Title
Adaptive Perceptual Learning in Electrocardiography:
The Synergy of Passive and Active Classification

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

Authors
Thai, Khanh-Phuong
Krasne, Sally
Kellman, Philip J

Publication Date
2015

Peer reviewed

Adaptive Perceptual Learning in Electrocardiography: The Synergy of Passive and Active Classification

Khanh-Phuong Thai (kpthai@ucla.edu)[1]
Sally Krasne (skrasne@mednet.ucla.edu)[2]
Philip J. Kellman (kellman@psych.ucla.edu)[1,3]
Departments of [1]Psychology and [2]Physiology and [3]Surgery, University of California, Los Angeles, CA, 90095, USA

Abstract

Recent research suggests that combining adaptive learning algorithms with perceptual learning (PL) methods can accelerate perceptual classification learning in complex domains (e.g., Mettler & Kellman, 2014). We hypothesized that passive presentation of category exemplars might act synergistically with active adaptive learning to further enhance PL. Passive presentation and active adaptive methods were applied to PL and transfer in a complex real-world domain. Undergraduates learned to interpret real electrocardiogram (ECG) tracings by either: (1) making active classifications and receiving feedback, (2) studying passive presentations of correct classifications, or (3) learning with a combination of initial passive presentations followed by active classification. All conditions showed strong transfer to novel ECGs at posttest and after a one-week delay. Most notably, the combined *passive-active* condition proved the most effective, efficient, and enjoyable. These results help illuminate the processes by which PL advances and have direct implications for perceptual and adaptive learning technology.

Keywords: perceptual learning; educational technology; active learning; passive learning; medical education

Introduction

Experts in many domains differ from novices in their ability to see patterns at a glance (Gibson, 1969). A radiologist can quickly recognize a tumor in an x-ray (Lesgold et al., 1988). A chess master can, at a glance, spot an impending checkmate multiple moves in advance (c.f., Chase & Simon, 1973). These important patterns – including relations that are quite abstract – are often invisible to novices; yet experts can recognize them rapidly and automatically. Such fluent pattern recognition characterizes experts in many domains of human expertise and largely develops from *perceptual learning*, defined by Gibson (1969) as experience-induced improvements in the extraction of information.

Until recently, perceptual learning (PL) has received little attention in instruction. Both familiarity with PL and suitable instructional methods have been lacking. Under unsystematic learning conditions, attaining expert pattern recognition may require many years of practice.

Recent research, however, has shown that PL can be systematically accelerated in real world learning domains (e.g., Goldstone, Landy, & Son, 2008; Kellman & Massey, 2013; Kellman, Massey & Son, 2009). In our work, PL methods are realized in perceptual and adaptive learning modules (PALMs). PALMs combine PL techniques with adaptive learning technology that uses both accuracy and speed to optimally sequence categories, determine mastery, and focus learning where it is most needed (e.g., Mettler & Kellman, 2014). These methods advance students' grasp of crucial structures and relations, develop fluency, and support transfer in mathematics (e.g., Kellman et al., 2009), medical learning (e.g., Krasne et al., 2013) and other domains (Mettler & Kellman, 2014).

Much is unknown about the cognitive components of adaptive PL and how these might be integrated to optimize the development of expertise in real-world settings. PALMs typically employ active classification practice, but does active classification better support PL than passive exposures to appropriate classifications? The little research done on this topic is not conclusive.

Active classification refers to learning tasks where the learners select a category label for a presented example and receive feedback that informs their perceptual, attentional and decision processes. Passive learning provides the same category membership information, but learners study the example and the category label without engaging in the choose-and-correct cycle.

Benefits of active retrieval have been well studied in the memory literature. The *testing effect* (Roediger & Karpicke, 2006) refers to the idea that when learners actively engage with the learning material by answering test questions, memory is improved. The improvement usually exceeds learning gains obtained from repeated passive study of the same information (e.g., Roediger & Karpicke, 2006). Most research on the testing effect has involved declarative memory, but similar principles may apply to PL.

Ordinary experience suggest that passive exposure alone can lead to discovery of relevant features and relations in PL. Children learn to tell dogs from cats by seeing a number of instances of dogs and cats. Novice wine drinkers can learn to discriminate between wines without any instruction (Hughson & Boakes, 2009). People can learn to recognize the styles of artists in new paintings by passive viewing of multiple samples of each artist (Kornell & Bjork, 2008).

In some cases, passive presentations may actually be better than active presentations. Passive presentation in the form of worked examples is the preferred mode of learning for novices (e.g., Recker & Pirolli, 1995), and is an effective instructional alternative to solving problems in a variety of domains. Paas and van Merrienboer (1994) studied student learning of geometrical problem solving skills and found that when students studied worked examples of problems (*passive*), they attained better accuracy on solving new problems than those who had to solve problems from scratch (*active*). Paas and van Merrienboer postulated that a considerable part of the mental effort in the active condition was allocated to processes that were irrelevant for learning. Those in the *passive* condition, on the other hand, could focus on the relevant aspects of problem structure and solutions, thus requiring less training time and less mental effort. *Passive* learning trials also offer error-free exposures to the classifications to be learned, eliminating residual effects of incorrect guesses that may occur in *active* learning. Conversely, Bodemer & Faust (2006) found that when asking students to make active connections between multiple representations of fractions, they were better able to understand the underlying structures of fractions than when they passively observed the correspondences.

Research on category learning suggests that passive and active processes have complementary benefits. An active task tends to encourage learners to focus on information that distinguishes categories, while a passive task tends to engage them with finding within-category regularities (e.g., Markman & Ross, 2003; Carvalho & Goldstone, 2014). In a recent article, Levering and Kurtz (2015) compared the category knowledge produced by an active classification task and a passive observational learning task. They trained participants to discriminate between two artificially created categories, each with 5 stimuli, in which a single feature determined category membership and other features correlated but did not perfectly predict category membership. They found that the active learning task biased learners toward more discriminative learning compared to the passive learning task. However, passive learning allowed for enhanced sensitivity to the features that were not perfectly predictive.

It is possible that combining passive and active learning may be most beneficial. This hypothesis accords with research on skill acquisition by Renkl, Atkinson, and colleagues under the ACT-R framework (e.g., Atkinson, Derry, Renkl, & Worthham, 2000; Renkl, Atkinson & Grobe, 2004), in which passive study of examples is valuable early in training. Much of this work focused on procedural problem solving domains, for which a smooth transition (fading) from study of worked-out examples to problem solving may be ideal. Initial passive presentations can reduce cognitive load early in training when it is highest by not having to engage in decision-making processes, resulting in fewer unproductive learning events (Renkl et al., 2004). Active learning, in contrast, forces guessing at the start, which might lead to cognitive overload. Wrong guesses or hypotheses may also tend to linger and impede later learning. In addition, being forced to produce responses without knowing much may be frustrating, undercutting motivation in some learners.

Most potential advantages of passive exposure can be realized by using passive trials only at the start of learning. Initial passive study in PL might focus learners' attention on specific features that define each category and in turn support the acquisition of the category representation. As the learning progresses, active learning can support discriminative processes needed for correct classification. Active learning after an initial stage may be especially valuable in an adaptive framework. We sought to test this hypothesis in a real-world, complex PL domain.

We trained undergraduates to classify seven diagnostic patterns in electrocardiography. The 12-lead electrocardiogram (ECG) is one of the oldest and most informative cardiac assessments available. Visual interpretation of ECGs, however, requires superior perceptual recognition skills ordinarily attainable only through years of practice (Wood, Batt, Appelboam, Harris & Wilson, 2013). One difficulty is discriminating relevant from irrelevant information in ECGs. For any category, some locations contain relevant information while some do not. Each category involves patterns of diagnostic features, but the features are variable across the ECG traces. Salient features of an ECG trace do not necessarily indicate an abnormality, and waveforms that indicate normality at one location may not be normal at another. Thus, learners have to know not only what to look for but also where to look.

We created three versions of an ECG PALM involving: (1) only *active* classification of ECGs for the underlying diagnostic pattern, (2) only *passive* presentations of the correct interpretations, (3) initial passive presentations followed by active classifications (*passive-active* condition). The *active* and *passive-active* conditions involved classification with feedback and were adaptive to the learner's performance, and the *passive* training involved study of the correct interpretations and was not adaptive. To compare learning across conditions, we examined participants' ability to correctly and quickly classify novel ECG traces into trained categories of diagnostic patterns. All *active* trials used an adaptive learning system – the ARTS (Adaptive Response-time-Based Sequencing) system (Mettler & Kellman, 2014).

Method

Participants

87 undergraduates from University of California, Los Angeles without any prior knowledge of ECG interpretation participated in the experiment for course credit.

Materials

The materials consisted of 250 unique 12-lead ECG traces from real patients, with 26 - 46 unique traces for each of seven categorical diagnostic patterns. The seven patterns were: Normal, Acute Anterior ST Segment Elevation Myocardial Infarction, Acute Inferior ST Segment Elevation Myocardial Infarction, Right Bundle Branch Block, Left Axis Deviation, Right Axis Deviation, Old Inferior Myocardial Infarction.

The training consisted of two phases: a brief *primer* on ECG interpretation (same for all conditions) and the PALM phase with either *active*, *passive*, or *passive-active* task formats. The *primer* was a PowerPoint slideshow consisting of a brief explanation of the ECG, how to measure widths and heights on the ECG grid, and one example of a typical

ECG trace for each diagnostic pattern. In each example, the relevant features were marked and described, similar to samples provided in textbooks. No other information about the heart anatomy, physiology, or other basics of ECG interpretation was provided in the primer.

In the *active* PALM, on each trial, participants chose among seven choices the diagnostic category for a given ECG trace. *Figure 1a* shows an example trial. Accuracy and speed were continually tracked; trial feedback was given after each response and block feedback was given after every 12 trials. The trial feedback played a sound corresponding to the correctness of the response, and displayed the correct answer, and response time when correct. It also marked relevant features on the ECG, along with a brief description of those features as seen in the *primer*. Block feedback provided average accuracy and speed by block and percentage of categories completed. Feedback screens were not timed. *Figure 1b* shows an example feedback screen following an incorrect response. Categories were adaptively sequenced based on both accuracy and response times as according to the ARTS sequencing algorithm (see Mettler & Kellman, 2014). Categories were dropped (retired) from the training set after reaching learning criteria (i.e., correctly identified consecutively in 4 out of 4 presentations, each in under 15 seconds). Participants completed the module when they had retired all 7 categories.

In the *passive* PALM, each trial was the same as the correct trial feedback screen for the *active* group (similar to *Figure 1b*). The correct label, the relevant features and their descriptions were provided, and participants were asked to pay attention and to study each correct diagnosis. The *passive* condition thus did not have classification feedback and was not adaptive. To equate the total number of trials across two groups, we yoked each participant in the *passive* training condition to the total number of trials seen by another participant in the *active* training condition. To determine how many items per category to show, we used the average proportions of trials per category that a pilot group of *active* participants needed to complete the module. These proportions were similar across *active* participants, so we used the same proportions for all *passive* participants. The duration of each *passive* trial was 13 seconds, determined from the average amount of time it took pilot participants in the *active* group to respond and view the trial feedback. After 13 seconds, the screen cleared. To keep the participants engaged and to equate the existence of a motor response with the *active* condition, participants had to click on a Next button to see the next trial, and there was a sound played to signal the beginning of each trial. There was an untimed break every 12 trials.

In the *passive-active* PALM, participants viewed a set of 14 *passive* trials (two examples from each category) as in the *passive* condition, in random order, before moving on to the adaptive *active* classification trials for which participants received the same feedback and learning criteria as those in the *active* classification condition. All three PALMs used the same pool of ECGs.

Three assessments, each consisting of 14 new ECG's (two from each category), were used in counterbalanced order as pretest, posttest, and delayed-posttest. None of the ECGs used in the assessments appeared in the PALM. Each assessment trial presented an ECG and seven answer choices (*Figure 1a*). No feedback was given after each trial.

Procedure and Design

Participants were given 20 minutes to study the *primer* followed by a quiz on which they were asked to match the descriptions of the diagnostic features to each of the seven heart patterns shown on the primer. This was to ensure that participants were familiar with the diagnostic features of each heart pattern. They checked their answers afterward.

After the quiz, participants took the pretest and were randomly assigned to learn with either the *active*, *passive*, or *passive-active* PALM. When participants finished the module (or after the 2-hour time allotted), they completed the posttest and a survey. The survey asked about their prior knowledge of ECG reading, amount of sleep they had the night before, and demographic information (age, gender, English fluency). Because *passive* and *active* training may differ not only in cognitive aspects of learning, but also in the motivational and engagement aspects, we asked participants to report their levels of engagement and enjoyment of the training experience, and to provide a judgment of learning and memory for the delayed test. Participants returned for the delayed-posttest one week later.

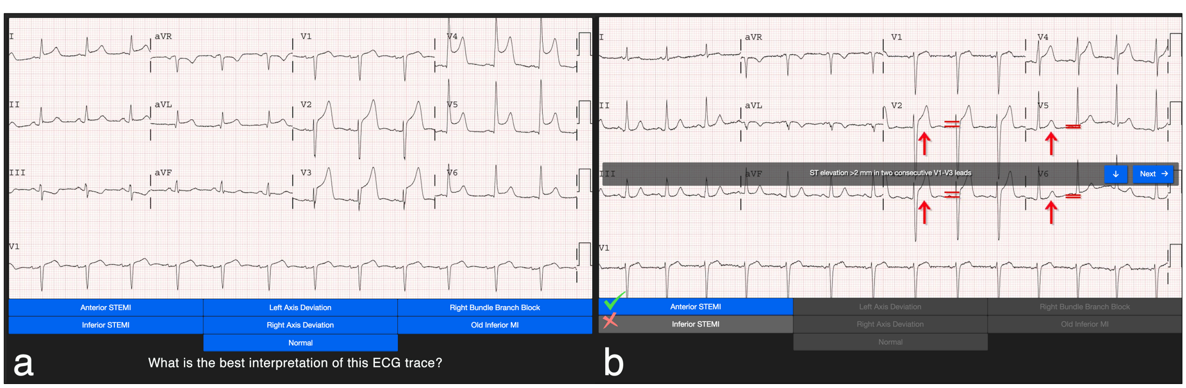

Figure 1. (a) Sample active classification trial; (b) feedback provided when incorrect.

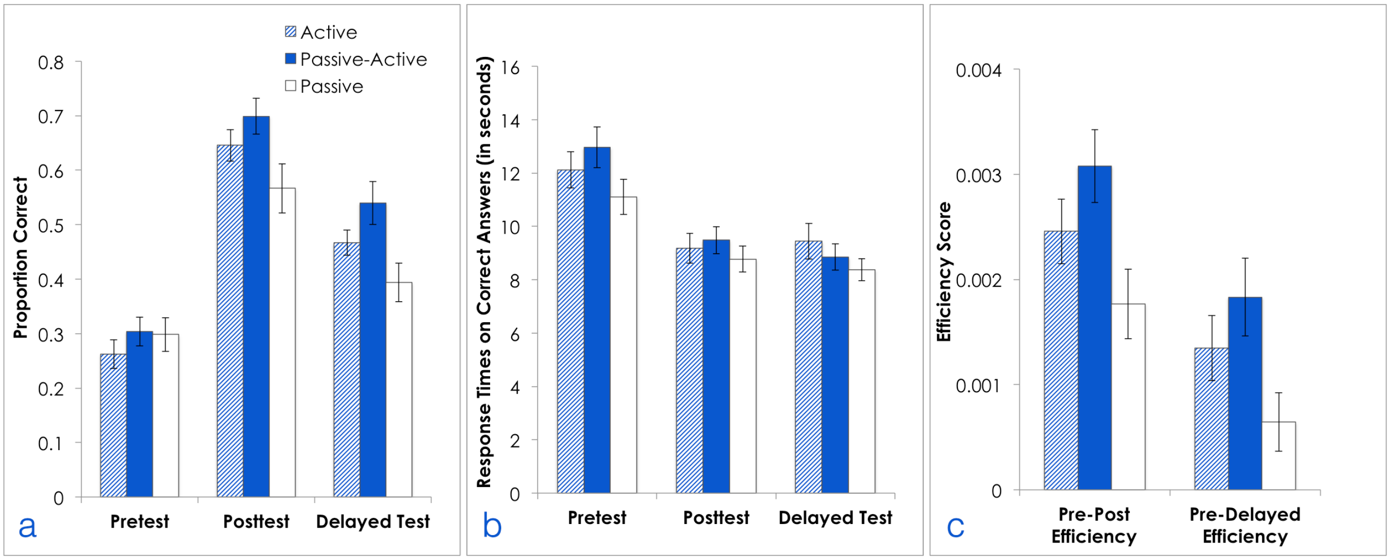

Figure 2. (a) Accuracy, (b) Response times on correct answers, (c) Training efficiency across conditions. Error bars ± 1SE.

Dependent Measures, Data Analyses and Hypotheses Based on prior work, we expected all PALMs to produce robust improvements in classification, and the *passive-active* group to produce the best results. Because we used learning to criterion, our primary measure was learning *efficiency*, defined as accuracy gain from pretest to posttest divided by minutes invested in the training. We expected the *active* group to have greater improvements in accuracy and/or response time (for correct answers -- RTc) than the *passive* group. We used analysis of covariance (ANCOVA) in analyzing differences among the groups in accuracy gain, RTc change, and Efficiency because of possible differences in pretest mean values between groups.[1] Participants in the *active* group on average retired 87.3% and the *passive-active* 89.9% of the categories. To compare the effectiveness of the training conditions, we report results from participants who completed the assigned modules (N = 23 per condition). The same patterns of results were found with all participants (N = 27 per condition). Yoking by number of trials was not perfect for 5 pairs of participants; however, we retained them in the analyses because (1) removing them did not change the results, and (2) total trials and training times were similar between the *active* and *passive* groups. The three groups did not differ on quiz performance or any other measures not reported here. Because we sought to compare differences across training conditions, we conducted planned comparisons across conditions. All statistical tests were two-tailed, with a 95% confidence level.[2]

Results and Discussion

Figure 2 shows the average accuracy, RTc at each test phase, and efficiency scores by condition. As expected, participants showed substantial learning gains from pretest to posttest and retained much of their learning at delayed posttest, regardless of condition. Participants were able to interpret ECGs they had never seen before and to do so with improved speed. The *passive-active* condition produced the greatest learning gain with the fewest training trials. The *active* condition also produced greater learning gains than the *passive* condition. *Table 1* contains the descriptive statistics from the training for each condition.

Accuracy

Accuracy Gain. We analyzed accuracy gain (posttest minus pretest) in a 2 phase (pre-post, pre-delayed) x 3 condition (*active*, *passive*, *passive-active*) repeated-measures ANCOVA with pretest accuracy as the covariate. The covariate, pretest accuracy, was significantly related to the posttest gains, $F(1,65) = 48.89$, $p < .001$, $\eta^2 = .43$. Indeed, better pretests predicted less improvement at posttest, $r = -.43$, $p < .001$, and delayed test, $r = -.55$, $p < .001$, suggesting that pretest variations were largely due to chance. After controlling for the effect of the pretest, there was a reliable effect of condition, $F(2, 65) = 6.00$, $p < .01$, $\eta^2 = .16$. The *active* and *passive-active* conditions produced higher gains than the *passive* condition, $t(44) = 2.18$, $p < .05$, $d = .64$; $t(44) = 2.41$, $p < .05$, $d = .71$, respectively. There were no reliable differences in accuracy gains between the *passive-active* and *active* conditions, $t(44) = .68$, $p > .05$, and no significant interactions (p's > .05).

There was a statistically significant main effect of phase, $F(1, 65) = 5.54$, $p < .05$, $\eta^2 = .08$. The pre-post accuracy

Condition	Total Trials Completed	Minutes on Training	Training Accuracy
Active	167.52 (11.82)	43.96 (3.37)	.49 (.02)
Passive-Active	137.78 (6.69)	37.70 (2.68)	.57 (.02)
Passive	159.74 (8.64)	47.91 (2.40)	--

Table 1. Average training performance across the three experimental groups (Standard errors in parentheses). Both passive and active trials were included in total trials completed for the *passive-active* condition.

[1] Assumptions for ANCOVA were met for all dependent variables, F's < 1, p's > .05.
[2] Due to small sample size, we followed the recommendations of Nakagawa (2004) and provided effect size estimates to evaluate the strength and direction of each relationship in our multiple tests.

gain was reliably higher than the pre-delayed gain (34% vs. 16%, respectively, $d = .81$).

Raw Accuracy. We also compared raw accuracy across groups. A 3 phase (pre, post, delayed test) x 3 condition ANOVA confirmed a main effect of condition, $F(2,66) = 4.61$, $p < .05$, $\eta^2 = .12$. The *passive-active* condition outperformed both the *active*, $t(44) = 2.15$, $p < .05$, $d = .62$, and *passive* conditions, $t(44) = 2.64$, $p < .05$, $d = .78$, on overall accuracy. *Active* and *passive* did not differ reliably, $t(44) = 1.23$, $p = .22$. There was a marginally significant phase x condition interaction, $F(4,132) = 1.99$, $p < .10$, $\eta^2 = .06$. There were no condition differences at pretest ($p > .10$), but the *passive-active* group outperformed the *passive* group at both posttest, $t(44) = 2.37$, $p < .05$, $d = .70$, and delayed test (50% vs. 39%), $t(44) = 2.75$, $p < .01$, $d = .82$. The *active* group also had a marginally higher delayed test accuracy than the *passive* group, $t(44) = 1.73$, $p < .10$, $d = .51$. There were no other differences among conditions.

Response Times

Generally, participants became faster at arriving at the correct answers at posttest and delayed test. However, there were no reliable effects of condition or phase (pre-post vs. pre-delayed post), p's $> .05$.

Efficiency

After controlling for the effect of pretest accuracy, there was a reliable main effect of condition, $F(2, 77) = 6.10$, $p < .01$, $\eta^2 = .14$. There were no reliable differences between *active* and *passive-active* groups in the average efficiency ($p = .11$), but both of the *active* and *passive-active* groups had better efficiency than the *passive* group, $t(44) = 2.34$, $p < .05$, $d = .69$, and $t(44) = 4.41$, $p < .001$, $d = 1.01$.

Since there were no condition differences in pretest accuracy, we also analyzed efficiency uncorrected for pretest variation (assuming that differences seen in pretest, before any experimental treatments, were random). *Passive-active* outperformed *active* at delayed test, $t(44) = 2.14$, $p = .04$, $d = .63$, and marginally at immediate posttest, $t(44) = 1.87$, $p = .07$, $d = .56$. *Passive-active* and *active* also had higher efficiencies than *passive* at both immediate and delayed posttest, $t(44) > 2$, $p < .03$, $d = .67$ to 1.31.

Progression of Learning

Figure 3 shows the average accuracy over the first 17 training blocks for the *active* and *passive-active* conditions. The *passive-active* group performed consistently better than the *active* group, $t(44) = 3.84$, $p < .001$, $d = 1.13$, after 3 blocks, $t(44) = 2.79$, $p < .01$, $d = .82$. This result suggests that initial passive exposure speeds learning relative to starting with active classification, despite the similar number of learning trials in the passive portion and the first active trial block. In the first few blocks, the abrupt change from passive to active introduced similar error rates as those in the *active* condition. However, after the first few blocks, as we expected, those in the *passive-active* group made fewer errors. These gains appear to be preserved through the course of learning and in posttests.

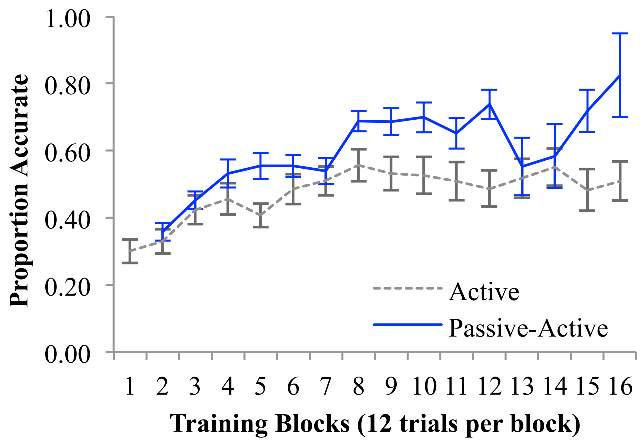

Figure 3. Average accuracy across training blocks. The *passive-active* group received 14 *passive* trials at block 1.

Self-Report Ratings

On the survey, participants differed marginally in how they responded to "How enjoyable was the training as a whole, on a scale from 1-6 (1 = not at all enjoyable, 6 = very enjoyable)",[3] $F(2,61) = 2.51$, $p < .10$. The *passive-active* PALM was reliably more enjoyable ($M = 4.55$, $SD = 1.23$) than the *passive* PALM ($M = 3.76$, $SD = 1.22$), $t(47) = 2.01$, $p = .05$, $d = .64$, and marginally more enjoyable than the *active* PALM ($M = 3.90$, $SD = 1.14$), $t(46) = 1.74$, $p < .10$, $d = .55$. Participants in the *passive-active* training condition also self-reported to be more highly motivated and engaged during the module (on a scale from 1-6, 1 = not at all, 6 = very much, $M = 4.90$, $SD = .72$) than the *active* ($M = 4.38$, $SD = 1.02$) and the *passive* groups ($M = 3.95$, $SD = 1.36$), $t(39) = 1.87$, $p = .07$, $d = .59$, and $t(39) = 2.77$, $p < .01$, $d = .87$. We found no differences in the reported level of engagement and motivation between the *active* and the *passive* condition, $p > .05$. There were no condition differences on the other self-report measures.

Conclusion

The *passive-active* condition in this study, consisting of initial passive exposure, followed by active adaptive learning, produced durable learning that was faster, more accurate, efficient, and enjoyable than *passive* learning for the same amount of time in this complex pattern recognition domain. *Passive-active* also outperformed *active* adaptive learning on some measures, especially comparisons during the course of learning (*Figure 3*), as well as in accuracies and efficiencies uncorrected for what were likely random pretest variations across groups. Effect sizes for learning differences between *passive-active* and *active* ranged from around .6 to .8, which are medium to large effect sizes. The *active* condition in this experiment, as well as the active part of the *passive-active* condition, utilized the ARTS adaptive learning algorithm previously found to be highly effective in earlier work. The *passive-active* condition here appears to markedly enhance a learning approach that has been

[3] The survey was implemented shortly after data collection began, so we did not have responses from the first 8 participants.

previously shown to outperform classic adaptive learning systems and a number of presentation schemes in adaptive PL (Mettler & Kellman, 2014).

The advantages of *passive-active* learning may have several explanations. Consistent with work on cognitive load and worked examples (e.g., Renkl et al., 2004), initial familiarization with category exemplars may allow relevant structure to be learned without imposing the additional task demands of active responding. Moreover, passive and active learning may complement each other in focusing attention on within-category similarities and between-category contrasts respectively (e.g., Carvalho & Goldstone, 2014). Specific advantages of passive exposure at the start of learning may include avoiding initial errors and persistence of incorrect guesses, as well as averting frustration that may arise in active learning from having to guess initially.

This work has clear practical implications. Incorporating passive-active training is an easily implemented technique that is likely to improve learning technology. The *primer* used in this study, modeled after textbook instruction, prepared undergraduates to benefit from the ECG PALMs, but it was clearly not sufficient for producing highly accurate or fluent interpretation of heart patterns (e.g., accuracy levels after the *primer* averaged around 30% (pretest scores in *Figure 2A*). Thus, the present results confirm the importance of PL interventions as a valuable complement to declarative and procedural components of instruction (Kellman & Massey, 2013). Our results also raise a number of new research questions. For example, does the combination of passive and active classification produce similar learning gains and efficiency in other domains, particular in those where the learner already has more prior knowledge? How many passive exposures are optimal, and what is the relationship between the optimal number of exposures and the complexity of the learning domain? Additional research will be needed to further understand and optimally utilize the passive-active synergy.

Acknowledgments

We thank Joel Zucker, Tim Burke, Rachel Older, and research assistants of the UCLA Human Perception Lab, and gratefully acknowledge support from NSF REESE Award No. 1109228 and the US Office of Naval Research to PJK.

References

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*(2), 181-214.

Bodemer, D., Faust, U. (2006). External and mental referencing of multiple representations. *Computers in Human Behavior, 22*, 27-42.

Carvalho, P. F., & Goldstone, R. L. (2014). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review*, 1-8.

Chase, W.G. & Simon, H.A. (1973). Perception in chess. *Cognitive Psychology, 4(*1), 55-81.

Gibson, E.J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Goldstone, R.L., Landy, D., and Son, J.Y. (2008). A well grounded education: The role of perception in science and mathematics. In M. De Vega, A. Glenberg, and A. Graesser (Eds.), *Symbols, Embodiment, and Meaning* (pp. 327–355). Oxford Press.

Hughson, A. L., & Boakes, R. A. (2009). Passive perceptual learning in relation to wine: Short-term recognition and verbal description. *The Quarterly Journal of Experimental Psychology, 62(*1), 1-8.

Kellman, P. J., & Massey, C. M. (2013). Perceptual learning, cognition, and expertise. *Psychology of Learning and Motivation, 58,* 117 – 165.

Kellman, P.J., Massey, C., & Son, J.Y. (2009). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science*, 1-21.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the enemy of induction? *Psychological Science, 19, 585-592.*

Krasne, S., Hillman, J. D., Kellman, P. J., & Drake, T. A. (2013). Applying perceptual and adaptive learning techniques for teaching introductory histopathology. *Journal of Pathology Informatics*, *4(34)*.

Markman, A.B., & Ross, B.H. (2003). Category use and category learning. *Psychological Bulletin*, *129*, 592- 613.

Mettler, E., & Kellman, P. J. (2014). Adaptive response-time-based category sequencing in perceptual learning. *Vision Research*, *99*, 111-123.

Mettler, E., Massey, C. & Kellman, P. (2011). Improving adaptive learning technology through the use of response times. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Boston, MA: Cognitive Science Society, 2532-2537.

Nakagawa, S. (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology, 15*, 1044-1045.

Paas, G., & van Merrienboer, J. (1994). Variability of worked examples and transfer of geometrical problem solving skills: a cognitive-load approach. *Journal of Educational Psychology, 86*, 122-133.

Pirolli, P., & Recker, M. (1994). Learning strategies and transfer in the domain of programming. *Cognition and Instruction, 12,* 235–275.

Renkl, A., Atkinson, R. K., & Grobe, C. S. (2004). How fading worked solution steps works – A cognitive load perspective. *Instructional Science, 32*, 59-82.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.

Wood, G., Batt, J., Appelboam, A., Harris, A., & Wilson, M. R. (2013). Exploring the impact of expertise, clinical history, and visual search on electrocardiogram interpretation. *Medical Decision Making*, *34*(1), 75-83.