Intelligent Tutoring Gets Physical: Coaching the Physical Learner by Modeling the Physical World

Benjamin Goldberg[✉]

U.S. Army Research Laboratory-Human Research and Engineering Directorate,
Orlando, FL, USA
`benjamin.s.goldberg.civ@mail.mil`

Abstract. Extending the application of intelligent tutoring beyond the desktop and into the physical world is a sought after capability. If implemented correctly, Artificial Intelligence (AI) tools and methods can be applied to support personalized and adaptive on-the-job training experiences as well as assist in the development of knowledge, skills and abilities (KSAs) across athletics and psychomotor domain spaces. While intelligent tutoring in a physical world is not a traditional application of such technologies, it still operates in much the same fashion as all Intelligent Tutoring Systems (ITS) in existence. It takes raw system interaction data and applies modeling techniques to infer performance and competency while a learner executes tasks within a scenario or defined problem set. While a traditional ITS observes learner interaction and performance to infer cognitive understanding of a concept and procedure, a physical ITS will observe interaction and performance to infer additional components of behavioral understanding and technique. A question the authors address in this paper is how physical interactions can be captured in an ITS friendly format and what technologies currently exist to monitor learner physiological signals and free-form behaviors? Answering the question involves a breakdown of the current state-of-the-art across technologies spanning wearable sensors, computer vision, and motion tracking that can be applied to model physical world components. The breakdown will include the pros and cons of each technology, an example of a domain model the data provided can inform, and the implications the derived models have on pedagogical decisions for coaching and reflection.

Keywords: Intelligent tutoring systems · Physical modeling · Psychomotor · Wearable sensors

1 Introduction

The concept of Augmented Cognition (AugCog) is based on the application of technology to impart information on a user that is not inherently perceivable in the natural task environment. This information is intended to assist a user in executing a task by enhancing an individual's cognitive function in support of meeting a specified objective. From a training and education perspective, AugCog practices are associated with adaptive instructional techniques that augment the path an individual takes in learning a topic or skill and the type of coaching they receive along the way. These management

decisions are based on models configured to inform interactions across domain, learner, and pedagogical representations of a training space. These applications are traditionally referred to in the literature as Intelligent Tutoring Systems (ITS).

To date, majority of ITSs are built around strictly cognitive problem domains, with notable successes seen across an array of academic and military applications [1]. What is a sought after capability, and more achievable now than ever with advances in wearable technologies, is extending these practices to the physical world in pursuit of training psychomotor skills. This involves tasks that associate cognition with physical interaction to meet an established goal, and incorporates a combination of hand-eye coordination, muscle memory, and behavioral techniques that dictate performance and assessed acquisition of skill.

While this can be considered a novel extension of traditional ITS methods, its implementation doesn't vary significantly from systems of the past. It utilizes models built on domain and learner information to inform a pedagogical decision. In this instance, the domain isn't informed solely by performance and procedural information communicated by a training application; it now requires methods to collect task relevant behavioral measures that can be used to capture a physical technique and assess performance against a set of specified standards. The important component here is that the information collected must be done in a task's natural environment, where the physical actions can be performed with zero hindrances.

An area of interest to the research community is identifying data types required to model psychomotor interactions at a hand-eye coordination level, and how best to utilize available sensor technologies to instrument the learner and the training environment with data streams that can accurately track behavior. In this paper we discuss the facets associated with the development of a psychomotor adaptive training capability. This includes reviewing theory surrounding psychomotor learning and skill acquisition, how to enable assessment and coaching from a physical problem space, and what commercial off-the-shelf sensors can provide valuable data to infer skill.

2 Learning a New Skill and the Role of Coaching and Feedback

There are common tenets expressed in the literature associated with learning a new skill (see Fig. 1 for a mind map of variables associated with psychomotor skill development [2]). The first and foremost is that experience and practice trumps all. However, simply practicing a skill over and over does not necessarily lead to expert performance. How individuals progress in skill development is based on a number of factors. Anders Ericson's theory of deliberate practice highlights the following attributes of an effective practice event: (1) the event is designed to improve performance; (2) the individual has the ability to repeat the application over multiple trials; (3) the task requires high mental engagement; and (4) feedback is continuously made available that is designed to serve in a coaching capacity [3]. The fourth factor is critical when determining the implications of using ITSs to replace human counterparts to train psychomotor skills.
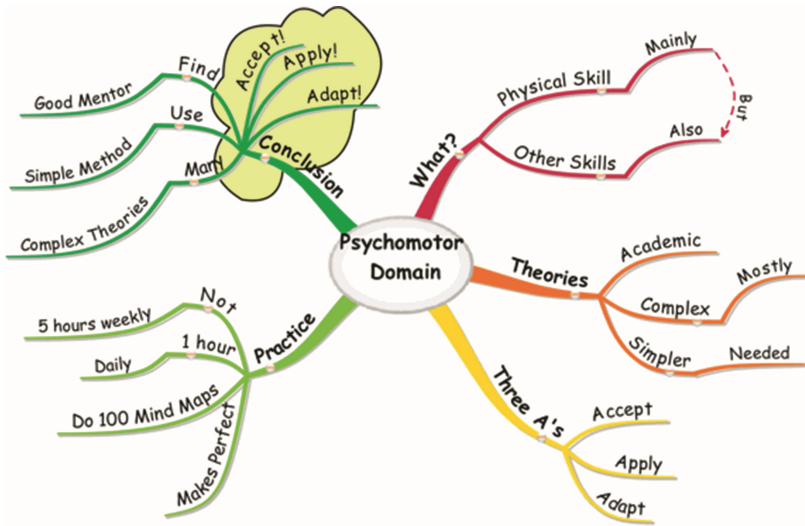
Fig. 1. The psychomotor domain as mind mapped by Faizel Mohidin [2]

Acquiring a new skill follows three primary phases of development, each building on top of the other: (1) beginner/novice phase where an individual tries to understand the cognitive and physical requirements of the activity to generate actions while avoiding errors; (2) the intermediate/journeyman phase where focused attention on task performance is no longer required and noticeable errors become increasingly rare; and (3) the expert phase where the execution of a skill becomes automated with minimal effort and exertion [4, 5]. How individuals progress through these three phases of skill acquisition and the rate at which they do so is dependent on the factors listed above.

From the coaching perspective, especially within the beginner/novice phase, how can an individual modify behavior if there is no way to effectively link actions to observed outcomes? During this phase of learning, behavioral tendencies are established and schemas are built in memory, making feedback to instill proper habits critical. In the traditional sense, a coach/instructor with knowledge in the domain will observe a learner, identify errors in their behavior as determined by a performance outcome, and provide feedback to correct errors and reinforce proper technique.

Utilizing technology to facilitate this inference procedure is challenging. It requires a machine to have the ability to consume perceptual information that associates with behaviors an expert human would assess, and models to determine how the captured data relates to a representation of desired behavior. This identified capability requires a representation of knowledge an expert works with to dictate coaching practices and warrants the utility of a deconstructed task analysis, breaking a domain down into its piece parts in a hierarchical structure of varying skills and applications.

2.1 Deconstructing a Psychomotor Domain

In terms of relating what's already been discussed to a real-world example, take the domain of basketball. When someone is attempting to learn basketball for the first time, the initial approach to instruction is focused on a set of fundamentals. These fundamentals set a foundation of required skills to successfully perform as an elite basketball player. In this instance you can decompose basketball into three physical fundamental skills: (1) dribbling, (2) passing, and (3) shooting. Each of these breakdown further into a set of sub-skills that ascend in complexity as you progress through practice opportunities (e.g., dribbling with your dominant hand, to dribbling with your non-dominant hand, to dribbling between hands, to dribbling between your legs, to dribbling behind your back, etc.). The desired end state is the development of muscle memory to automatically perform a task without dedicating cognitive function to make it happen. When you establish automated execution of fundamental behaviors, then an individual can progress to more complex scenarios requiring advanced application of a skill (e.g., dribbling while being defended). This is followed by practice opportunities to combine the application of skills to perform a higher level task.

This analogy can associate with almost all psychomotor domains of instruction, regardless if its association with job-related activities or athletics. Each domain can be deconstructed into a set of fundamental components that are performed when a situation warrants their execution. The goal of an automated ITS is to establish models of fundamental behaviors to make the assessment space manageable. While the assessment space of a domain is defined around a set of concepts and objectives, it is inherently dictated by the data one can collect.

3   Modeling the Physical World

Modeling a physical task requires an understanding of the physical environment that task is being performed within. The environment will determine the granularity level of data a model can be built from, and can range from a highly customized room built with sensing technologies to detect specific data feeds, to an open warehouse or gymnasium, to an open field in the wild. The ideal situation involves a task environment instrumented to inform both performance and behavior metrics that can assess causal relationships. But the ideal environment to support this methodology is rare. That is why establishing tools for collecting relevant information in a less controlled space is critical to the success of ITSs being used in the wild. In the following subsections, we will describe three modeling scenarios: (1) modeling a psychomotor task in a highly sensorized environment that is tightly-coupled, (2) modeling a psychomotor task in a confined space with no custom sensors that is loosely-coupled, and (3) modeling a psychomotor task in an unrestricted space out in the open.

In each scenario, sensor inputs will be identified. While the first scenario is based on an actual research project being conducted at the U.S. Army Research Laboratory, the latter two are presented as hypothetical applications. In this instance, we aim to identify notional applications of sensor technologies to monitor physical interaction and behaviors that can personalize training in physical spaces.

3.1 Modeling a Physical Task in a Highly Sensorized Environment

Highly sensorized training environments used to develop physical skills provide excellent opportunities to produce initial psychomotor ITS applications. In these scenarios, an environment is built with components to track predefined behaviors. These behaviors are tracked to allow instructors to better inform their decisions on what aspects of a skill to instruct, or in this instance use sensors to model task behaviors to automatically assess skill and trigger coaching feedback.

An example of this ideal scenario can be seen in work we're performing on the development of an adaptive marksmanship training capability [6]. For this project, we are working with the U.S. Army's Engagement Skills Trainer (EST; see Fig. 2). The EST is a simulated firing range that recreates the tasks executed on a live range in a safe/cost-effective setting. In our effort, we're focused on building an ITS to support the development of Basic Rifle Marksmanship (BRM) skills and fundamentals.
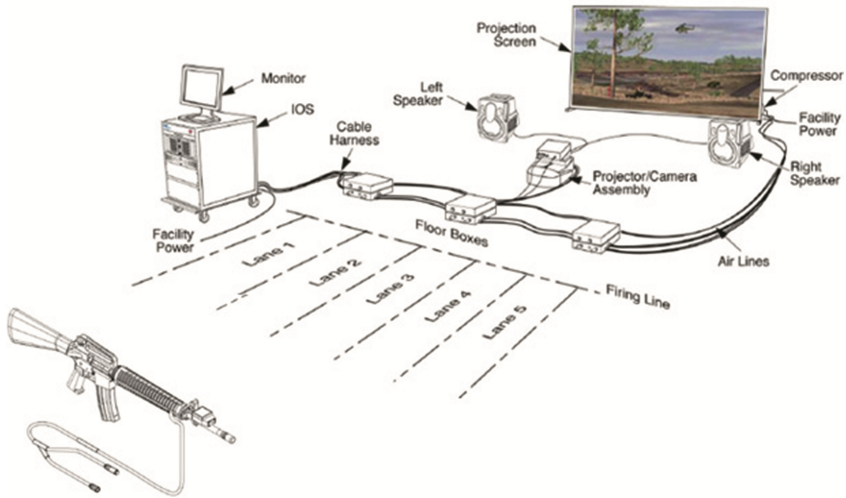

Fig. 2. The U.S. army engagement skills trainer

The EST offers an excellent testbed for this use case based on the features it provides. The system was developed to log data linked to both performance and behavioral metrics. In terms of BRM, the EST tracks performance across a set of grouping exercises that gauge metrics on group size and group accuracy. In addition to the performance metrics, the EST weapons are outfitted with sensor technologies that associate with behavioral properties. These include sensors to track: (1) the aim trace of the weapon barrel, (2) the distance the trigger travels in relation to time, (3) the cant angle of the rifle, and (4) the amount of pressure applied to the buttstock of the rifle.

With all of this contextual data, we were able to build models of expert behavior across a set of performers within the Army Marksmanship Unit's Service Rifle Team. The models were based on the four fundamentals of BRM outlined in the training field manual: (1) breathing, (2) trigger control, (3) body position, and (4) sight alignment.

The models are used as benchmarks to assess a trainee's behavior against to determine if they are properly performing the fundamentals of BRM procedures as deemed by a field of experts. To build out this prototype, we utilized the Generalized Intelligent Framework for Tutoring (GIFT).

Generalized Intelligent Framework for Tutoring (GIFT). GIFT is a domain-independent framework established to author, deliver, and evaluate ITS technologies [7]. It provides a set of standards to follow in creating models across a domain, learner, and pedagogical schema. In special instances, GIFT supports the consumption of sensor information to inform states not linked directly to system interaction data and can extend modeling techniques to monitor affective and physical attributes of a learner. For the adaptive marksmanship ITS, GIFT is configured to take in the EST data streams through the sensor module, where the data is filtered and processed in real-time to assess against the represented expert behaviors. This assessment is used to select a performance state for each of the BRM fundamentals, where those skills assessed as below expectation are used to guide the selection of coaching feedback to deliver. While we are making good progress on developing an ITS for teaching BRM, and the EST provides an ideal set of behavioral information, how to extend these methods into more advanced skills of marksmanship execution (i.e., hitting moving targets) need to be further conceptualized.

3.2   Modeling a Physical Task in a Confined Space with no Custom Sensors

In this scenario, psychomotor activity is believed to be performed in a controlled space designed to support task execution, but lacks the inclusion of customized sensors built to collect task relevant information. In this instance, we are interested in identifying technologies that can extend the assessment space of these environments to inform contextually relevant variables. This includes identifying what variables matter in consistently gauging performance and behavior across a wide range of tasks.

While each domain has unique assessment requirements, many can share similar data to infer completely different skill applications. This would depend on how the data is represented in a model and how that model is linked to a component and/or fundamental of a skill. The hope is to avoid building a set of custom sensors and to utilize commercial off-the-shelf sensing technologies. In a controlled space, the following technologies are believed to provide valuable information for modeling the physical world: (1) motion tracking and (2) wearable sensors.

In most instances, this scenario may be the most complex of the three. With the task being performed in a confined space, the characteristics of the task actions most likely associate with fine motor control over a set of environmental objects. For this reason, it is important to understand the current state of the art of what information can be made available for modeling purposes, and how that impacts the type of domains an ITS can currently support under these environmental conditions. An overarching assumption with all selected technologies is that they can communicate data in real-time to an architecture that can process and model its outputs, such as GIFT.

Motion Tracking. In many psychomotor tasks, monitoring the physical motions of an individual's movement and skeletal structure can go a long way into assessing their behavior. This is especially true in fitness training, athletics, and task-oriented domains that require precise movements to meet standard performance. Depending on the domain being trained, the body is constrained to certain movements and actions that can be executed in support of meeting a task objective. It is assumed under these conditions that behavioral characteristics can be modeled for determining proper execution. In this instance, logging data over a window of time as you observe a set of experts perform a task can be used to determine if there are trends in behavior that influence outcomes and proper techniques. If trends can be identified through statistical inference procedures, then models can be established to compare trainee data against in real-time to diagnose performance on a set of behaviors.

The challenge is using motion based information, which can be noisy in nature and not of appropriate validity in measuring precise movements, to build assessment models that operate in an ITS. Outfitting the environment with motion tracking technology can be used to quantifiably monitor user actions, which is a start. For the context of this paper, we associate motion tracking technology as a free standing system of cameras and sensors that can be placed throughout a confined space. The issue is that a technology of this nature limits the amount of space a task can be conducted within. That's why it's important to understand the characteristics of a domain to determine if this type of modeling technique is viable.

For implementation purposes, systems can be as complex as cameras placed throughout a training space that are designed to locate and track a set of reflective markers that can be placed on a number of items (e.g., placed on a bodysuit worn by a user or on interaction environment elements in an environment, like a baseball bat seen in the left image of Fig. 3). Motion tracking can also be supported by commercial products like the Microsoft Kinect 2, where no markers are required to capture and track an individual's skeletal structure, producing an image like the one seen on the right of Fig. 3. Regardless of the approach selected, motion tracking can be a nice option when the task environment is confined. There is still much work to.
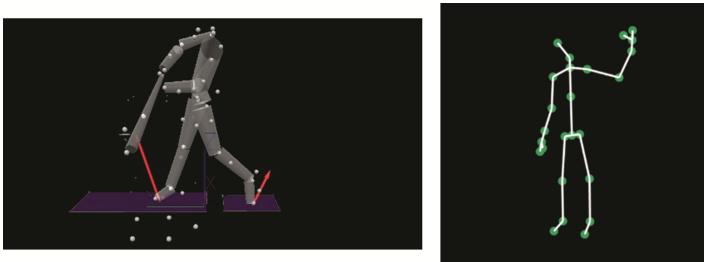

Fig. 3. Motion tracking technologies: on left is motion tracking resulting from wearable bodysuit with mounted reflectors; on right is motion tracking resulting from Microsoft Kinect.

Wearable Sensors.   Recent advancements in wearable technologies have made them a viable data source when considering inputs for informing models to train psychomotor skills. In the context of this paper, wearable sensors are any technology that can be unobtrusively attached to a user that logs physiological and behavioral measures. Common metrics collected by these devices include electrocardiogram/heartrate information, accelerometer data, gyroscope information, galvanic skin response, breathing patterns in some instances, and location data if GPS compliant.

The current application of these sensors within the commercial world is primarily for health tracking purposes. The combination of data channels can output metrics related to activity levels, stress, and sleep patterns. The market is very competitive, with a near endless selection of options ranging from the data they provide and the style of which they are worn.

From a training perspective, the research goal is to identify how best to use these technologies to collect information that can be used to guide skill acquisition of a physical task. Majority of the current products involve sensors that wear either around the wrist or ankle. These typically record a comprehensive set of physiological markers, behavioral movement data, and environmental factors such as temperature and UV exposure. While many of these provide valuable information to monitor activity and affective variables such as stress, they lack granular data sources linked to precision of movements from a motor-control angle.

However, that's not true for all wearable sensors. Products like the MOOV Now and Zepp Sports sensors provide real-time 3D motion tracking of a joint/limb on the human body (see Fig. 4). This information is logged and visualized for replay purposes. Current applications like Zepp allow you to replay your data feed side-by-side with data collected

Fig. 4. Wearable sensing technologies: upper left is the Microsoft Band 2 [http://www.microsoft.com/microsoft-band]; lower left is the MOOV now activity bracelet that supports 3D motion tracking [http://welcome.moov.cc/] and the Zepp Baseball sensor output [http://www.zepp.com]; right is Zephyr Technologies' BioHarness 3 wearable sensor [http://www.zephyranywhere.com/].

from an expert for a comparative evaluation, but there is no coaching beyond that. Exploring modeling techniques that take these three dimensional tracking feeds and associate behavioral parameters in association with a task fundamental or objective is essential for ITSs supporting this interaction environment.

3.3   Modeling a Physical Task Out in the Open

In considering a physical domain performed in a boundless open environment, sensing technologies play a different role than seen in the confined space. In this instance, tasks are performed that require coordination of movements and activities over possible large distances, with factors of location, speed, acceleration, and terrain playing a role in how the task is performed. For an excellent review of this conceptual environment with a use case centered on land navigation, see Sottilare & LaViola, 2015 [8]. In their review, the authors present a set of 'smart glasses' and a feasibility analysis of their application in a live land navigation training scenario.

As the majority of smart glasses sync to a cellular device for processing purposes, the smart glasses themselves primarily serve as tools to present information to a user, with many options including simple text message overlays, objects placed to augment the visual environment, or videos containing instructional material. From a behavioral sensing standpoint, these smart glasses can utilize forward facing cameras to assess an individual's orientation within an environment, with the ability to make inferences on what someone should see based on a calculation of their visual field of view from a specified Global Positioning System (GPS) coordinate.

Beyond the smart glass inputs and outputs, the real sensing taking place in this training environment is provided through the phone, with this example providing GPS location data as tracked over a cellular network. With this information, assessment rules can be built in GIFT based on the data factors listed above. Zones of interest can also be established that can trigger situational awareness oriented tutorial interventions, forcing an individual to reflect on the situation and respond to a prompt that can be used to assess competency and trigger coaching interactions.

In these open environment training events, there can be a combination of open and confined task characteristics, where a sensing technology can be applied to track both location, as well as fine motor-movements. In this instance, computation on wearable sensors providing 3-D motion tracking, like the MOOV Now sensor described above, must be done on a mobile device. This requires GIFT modules to run locally, as the range on wearable sensors doesn't support long range distances. Computations must be performed on the cellular device, with behavioral state information communicated through the network in support of GIFT's learning effect chain. This approach to physical modeling is also critical for team-oriented training tasks. From this perspective, multiple entities can be tracked in a single environment. Formations can be monitored, and team oriented behaviors can be modeled to establish boundaries of acceptable performance.

4 Conclusion

In this paper, we present high-level hypothetical considerations that can be used to guide requirement discussions in the development of a psychomotor-based ITS. Modeling the physical world to support automated coaching of psychomotor skills is not by any means a simple task. As evident by the described modeling use cases, capturing data granular enough to inform accurate assessments is limited, even when customized environments are established. In addition, there are multiple architectural considerations that must be addressed to consume, process, and act upon behavioral information linked to a skill fundamental.

References

1. Kulik, J.A., Fletcher, J.: Effectiveness of intelligent tutoring systems: a meta-analytic review. Rev. Educ. Res. **86**(1), 42–78 (2016)
2. Mohidin, F.: Blooms Taxonomy – The Psychomotor Domain and Mind Mapping (No Date). http://www.mindmaptutor.com/blooms-taxonomy-the-psychomotor-domain-and-mind-mapping. Accessed 8 Jan 2016
3. Ericsson, K.A.: The influence of experience and deliberate practice on the development of superior expert performance. In: Cambridge Handbook of Expertise and Expert Performance, pp. 683–703 (2006)
4. Ericsson, K.A., Krampe, R.T., Tesch-Romer, C.: The role of deliberate practice in the acquisition of expert performance. Psychol. Rev. **100**(3), 363–406 (1993)
5. Fitts, P.M., Posner, M.I.: Human Performance. Brooks/Cole Publishing, Belmont (1967)
6. Goldberg, B., Amburn, C.: The application of GIFT in a psychomotor domain of instruction: a marksmanship use case. In: Proceedings of 3rd Annual GIFT Users Symposium, Orlando, FL (2015)
7. Goldberg, B., Sottilare, R., Brawner, K., Sinatra, A., Ososky, S.: Developing a generalized intelligent framework for tutoring (GIFT): informing design through a community of practice. In: Workshop at the 2015 International Conference on Artificial Intelligence in Education (AIED). Madrid, Spain (2015)
8. Sottlare, R., LaViola, J.: Extending intelligent tutoring beyond the desktop to the psychomotor domain. In: Proceedings of the Interservice/Industry Training Simulation and Education Conference (I/ITSEC), Orlando, FL (2015)