PROCEEDINGS

ICMPC14

JULY 5 - 9, 2016
Hyatt Regency Hotel
San Francisco, California

International Conference on Music Perception and Cognition

14th Biennial Meeting

Perceptual Learning of Abstract Musical Patterns: Recognizing Composer Style

Carolyn A. Bufford,[*1] Khanh-Phuong Thai,[2] Joselyn Ho[*], Cindy Xiong[*], Carly A. Hines[*], Philip J. Kellman[*]
[*]*Department of Psychology, University of California, Los Angeles, USA*
[1]cabufford@g.ucla.edu

ABSTRACT

How can we improve abstract pattern recognition in music? Can principles enhancing visual learning be extended to auditory stimuli, such as music? Perceptual learning (PL), improvements in the pickup of information from experience, is well-established in both vision and audition. Most work has focused on low-level discriminations, rather than extraction of relations in complex domains. In vision, recent research suggests that PL can be systematically accelerated in complex domains, including aviation, medicine, and mathematics, through perceptual and adaptive learning modules (PLMs). PLMs use brief interactive classification trials along with adaptive spacing to accelerate perceptual category learning. We developed a Composer PLM to apply these techniques to recognizing composers' styles, a challenging, abstract, PL task involving complex relations over time. We investigated whether (1) participants could learn composers' styles, and (2) whether the PLM framework could be successfully extended to rich auditory domains. On each PLM trial, participants listened to a 15-second clip of solo piano music by one of 4 composers (2 Baroque: Bach, Handel; 2 Romantic: Chopin, Schumann), attempted to identify the composer, and received feedback. Before and after multiple PLM sessions, we tested participants' ability to distinguish clips from composers in the PLM, unlearned composers in trained periods (Baroque: Scarlatti; Romantic: Mendelssohn) and composers in untrained periods (Renaissance: Byrd; post-Romantic: Debussy). At pretest, participants' sensitivity was at chance, but they showed robustly improved classification at posttest on both composer styles ($p < .01$) and period styles ($p < .001$). Results indicate that PLM training can improve participants' recognition of composers' styles, demonstrating that (1) composer style can be learned, and that (2) PL-based interventions are effective in complex auditory domains.

I. INTRODUCTION

Humans have impressive abilities to recognize structure and patterns in the world, in multiple modalities and many domains. For example, we can recognize familiar voices on the phone and identify objects at unusual sizes and angles. Music, in particular, is a useful domain in which to study pattern recognition because of its many patterns at many levels of complexity and abstraction which allow research into auditory perception across the spectrum from basic processing to high-level real-world tasks. The simplest patterns in music are specific note durations and pitches. Themes are more complex patterns, but even untrained listeners can recognize a theme at its restatement in a piece they have not heard before (Java, Kaminska, & Gardiner, 1995). The styles of composers are perhaps the most abstract and complex patterns that humans can detect in music, because these patterns transcend individual pieces or works. Research on perceptual learning as well as research specifically on composer style contribute insights into our ability to recognize composer styles, which are abstract auditory patterns.

A. Perceptual Learning

Perceptual learning (PL) refers to improvements in the pickup of information as a function of experience (Gibson, 1969). In particular, PL has been shown to accelerate the development of expert pattern recognition, contributing to the development of expertise. PL is well-established in both vision (Sagi, 2010) and audition (Sabin, Eddins, & Wright, 2012). Much work in both modalities has focused on expertise in low-level discriminations. For example, many studies have demonstrated PL in studies of pitch duration (e.g. Karmarkar & Buonomano, 2003) or orientated visual gratings (e.g. Song, Peng, Lu, Liu, & Li, 2007). However, perceptual learning also supports the extraction of relations in complex domains. In vision, recent research suggests that PL can be systematically accelerated in aviation (Kellman & Kaiser, 1994), medicine (Thai, Krasne, & Kellman, 2015), mathematics (e.g. Kellman & Massey, 2013; Kellman, Massey, & Son, 2010; Landy & Goldstone, 2007), and other complex domains. Perceptual and adaptive learning modules (PALMs) accomplish this acceleration by using brief, interactive classification trials (e.g. Mettler & Kellman, 2014; Mettler, Massey, & Kellman, 2011). PALMs have only been created for visual stimuli thus far, which do not share the temporal nature of auditory stimuli. Can principles enhancing visual learning be extended to auditory stimuli, such as music?

B. Composer Style

Psychologists studying musical style perception have focused primarily on documenting human sensitivity to period styles (e.g. Hasenfus, Martindale & Birnbaum, 1983), leading to the development of hundreds of machine learning programs that simulate this impressive human ability (e.g. Widmer, 2005). However, while our human ability to recognize patterns in period styles has been well-documented, psychologists have done very little empirical work on whether humans have a similar sensitivity to composer styles, a topic that musicologists have long been interested in and discussed. Tyler (1946) played the three movements of a Mozart sonata and selected and played three movements each of trios by Beethoven and Schubert to students in a music appreciation course and found that students could distinguish between the composers better than chance, but she did not control for musical form (e.g. sonata or ballad) nor did she study learning. Crump (2002) chose J. S. Bach and Mozart because the composers were "stylistically distinct", and then only used the Goldberg variations for Bach and Minuets for Mozart. Participants and a machine learning algorithm successfully learned to distinguish these composers, but confounding musical form with composer limits the generalizability of these results. Due to the limitations of these studies, the empirical proof of composer style and of the learnability of composer style is uncertain.

C. Current Study

We developed a Composer PALM to apply perceptual learning technology to recognizing composers' styles, a challenging, abstract, perceptual learning task involving complex relations over time. We investigated whether (1) participants could learn composers' styles, and (2) whether the PALM framework could be successfully extended to a rich auditory domain.

II. METHOD

D. Participants

Forty-three (19 men, 24 women) undergraduate students at the University of California, Los Angeles participated for course credit in psychology or linguistics courses. Two participants were excluded for non-completion of the posttest. Included participants have learned to play an instrument for on average 7.17 years, and specifically, piano, for on average 6.88 years. Seventeen participants did not have any prior musical training.

E. Materials

1) *Music Stimuli.* There were clips from eight different composers from four different time periods (Renaissance: Byrd; Baroque: Bach, Handel, and Scarlatti; Romantic: Chopin, Mendelssohn, and Schumann; and post-Romantic: Debussy). The composers were chosen partially based on the size and availability of their musical work. The stimuli were 15-second clips of classical piano music collected from Youtube.com or the UCLA Music Library and were balanced with respect to stylistic features such as tempo and form. We chose to only use piano music so that instrumentation would not be confounded with composer or musical period. Thus, timbral information was relatively constant across recordings, periods, and composers. Participants, therefore, had to rely on more temporally extended patterns to learn to distinguish composers. We used only modern solo piano recordings (no harpsichord or four-hand recordings).

1) *Assessments.* The pretest and posttest assessments consisted of twenty-four clips from four composers trained in the Composer PALM (Bach, Handel, Schumann, and Chopin) and four untrained composers (Byrd, Scarlatti, Mendelssohn, Debussy).

Sixteen of the clips were from the trained composers (four each), and for each of these trained composers, about two clips were included in the PALM, and about two clips were new, novel clips that were not included in the PALM, to test both for learning in the PALM and for transfer. For each composer, the clips were divided among the composer's typical and atypical composition style. Experienced listeners determined typicality of composer style by listening extensively to all of the collected clips and subjectively categorizing them as typical or atypical of each composer's style.

The other eight clips were from the untrained composers (two each) - four from trained periods (Baroque and Romantic) and four from untrained periods (Renaissance and post-Romantic).

In each of 24 trials, the participants listened to a clip and chose who they think composed the music clip. They were given seven response options, clustered by period. Four of the options were the four PALM-trained composers, and the remaining three were "Other Baroque" (Scarlatti), "Other Romantic" (Mendelssohn), and "Other Period" (Byrd or Debussy) (see Figure 1).

There were three versions of the assessment (A, B, C), all containing unique clips. The amount of clips per composer in each version was held constant. The participants were randomly assigned versions in their pretest and posttest, without replacement, such that no participant received the same version in posttest as they did in pretest.

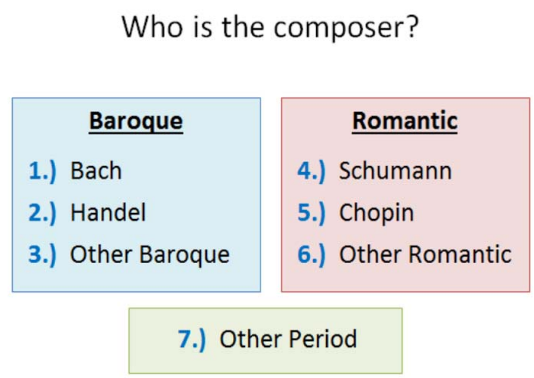

Figure 1. Response options for assessment. The "Other Baroque" composer was Scarlatti, the "Other Romantic" composer was Mendelssohn, and the "Other Period" composers were Byrd (Renaissance) and Debussy (post-Romantic).

2) *Intervention.* The Composer PALM consisted of 400 clips - 100 from each of four composers from two periods (Baroque period: Bach and Handel, Romantic period: Chopin, Schumann). Participants were presented with many trials, each containing one of the fifteen-second clips. In each trial, participants listened to a music clip and attempted to choose the correct composer from the four given composers (Bach, Handel, Chopin, and Schumann). If they answered correctly, the PALM showed that their answer was correct and let them proceed to the next trial (see Figure 1B). If their response was incorrect, the participants were shown the correct answer and listened to the remainder of the music clip (see Figure 1C).

For every twenty trials they completed, they were given a feedback summary on their average accuracy and average response time over the block of trials. Participants were also given feedback on mastery, using objective criteria. A participant reached mastery for a given composer when they correctly answered four of five consecutive trials of that composer, with a response time of less than 21s on those correct trials. Mastery level, the number of composers mastered, was indicated in the PALM by four circles on the bottom of the screen; for each composer mastered, one circle was filled in green. The module was delivered online, via a web browser.

F. Procedure

Participants took the pretest at the laboratory. Afterwards, they were instructed to undergo PALM training on their own time, using their own computer, for 45 minutes every day for seven days or until they completed the PALM by reaching mastery for all four of the composers. We sent daily emails reminding them to complete their training for the day as well as updating them on their progress. Following completion of the

PALM or a maximum of seven days of training, the participants returned to the laboratory to take the posttest.

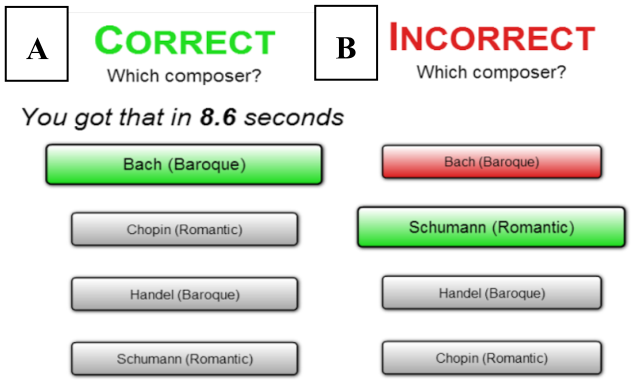

Figure 2. Screenshots of PALM trials showing feedback. Panel A shows a correctly answered PLM trial. Each trained composer is a response option, and their period is noted in parentheses. Feedback on accuracy was given on every trial. Panel B shows an incorrectly answered PLM trial. Participants saw which composer they had selected and which composed the music in the clip. After incorrect answers, participants listened to the rest of the clip.

G. Dependent Measures

This study used sensitivity as a dependent measure, with correction $1/(10n + 1)$, where n is the number of signal trials for hit rates and n is the number of noise trials for false alarm rates. For example, out of the twenty-four assessment trials, four were music clips by Bach. This means Bach had signal trials correction $1/(10(4) + 1) = 1/41$, and noise trials correction $1/(10(20) + 1)=1/201$.

III. RESULTS

H. PALM

On average, participants completed 482.10 (range: 94 - 979, *SE* = 34.83) trials in 4.25 (range: 0.58 - 13.32, *SE* = 0.45) hours over 6.95 (range: 1 - 16, *SE* = 0.51) PALM sessions before returning to the lab.

I. Pretest Assessment

At pretest, we evaluated participants' initial sensitivity to each composer and period style by testing whether their pattern of sensitivities differed from the hypothesis of no sensitivity (d' = 0). Specifically, we conducted a one-way repeated-measures analysis of variance (ANOVA) of Composer (Bach, Chopin, Handel, Mendelssohn, Scarlatti, Schumann, Other Period [Byrd+Debussy]) on sensitivity. Our custom hypothesis test of the insensitivity hypothesis confirmed that participants were insensitive to composer styles (M = -0.11, SE = 0.17) at pretest, $F(1,40) = 0.42$, $p = .52$, *partial-eta-squared* = 0.01. A second one-way repeated-measures ANOVA of Period (Baroque, Romantic, Other Period [Renaissance + post-Romantic]) on sensitivity with a custom hypothesis test confirmed that participants were also insensitive to period styles (M = 0.21, SE = 0.15), $F(1,40) = 2.08$, $p = .16$, *partial-eta-squared* = 0.05.

We also conducted a manipulation check of the three assessment versions via a mixed ANOVA of sensitivity with Assessment (A, B, C) as a between-subjects variable and Period as a within-subjects variable. There was no interaction, but there was a statistically significant main effect of Assessment such that performance on version A (N = 12, M = -0.26, SE = 0.14) was significantly worse than performance on versions B (N = 12, M = 0.29, SE = 0.14) and C (N = 17, M = 0.24, SE = 0.12), Pillai's Trace $F(1,40) = 4.91$, $p = .01$, *partial-eta-squared* = 0.21. Because of this performance difference, we analyzed improvement from pretest to posttest using analyses of covariance (ANCOVAs) of sensitivity change (d' change = posttest d' - pretest d'), controlling for pretest sensitivity.

J. Composers

One of our primary goals was to empirically investigate whether composer style could be identified and learned, by testing sensitivity to clips' composers. We conducted a custom hypothesis test to test the hypothesis of no sensitivity change (d' change = 0) in a repeated-measures ANCOVA for Composer (Bach, Chopin, Handel, Mendelssohn, Scarlatti, Schumann, Other Period [Byrd+Debussy]) on sensitivity change, covarying out pretest sensitivity. The custom hypothesis test confirmed that participants significantly improved their sensitivity to composers (see Figure 3) from pretest to posttest (M = 0.61, SE = 0.21), $F(1,33) = 8.29$, $p < .01$, *partial-eta-squared* = 0.20. We also found a significant main effect of Composer: participants improved on all composers (Bach M = 0.26, SE = 0.14; Chopin M = 0.36, SE = 0.14; Mendelssohn M = 0.43, SE = 0.10; Scarlatti M = 0.29, SE = 0.10; Schumann M = 0.59, SE = 0.14; Byrd and Debussy M = 0.38, SE = 0.12) except Handel (M = -0.17, SE = 0.14), Pillai's Trace $F(6,28) = 3.05$, $p = .02$, *partial-eta-squared* = 0.40.

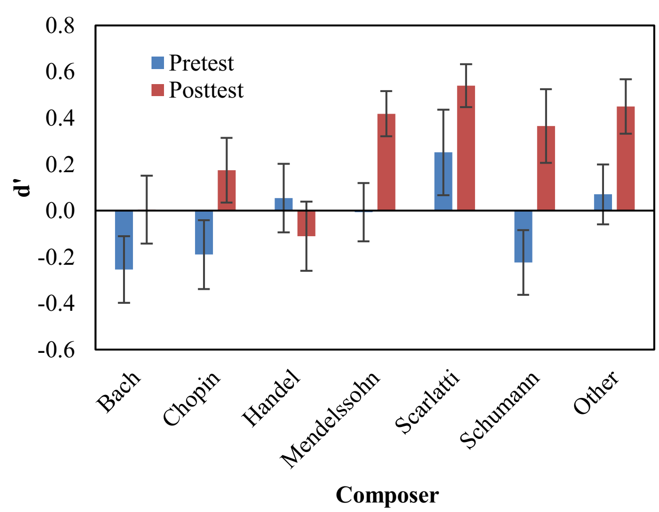

Figure 3. Participants' sensitivity at pretest and posttest for the seven possible composer options in the assessment. The trained composers (included in the PLM) were Bach, Chopin, Handel, and Schumann.

K. Periods

Given that our composers were clustered in the Baroque and Romantic periods and that sensitivity to period style has been well-documented (e.g. Hasenfus et al., 1983), we were interested to see if our participants would incidentally learn period style through PALM training on composers. We included a custom hypothesis test of the hypothesis of no change in sensitivity (d' change = 0) in a repeated-measures ANCOVA of Period (Baroque, Romantic, Other Period [Renaissance + post-Romantic]) on sensitivity change, covarying out pretest sensitivity to periods. The custom hypothesis test demonstrated that participants significantly improved their sensitivity to periods (*M* = 1.18, *SE* = 0.17) through incidental exposure during the PALM (see Figure 4), $F(1,37) = 40.06$, $p < .001$, *partial-eta-squared* = 0.52. The ANCOVA revealed no main effect of Period: participants performed equally well on the Baroque (*M* = 0.59, *SE* = 0.13) and Romantic (*M* = 0.78, *SE* = 0.15) periods and the combination of the Renaissance and post-Romantic periods (*M* = 0.38, *SE* = 0.12).

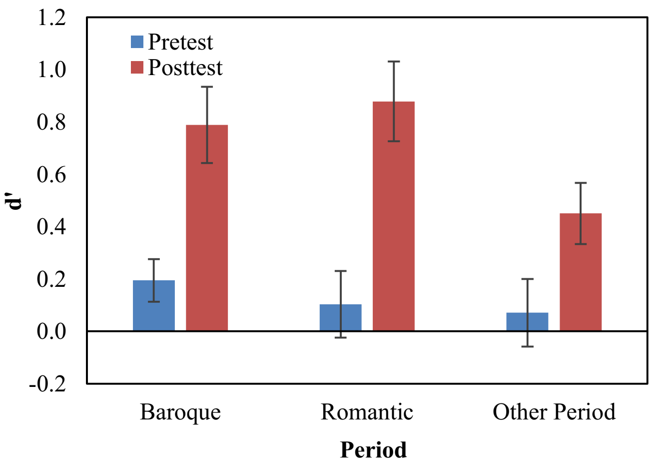

Figure 4. Participants' sensitivity at pretest and posttest for the two trained periods (Baroque and Romantic) and the untrained periods ("Other Period" = Renaissance and post-Romantic).

L. Familiarity

The goal of perceptual learning is not memorization of particular instances, but learning of structural regularities that facilitate accurate classification of new instances. If learning is primarily of this type, we would expect that posttest performance on familiar and novel examples should be similar. We included both familiar clips from the PALM and novel clips in each version of the assessment. To test for learning of the training music and for transfer to novel clips, and assess memorization of familiar clips as an explanation for our results, we compared sensitivity at posttest for familiar and novel clips. On period sensitivity at posttest, a two-way repeated-measures ANOVA of Period (Baroque, Romantic) and Familiarity (Familiar, Novel) confirmed no significant main effects or interactions. Participants showed no reliable differences for familiar Baroque (*M* = 1.14, *SE* = 0.26), novel Baroque (*M* = 0.80, *SE* = 0.20), familiar Romantic (*M* = 1.19, *SE* = 0.26), and novel Romantic (*M* = 0.96, *SE* = 0.19) clips. These results indicate that improvements were based on the pickup of relevant structure through perceptual learning rather than clip memorization.

Similarly, for composer sensitivity at posttest, we found little evidence of familiarity for each composer. We conducted a two-way repeated-measures ANOVA of Composer (Bach, Handel) and Familiarity (Familiar, Novel). We found no interaction or main effect of Composer, but we did find a main effect of Familiarity, such that sensitivity was higher on novel (*M* = 0.54, *SE* = 0.14) than familiar (*M* = -0.16, *SE* = 0.14) clips, $F(1,40) = 12.81$, $p < .001$, *partial-eta-squared* = 0.24. Participants were actually more sensitive to novel Bach (*M* = 0.38, *SE* = 0.18) and novel Handel (*M* = 0.72, *SE* = 0.22) clips than to familiar Bach (*M* = -0.23, *SE* = 0.18) and familiar Handel (*M* = -0.09, *SE* = 0.17) clips. This contradicts the explanation of participants memorizing clips because memorization predicts better performance on learned clips, not better performance on novel clips. There is no obvious explanation for the superiority of novel clips, but one possibility is that the clips used in training were somewhat less representative of the composer, or more difficult in some way, than the clips used in the assessments.

We could not include Chopin and Schumann in the above analysis because a programming error had caused assessment version B to have only novel clips (no clips included in the PALM) for Schumann, and assessment version C to have only novel clips for Chopin. To analyze familiarity for these composers, we conducted paired-samples *t*-tests on the data of participants not impacted by the error. For Chopin, participants (*N* = 27) were equally sensitive to familiar (*M* = 0.33, *SE* = 0.25) and novel (*M* = 0.38, *SE* = 0.30) clips, $t(26) = -0.12$, $p = .90$, *Cohen's d* = -0.02. For Schumann, participants (*N* = 26) were more sensitive to familiar (*M* = 0.91, *SE* = 0.32) than novel (*M* = 0.19, *SE* = 0.18) clips, $t(25) = 2.18$, $p = .04$, *Cohen's d* = 0.43. Of all the trained composers and periods, only Schumann lends any support to the explanation of familiarity or clip memorization. Overall, the familiarity results gave evidence that participants learned composers' and period styles.

IV. DISCUSSION

We used perceptual learning technology to investigate empirically human learning of composer styles. We found that participants successfully learned composer styles and incidentally learned period styles. Perceptual learning technology facilitated perceptual learning with music, complex auditory stimuli.

Gains in sensitivity were equal for trained and untrained composers. The improved sensitivity to untrained composers in trained periods (i.e. Baroque: Scarlatti, Romantic: Mendelssohn) and composers in untrained periods (Renaissance: Byrd, post-Romantic: Debussy) must then be due to learning of both trained composers' styles and of the Baroque and Romantic period styles.

Familiarity with the clips included in both the PALM and the assessments does not explain these results. Participants averaged 482.10 (*SE* = 34.83) trials in training with a 400-clip training set, so repeating clips, particularly clips included in the assessments, was a concern. Tests of sensitivity to familiar versus novel clips, however, generally revealed no effect of familiarity. For Bach and Handel, participants actually showed significantly greater sensitivity on novel clips!

In the absence of timbral cues and instrumentation, we found that participants could learn composer styles and period styles with PL technology. Because composers often favor particular forms, musical form is another potential confound, and prior work had confounded musical form and composer (Crump, 2002; Tyler, 1946). We selected a range of musical forms for each composer. Our participants demonstrated learning and sensitivity under more controlled and demanding conditions than those used in prior work. Thus, our results are strong empirical evidence for the existence and learnability of composer style in humans.

Perhaps psychologists have rarely previously studied composer style because of its complexity and difficulty to learn. Undergraduates who have learned to play an instrument (e.g. piano) for on average 7.17 years showed no sensitivity to composer or period styles at pretest; with PALM training they showed significant sensitivity to both composer styles and period styles. Seven years of musical training was insufficient to learn composer and period; but with only four and a half hours of PALM training, they gained expertise in these composer and period styles that would otherwise take years to learn. This is remarkable.

Our results demonstrate that perceptual learning technology is effective for learning auditory stimuli, even in complex and challenging domains. This suggests that perceptual learning interventions could accelerate learning in other complex auditory domains. Perhaps future work can use perceptual learning technology to investigate these domains, including important domains such as language.

ACKNOWLEDGMENT

The authors gratefully acknowledge support from NSF REESE award #110922 and the US Department of Education, Institute of Education Sciences (IES), Cognition and Student Learning (CASL) Program awards R305A120288 and R305H060070. We thank Timothy Burke, Joel Zucker, Rachel Older, and the Human Perception Lab research assistants for expert assistance.

REFERENCES

Crump, M. (2002, May). A principal components approach to the perception of musical style. In *Paper at the Banff Annual Seminar in Cognitive Science (BASICS),* Banff, Alberta.

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

Hasenfus, N., Martindale, C., & Birnbaum, D. (1983). Psychological reality of cross-media artistic styles: Human perception and performance. *Journal of Experimental Psychology, 9*, 841–863.

Java, R. I., Kaminska, Z., & Gardiner, J. M. (1995). Recognition memory and awareness for famous and obscure musical themes. *European Journal of Cognitive Psychology, 7*, 41–53.

Karmarkar, U. R., & Buonomano, D. V. (2003). Temporal specificity of perceptual learning in an auditory discrimination task. *Learning & Memory*, *10*(2), 141-147.

Kellman, P. J., & Kaiser, M. K. (1994, October). Perceptual learning modules in flight training. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 38, No. 18, pp. 1183-1187). SAGE Publications.

Kellman, P. J., Massey, C. M., & Son, J. Y. (2010). Perceptual learning modules in mathematics: Enhancing students' pattern recognition, structure extraction, and fluency. *Topics in Cognitive Science, 2*(2), 285–305.

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. J. (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 (pp. 2532-2537). Boston, MA: Cognitive Science Society.

Sabin, A. T., Eddins, D. A., & Wright, B. A. (2012). Perceptual learning evidence for tuning to spectrotemporal modulation in the human auditory system. *The Journal of Neuroscience*, *32*(19), 6542-6549.

Sagi, D. (2011). Perceptual learning in vision research. *Vision Research*, *51*(13), 1552-1566.

Song, Y., Peng, D., Lu, C., Liu, C., Li, X., Liu, P., et al. (2007). An event-related potential study on perceptual learning in grating orientation discrimination. *Neuroreport*, *18*(9), 945–948.

Thai, K. P., Krasne, S., & Kellman, P. J. (2015). Adaptive perceptual learning in electrocardiography: The synergy of passive and active classification. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Tyler, L. E. (1946). An exploratory study of discrimination of composer style. *The Journal of General Psychology*, *34*(2), 153-163.

Widmer, G. (2005). Studying a creative act with computers: Music performance studies with automated discovery methods. *Musicae Scientiae*, *9*(1), 11-30.