Learn Piano with BACh: An Adaptive Learning Interface that Adjusts Task Difficulty based on Brain State

Beste F Yuksel [1], Kurt B Oleson [1], Lane Harrison [1,2], Evan M Peck [3], Daniel Afergan [1,4], Remco Chang [1] and Robert JK Jacob [1]
[1]Department of Computer Science, Tufts University, Medford, MA
[2]Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA
[3]Department of Computer Science, Bucknell University, Lewisburg, PA
[4]Google Inc., Mountain View, CA
{beste.yuksel, kurt.oleson}@tufts.edu, ltharrison@wpi.edu, emp017@bucknell.edu, afergan@google.com, {remco, jacob}@cs.tufts.edu

ABSTRACT
We present Brain Automated Chorales (BACh), an adaptive brain-computer system that dynamically increases the levels of difficulty in a musical learning task based on pianists' cognitive workload measured by functional near-infrared spectroscopy. As users' cognitive workload fell below a certain threshold, suggesting that they had mastered the material and could handle more cognitive information, BACh automatically increased the difficulty of the learning task. We found that learners played with significantly increased accuracy and speed in the brain-based adaptive task compared to our control condition. Participant feedback indicated that they felt they learned better with BACh and they liked the timings of the level changes. The underlying premise of BACh can be applied to learning situations where a task can be broken down into increasing levels of difficulty.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation : Misc.

Author Keywords
Learning; Brain-Computer Interface (BCI); functional Near Infrared Spectroscopy (fNIRS); Adaptive; Cognitive Workload; Music; Piano

INTRODUCTION
Good teachers, whether they realize it or not, often guide their students into the *zone of proximal development*. That is, they help their students fulfill a learning potential with guidance that the students could not have reached by themselves [51]. However, this is not always an easy task as learning is a complex activity affected by factors such as difficulty of the task to be learned, learner cognitive ability, instructional design, and learner motivation.

CHI'16, May 7–12, 2016, San Jose, CA, USA.

ACM 978-1-4503-3362-7/16/05$15.00
DOI: http://dx.doi.org/10.1145/2858036.2858388

Intelligent tutoring systems (ITS) and other computer-based education (CBE) systems have been developed in the last decades to aid this learning process. There have been promising results; meta-analyses revealed that students tend to learn faster and more accurately, and have increased enjoyment of their learning material [27].

Cognitive learning theory (CLT) plays an important role in the designing of instructional learning material and systems. It is based on idea that there is only a finite capacity for working memory available [39, 46] so designers of learning material should avoid overloading students. One approach has been to present learners with tasks of increasing difficulty. Interestingly, meta-analyses have revealed that this is only helpful if the levels of difficulty are increased *adaptively* to learners [52, 53].

Adaptive, interactive learning tasks generally tend to use a mixture of performance and cognitive workload of learners measured by self-reporting on a one-item scale (e.g. [40]). While these studies have made progress one factor is repeatedly highlighted as the weak link in CLT studies and learning studies in general: the measurement of cognitive workload [10, 31, 24, 7]. An adaptive learning system that could adjust task difficulty based on the learner's actual cognitive state, rather than their task performance, would address this issue.

Brain sensing using functional near-infrared spectroscopy (fNIRS) has recently been demonstrated by HCI researchers to measure user cognitive workload in adaptive systems [4, 44, 1] and is resilient to subject movement as might be seen while playing a piano [43]. In this paper, we use brain sensing as the *determining factor* in a learning task by increasing task difficulty adaptively when the learner's cognitive workload falls below a certain threshold, demonstrating that they can handle more information.

We present **Brain Automated Chorales (BACh)**, an adaptive musical learning system that helps users learn to play Bach chorales on the piano by increasing the difficulty of the practice sessions based on cognitive workload using brain sensing. BACh aims to guide learners into the zone of proximal development by measuring when they can cognitively handle more information and providing them with the next stage of learning at the right moment.

We use a musical task on the piano keyboard because it lends itself well to task segmentation with increasing levels of difficulty and to high element interactivity due to concurrency. A musical task is also easy to evaluate in terms of accuracy and speed when compared with a control condition. Research has shown that cognitive workload can indeed be measured with fNIRS brain sensing while playing the piano [55].

By analyzing performance data, we found that participants played the pieces faster and with increased accuracy when using our BACh system compared to how they would normally learn a piece of music. Participants also subjectively reported that they played the pieces better, and liked the timings of the level changes.

Thus, the contributions of this paper are as follows:

1. Build an adaptive, interactive musical learning system, BACh, that uses objective measurement of cognitive workload to dynamically increase difficulty levels of Bach chorales.
2. Demonstrate that users could play the musical pieces faster and with more accuracy using BACh than a control condition.
3. Demonstrate through questionnaire and interview data that participants felt they learnt the pieces better with BACh and they liked timings of the level changes.

RELATED WORK

Computer-Based Education

Computer-based education (CBE) is a generic term that is frequently used to encompass a broad spectrum of interactive, adaptive computer applications used in teaching [21]. We will use the term CBE as the broadest umbrella term in this paper, covering all methods of computer aided pedagogy.

A more specific sub-set of computer applications used in teaching is covered by the term intelligent tutoring systems (ITS). These systems are delineated by having knowledge of: (a) *the domain* by having the expert model for the subject, (b)*the student* and following their progress, (c) *tutor model* by making teaching choices based on the student and the material or (d) *user interface model* to tie all models together [33, 32].

CBE has shown that students tend to learn more, take around 30% less time and have increased enjoyment of their subject [27]. This results in students feeling more successful leading to greater motivation to learn and increased self-confidence [6].

In this paper, we develop BACh, based on measuring and adapting to learners' cognitive workload using brain sensing, under the broad category of CBE. We do not profess to our system meeting all the requirements of the definition of an ITS, instead, however, we think that our system can certainly bring benefits to the field of ITS by providing knowledge about the learner that can sway an ITS's teaching choices.

We now discuss the foundations of cognitive load theory and how this fits into our system.

Finite Cognitive Workload and Learning

The fundamental idea behind both Baddeley's working memory model [39] and CLT [46] is that cognitive workload is limited in its capacity to handle information. CLT applies this finite, limited cognitive capacity of working memory to instructional design in order to minimize any unnecessary burdens on working memory and maximize the formation of automated schemas in long-term memory (i.e. learning).

The degree of element interactivity is seen to be the most important characteristic that determines the complexity of learning material. CLT says that an expert differs from a novice because they have previously acquired schemata that have built complex elements into fewer elements. In the example of a pianist, a skilled musician can cognitively group together notes to form chords in their schemata, whereas a beginner has to process each note individually, suggesting that the beginner will have a higher cognitive workload while learning the same piece as the skilled pianist.

Part-task approaches have been investigated to generate instructional design that does not overload the learner. However, artificially reducing cognitive workload by making the learning material easier also reduces understanding therefore it is important to know when to integrate all the parts that have been learned separately. Meta-analyses have found part-task training is only helpful with sequential material, such as playing sequential notes on a piano (e.g. [3]), rather than concurrent tasks, such as playing two hands on a piano at the same time [52, 53]. The missing element in part-task training strategies for this type of learning task is thought to be the opportunity to practice time-sharing skills [9, 53, 52] to allow for the multitask fluency. However, it has been found that increasing-difficulty task *were* helpful if the tasks adapted to the learners individually and not with fixed scheduling [52].

These findings highlight the importance of adaptive learning tasks, especially when using increasing-difficulty strategies. In order for increasing-difficulty training to be adaptive, there has to be some measurement of when the user is ready to move on to the next level of difficulty.

Measurement of Cognitive Workload in Learning Studies

The method of cognitive workload measurement in learning studies have come under criticism and scrutiny in recent years [7, 10, 31, 24]. Moreno [31] has stated that "insufficient attention has been given to the rigorous development of [cognitive load] measures".

Cognitive workload in CLT studies have typically been measured in one of three ways: self-reporting, secondary-task methodology, and physiological measures.

The most frequently used measure is the self-reporting questionnaire created by [35]. De Jong [10] and Moreno [31] raise several issues with the scale's sensitivity, reliability and validity in measuring cognitive workload. Firstly, it is not a direct measurement of cognitive workload, it is measured, instead by interpretation of post-tests [10]. The need for direct measurement of cognitive workload has been highlighted [30]. In fact, more objective measures of cognitive workload using brain sensing have shown discrepancies between self-reported levels and brain activation levels [19].

Another important factor is the timing of the questionnaires, which varies across studies [10]. Many studies only present the questionnaire after learning has taken place, so it is not clear if learners provide an average estimate for the whole task, and it leaves reports open to memory and consciousness effects. However, cognitive workload is a dynamic measure that fluctuates during a learning task [54].

A second method used to measure cognitive workload in CLT studies has been the use of secondary-task or dual-task measures (e.g. [45, 7]). An important disadvantage to this methodology is that the secondary task can interfere considerably with the primary task, especially if the primary task is complex and requires much of the learner's limited cognitive capacity [34, 7]. Furthermore, Brunken et al. [7] points out that the sensory modality of the dual tasks will affect the outcome. According to Baddeley's working memory model [39] different components of working memory use different sensory inputs. However, the different components are not completely independent and still affect each other.

Another set of methods that have been used to measure cognitive workload in learning studies are physiological measures. However, most physiological measures of cognitive workload have only been analyzed offline. This misses the advantage that physiological measures provide which is the continuous availability of data which is especially important for adaptive tasks that can react to learners' needs in real-time.

Heart rate variability (HRV) and electrodermal activity (EDA) (previously called galvanic skin response (GSR)) have been shown to be affected cognitive workload [14, 28, 41]. However, changes physiological measures such as HRV and EDA can also be mapped to other phenomenon [12] such as changes in emotional state. Task-evoked pupillary responses (TEPRs) have also been associated with changes in cognitive workload [49, 20]. However, pupil size is sensitive to a number of factors that are not relative to cognitive workload such as the ambient light in the environment [42].

Brain sensing or neuro-imaging techniques have been discussed in many studies as a promising field for cognitive workload measurements during learning [34, 7, 42, 10]. Brain sensing provides a more direct measurement of the physiological changes occurring in the brain during higher cognitive workload and can be measured continuously which can be used as a real-time determinant in adaptive learning studies.

EEG has been shown to measure cognitive workload via changes in theta and alpha range signals (e.g. [18, 17, 16]). Szafir and Mutlu [48] monitored student attention with EEG and determined which lecture topic students would benefit from reviewing. Szafir and Mutlu (2012) [47] used EEG to detect and recapture dropping attention levels in real-time. They found this improved recall abilities. This is a rare example of how real-time adaptations can help learning using physiological sensing. Instead of measuring attention levels, however, in this paper we will be measuring user cognitive workload.

fNIRS and the Prefrontal Cortex

Functional magnetic resonance imaging (FMRI) and positron emission tomography (PET) studies have established that cognitive workload can be measured in the prefrontal cortex [11, 29]. More recently, brain imaging studies have examined changes in brain function associated with improvements in performance of tasks [15, 38, 5, 25]. Decreases in activation in the prefrontal cortex as users learn have also been found using functional near-infrared spectroscopy (fNIRS). FNIRS is non-invasive imaging technique that measures levels of activation in the brain due to the hemodynamic response. Increased activation in an area of the brain results in increased levels of oxyhemoglobin [13]. These changes can be measured by emitting frequencies of near-infrared light around 3 cm deep into the brain tissue [50] and measuring the light attenuation caused by levels of oxyhemoglobin.

We can therefore use fNIRS to measure levels of cognitive activation in the anterior prefrontal cortex (PFC) by placing sensors on the forehead. The PFC is seat of higher cognitive functioning such as complex problem solving and multitasking [26]. In this paper, we measure activation in the PFC with fNIRS to analyze and respond to differences in cognitive activity while users are engaged in a musical learning task.

The fNIRS signal has been found to be resilient to respiration, heartbeat, eye movement, minor head motion, and mouse and keyboard clicks [43]. It is generally more tolerant of motion than EEG and has a higher spatial resolution. However it does have a slower temporal resolution than EEG with a delay of 5-7 seconds due to the hemodynamic response of blood flow to the brain. Due to its general ease in setting up with users and its relative tolerance of minor motion, fNIRS is an increasingly popular method of brain sensing in the HCI community [43, 4, 44, 36, 1, 2, 19].

Studies in HCI have successfully used fNIRS to measure and adapt to cognitive workload in real-time [44, 1, 2, 55]. Such studies have been aimed at increasing task efficiency [1, 2] and automation of systems [44, 55]. However, *real-time* monitoring of cognitive workload to act as determinant in a *learning* task has not yet been explored.

Harrison et al. [19] examined whether cognitive workload could be accurately monitored for incrementally changing task difficulty levels using fNIRS in an *offline analysis*. In an air traffic control simulation, they found oxygenation levels measured by fNIRS increased along with self-reports of increased cognitive workload as the number of aircrafts under operator control increased [19]. Interestingly, they also found a learning effect with lower levels of mean oxygenation values in the last two days compared to the first day [19]. However the authors did not measure task performance such as error rate or task length, so while it is assumed the operators were learning, there are no performance metrics to compare day differences in terms of learning effect. While their

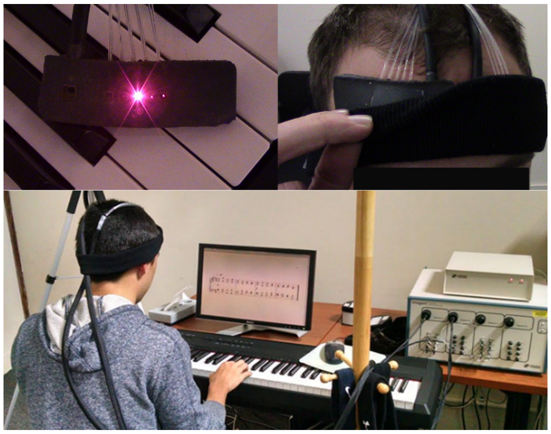

Figure 1. fNIRS equipment and setup. A fNIRS sensor (top-left) that can be placed on the forehead and held in place with a headband (top-right). Experimental setup with participant wearing fNIRS playing at piano keyboard (bottom). The Imagent is visible on the right.

study illustrates the possibility of using fNIRS as a measurement device for cognitive workload during an increasingly difficult air traffic control task, it does not use this signal for the purposes of adaptive learning.

In this paper, we use fNIRS to measure learners' cognitive workload and adapt in real-time to increase levels of difficulty in a musical learning task.

EXPERIMENTAL DESIGN

Sixteen participants (8 female, mean age of 21, SD of 2.4) took part in a within-subject design. All participants first undertook a *training task* where they played 15 easy and 15 hard pieces on the piano. Each piece was 30 seconds long with a 30-second rest between each piece.

After completion of the training task participants received a break and then moved onto the *learning task*. During the learning task, each participant learned two Bach chorales. They had 15 minutes to learn each chorale. One chorale was presented in normal form and participants were instructed to learn the piece the way they normally would. The other chorale was presented with our adaptive interface BACh. The order of the conditions alternated between participants. At the end of the allotted learning time, participants were asked to play the piece once all the way through the best they could. Performance data from both conditions were recorded and evaluated based on the performance where they played the piece all the way through as best they could. Figure 1 shows the fNIRS equipment and experimental setup. Participants wore the fNIRS equipment throughout the entire experiment, including both the control and experimental conditions.

At the conclusion of the learning task, participants were given a questionnaire on how they felt they had learned the two pieces. We also gave participants a short interview on what they thought of the adaptations made by BACh.

All participants were compensated $20. While eleven out of 16 participants rated themselves as beginners, out of the remaining five who rated themselves as intermediate, 3 of them no longer played piano. The median time our participants had played piano was 3 years.

The details of the materials, tasks, and data analysis are given below.

EXPERIMENTAL SETUP

Equipment

We used a multichannel frequency domain Imagent fNIRS device from ISS Inc. (Champaign, IL) for our data acquisition. Two probes were placed on a participant's forehead to measure data from the two hemispheres of the prefrontal cortex. Each probe contains four light sources, each emitting near-infrared light at two wavelengths (690 and 830 nm) and one detector; thus we had sixteen data channels (2 probes x 4 source-detector pairs x 2 wavelengths). The source-detector distances ranged from 1.5 and 3.5 cm, and the sampling rate was 11.79 Hz. The signals were filtered for heart rate, respiration, and movement artifacts using a third-degree polynomial filter and low-pass elliptical filter.

Training Task

All participants carried out a training task so that BACh could classify their high and low cognitive workload. Participants were given 15 easy and 15 hard pieces of music that were chosen by a musicologist to play in random order on the piano. They were given 30 seconds to sight-read each piece (i.e. play a previously unseen piece) followed by a 30 second rest period. A metronome was played at the start of each piece for 4 seconds at a speed of 60 beats per minute. Participants were asked to try to play at this speed but told they could go slower if they needed to.

The criteria for the easy pieces were that a) all notes were in C major (i.e. no sharps or flats), b) there were only whole notes (very slow, long notes), c) there were no accidentals (i.e. no additional sharps, flats, or naturals that are not part of the scale), d) all notes were within the C to G range so that participants did not need to move their hands e) there were no dynamics (i.e. volume of a note or stylistic execution). The hard pieces were chosen by a musicologist and the criteria consisted of pieces that a) had a harder key signature (most pieces had a key signature of at least 3 sharps or flats), b) contained accidentals , c) contained mostly eighth and sixteenth notes (i.e. short, fast notes), d) required some moving of the hands but not too excessively, and e) contained dynamics.

Modeling Brain Data in the Training Task

The easy and hard musical pieces were used to train the system for each individual user's cognitive activity during low and high cognitive workload, respectively. During each piece, the system calculated the change in optical intensity compared to a baseline measurement for each of the sixteen channels. Markers sent at the beginning and end of each trial denoted the segments for each piece. The mean and linear regression slope were calculated for each 30 second trial for each channel resulting in 32 features (16 channels x 2 descriptive features). These features were inputted into LIBSVM, a support vector machine classification tool, with a linear kernel [1].

Learning Task

Music for Learning Task

For the learning task, we chose two Bach chorales. These chorales were chosen because of their similarity in style and difficulty. It is often hard to find two different pieces of music that can objectively said to be similar in level of difficulty. These two chorales, however, were composed using the principles of musical counterpoint. Counterpoint is a set of rules and guidelines that dictate how music can be composed. It is a relationship between the voices of a piece that is interdependent harmonically but independent in contour and rhythm. The two chorales we chose are standardized based on this principle. While they are different pieces of music, they ought to be highly similar in musical difficulty level. They also have an equal number of accidentals (sharps, flats or naturals in the score). Both pieces are the same length, and each has four voices bass, tenor, alto, and soprano. This allowed for an easy way to segment the music in the adaptive condition in order to progressively increase the difficulty. Due to experimental constraints and skill level of participants, the pieces were slightly altered. They were transposed into the same key (G Major) and the eighth notes were removed to maintain rhythmic consistency. We note that by removing some of the notes, some minor counterpoint rules are no longer met. However, the underlying structure and form of both pieces is still intact.

Figure 2 shows the musical scores for the BACh and normal conditions. In the BACh condition, each level adds a full line of music until it reaches the full piece at Level 4. In the normal condition, learners were told they could learn the piece the way they normally would. Piano players are typically classically taught to play pieces first with their right hand, then their left hand, and finally both together. During that process, it is not uncommon for learners to play one line of the right or left hand, and add further lines as competency grows. As a result, all participants segmented the pieces while they learnt during the normal condition. Most participants started learning with their right hand only, one participant who was very good with his left hand started with his left hand first. Participants then incorporated the other hand as they progressed while learning. This can be viewed as akin to self-judgment of when to increase task complexity while learning.

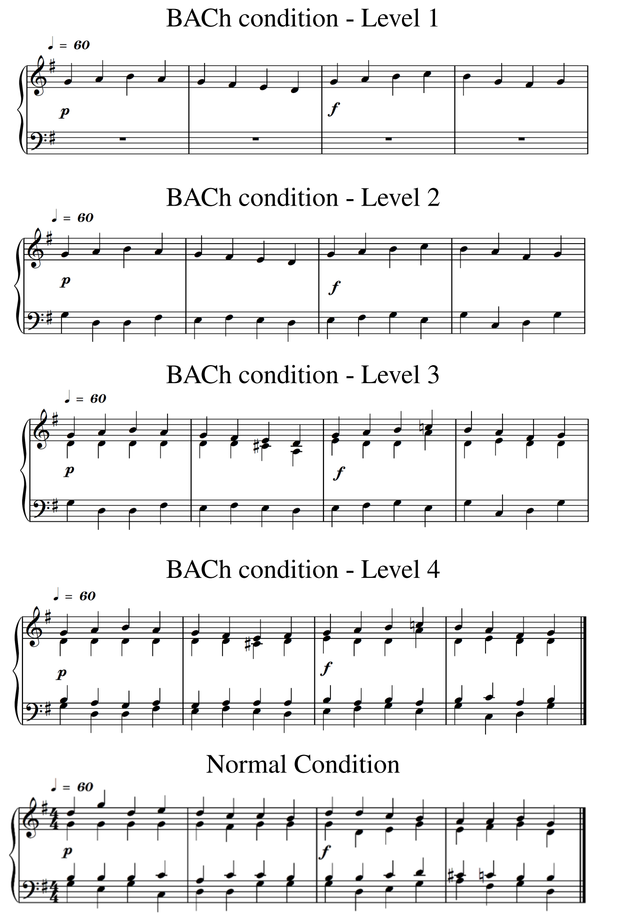

Figure 2. Scores used for BACh and Normal conditions. BACh increased the difficulty level by adding a full line of music when users' cognitive workload fell below a threshold, indicating that they could handle more information. Learners were told that they could learn the normal condition in any way they wished as they normally would. As a consequence, all learners segmented the pieces while they learnt during the normal condition.

MIDI Keyboard and Bitwig

Participants completed tasks on a full-sized Yamaha keyboard with weighted keys. The keyboard was transmitting MIDI data via USB to a computer running Bitwig studio, a digital audio workstation that allows for MIDI recording and playback as well as visual displays of recordings. Participants' musical data was recorded with this equipment for later analysis.

Learning Task Real Time Classification

BACh analyzed the last 30 seconds of real-time fNIRS data in order to calculate a prediction and confidence interval based on the LIBSVM model created by the training task. However, we found during pilot studies that this was not sensitive enough by itself for the varying levels of difficulty presented in BACh. We therefore used an algorithm that took in the LIBSVM model from the training task *and* learners' cognitive workload during each level of difficulty while they were learning the Bach chorales in order to decide when to increase task difficulty.

For each level of difficulty, we measured each learner's cognitive workload for a period of time determined by pilot studies. We started measuring their workload after they started playing *each line of the level that they were on*. For example, if the learner was on Level 3, we started measuring their cognitive workload after they were playing all 3 lines of music. We found that if we started measuring their cognitive workload earlier before the full level of difficulty, BACh would quite correctly think that they were ready to move on earlier as it was reading their cognitive workload for an easier difficulty level.

BACh would then set an individual threshold for each learner for each level based on a combination of brain data from the training task and brain data from the level that the learner was currently on. We ran many pilot studies to create an algorithm that worked for beginners.

EVALUATION OF DEPENDENT VARIABLES

Dependent Variables

We compared the following dependent variables between the BCI and control conditions:

- *Correct notes*: This is the number of correct notes on the musical score given to participants.
- *Incorrect notes*: When an incorrect note is played in the place of a correct note (Figure 3). This is a measure of precision or quality of learning.
- *Extra notes*: A note that is a correct repeat or incorrect extra note. This is different from an incorrect note. Often when learners make a mistake they will play the beat again to correct it, this will result in extra notes (see Figure 3) and is indicative of incomplete learning.
- *Errors*: A temporal group of notes that is a mistake which includes incorrect or extra notes. This is a more lenient measure as it differs from incorrect or extra notes which are counted individually. For example, a user could play several incorrect notes in one temporal group, this would be counted as one error (Figure 3).
- *Missed notes*: This measure is to account for recall (vs. precision which can be measured by the number of correct and incorrect notes). Some people may play very few notes and prioritize precision (i.e. a high number of correct notes out of total notes) but this would not be an accurate representation of learning. The number of missed notes (Figure 3) will account for recall and acts as a measure of completeness.
- *Total time played*: Learners had an unlimited length of time to play the piece once all the way through at the end. The total time it took to play the piece is a good indicator of how well they learned the piece by fast they could play it.
- *Gap between notes*: This is an indicator of well a piece is learned by how long it takes to move from beat to beat. Incomplete learning often involves hesitation and variance between notes as learners try to move from one group of notes to the next.
- *Beats per minute (bpm)*: A faster tempo indicates increased learning as players can move with more ease from beat to beat. This is different from gap between notes which can illustrate greater variance while bpm shows overall speed.

Musical Assessment

The dependent variables total time played, gap between notes and bpm were assessed computationally from the performance data. The others had to be compared to the groundtruth by hand as participants all played at different speeds and score following is an open research problem and beyond the scope of this paper. An expert musician who had taught piano in the past rated participants' learning compared to the groundtruth piano rolls. Figure 3 shows piano rolls from an example of a participant's performance compared to a computer generated groundtruth of how the piece should be played. The figure shows how incorrect notes, missed notes, extra notes and errors were identified and counted.

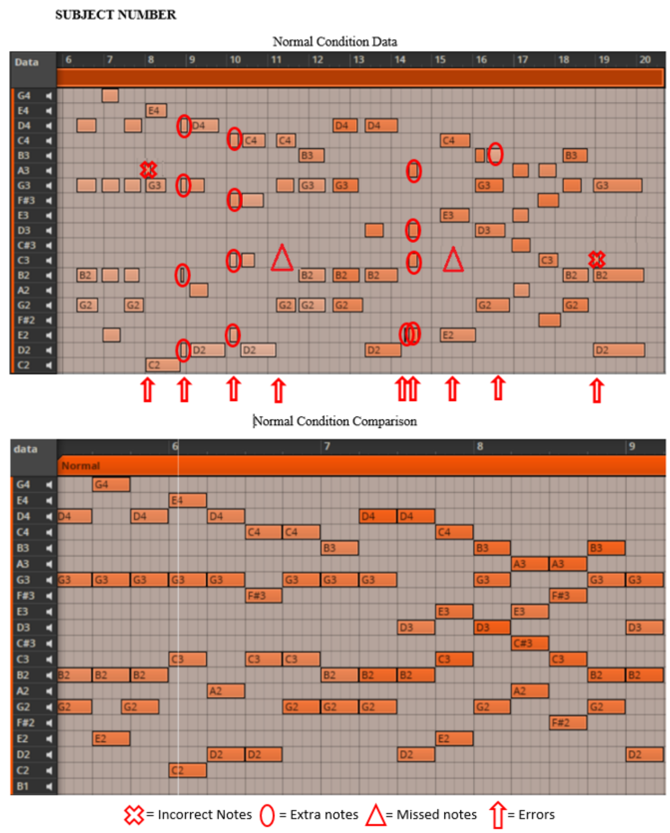

Figure 3. Examples of piano rolls from a participant's performance (top) compared with a computer generated groundtruth of the same piece (bottom). Incorrect notes are denoted by a cross, extra notes by a circle, and missed notes by a triangle. An error is a temporal group of notes and is denoted by an arrow pointing up at the column of notes containing an error.

RESULTS

fNIRS Data

We first verify that the brain data demonstrated effective classification of high and low cognitive workload while users played easy and hard pieces on the piano.

Figure 4 shows the mean and standard error in the oxygenated hemoglobin (oxy-Hb) of participants while they played easy (blue) versus hard (green) pieces on the piano. We present the mean findings across all participants across all 30 trials in Figure 4 to illustrate this general trend.

To investigate differences between hard and easy pieces, we performed a t-test on the mean change in oxygenated hemoglobin. This revealed a significant difference between conditions when participants played an easy piece ($\mu = -0.068, \sigma = 0.077$) versus a hard piece ($\mu = -0.00005, \sigma = 0.124$) on the piano ($t(29) = -2.42, p = .01$). Means and standard errors are shown in Figure 4.

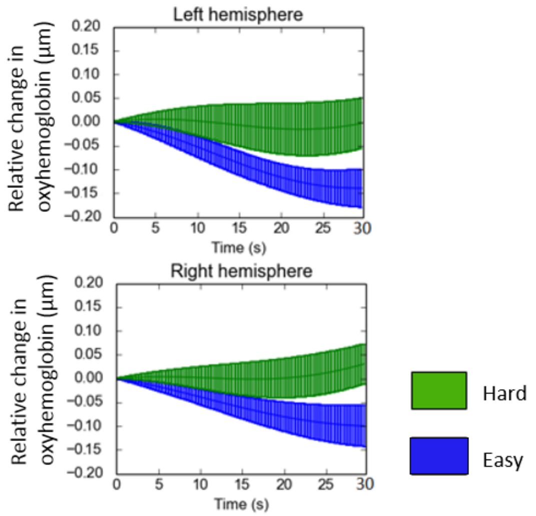

Figure 4. Mean change and standard error in oxy-Hb across all participants. Although each participant was modeled individually, the fNIRS signal exhibited a general trend with higher levels of oxy-Hb corresponding with hard pieces. The mean change in oxy-Hb was significantly higher in participants when they played an hard piece than an easy piece (p = .01).

The significantly higher levels of oxy-Hb when participants were playing harder pieces on the piano correspond with the hemodynamic literature, whereby, when there is increased cognitive activity in an area of the brain, excess oxygen is provided to that area. The increase in oxygen consumption is less than the volume of oxy-Hb provided, resulting in more oxy-Hb [13].

Performance Data

We evaluated the performance data of participants using several dependent variables (Table 1). We carried out Shapiro-Wilk tests that verifed that the data was non-parametric, followed by Wilcoxon Signed-rank tests on all dependent variables (Table 1).

Results indicated a pattern for two traits: 1) *Increased Accuracy*, and 2) *Faster speed* with BACh. We present the results below.

Increased Accuracy

Results showed that participants played significantly more correct notes ($Z = -1.9689, p = .05$), missed significantly fewer notes ($Z = 2.3151, p = .019$), played significantly fewer incorrect notes ($Z = 2.4401, p = .015$) and made fewer errors ($Z = 3.0351, p = .004$) with BACh (Table 1). All of these findings are indicative of higher accuracy with BACh.

The first two graphs in Figure 5 represent the number of incorrect notes and number of errors for each participant. The upward sloping lines (blue) are indicative of better performance with BACh with fewer incorrect notes and errors. Downward sloping lines (green) indicate better performance in the normal condition. Horizontal lines that indicate equality are also indicated in green. Only three participants played fewer incorrect notes in the normal condition, and only one participant played fewer errors in the normal condition.

Dependent Variable	Z	p	effect size
Number of correct notes	-1.9689	**0.05202**	0.304
Number of incorrect notes	2.4401	**0.0153**	0.377
Number of missed notes	2.3151	**0.01911**	0.357
Number of errors	3.0351	**0.003793**	0.468
Number of extra notes	0.8796	0.3633	–
Total time played	2.5337	**0.009186**	0.391
Mean gap between notes	2.482	**0.01099**	0.383
Average BPM	-2.719	**0.00525**	0.419

Table 1. Results from Wilcoxon Signed-rank test. Significant results are highlighted in bold. Findings indicate a pattern for increased accuracy and speed in BACh over the normal condition.

Figure 5 reveals that players who are less skilled (i.e. play more incorrect notes or make more errors) tend to benefit more greatly from BACh with steeper slopes demonstrating that there was a larger difference between the two conditions.

Faster Speed

Results also showed a propensity towards faster playing speed with BACh. With BACh, participants had a faster overall time in the playing the whole piece ($Z = 2.5337, p = .009$), had a lower time in the mean gap between notes ($Z = 2.482, p = .01$), and had a higher beats per minute (bpm) speed ($Z = -2.719, p = .005$) than in the normal condition.

The last two graphs in Figure 5 represent the total time it took to play the whole piece and the mean time in between notes. The upward sloping lines (blue) are indicative of better performance with BACh with faster overall playing time and less time between notes. Downward sloping lines (green) indicate better performance in the normal condition. Horizontal lines that indicate equality are also indicated in green. Only three participants played a faster time in the normal condition and only one participant had less gaps between notes in the normal condition.

Taken together, these results suggest that BACh helps piano players to play with significantly higher accuracy and speed than the normal condition.

Questionnaire Data

Participants' subjective ratings correspond with their performance. Figure 6 shows that participants showed noticeable preferences for BACh in how well they felt they mastered the piece, how correctly they felt they played, and how easy they felt the piece was to learn.

Timings of Level Changes

One of the key contributions of this paper is a learning task where difficulty is increased based on participant cognitive workload falling below a certain threshold. We investigated the timing of these changes while users' were learning the pieces in further detail in two ways: 1) We interviewed participants and asked them what they thought of the changes in the levels of difficulty, and 2) We investigated the variance in timing data to see if there were individual differences in learning times. We present the results below.

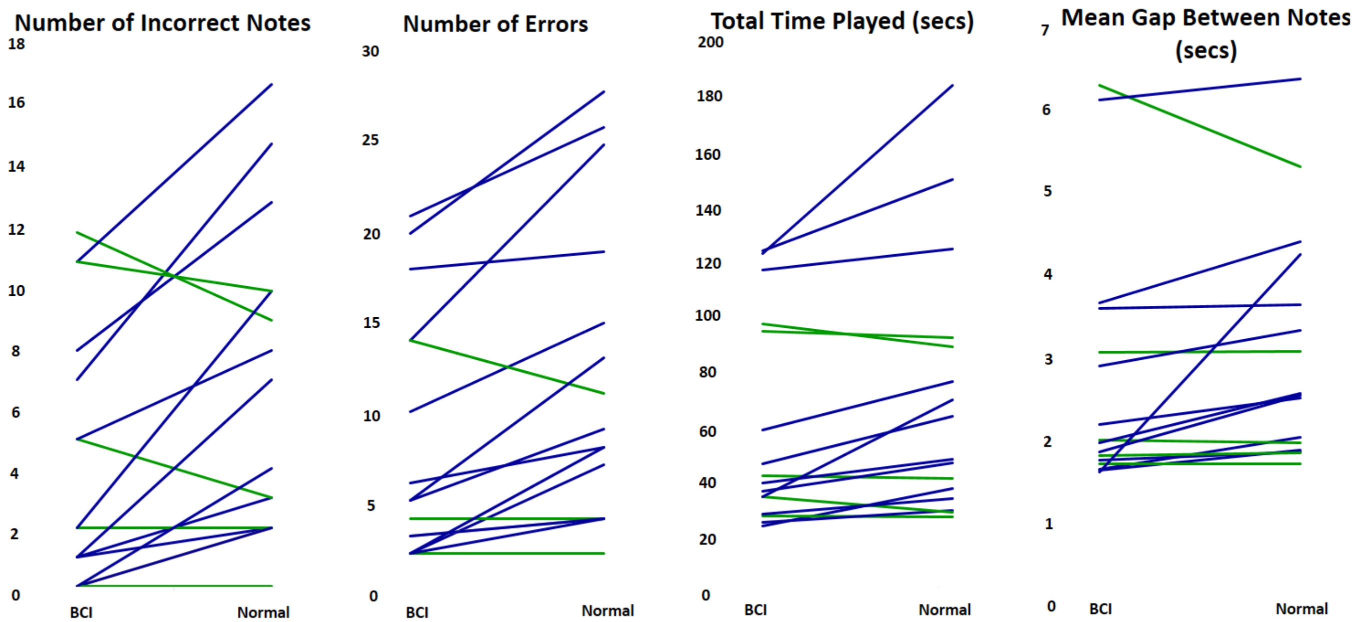

Figure 5. Slopegraphs showing the effects of BACh for each participant. Upward sloping lines (blue) are indicative of better performance with BACh. Downward sloping lines (green) indicate better performance in the normal condition. Horizontal lines that indicate equality are also indicated in green. Participants showed significantly better performance with BACh in all four conditions ($p < .01$). Interestingly, participants who were less skilled seemed to benefit more greatly from BACh.

Interview Data

After the experiment was over we gave participants a recorded interview where we asked them what they thought of the timing of the changes in difficulty levels. The goal of BACh was to change the level of difficulty at a time when the participants' cognitive workload was low enough to handle more information and a higher level of difficulty. When participants were asked about the timings of these changes, which were controlled entirely by the system, their feedback was generally positive:

> *"I thought it was good timings because by the time I learned, it gave me enough time to learn the individual lines, one by one."*

> *"I thought they were good times for changes, all of them."*

> *"Having a timing system can be jarring; you should only add new things when you know that the person has completed the existing part, but these timings were fine."*

One participant even seemed convinced that the experimenters were triggering the changes:

> *"I wasn't sure if you were controlling it or not because when it was added was a pretty appropriate time for me to add on to a part. Especially because in the beginning, one line, for me at least, is very easy to sight read so just getting that melody in my head and figuring out fingering for that one line and then adding on to it very quickly afterwards was helpful. I felt the timing was pretty good. I wasn't sure if it was timed or if you were like, oh she's done with this part, so add on to the second part."*

Such comments suggest that BACh was able to effectively support participants in their learning process by increasing difficulty levels when they were ready and able to handle more information. Some participants were not completely satisfied but in cases like these comments were similar to phrases like this:

> *"Sometimes I wouldn't notice it would change until I would look at the screen; it was a little confusing when I would look up. Yeah, it changed when I had learnt pretty much what I could learn before it changed; it was enough time to learn it."*

or

> *"I thought they [the timings] were pretty good. I think it seemed pretty good overall; the only thing would be the first one was a lot easier to learn because there was only one line, but it wasn't that bad."*

The comment about the levels changing without being noticed can be easily fixed with a small notification sound. The comment about the easiest level taking too long can also remedied by increasing the cognitive workload threshold when levels are very easy. Overall, there were no standalone or overtly negative comments about the timings of BACh. Participants seemed to feel the changes were happening at appropriate times, suggesting that we could help manage their cognitive workload as they were learning by altering the difficulty of a given stimulus.

Individual Differences in Level Changes

While participants spent less time on easier levels and longer lengths of time on levels with increasing difficulty, we did find individual differences in time spent *within* levels. The

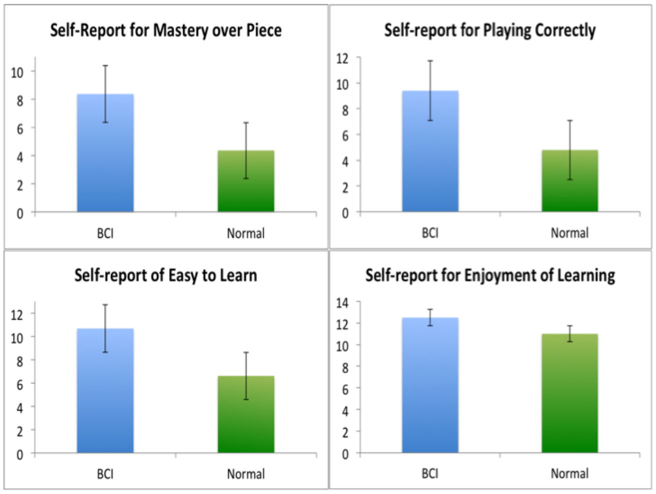

Figure 6. Mean and standard error of participants' ratings of BACh and the normal condition. Participants showed noticeable preferences for BACh in how well they felt they mastered the piece, how correctly they felt they played, and how easy they felt the piece was to learn.

length of time (in seconds) spent on level 1 ($\mu = 58.58, \sigma = 33.47$), level 2 ($\mu = 185.60, \sigma = 64.62$), level 3 ($\mu = 248.49, \sigma = 151.59$), and level 4 ($\mu = 407.32, \sigma = 179.80$) had high standard deviations, suggesting that BACh was responding to individuals' cognitive workload and learning abilities.

Figure 7 shows the variation in time within levels for each participant and the general pattern of longer time spent on more difficult levels.

DISCUSSION

We presented an adaptive brain-computer interface, BACh, that increases difficulty in a musical learning task when learners' cognitive workload fell below a certain threshold. We showed that BACh significantly increased speed and accuracy compared to a control condition where participants learned pieces the way they normally would. Participant feedback also demonstrated that they felt that they played better and it was easier to learn with BACh. Participants felt that the timings of the difficulty level changes were accurate. Results showed that while there was a general pattern of participants spending longer time on the more difficult levels, there was individual variation on time spent *within* levels, indicating that BACh was responding to individual differences.

We now discuss the challenges of modeling and adapting to learner cognitive workload using brain sensing and designing adaptations accordingly, the importance of responding to learners individually, and designing learning systems based on learner expertise.

Modeling and Adapting to Cognitive Workload

When designing BACh, we were extremely careful to first do no harm to the learning process. We did not want the adaptations made by BACh to impede or interrupt periods of learning in the user. We did this by avoiding disruptive adaptations during periods of cognitive workload that indicated possible learning.

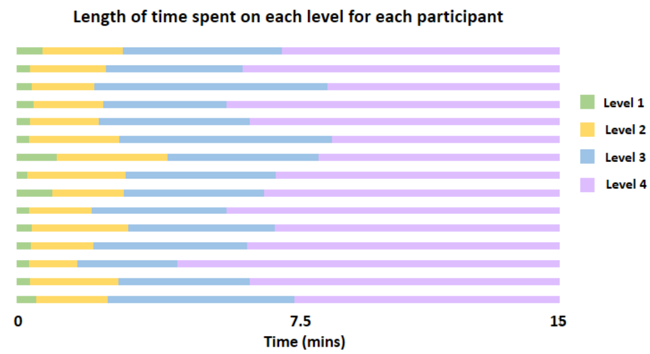

Figure 7. Length of time spent participants spent on each level of difficulty. While participants spent increasing lengths of time as level difficulty increased, there were individual variation within levels, suggesting that they system was responding to user cognitive workload for each participant individually.

One of the limitations of physiological computing is that of mapping from physiological signals to psychological states [12]. For example, if the fNIRS data signals that the learner's cognitive workload is high, this could be indicative of either a phase of learning where the user is being pushed cognitively in a constructive way, or it could indicate that the user is overloaded or overwhelmed and is not able to learn in that state.

We purposefully designed BACh not to interfere when learners' cognitive workload was high in order to avoid disrupting a period of possible learning. This is the reason why the adaptations never *decreased* the difficulty level during periods of high cognitive workload, as this could have proven both disruptive and frustrating to the learners. However, using high cognitive workload as the determining factor in adaptive learning systems would be a very interesting topic to investigate further. The discussion on whether high cognitive workload is 'good' or 'bad' during learning is still an open research problem. It may be that the solution lies in identifying learner emotion, or affect, in conjunction with cognitive workload, to reveal a higher dimensional mapping of the learner's state.

By examining the performance data and participant feedback, it seems that BACh was able to correctly respond to learner cognitive workload to improve playing speed and accuracy. This suggests that we were able to accurately model and adapt to learner cognitive workload in a learning task and guide learners into the zone of proximal development.

Responding to Learners Individually

BACh responds to each learner individually. In the early stages of development, BACh was originally designed to increase difficulty levels during learning when learners' cognitive workload fell below a *fixed percentage threshold*. Early studies showed us very quickly that this was not going to work. Each learner was very individual in their brain measurements, and what worked very well for some, would not even trigger a fixed percentage threshold for others. Furthermore, even if it worked well for one level, it certainly was not guaranteed that it would work for levels of varying difficulty.

Through a series of iterative design techniques, extensive participant feedback, and many pilot studies, we came to an algorithm that would assess learner cognitive workload using both the learner's brain data from the training task and brain data from the *current* level of difficulty that the learner was on to create a threshold for that individual on that level of difficulty.

While there is room for further improvement, our algorithm significantly increased learner speed and accuracy compared to how learners would normally learn a piece of music. Some of this individual responsiveness of BACh can be seen in the variation of the length of time taken by learners on each level. While there was a general pattern of longer periods of time spent on levels with increasing difficulty, there were still individual differences *within* each level. Individual differences in learning is a significant and complex topic and it is important for any CBE system to take it into account.

Expertise of Learners

Learning and CLT literature has explored a phenomenon called *the expertise reversal effect* whereby instructional techniques that are beneficial to beginners can have the reverse effects on more experienced learners [22, 23]. This difference between expert and novice is thought to be due to experts' previously acquired schemata that have built complex elements into fewer elements. In a task, the novice will have to use more of their limited working memory to access and process the individual elements than the expert, reflecting in the difference in skill.

Our results actually suggested that less skilled piano players benefited more from learning with BACh, as demonstrated by the steeper lines in Figure 5 for players who made more mistakes. BACh was designed for beginner piano players. We had originally recruited both beginner and intermediate piano players during our pilot studies, but quickly found that both the musical pieces and the algorithmic parameters of BACh were too easy and frustrating for intermediate players.

We foresee a similar system being of use to intermediate players with harder pieces to learn and different parameter settings for threshold changes. We can also see such a system being frustrating to even more advanced players who might prefer to learn their own way and have access to the whole piece from the beginning. We, therefore, simply suggest that our system is targeted at and useful for *beginner* piano players.

FUTURE WORK

This work takes a first step in using cognitive workload, measured by brain sensing, as a determining factor in a user interface for adaptive learning. While meta-analyses have shown that fixed scheduling when increasing task difficulty is not helpful [52, 53], now that we have sufficient data for the length of time participants spent on each level, we can also investigate and compare other conditions such as using random intervals to increase difficulty level based on means or ranges of the times spent on the different levels in this study.

We can also look at increasing task difficulty based on judgment from an expert such as a piano teacher or self-judgment from the player of when to move onto the next level. The latter is very similar to the normal condition in this paper as learners were encouraged to learn how they normally would. They therefore segmented the piece between the left and right hands and increased the complexity as their competency grew.

We are very enthusiastic about adding the detection of emotion to BACh. There has been much work carried out on the detection of emotion using physiological sensing and facial expression recognition in the field of affective computing for several decades [8]. Such measurements could be used in conjunction with fNIRS or other brain sensing devices that are relatively resilient to motion artifacts. Emotion and learning are very closely tied together [37], with frustration often preceding giving up. If a learning system could detect both cognitive workload *and* affective state, it could be very powerful learning tool indeed.

Lastly, we address the topic of generalizability to other fields of learning. There has been a wealth of learning literature on other subjects such as mathematics or electrical engineering. The underlying premise of BACh is to increase the level of difficulty in a task when cognitive workload falls below a certain threshold measured by brain sensing. This can be applied, or at least investigated, in any field where information can be presented in increasing steps of difficulty.

CONCLUSION

We presented a brain-computer interface, BACh, that measured cognitive workload using brain sensing and increased difficulty levels when learners' cognitive state indicated that they could handle more information. We showed that participants learned musical pieces with increased accuracy and could play with faster speed with BACh compared to a control where they learned the way they normally would. Participants also commented that they felt that they played better and they liked the timings of the changes. We also found individual variations in the time spent within each level of difficulty, suggesting that it is important for tutoring systems to respond to individual differences and needs.

We designed BACh for beginner piano players, however, the underlying premise of BACh can be applied to any learning situation for a complex task that can be broken down into steps of increasing difficulty. By using brain sensing, BACh is able to acquire an objective view into the learner's cognitive state and adjust the learning task to best guide the learner into the zone of proximal development.

ACKNOWLEDGMENTS

We thank Sergio Fantini and Angelo Sassaroli from the Biomedical Engineering Department at Tufts University for their expertise on the fNIRS technology; Erin Solovey from Drexel University for her expertise on processing the fNIRS signal; Kathleen Kuo and Paul Lehrman from the Music Engineering Program and Garth Griffin who is now graduated from the HCI Lab at Tufts University for their musical expertise; and Galen McQuillen from Harvard University for his expertise on pedagogy. We thank the National Science Foundation (grant nos. IIS-1065154, IIS-1218170) and Google Inc. for their support of this research.

REFERENCES

1. Daniel Afergan, Evan M. Peck, Erin T. Solovey, Andrew Jenkins, Samuel W. Hincks, Eli T. Brown, Remco Chang, and Robert J.K. Jacob. 2014a. Dynamic difficulty using brain metrics of workload. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (2014), 3797–3806.
2. Daniel Afergan, Tomoki Shibata, Samuel W. Hincks, Evan M. Peck, Beste F. Yuksel, Remco Chang, and Robert J.K. Jacob. 2014b. Brain-Based Target Expansion. *Proc. UIST 2014* (2014).
3. Daniel W. Ash and Dennis H. Holding. 1990. Backward versus forward chaining in the acquisition of a keyboard skill. *Human Factors* 32 (1990), 139–146.
4. Hasan Ayaz, Patricia A Shewokis, Scott Bunce, Kurtulus Izzetoglu, Ben Willems, and Banu Onaral. 2012. Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 1 (2012), 36–47.
5. Miriam H. Beauchamp, Alain Dagher, John A. D. Aston, and Julien. Doyon. 2003. Dynamic functional changes associated with cognitive skill learning of an adapted version of the Tower of London task. *NeuroImage* 20 (2003), 1649–1660. DOI: http://dx.doi.org/10.1016/j.neuroimage.2003.07.003
6. Ellen R. Bialo and Jay Sivin-Kachala. 1996. The effectiveness of technology in schools: A summary of recent research. (1996).
7. Roland Brünken, Jan L Plass, and Detlev Leutner. 2003. Direct Measurement of Cognitive Load in Multimedia Learning. *Educational Psychologist* 38, 1 (2003), 53–61. DOI: http://dx.doi.org/10.1207/S15326985EP3801_7
8. Rafael A. Calvo, Sidney D'Mello, Jonathan Gratch, and Arvid Kappas (Eds). 2014. *The Oxford Handbook of Affective Computing.* Oxford University Press, 2014.
9. Diane Damos and Christopher D. Wickens. 1980. The identification and transfer of time-sharing skills. *Acta Psychologica* 46 (1980), 15–39.
10. Ton de Jong. 2010. Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science* 38 (2010), 105–134. DOI: http://dx.doi.org/10.1007/s11251-009-9110-0
11. Mark D'Esposito, Bradley R. Postle, and Bart Rypma. 2000. Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. *Experimental brain research* 133, 1 (2000), 3–11. http://www.ncbi.nlm.nih.gov/pubmed/10933205
12. Stephen H. Fairclough. 2009. Fundamentals of physiological computing. *Interacting with Computers* 21, 1-2 (Jan. 2009), 133–145. DOI: http://dx.doi.org/10.1016/j.intcom.2008.10.011
13. Peter T Fox, Marcus E Raichle, Mark A Mintun, and Carmen Dence. 1988. Glucose During Focal Physiologic Neural Activity Nonoxidaive Consumption. *Science* 241 (1988), 462–464.
14. Tycho K. Fredericks, Sang D. Choi, Jason Hart, Steven E. Butt, and Anil Mital. 2005. An investigation of myocardial aerobic capacity as a measure of both physical and cognitive workloads. *International Journal of Industrial Ergonomics* 35 (2005), 1097–1107. DOI: http://dx.doi.org/10.1016/j.ergon.2005.06.002
15. Hugh Garavan, Dan Kelley, Allyson Rosen, Steve M. Rao, and Elliot A. Stein. 2000. Practice-related functional activation changes in a working memory task. *Microscopy research and technique* 51, 1 (2000), 54–63.
16. Alan Gevins, Michael E. Smith, Harrison Leong, Linda McEvoy, Susan Whitfield, Robert Du, and Georgia Rush. 1998. Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 40, 1 (1998), 79–91.
17. Alan Gevins, Michael E. Smith, Linda McEvoy, and Daphne Yu. 1997. High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cerebral Cortex* 7 (1997), 374–385. DOI: http://dx.doi.org/10.1093/cercor/7.4.374
18. Alexander Gundel and Glenn F. Wilson. 1992. Topographical changes in the ongoing EEG related to the difficulty of mental tasks. *Brain topography* 5, 1 (1992), 17–25.
19. Joshua Harrison, Kurtulus Izzetoglu, Hasan Ayaz, Ben Willems, Sehchang Hah, Ulf Ahlstrom, Hyun Woo, Patricia Shewokis, Scott C. Bunce, and Banu Onaral. 2014. Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. *IEEE Transactions on Human-Machine Systems* 44, 4 (2014), 429–440.
20. Shamsi T Iqbal, Xianjun Sam Zheng, and Brian P Bailey. 2004. Task-evoked pupillary response to mental workload in human-computer interaction. *Extended abstracts of the 2004 conference on Human factors and computing systems CHI 04* (2004), 1477. DOI: http://dx.doi.org/10.1145/985921.986094
21. Chen-Lin C. Kulik James A. Kulik and Robert L. Bangert-Drowns. 1985. Effectiveness of Computer-Based Education in Elementary schools. *Computers in Human Behavior* 1 (1985), 59–74.
22. Slava Kalyuga. 2003. The expertise reversal effect. *Educational psychologist* 38, 1 (2003), 23–31.
23. Slava Kalyuga. 2007. Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review* 19, 4 (2007), 509–539.
24. Slava Kalyuga. 2011. Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educational Psychology Review* 23 (2011), 1–19. DOI: http://dx.doi.org/10.1007/s10648-010-9150-7

25. Clare Kelly and Hugh Garavan. 2005. Human functional neuroimaging of brain changes associated with practice. *Cerebral Cortex* 15, August (2005), 1089–1102. DOI: http://dx.doi.org/10.1093/cercor/bhi005
26. Etienne Koechlin, Gianpaolo Basso, Pietro Pietrini, Seth Panzer, and Jordan Grafman. 1999. The role of the anterior prefrontal cortex in human cognition. *Nature* 399 (May 1999), 148–51. DOI: http://dx.doi.org/10.1038/20178
27. James A. Kulik. 1994. *Technology Assessment in Education and Training*. Eva Baker and Harold F. O'Neil Jr, Chapter Meta-analytic Studies of findings on computer-based instruction., 9–34.
28. Taija M M Lahtinen, Jukka P. Koskelo, Tomi Laitinen, and Tuomo K. Leino. 2007. Heart rate and performance during combat missions in a flight simulator. *Aviation Space and Environmental Medicine* 78, 4 (2007), 387–391.
29. Dara S. Manoach, Gottfried Schlaug, Bettina Siewert, David G. Darby, Benjamin M. Bly, Andrew Benfield, Robert R. Edelman, and Steven Warach. 1997. Prefrontal cortex fMRI signal changes are correlated with working memory load. *Neuroreport* 8, 2 (1997), 545–549. http://www.ncbi.nlm.nih.gov/pubmed/9080445
30. Richard E. Mayer. 2005. Cognitive theory of multimedia learning. (2005).
31. Roxana Moreno. 2010. Cognitive load theory: more food for thought. *Instructional Science* 38, November 2009 (2010), 135–141. DOI: http://dx.doi.org/10.1007/s11251-009-9122-9
32. Roger Nkambou, Riichiro Mizoguchi, and Jacqueline Bourdeau. 2010. Advances in intelligent tutoring systems. (2010).
33. Hyacinth S. Nwana. 1990. Intelligent tutoring systems: an overview. *Artificial Intelligence Review* 4, 4 (1990), 251–277. DOI: http://dx.doi.org/10.1007/BF00168958
34. Fred Paas, Juhani E. Tuovinen, Huib Tabbers, and Pascal W.M. Van Gerven. 2003. Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist* 38, 1 (2003), 63–71.
35. Fred G Paas. 1992. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology* 84, 4 (1992), 429.
36. Evan M. Peck, Beste F. Yuksel, Alvitta Ottley, Rob J.K. Jacob, and Remco Chang. 2013. Using fNIRS Brain Sensing to Evaluate Information Visualization Interfaces. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (2013), 473–482.
37. Rosalind Picard. 1997. *Affective computing* 252 (1997).
38. Russell A Poldrack. 2000. Imaging brain plasticity: conceptual and methodological issues a theoretical review. *Neuroimage* 12, 1 (2000), 1–13.
39. Grega Repovs and Alan Baddeley. 2006. The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience* 139, 1 (2006), 5–21.
40. Ron J.C.M. Salden, Fred Paas, Nick J. Broers, and Jeroen J.G. van Merrinboer. 2004. Mental effort and performance as determinants for the dynamic selection of learning tasks in air traffic control training. *Instructional Science* 32, 1-2 (2004), 153–172.
41. Wolfgang Schnotz and Christian Kürschner. 2007. A eeconsideration of cognitive load theory. *Educational Psychology Review* 19 (2007), 469–508. DOI: http://dx.doi.org/10.1007/s10648-007-9053-4
42. Holger Schultheis and Anthony Jameson. 2004. Assessing cognitive load in adaptive hypermedia systems: Physiological and behavioral methods. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 3137 (2004), 225–234. DOI: http://dx.doi.org/10.1007/b99480
43. Erin Treacy Solovey, Audrey Girouard, Krysta Chauncey, Leanne M Hirshfield, Angelo Sassaroli, Feng Zheng, Sergio Fantini, and Robert J K Jacob. 2009. Using fNIRS Brain Sensing in Realistic HCI Settings : Experiments and Guidelines. *Proc. UIST 2009* (2009).
44. Erin Treacy Solovey, Paul Schermerhorn, Matthias Scheutz, Angelo Sassaroli, Sergio Fantini, and Robert J K Jacob. 2012. Brainput : Enhancing Interactive Systems with Streaming fNIRS Brain Input. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (2012).
45. John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive science* 12, 2 (1988), 257–285.
46. John Sweller. 1999. *Instructional Design in Technical Areas*. ACER Press.
47. Daniel Szafir and Bilge Mutlu. 2012. Pay attention!: designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 11–20.
48. Daniel Szafir and Bilge Mutlu. 2013. ARTFul: adaptive review technology for flipped learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1001–1010.
49. Pascal W M Van Gerven, Fred Paas, Jeroen J G Van Merriënboer, and Henk G. Schmidt. 2004. Memory load and the cognitive pupillary response in aging. *Psychophysiology* 41 (2004), 167–174. DOI: http://dx.doi.org/10.1111/j.1469-8986.2003.00148.x
50. Arno Villringer and Britton Chance. 1997. Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences* 20 (1997). http://www.ncbi.nlm.nih.gov/pubmed/9347608

51. Lev Vygotsky. 1978. Interaction between learning and development. *Mind and Society* (1978), 79–91.
52. Christopher D. Wickens, Shaun. Hutchins, Thomas. Carolan, and John. Cumming. 2013. Effectiveness of Part-Task Training and Increasing-Difficulty Training Strategies: A Meta-Analysis Approach. *Human Factors: The Journal of the Human Factors and Ergonomics Society* (2013). DOI: http://dx.doi.org/10.1177/0018720812451994
53. Dennis C. Wightman and Gavan Lintern. 1985. Part-Task Training for Tracking and manual control. *Human Factors* 27, 3 (1985), 267–283.
54. Bin Xie and Gavriel Salvendy. 2000. Prediction of mental workload in single and multiple task environments. *International journal of cognitive ergonomics* 4 (2000), 213–242.
55. Beste F. Yuksel, Daniel Afergan, Evan M. Peck, Garth Griffin, Lane Harrison, Nick Chen, Remco Chang, and Rob J.K. Jacob. 2015. BRAAHMS: A Novel Adaptive Musical Interface Based on Users' Cognitive State. *Proc. NIME 2015* (2015).