbols. These sensed object attributes are required so that the learning application can provide feedback. For example, the system can determine if the correct object was placed. These attributes are also required so that the system can use AR to augment physical objects with digital content (see REQ4). For example, to augment a physical object with a colour overlay, the system must be able to register the digital overlay to the physical object using sensed attributes of position, orientation, shape, size of the physical object.

Tracking involves determining relevant attributes of an object over time (e.g., position, orientation). Moreover, the tracking should be persistent (i.e., maintained over time) so objects that are unchanged remain in the system, and those that are changed (e.g., adding, removing, relocating) can be detected without noticeable system lag. Since children often do not have precise motor control in putting multiple objects in a straight line, the system should be able to account for slightly haphazard placements. For example, it should be able to detect several letters placed in proximity (but not perfectly aligned) and group them into a word. Many early learning applications involve learning about relationships between multiple symbols (e.g., spelling, order). As such the system should be able to both sense and track multiple objects (typically 5-10) in real time.

REQ4: Digital Augmentation of Physical Objects

The fourth requirement is a key feature of AR. Digital content that augments a physical object must be displayed on the tablet registered to the location of the physical object. That is, digital overlays should be "attached" to physical objects and move with them. In the learning context, this requirement means that applications should be designed to reinforce the association between the digital and physical forms of representation. It is common in early learning to connect abstract and concrete representations [10]. For example, a digital overlay of the letter 'a' could be directly positioned on the tablet screen over the physical letter 'a' to show the association. The overlay could contain a dynamic colour, a pattern, a word containing 'a', a picture related to the letter 'a' or even a touch point, so that touching the digital 'a' creates the 'a' sound. For some of these features, it is important that the digital and the physical object have the same size, shape, position and orientation (i.e., are registered, as mentioned in REQ3) to show the association of the two.

As objects are moved, the system should use the tracking information of the objects to properly align their digital augmented contents on the tablet display. That is, as objects are moved, their digital representations should move with them as a consistent overlay. Consideration must be given to the placement of overlays with other displayed digital content as objects and their augmented overlays relocate.

THREE HYBRID TANGIBLE-AR TECHNIQUES

In this section, we report our development and analysis of three techniques that can be used to create systems enabling tangible interactions with tablet-based augmented reality (AR), specifically targeted to early education apps. We chose these techniques because of their variety and technical maturity, and their use of readily available features of most tablets (camera and touchscreen) and craft materials (wood/paper/cardboard/plastic) (REQ1); their compatibility with easily customizable objects (REQ2); and their capability in providing AR experiences (REQ3&4). To facilitate our discussion, we abbreviate the techniques as: VMB (Visual Markers-based), CVML (Computer-Vision & Machine-Learning), and CT (Capacitive Touches).

For each technique, we describe how it works, how we implemented it, and how it operates using a simple word spelling learning scenario as a theoretical use case scenario.

Use Case Scenario

The use case scenario is based on the PhonoBlocks system [3], a three-year-old tangible reading system (Figure 1). We chose it because it was one of the first tangible systems using physical letters as part of the learning activities that was shown to be effective, with significant learning gains in two case studies. As such it serves as a strong exemplar of a foundational interaction strategy for early childhood learning applications with physical symbols or shapes.

In contrast to the original system using a tablet PC and a platform holding 3D plastic physical letters with embedded electronics, our envisioned tangible-AR system has only two main components: (a) a tablet with rear-facing camera and touchscreen running the education app; and (b) a set of physical objects in the shape of English letters, which are easily available in toy stores as alphabet sets, everyday stores as fridge magnets, or just cutouts from cardboard.

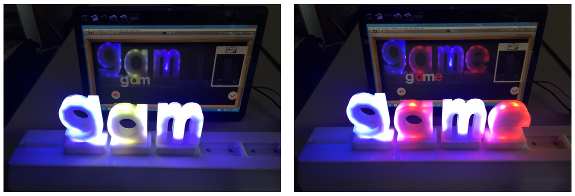

Figure 1. The PhonoBlocks system on which our use case scenario is based, where physical objects in letter shapes light up according to English alphabetic principles, with digital representations displayed in a close-by monitor.

The rules of the English alphabetic principles (e.g., blending consonant sounds, consonant-vowel-consonant patterns) are presented in the education app as activities such as lessons or mini-games, during which the user arranges the physical letters as instructed. The tablet then identifies these letters, tracks their locations, and augments them with overlays such as colour cues, as well as corresponding animations and sounds as rewards.

Technique 1: Visual Markers-based

The Visual Markers-based (VMB) technique is a common approach for tracking 2D objects, on which special patterns (markers) are placed for optical identification, often in real-time. The markers can be simple colours [53], or more sophisticated ones like Fiducial markers [42], allowing tracking of orientations in addition to locations.

Implementation

To leverage the use of currently available physical letters that learners could easily get access to, we first purchased two sets of commonly-used physical letters (see Figure 2 left for an example) from an online store (Alibaba). We then measured the size and shape of the letters and used them as baselines to create the initial design of the markers. We tested each letter one by one, and found that the detection results were unstable, particularly in various lighting conditions. Sometimes the system was not able to detect similar letter shapes such as *t* and *f*. We suspected that the results were caused by (1) the drawing lines in the markers were too thin, and (2) the letter size was too small, and thus it was difficult for the tablet camera to detect those lines from a distance, particularly in a dim environment.

Therefore, we decided to fabricate larger-sized (around 6.5*4.5*1cm) wooden letters ourselves using a laser cutter and iteratively tested the patterns with them (Figure 2 right). For each iteration of the patterns, we glued them onto the letters and tested their reliability using the Vuforia plugin for Unity [51] (an AR software development kit for mobile devices, for its popularity and compatibility with Unity) and the education app running on a tablet.

Figure 2. A set of physical letters we purchased from an online store (Alibaba), size around 3.4*3.6*0.5 cm (Left); and a set of physical letters we fabricated ourselves, with markers glued onto them, size around 6.5*4.5*1cm (Right).

1) 2) 3) 4)

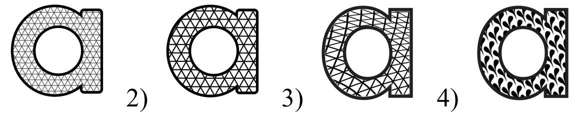

Figure 3. Iterations of the marker pattern for the letter 'a'. 1) first design, 2) thicker lines, 3) randomly distorted lines, and 4) different pattern.

Figure 4. A demo of the Visual Markers-based technique overlaying physical letters with their digital representations.

We also revised our marker design by increasing the stroke thickness of our drawing lines and printed out larger sized physical letters ourselves (Figure 3-2), yet the testing results with similar letters were still not very accurate. We then maximized the differences between each letter by adding more randomly distorted patterns (Figure 3-3) and used various patterns for each letter (Figure 3-4). This led the detection of the patterns to an acceptable level.

Using It for Word Spelling Learning

To use the education app implemented with VMB, the adult user, such as teacher or parent, has to acquire a set of English alphabet physical letters and a corresponding set of pre-defined patterns (shaped like the letters). The patterns are then attached to the letters to be recognized by the app.

During a learning activity, the child user points the tablet's camera to a surface on which they arrange the letters as instructed (Figure 4). The tablet then identifies these letters, track their locations, and augment them with digital content as an overlay in the live video feed captured by the camera.

Technique 2: Computer-Vision & Machine-Learning

The Computer-Vision & Machine-Learning (CVML) technique is based on the widely-used Optical Character Recognition (OCR) technique, where images of text are scanned and converted into machine-encoded text. This is mainly achieved by extracting features of the scanned images and comparing to a pre-existing feature model for the closest match. Some examples include Microsoft Lens and Adobe Scan, which are mobile apps converting pictures of documents into text files for reading and editing.

Implementation

We explored this technique by first using a few existing OCR tools, including open source OCR models and commercialized OCR tools, such as tesseract [56] and ABBYY OCR SDK [36], on the physical letter set we purchased (Figure 2-1). We took pictures of the letters and passed the images to the OCR tools for testing. However, performances were not satisfactory. We believed there were two main reasons. First, instead of 2D letters used in the traditional OCR tasks, our letters were 3D. Therefore, if the light is strong or the light condition changes, the resulting shadows will interfere with the detection of the letters. In most cases, training sets of previous OCR tools do not include letters with shadows. Second, the fonts of the 3D letter set we used were different from the standard fonts in newspapers and books. Therefore, it was difficult for previous models to perform satisfactorily.

Thus, we decided to curate a training set based on our purchased letter set and built a machine-learning model from scratch. We curated the training set by processing all the pictures we took for the 26 characters, each 25 times under different lighting conditions. On each image, we performed image binarization by setting up a threshold of gray level to convert a color image into a black and white image. Next, we segmented the image based on connected component analysis to detect the boundary of each letter. Then, we rescaled each sub-image to the size of 50 × 50 (width × height) for passing it to the Convolutional Neural Network (CNN) [27]. To avoid overfitting and improve the generalization of the deep learning model, we augmented the dataset by adding distortions and rotations to each image. This resulted in 280 variations per picture, that is 7000 images per letter. We ended up having 182000 images in total in our training set, and 6325 images in our testing set. The procedure is described in Figure 5.

To build the CNN model, we applied grid search to find the best hyperparameters including the number of kernels in each layer, kernel size, learning rate, and decay. Our CNN is composed of two convolutional layers and one fully connected layer (refer to ANNEX for details of the layers).

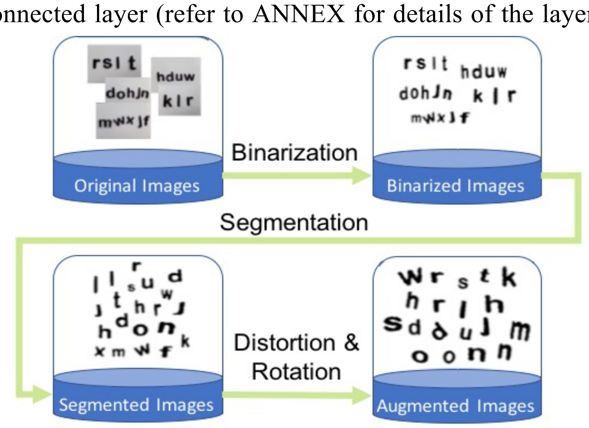

Figure 5. Procedure of binarization, segmentation, and augmentation of the letter pictures for training the Convolutional Neural Network (CNN).

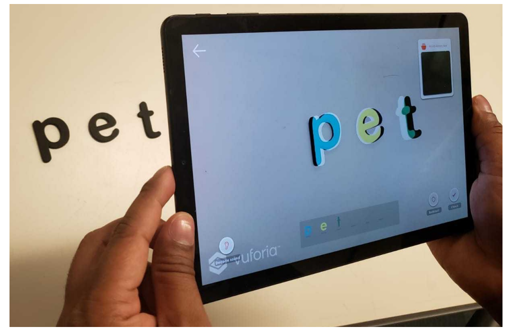

Figure 6. A demo of the CVML technique overlaying physical letters with their digital representations.

We trained 50 epochs before testing the model and built five models based on the previous approach. The average accuracy of the model is 98.73%. When using all the five models together by averaging the results, the system could achieve 99.11% accuracy. Once the models were built, we transferred them to the Unity-based app to recognize the letters via the TensorFlowSharp wrapper API [23].

Considering computational resources, we down-sampled the video captured by a tablet (Samsung Galaxy Tab S4) to 3 frames per second and input them to the trained CNN model after extracting and preprocessing the letters into segments. Each letter was recognized in approximately 0.02 second, with the overlay generated in a similar time.

Using It for Word Spelling Learning

To use the education app implemented with CVML, the adult user, such as teacher or parent, only has to acquire a set of English alphabet physical letters. However, this set needs to be from the same font family used to curate the machine learning model. This issue can however be mitigated by curating the model with more typefaces that are commonly used or including a "training mode" in the app for the model to learn on-site.

A learning activity with this technique is the same as that of VMB: the child user points the tablet's camera to a surface on which they arrange the letters as instructed (Figure 6). The tablet then identifies these letters, tracks their physical locations, and augments them with digital content as an overlay in the live video feed captured by the camera.

Technique 3: Capacitive Touches

The Capacitive Touches (CT) technique builds on the fact that most tablets detect capacitive touches directly on the display. By arranging multiple conductive nibs (commonly used as the tip of a capacitive stylus), or more recently 3D printed conductive materials (e.g., [18, 24]) in a pre-defined configuration, it is possible for the tablet to recognize the arrangement and thus the object to which it is associated. Tiggly [47] and Marbotic [32] are two commercially available products using this approach associating different arrangements to objects that are shaped like letters. However, in their product description both products are designed to accept one letter at a time that acts like a stamp (this action also results in the person touching the letter to provide the necessary electrical ground for the touches to be detected). We are instead interested in a system which simultaneously recognizes multiple letters (e.g., up to five) to support spelling of words, resulting in some letters not being touched and yet still be recognized and tracked.

Implementation

We based our own implementation of the CT approach on the Passive Untouched Capacitive Widgets (PUCs) developed by Voelker et al. [50]. In PUCs, two or more conductive round pads are connected using a conductive bridge, which allowed capacitive coupling to occur and form an electrical ground within the pads. This approach allows the object to be detected even when no one is touching it, and thus frees the person's hands to handle other objects (letters). The authors further investigated the sizes of and the distances between the conductive pads on various devices and concluded that a diameter of 7mm (of each pad) and a distance of 20mm (between pads) to be the minimal for a 90-100% detection rate. These parameters formed the basis of our implementation.

To comply with REQ1&2 of being accessible, we decided to fabricate the letters in-house using readily available equipment in the maker community, including 3D printers and laser cutters, so others can replicate our implementation relatively easily. After several design iterations from printing both the letter and conductive pads with a 3D printer (Figure 7 left) to offloading the letter fabrication to laser cutting/engraving (Figure 7 right), we developed a process of 3D printing the connected conductive pads, converting a font family into a template for laser cutting, and carving out the cavity to embed the conductive pads into the letters. In addition, we used transparent acrylic plastics for the letters to allow the tablet display to augment them from underneath. We also developed an application that recognized the letters based on the distance between the conductive pads (Figure 8).

Figure 7. Iterations of capacitive letters (back side of 'a' and 'b'). 3D printed letters and pads (Left), laser-cut & engraved letters and 3D printed pads (Right).

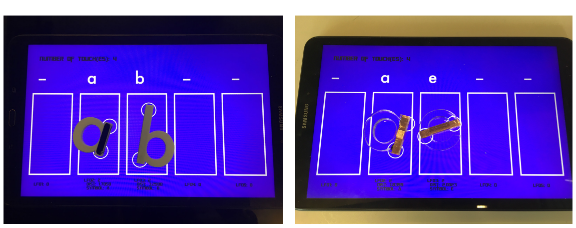

Figure 8. An app recognizing the letters when put on the displayed slots. 3D printed letters & pads (Left), laser-cut letters & 3D printed pads connected by copper tapes (Right).

However, we soon realized that the recognition relies heavily on the hardware and filtering thresholds being used in the tablets to detect touches. When testing this technique on a Samsung Galaxy Tab A SM-T580, the conductive pads were detected as touches and the associated letters were recognized (Figure 8 left & right); but not on an Apple iPad Mini 2. The authors of the PUCs paper [50] also reported various detection behaviours and durations in their tested devices. We have yet to find a way for the letters to be consistently recognized across tablets without significant modifications to the letters (e.g., embedding electronics [49]) or the tablets (e.g., editing the filtering thresholds).

Moreover, we faced two main challenges when adopting PUCs' technique. First, as PUCs were only presented as widgets of generic shapes (bridge and ring), we had to design 26 different configurations to represent the 26 letters in the English alphabet. Second, a typical tablet can detect up to 10 touches simultaneously, meaning that if we want to have up to five letters detected, each letter can only contain two conductive pads, which happens to be the minimal number for the app to use distance as an identifying feature (and still without the orientation information). We also had to limit where each letter could be (shown as white bounding boxes in Figure 8) so touches from different letters would not interfere with each other.

Using It for English Spelling Learning

To use the education app implemented with CT, the adult user, such as teacher or parent, needs to acquire a set of transparent English alphabet physical letters, conductive pads, and materials connecting the pads (we used copper tapes found in hardware stores). The conductive pads are then attached to the letters and connected by the copper tape according to a pre-defined set of configurations (distances).

During a learning activity, the child user puts the letters on the surface of the tablet on which they arrange the letters as instructed. The tablet then identifies these letters, tracks their physical locations, and augments them with digital content. But instead of as an overlay, the digital content will be displayed underneath the letters. This is the reason why the letters should be transparent so content can be seen.

COMPARATIVE ANALYSIS: BENEFITS & LIMITATIONS

We present our comparative analysis of the benefits and limitations of each technique relative to our requirements for a valid system. For each requirement, two experts in tangibles and mobile development rated each technique using a three-level scale (H=high compliance to requirements, M=medium, L=low). Through discussion the raters reached agreement on all items. We also provide details on how well each technique fulfills the requirements based on what we learned from our implementations. Our accuracy and lag testing of each proof-of-concept system was done using a "perceivable latency" criterion (e.g., [26]) and we acknowledge that it may be tablet-specific, in particular for lag (the time between moving a letter and perceiving its digital representation move on the display).

REQ1: Readily Available Physical Objects

All techniques required no more than the tablet and the letter set to function, as compared to commercial products such as Osmo [40], where a mirror is needed for the front-facing camera to see the physical objects; and to research prototypes such as Phonoblocks [3], where a platform with pogo pins is needed to detect the physical letters.

Both VMB and CVML used computational algorithms that were hardware independent to recognize the physical objects, and thus available for both iOS and Android. On the other hand, CT relied on the capacitive sensing hardware and the way the operating system registers touches, which from our testing was not viable in all tablets.

CVML used physical objects as they were, whereas both VMB and CT required modifications for recognition. VMB required a specific set of patterns attached to the objects, but was easier than attaching and connecting conductive pads to the objects at specific distances in CT.

To sum up, CVML fulfilled REQ1 the best because of its ease of acquisition of all the components, followed by VMB. CT falls short here as it was not viable in all tablets.

REQ2: Hands-On Interaction with Physical Objects

We managed to purchase/fabricate all the physical objects in our implementations, thus were able to control their shapes for matching representation, as well as their sizes for easy handling. CVML recognized physical objects as they were, hence posed no size constraints as long as they were visibly distinguishable, and thus made them easy to handle. On the other hand, from our testing, there was a minimum size limit on the patterns in VMB and the distances between connected capacitive pads in CT, which was about 2 times larger than the expected 1-2.5 cm height and width.

As CVML utilizes physical objects without modification, the objects are thus as durable as their composite material. The modifications required for both VMB and CT might wear over time. However, VMB used stick-on markers that could easily be replaced. In contrast, CT required reapplying and reconnecting capacitive pads, making it the least durable and hardest for maintenance.

To sum up, similar to REQ1, CVML fulfilled REQ2 the best because of its use of non-modified physical objects, followed by VMB. CT fell short due to the extra procedures in making the objects conductive.

REQ3: Sensing and Tracking of Physical Objects

Both VMB and CVML could identify multiple objects (5-10, as patterns in VMB and as islands in CVML) in real-time. In contrast, due to the limit of 10 capacitive touch points detection in most tablets, CT was not able to identify more than 3 objects in arbitrary orientations.

Nevertheless, CT used a simple mapping between touch configurations and objects, so the recognition had no perceivable lag. As each object in VMB was recognized via its pattern with multiple features, its presence and position were tracked with a slight lag (within seconds). CVML required a separate step to determine the location of individual objects after recognition, thus resulted in an observable lag in our testing (approximately 1-2 seconds). Its tracking was also affected by occlusion. While we acknowledge the lags are tablet-specific, and will likely be improved with technologies, the relative time difference in recognition and tracking will remain similar.

Both VMB and CVML used images taken by the built-in camera as input for object recognition and were sensitive to poor lighting conditions (sensing and tracking accuracy decreases). VMB performed slightly better due to the use of multiple features being printed in a discernable manner, while CVML suffered from arbitrary shadows due to light source variations and occlusions. CT was unaffected by any lighting condition as it used touch points as input.

Both VMB and CVML recognized objects individually along with their physical location relative to each other, hence the order of the objects could be deduced. However, as there was no distinction between touch points; objects in CT had to be placed in sufficient separation for the software to correctly isolate them. In comparison, many tangible learning systems utilize physical constraints to limit how objects are placed (e.g., sides must match in Tiblo [41], letters must fit in a platform in PhonoBlocks [3]).

As an additional observation, all techniques recognized objects without continuous touch from a hand. However, in CT the touch points eventually disappeared, as explained as adaption of the capacitive touch filtering algorithms in [50], and required a touch to reappear.

To sum up, VMB and CVML had similar strengths in multi-object sensing and tracking but were not as responsive and as resilient against poor light conditions and occlusions as compared to CT, resulting in varying degrees of fulfillment of REQ3.

REQ4: Digital Augmentation of Physical Objects

All techniques were capable of recognizing objects individually and in real time, displaying their digital representations (e.g., images with same outlook, associate sounds) within the app; and in case of representations visually over (in VMB/CVML) or under (in CT) them with proper alignments.

To sum up, as the digital representations were controlled by the system, all techniques provided support for effective learning through AR, hence all fulfilled REQ4 equally well.

DISCUSSION

Table 1 summarizes the comparison of the three techniques we implemented, with detailed breakdowns under each requirement. The breakdowns were created by an expert in tangibles and mobile development using a High-Medium-Low scale and were verified by a second expert not involved in building the prototypes. Overall, we found that they all fulfilled the requirements to varying degrees, resulting in trade-offs that must be considered when developers choose one technique over the other. There is no one best solution at this moment in time, but we believe a CVML-based implementation is the most promising one.

Trade-offs

In terms of recognizing symbols, CT was the simplest to implement as it was a direct conversion from touch configurations (distances) to symbols, whereas VMB required a dataset of patterns and CVML required a learning model. However, CT required the most effort to fabricate the physical objects (attaching conductive pads to laser-cut/3D-printed objects), followed by VMB which required printing out pre-made patterns. CVML was the easiest as it required no fabrication, since the objects could be purchased and used without any modification.

In terms of multiple sensing – that is, the maximum number of objects detected and tracked simultaneously – CVML could, in theory, sense any number as long as the objects were visually separated from each other. Currently, the Vuforia SDK that we used for VMB had a limit of five simultaneous active targets. While we expect the limit will increase as the SDK improves, we do not anticipate it to be more than that of CVML, due to the need of multiple feature points for each pattern to be processed. As CT relied on the number of maximum touches detected by the tablet, we believe it will remain sensing the lowest number of objects amongst all three techniques. However, CT required the least amount of sensing time, and was impervious to varying or poor lighting conditions. In contrast, both VMB and CVML's accuracy degraded when the patterns were not clear, or the shadows were too prominent.

Requirements	**VMB**	**CVML**	**CT**
REQ1: Readily Available			
Composing components	High	High	High
Hardware independence	High	High	Low
Easy of object acquisition	Medium	High	Low
REQ2: Repeated Hands-On Use			
Sizing freedom	Medium	High	Medium
Ease of object maintenance	Medium	High	Low
REQ3: Persistent Multi-Object Sensing & Tracking			
Multiple sensing	High	High	Low
Recognition simplicity	Medium	Low	High
Environmental resilience	Medium	Low	High
Object placement resilience	High	High	Medium
Sensing persistence	High	High	Medium
REQ4: Associated Augmentation			
Matching outlook	High	High	High
Support of alignment	High	High	High

Table 1. Tabular comparison of the techniques in terms of the requirements, shaded for ease of reference.

In terms of augmenting the physical objects, both VMB and CVML functioned very similarly: objects were placed on a surface with the tablet held above (or leveled if the objects are magnetically attached to a vertical surface), digital content was shown on the tablet superimposing the live video feed from the built-in camera. For the camera to capture all the objects, the tablet had to be held steadily at an adequate distance from the surface, which might cause fatigue and might not be easy for young children when their motor skills are developing. To address this issue, we have explored using a tablet stand for stability and/or had one person holding the tablet, while another worked with the letters [16]. In contrast, with CT, objects were placed on the tablet, with digital content shown under/around them, thus posing little space constraint. The tablet could also be placed on a surface to reduce fatigue.

Recommended Implementation

Based on the results of our comparative analysis, we recommend that at this time the best technique for HCI researchers and educational developers to use is CVML. In this case, once deployed, an adult user would purchase a letter set, regardless of what font it is in, and run the app in a "training mode" which guides them through the building process of the learning model by capturing multiple images of the letters (as simple as taking multiple photos), or download a pre-trained model that recognizes the font. The education app can then be passed to the child user to be used as designed without any further setting up.

As mentioned, we used the English language learning scenario as a foundational interaction strategy. However, our requirements and the techniques we explored are general enough to accommodate many other symbols. By expanding the physical objects set to include symbols from other areas such as arithmetic, music, and geometry, children would be able to enjoy a wide range of hands-on learning experience in topics and interactivity.

Limitations & Future Work

During our development and analysis of the three tangible-AR techniques we identified benefits and limitations, as well as trade-offs. Some of these findings arose from current technological constraints, such as sophistication of learning models (and quantity of available data to train one), touch point detection; or our own knowledge deficits in fabrication and programming, such as working with conductive materials and implementing more efficient sensing and tracking algorithms.

As knowledge and technologies advance, researchers can take advantage of improvements and create systems that can better meet the requirements listed above. One promising direction is the CVML technique, which can recognize unmodified physical objects. The biggest challenge, however, is to make CVML less prone to poor light conditions (e.g., varying light sources) and faster in tracking objects. These challenges could be addressed by improvements in camera sensors and algorithm efficiency.

Moreover, over time there might be new techniques to sense and track physical objects. However, it is likely that they will be variations of those we have listed. Our goal was to introduce these techniques and show their feasibility to educational developers and HCI researchers, so that they can then implement and conduct further research in the area of early childhood learning; using our requirements to evaluate new techniques and select the best for the job. With this platform they will have access to the core functionality required to create diverse hybrid tangible-AR systems, and be able to conduct more rigorous studies on the effectiveness of such systems with child learners.

GENERALIZATION: NUMBER USE CASE

We used language learning and a lower-case Latin letter set to illustrate how our technology platform can be implemented through three selected hybrid tangible-AR techniques. To demonstrate its generalizability, we provide a second use case using Arabic numbers (i.e., 0-9), which can be accommodated by all of the three techniques (REQ1&2). This is because the core strategy of VMB and CVML was to recognize visual patterns and CT to recognize configurations of conductive materials. Specifically, we demonstrate generalizability through sensing, tracking, and augmenting number objects using the CVML technique. In Figure 9 we show the most demanding test case in which multiple physical numbers were duplicated exactly in their position, orientation, and size; and augmented with a precisely registered colour overlay.

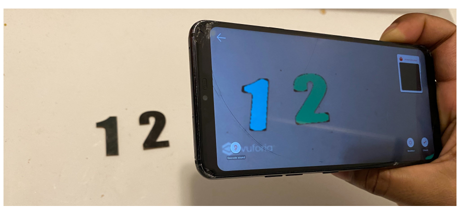

Figure 9. A demo of the CVML technique digitally augmenting the numbers exactly to their shapes.

In order to sense and track numbers we adapted the CVML technique using transfer learning to train the CNN model to recognize the new number objects [39]. Transfer learning is a machine learning technique in which information from previously learned tasks is transferred to the task of learning a new, similar task. By using existing information from a previous task, the efficiency of learning a new symbol set is significantly improved compared to the initial work to train the system. To train our CVML program to sense and track number objects, we added pictures of the number set to the existing model as a new task (refer to ANNEX for details). To use our platform to augment numbers, we simply used our existing code without modification, noting that developers can adapt the platform to use other forms of overlays besides number colour. In summary, we demonstrate the ease of generalizability of the CVML technique to sense, track and augment numbers, another reason we recommend this approach.

CONCLUSION

In this paper, we demonstrated and analyzed three techniques that enable physical learning materials to be integrated with tablet-based applications for early childhood education using augmented reality. To inform our implementations, we established four requirements from analyses of existing systems and our own previous work in the field of tangible computing for early childhood education. Our analysis focused on comparing how each approach met these requirements, explored through a use case derived from a previously validated tangible reading system. Results from this analysis revealed current limitations, trade-offs, as well as opportunities for further development and deployment. Lastly, we demonstrated generalizability of our recommended CVML technique through a second use case involving 3D numbers, which could be used to create arithmetic learning applications.

This work moves us one step closer to a development platform that enables educational developers, and HCI and learning sciences researchers to create scalable hybrid tangible-AR applications for early childhood education, with more complex, creative scenarios and use cases, as well as real world testing in effectiveness and robustness.

ANNEX

We include details of our CNN model in the CVML technique for replicability. The CNN model is composed of two convolutional layers and one fully connected layer. The first convolutional layer filters the $50 \times 50 \times 1$ input features with 64 kernels of size $3 \times 3 \times 1$ and a stride of 1. The second convolutional layer uses 128 kernels of size $3 \times 3 \times 64$. The third convolutional layer uses 256 kernels of size $3 \times 3 \times 128$. We used max pooling (2×2) with a stride of 2 for the outputs of convolutional layers, and added one dropout layer after each convolutional layer, with a dropout rate of 0.4. The fully connected layer, connected to the third convolutional layer, is composed of 1024 neurons. The output layer is composed of 26 neurons (one per letter). We used the ReLU non-linearity activation function [39] for all convolutional layers and the fully connected layer. For the output layer, we used softmax activation to predict the category. All the weights were initialized based on a Xavier uniform. The CNN was trained using the Adam optimizer with a batch size of 64 examples, learning rate of 1×10^{-4}.

To train the existing CNN model to recognize numbers (our second use case) using transfer learning, we created a training set for detecting number by making a small video (approx. 15s) for each of the numbers (0-9) and extracting images from it (or taking snapshots). We then froze the last layer of the existing model (which classifies the images into 26 letters, i.e., 26 output neuron) and replaced it with the 10 output neuron. This process is true for any kind of shapes (e.g., other letters, arithmetic symbols, music notes).

ACKNOWLEDGMENTS

This research study funded by the Social Sciences and Humanities Research Council (SSHRC) of Canada.

SELECTION AND PARTICIPATION OF CHILDREN

No children participated in this work.

REFERENCES

[1] Alakärppä, I. et al. 2017. Using Nature Elements in Mobile AR for Education with Children. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (New York, NY, USA, 2017), 41:1–41:13.

[2] Antle, A.N. 2013. Exploring how children use their hands to think: an embodied interactional analysis. *Behaviour & Information Technology*. 32, 9 (Sep. 2013), 938–954. DOI:https://doi.org/10.1080/0144929X.2011.630415.

[3] Antle, A.N. et al. 2015. PhonoBlocks: A Tangible System for Supporting Dyslexic Children Learning to Read. *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction* (New York, NY, USA, 2015), 533–538.

[4] Antle, A.N. and Wise, A.F. 2013. Getting Down to Details: Using Theories of Cognition and Learning to Inform Tangible User Interface Design. *Interacting with Computers*. 25, 1 (Jan. 2013), 1–20. DOI:https://doi.org/10.1093/iwc/iws007.

[5] Apple: most popular app store categories 2019 | Statista: *https://www.statista.com/statistics/270291/popular-categories-in-the-app-store/*. Accessed: 2019-09-11.

[6] ARToolKit Home Page: *http://www.hitl.washington.edu/artoolkit/*. Accessed: 2019-09-19.

[7] Barreira, J. et al. 2012. MOW: Augmented Reality game to learn words in different languages: Case study: Learning English names of animals in elementary school. *7th Iberian Conference on Information Systems and Technologies (CISTI 2012)* (Jun. 2012), 1–6.

[8] Boonbrahm, S. et al. 2015. Using augmented reality technology in assisting english learning for primary school students. *Learning and Collaboration Technologies* (Los Angeles, CA, USA, 2015), 24–32.

[9] Brosterman, N. 1997. *Inventing Kindergarten*. Harry N. Abrams.

[10] Bujak, K.R. et al. 2013. A psychological perspective on augmented reality in the mathematics classroom. *Computers & Education*. 68, (Oct. 2013), 536–544. DOI:https://doi.org/10.1016/j.compedu.2013.02.017.

[11] Chiang, T.H.C. et al. 2014. An Augmented Reality-based Mobile Learning System to Improve Students' Learning Achievements and Motivations in Natural Science Inquiry Activities. *Journal of Educational Technology & Society*. 17, 4 (2014), 352–365.

[12] Dewey, J. 1938. *Logic: The Theory of Inquiry*. Holt and Co.

[13] Di Serio, Á. et al. 2013. Impact of an augmented reality system on students' motivation for a visual art course. *Computers & Education*. 68, (Oct. 2013), 586–596. DOI:https://doi.org/10.1016/j.compedu.2012.03.002.

[14] Fan, M. et al. 2018. From Tangible to Augmented: Designing a PhonoBlocks Reading System Using Everyday Technologies. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), LBW555:1–LBW555:6.

[15] Fan, M. et al. 2017. Why Tangibility Matters: A Design Case Study of At-Risk Children Learning to Read and Spell. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2017), 1805–1816.

[16] Fan, M. and Antle, A.N. 2020. An English language learning study with rural Chinese children using an aygmented reality app. *Proceedings of the 19th International Conference on Interaction Design and Children* (Jun. 2020), in press.

[17] Gibbs, R.W. 2005. *Embodiment and Cognitive Science*. Cambridge University Press.

[18] Götzelmann, T. and Schneider, D. 2016. CapCodes: Capacitive 3D Printable Identification and On-Screen Tracking for Tangible Interaction. *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (New York, NY, USA, 2016).

[19] Hengeveld, B. et al. 2008. The Development of LinguaBytes: An Interactive Tangible Play and Learning System to Stimulate the Language Development of Toddlers with Multiple Disabilities. *Adv. in Hum.-Comp. Int.* 2008, (Jan. 2008), 1:1–1:13. DOI:https://doi.org/10.1155/2008/381086.

[20] Horn, M.S. et al. 2012. Tangible interaction and learning: the case for a hybrid approach. *Personal and Ubiquitous Computing*. 16, 4 (Apr. 2012), 379–389. DOI:https://doi.org/10.1007/s00779-011-0404-2.

[21] Hu, F. et al. 2015. Strawbies: Explorations in Tangible Programming. *Proceedings of the 14th International Conference on Interaction Design and Children* (New York, NY, USA, 2015), 410–413.

[22] Hudson, S.E. and Mankoff, J. 2014. Concepts, Values, and Methods for Technical Human–Computer Interaction Research. *Ways of Knowing in HCI*. J.S. Olson and W.A. Kellogg, eds. Springer New York. 69–93.

[23] Icaza, M. de 2019. *TensorFlow API for .NET languages*. https://github.com/migueldeicaza/TensorFlowSharp.

[24] Ikematsu, K. and Siio, I. 2018. Ohmic-Touch: Extending Touch Interaction by Indirect Touch through Resistive Objects. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018).

[25] Jorge Bacca et al. 2014. Augmented Reality Trends in Education: A Systematic Review of Research and Applications. *Journal of Educational Technology & Society*. 17, 4 (2014), 133–149.

[26] Jota, R. et al. 2013. How Fast is Fast Enough? A Study of the Effects of Latency in Direct-Touch Pointing Tasks. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2013), 2291–2300.

[27] Krizhevsky, A. et al. 2012. ImageNet Classification with Deep Convolutional Neural Networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1* (Red Hook, NY, USA, 2012), 1097–1105.

[28] Lin, V. and Blikstein, Paulo 2019. Project Bloks: Embodied and Collaborative Learning with Tangible Interfaces for Young Children. *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings* (Lyon, France, Jun. 2019), 1021–1024.

[29] Lo, W.-T. and Quintana, C. 2013. Students' Use of Mobile Technology to Collect Data in Guided Inquiry on Field Trips. *Proceedings of the 12th International Conference on Interaction Design and Children* (New York, NY, USA, 2013), 297–300.

[30] Major, L. et al. 2017. Tablet Use in Schools: Impact, Affordances and Considerations. *Handbook on Digital Learning for K-12 Schools*. A. Marcus-Quinn and T. Hourigan, eds. Springer International Publishing. 115–128.

[31] Maldonado, H. and Zekelman, A. 2019. Designing Tangible ABCs: FrÖBel's Sticks and Rings for the 21st Century. *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (New York, NY, USA, 2019), 326–333.

[32] Marbotic | Wooden connected toys & educational apps for tablets: *https://www.marbotic.com/*. Accessed: 2019-08-19.

[33] Melcer, E.F. and Isbister, K. 2018. Bots & (Main)Frames: Exploring the Impact of Tangible Blocks and Collaborative Play in an Educational Programming Game. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), 266:1–266:14.

[34] Milgram, P. and Kishino, F. 1994. A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Information Systems*. E77-D, 12 (1994), 1321--1329.

[35] Minogue, J. and Jones, M.G. 2006. Haptics in education: Exploring an untapped sensory modality. *Review of Educational Research*. 76, 3 (Jan. 2006), 317–348. DOI:https://doi.org/10.3102/00346543076003317.

[36] Mobile OCR with ABBYY Cloud SDK for Android, iPhone, Windows phone: *https://www.ocrsdk.com/features/platforms/*. Accessed: 2019-08-19.

[37] Montessori, M. 1966. *The Secret of Childhood*. Ballantine Books.

[38] Nair, V. and Hinton, G.E. 2010. Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the 27th International Conference on International Conference on Machine Learning* (USA, 2010), 807–814.

[39] Olivas, E.S. et al. 2009. *Handbook Of Research On Machine Learning Applications and Trends: Algorithms, Methods and Techniques - 2 Volumes*. Information Science Reference - Imprint of: IGI Publishing.

[40] Osmo – Award-Winning Educational Games System for iPad: *https://www.playosmo.com/en-ca/*. Accessed: 2019-08-19.

[41] Pandey, S. and Srivastava, S. 2011. Tiblo: A Tangible Learning Aid for Children with Dyslexia. *Procedings of the Second Conference on Creativity and Innovation in Design* (New York, NY, USA, 2011), 211–220.

[42] reacTIVision - a toolkit for tangible multi-touch surfaces: *http://reactivision.sourceforge.net/*. Accessed: 2019-08-22.

[43] Rogers, Y. et al. 2004. Ambient Wood: Designing New Forms of Digital Augmentation for Learning Outdoors. *Proceedings of the 2004 Conference on Interaction Design and Children: Building a Community* (New York, NY, USA, 2004), 3–10.

[44] Schroth, S. et al. 2019. An Exploratory Study of Osmo Tangram and Tangram Manipulative in an Elementary Mathematics Classroom. *Journal of Educational Technology Development and Exchange (JETDE)*. 11, 1 (2019).

[45] Shuler, C. et al. 2012. *iLearn II An Analysis of the Education Category of Apple's App Store*. The Joan Ganz Cooney Center at Sesame Workshop.

[46] Tablet Operating System Market Share Worldwide | StatCounter Global Stats: *https://gs.statcounter.com/os-market-share/tablet/worldwide*. Accessed: 2019-08-19.

[47] Tiggly | Interactive Toys & iPad Learning Apps for Kids: *https://tiggly.com/*. Accessed: 2019-08-19.

[48] Ullmer, B. and Ishii, H. 2000. Emerging Frameworks for Tangible User Interfaces. *IBM Syst. J.* 39, 3–4 (Jul. 2000), 915–931. DOI:https://doi.org/10.1147/sj.393.0915.

[49] Voelker, S. et al. 2015. PERCs: Persistently Trackable Tangibles on Capacitive Multi-Touch Displays. *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (New York, NY, USA, 2015), 351–356.

[50] Voelker, S. et al. 2013. PUCs: Detecting Transparent, Passive Untouched Capacitive Widgets on Unmodified Multi-touch Displays. *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces* (New York, NY, USA, 2013), 101–104.

[51] Vuforia Developer Portal: *https://developer.vuforia.com/*. Accessed: 2019-08-19.

[52] Wise, A. et al. 2017. Explanation-giving in a collaborative tangible tabletop game: Initiation, positionality, valence and action-orientation. (Philadelphia, PA, 2017).

[53] Yilmaz, A. et al. 2006. Object Tracking: A Survey. *ACM Comput. Surv.* 38, 4 (Dec. 2006), 13–es. DOI:https://doi.org/10.1145/1177352.1177355.

[54] Yilmaz, R.M. 2016. Educational magic toys developed with augmented reality technology for early childhood education. *Computers in Human Behavior*. 54, (Jan. 2016), 240–248. DOI:https://doi.org/10.1016/j.chb.2015.07.040.

[55] Zaman, B. et al. 2012. The evolving field of tangible interaction for children: The challenge of empirical validation. 16, 4 (2012). DOI:https://doi.org/10.1007/s00779-011-0409-x.

[56] 2019. *Tesseract Open Source OCR Engine*. tesseract-ocr. https://github.com/tesseract-ocr/tesseract.