2015

Cognitive Factors Predicting Expressive Music Synchronization: Roles for Auditory Imagery and Working Memory

Ian Daniel Colley
idc001@bucknell.edu

Recommended Citation
Colley, Ian Daniel, "Cognitive Factors Predicting Expressive Music Synchronization: Roles for Auditory Imagery and Working Memory" (2015). *Master's Theses*. 139.
https://digitalcommons.bucknell.edu/masters_theses/139

COGNITIVE FACTORS PREDICTING EXPRESSIVE MUSIC SYNCHRONIZATION:

ROLES FOR AUDITORY IMAGERY AND WORKING MEMORY

by

Ian D. Colley

A Thesis

Presented to the Faculty of Bucknell University In Partial Fulfillment of the Requirements for the Degree of Master of Science in Psychology

Approved:

Andrea R. Hal

Adviser: Andrea Halpern

J. T. Ptacek

Department Chairperson: John Ptacek

April, 2015___________
(Date: Month and Year)

I, Ian Colley, do grant permission for my thesis to be copied.

Table of Contents

List of Tables

List of Figures

Abstract

Sensorimotor synchronization (SMS) is especially apparent—and therefore readily studied—in musical settings, as most people are naturally able to perceive a musical beat and synchronize to it (e.g. by tapping a finger). SMS processes have been tested extensively using pseudo-musical pacing signals, so we chose to extend this by using naturalistic, expressively timed piano music, characterized by slight tempo fluctuations for artistic interpretation. Previous research has also shown that people vary greatly in their SMS abilities. Given the dynamic nature and variability of SMS, we hypothesized that individual differences in working memory and auditory imagery—both fluid, cognitive processes—would predict SMS at two levels: 1) asynchrony (a measure of synchronization error), and 2) anticipatory timing (i.e. predicting, rather than reacting to beat onsets). In Experiment 1a, participants (*N* = 36) completed two working memory tests, a tempo imagery test, a pitch imagery test, and a self-report test of auditory imagery with separate subscales for vividness (clarity of an image) and control (ability to alter an image). They were then tested in a SMS-tapping task. In Experiment 1b, the same set of tasks was given to highly trained musicians. In Experiment 2, participants were given an expressive timing perception test to see the extent to which the cognitive variables related to perception without action. Hierarchical regression models were used to assess the contribution of the cognitive variables to SMS. Results showed dissociations among imagery types as they relate to asynchrony, perception, and suggest a role for working memory in anticipatory timing. Musicians performed better on the SMS task, but showed fewer correlations between the cognitive variables and SMS. These results suggest that in nonmusicians imagery for pitches and temporal patterns is important for synchronizing to an auditory stimulus, but working memory is implicated in strategically synchronizing via anticipation of beat onsets.

Introduction

Human behavior is marked at its highest level by the union of many basic psychological and motor processes. The leisurely, and seemingly frivolous activities that we have developed—for instance, competitive athletics, dramatic theater, or language arts—are examples of how cognition, perception, emotion, and action are synthesized to produce phylogenetically novel behaviors (McDermott & Hauser, 2007). One of the most common—and possibly oldest (Patel, 2008; Hodges, 1996)—of these high level behaviors is music. The human fascination with music seems to appear with very little learning at an early age (Ilari, 2004), and is persistent throughout life (Halpern & Bartlett, 2002). There are rare exceptions, as in cases of amusia (Ayotte, Peretz, & Hyde, 2002), but generally people readily and repeatedly engage in musical behaviors, as in listening to the radio during a commute, or learning an instrument as a hobby. Interestingly, despite the apparent ease with which most people perceive and appreciate music, there are numerous individual differences in natural and trained musical abilities (Prior & Troup, 1988; Bayner, 1998; Naoumenko, 1982; Sloboda, 2000).

Fortunately, individual differences in psychological processes can be studied using many validated laboratory tasks and highly controlled stimuli. Why study individual differences in musical contexts? If musical abilities are in fact the product of low-level mechanisms, why not simply study those mechanisms? There are several reasons to examine music in scientific inquiry. For example, the apparent individual differences in musical abilities help us understand how the brain makes sense of, and derive meaning from, artistically arranged sounds. This approach can then be applied to certain populations. Musical settings promote social bonding (Tarr, Launay, & Dunbar, 2014) and the associated benefits of a sense of belonging (Rentrow, 2012). Music is also used in anxiolytic therapies (Lin, 2014; Croom, 2015), and seems to offer some mild therapeutic benefits in neurodegenerative disorders such as Parkinson's (Nombela et al, 3013). By understanding how aspects of music fit into models of cognition and neural systems, such therapies can become more focused and effective, especially as we understand how individual cases of pathology respond differently to treatments.

From a scientific perspective, music can inform our understanding of general psychological theories. That is, music as a medium for conducting empirical tests can expand our knowledge of many human cognitive processes, given that it is a complex and uniquely human behavior. For example, memory is widely studied using recall or recognition paradigms that require participants to remember lists, or identify items that were not in a list. This sort of memory test does not capture people's remarkable memory for melodies; even years after hearing a well-liked tune people can easily recognize it, even in cases of Alzheimer's disease (Halpern & Bartlett, 2010). However, tunes are very difficult to learn within a single laboratory session (Halpern & O'Connor, 2000). Lists, on the other hand, can be learned very quickly, but can decay after just a few hours. Thus, there may be certain qualities of memory that are best studied in music or related media.

Similarly, action planning and motor learning are frequently studied using standardized paradigms such as grip force and optical following response (Wing, 1997; Flanagan et al, 2003). But music offers insight into the precise timing mechanisms of action as they relate to perception. This is because music is a dynamic perceptual and motor process. Trained pianists execute elaborate sequences of finger movements that are nearly perfectly coordinated between the two hands. In an ensemble setting, similar action sequences are adjusted and updated as performers receive auditory feedback from their partners. Even non-musicians readily tap a finger or sway along to passages of music. Thus, the dynamic and integrative nature of music perception and action makes music an excellent tool for empirical investigation. This is especially true for studies of how motor behavior is connected to sensory systems. One way of describing this connection is through theories of sensorimotor synchronization, which is closely related to musical behaviors.

Music as Sensorimotor Synchronization. The process of coupling sensory inputs with motor outputs is called sensorimotor synchronization (SMS). Theories of the mechanisms underlying SMS often implicate two theoretical dynamic, internal models (Wolport et al, 1998) that in recent years have been supported by neural investigations of motor control (Ito, 2008). These models describe how perceptual information might be integrated and translated into appropriate motor output.

The *forward model* describes a mechanism through which motor commands innervate an effector, resulting in goal-directed motion. The sensory feedback from the actual position outcome is compared to an efferent copy (the expected outcome), enabling error correction. The *inverse model* begins with an estimate of the motor commands needed for a particular outcome, and maps that estimate to efferent signals. In this case, accurate estimates produce accurate behavior. The inverse model also receives feedback, though it is incorporated over repeated use of a motor skill as the demands of that skill become more familiar. Another way of distinguishing these models is that forward models predict the consequences of an action and success depends on feedback, whereas inverse models predict the signals needed for an action and success depends on good signal estimates. Importantly, each model has limitations that are accounted for by the other, and so it is likely that both systems work in tandem depending on the situation (Wolport, Doya, & Kawato, 2003). The inverse model, for instance, may be slower to receive feedback than the forward model, but it facilitates online action output because it anticipates the required resources. Based on these theories, researchers have established a foundational explanation for how actions are planned and subsequently realized. However, because these models must operate constantly, and with great speed, accuracy, and precision, timekeeper mechanisms are essential, and must be a part of SMS. Thus, studying the inherently rhythmic qualities of music is useful.

The emergence of musical rhythm comes from the fact that most music is built on a periodicity: the beat. Perceptual and motor systems respond strongly to periodicities (Haueisen & Knösche, 2001; Grahn, 2012; Schmidt-Kassow et al, 2013; Merker, 2014). By tampering with the regularity of the beat, one can measure perceptual sensitivity to such errors (Iverson & Patel, 2008; Grahn & Rowe, 2012). One can also examine how disrupting the regularity affects motor output by having people try to tap along to the beat

(Repp, 2002). Thus, music becomes a very useful research tool for testing sensorimotor processes: the readiness with which people extract temporal regularities, especially in music (Phillips-Silver et al, 2011), is accounted for by an adjustable independent variable (the beat) and measurable outcomes (tapping accuracy). The most commonly studied dependent variable in SMS tapping tasks, therefore, is the absolute value of asynchrony. This value indicates how far off in time taps are from actual beats in a stimulus, with higher numbers indicating poorer performance. Tap timing is measured by the inter-tap interval (ITI), and the corresponding beat times are defined by their inter-onset intervals (IOI), most often quantified in milliseconds.

Tapping tasks are also used to explore two other aspects of SMS: anticipatory timing patterns, and adaptive timing patterns. The former represents the degree to which an individual anticipates or responds to the onset of a beat (Repp, 1998; Repp, 1999). Such a measure is informative because asynchrony alone does not reliably indicate which strategy (predicting or tracking) an individual is using to match their action with the stimulus. Anticipatory timing is important because it allows actions to be planned in advance and minimize errors (Keller, 2012; Van der Steen & Keller, 2013).

Adaptive timing refers to the use of error correction, and has two main aspects in SMS tapping tasks. The first is phase correction. This refers to local timing adjustments at each beat to account for tapping errors (e.g. shortening the time between two taps if the previous tap was late, thereby accounting for the added time). Without phase correction, errors would compound so that the cumulative times of the tap and beat sequences would not match. This is unlikely to happen, as phase correction is thought to be an automatic process (Vorberg & Wing, 1996; Large, 2008). The second adaptive timing measure is period correction. Unlike phase correction, this is consciously controlled (Repp, Keller, & Jacoby, 2012), and involves intentionally, and gradually lengthening or shortening inter-tap intervals (ITIs, i.e. the time between taps). This reflects accelerating or decelerating the tempo in music.

SMS and Expressive Timing. Musicians often employ expressive timing, which is characterized by intentional fluctuations in beat regularity for artistic purposes. For example, a performer might slow down briefly to highlight a particular chord, or speed up to convey a sense of high arousal. However, SMS processes are typically studied using pacing signals presented as simple clicks or beeps (reviewed in Repp & Su, 2013), devoid of realistic expression. Few studies have looked at SMS and related processes in the context of actual music. Furthermore, studies that have employed real music used mostly contemporary styles that are isochronous (i.e. built on evenly spaced beats) (Iverson & Patel, 2008). This is informative, as results have shown that people vary in their ability to phase lock (i.e. match the isochronous beat sequence), but it overlooks the fact that music is rarely performed isochronously.

Thus, SMS should also be studied in real, expressively timed music. Even in the event of such beat fluctuations, people are generally able to synchronize fairly well. This is seen anecdotally in successful ensemble performances that require musicians to play together amidst expressive timing patterns. Empirically, studies of dyadic tapping show how individuals can learn to adjust tapping patterns to match a partner (Konvalinka et al, 2010; Keller et al, 2011; Nowicki et al, 2013). Also, individual tapping tasks with expressive music have shown that people are able to synchronize with the music, even though it is not automatic (Repp, 1999; Repp, 2002).

These findings suggest that processing and engaging expressively timed music requires both phase and period correction. A person will tend to automatically correct errors during sections of beat regularity (phase correction), and consciously adjust his or her ITIs when a beat deviates from the established regularity (period correction) (Keller, 2001). This implicates the use of both forward and inverse models. The forward model provides fast feedback for correcting mistakes, while the inverse model generates a plan for when a tap should occur in the event of beat fluctuation. However, research on the role of SMS during dynamic timing as seen in expressive music is sparse, which motivates the use of expressively timed music in the present study. Furthermore, expressive timing has been tested primarily in musicians (especially pianists, Repp, 1999; Repp, 2002) who presumably approached this inherently musical task (synchronizing to music) differently than non-musicians would. Therefore, there is interest in examining group differences between musicians and non-musicians.

Shared Cognition in Perception and Action: A Role for Auditory Imagery. The factors that facilitate and enable the internal models of perception-action links must, by definition, include peripheral variables (e.g. different sensitivities in the neuromuscular junction) and perceptual abilities (e.g. audition). But cognitive variables must be important in synchronization, especially with expressively timed music. Evidence for this comes from the fact that synchronization to dynamic time series is effortful (Nowicki et al, 2013). Further support comes from Common Coding Theory, which posits that perception and action have common substrates (Prinz, 1984). An important supposition in this theory is that perception and action share cognitive processes, such that one will implicitly prime the other.

Indeed, several lines of research suggest that auditory imagery might be necessary for accurate SMS. Auditory imagery is the way in which one thinks about sounds in the absence of an exogenous auditory stimulus. It is used when one replays a song in his or her mind, or recalls the voice of a friend. Even though it is an internal process, it can be probed using mental scanning tasks (Halpern, 1988a) and comparisons of perceived and imagined tempos in music (Halpern, 1988b). It is also evidenced by functional magnetic resonance imaging (fMRI) studies which have shown that parts of the auditory cortex are active both when listening to a sound and when imagining a sound (Halpern et al, 2004; Zatorre & Halpern, 2005), suggesting that imagery is akin to actual perception, but under the control of the individual.

Another way of measuring auditory imagery is through self-report. A reliable self-report test of auditory imagery—the Bucknell Auditory Imagery Scale (BAIS) (Pfordresher & Halpern, 2013; Halpern, in press)—measures how vividly an individual can imagine sounds, and also how easily one can control the image by manipulating its contents. Behavioral tests of auditory imagery have validated the BAIS by showing direct relationships between task performance and self-report (Gelding, Thompson, & Johnson, 2015). Tests of auditory imagery also show that it varies greatly among individuals (Hubbard, 2013; White, Ashton, & Brown, 1977). This could partially account for the variability in SMS abilities, if imagery contributes to the contents of the inverse model that links perception and action.

There is considerable evidence that imagery is part of planning and executing actions. That is, in order to predict the necessary commands for an action, one must be able to first imagine the action or related outcomes (James, 1890). For example, being able to imagine a sound and map it to a specific action sequence produces more efficient action execution (Keller & Koch, 2008). Similarly, anticipating a sound that has been learned as being incompatible with an action slows down that action relative to anticipating a sound that has been learned as compatible (Keller & Koch, 2006). Additionally, imagery for pitches is significantly correlated with better SMS (Pecenka & Keller, 2009a). Interestingly, imagery for rhythm is an even better predictor of SMS (Pecenka & Keller, 2009b). Another form of SMS is singing, in which one must adjust laryngeal movements to produce correct pitches without visual feedback. Better pitch matching in a sample of intermediate singers was related to higher self-reported imagery using the BAIS, implicating the use of imagery in an inverse model that maps perception to laryngeal action (Pfordresher & Halpern, 2013).

A Role for Working Memory. Considering the rapidity with which SMS is accomplished—especially in musical contexts—some form of executive function is probably needed to incorporate predictive images into perception-action links (Maes et al, 2014). A likely candidate is working memory, as described by Baddeley and Hitch (1986). In their multicomponent model, a “central executive” system manages the effortful maintenance, manipulation, and synthesis of information from multiple sources.

Information could be incoming and temporarily stored in the visual or auditory modality, or retrieved from long-term memory. Robust working memory is characterized by successful maintenance amidst distraction and quick manipulation. However, working memory is subject to rapid decay and limited capacity (Baddeley & Hitch, 1986). Thus, in any given task it is used briefly and its cache is constantly updated. In SMS, working memory could enable the maintenance and updating of the auditory image to be used in the inverse model (which predicts necessary motor signals) while incorporating the feedback of the forward model (which predicts outcomes to be compared to actual results) to adjust motor output.

Indeed, Baddeley and Andrade (2000) found that self-reported vividness of both visual and auditory images was positively correlated with working memory, suggesting that distraction-resistant maintenance is needed for effective imagery. In a study of the "cocktail party effect," people with lower working memory spans were more prone to distraction by irrelevant auditory information (Conway, Cowen, & Bunting, 2001). A study of musical stream segregation found that polyphonic music (music with more than one melody line) but not monophonic music (one melody) recruited domain general, cortical working memory areas (Janata, Tillmann, & Bharucha, 2002). Thus, even just attending to complex auditory sequences implicates working memory.

While these studies show that effortful auditory processing is related to working memory, they do not speak to the ensuing connection to action. A few studies have explored the role of working memory in linking perception to action. For example, in cellists, a bowing accuracy exercise paired with a concurrent working memory task impaired the regularity of performed rhythms (Maes, Wanderley, & Palmer, 2015). In a rhythm reproduction task, participants with higher verbal short-term memory maintenance (a measure closely related to working memory) were better at replicating rhythms (Grahn & Schuit, 2012). The reproduction task was reminiscent of an expressive timing synchronization task, in that the pattern to be tapped was not simply a regular series of beats. This seems to necessitate effortful information processing, mediated by working memory. On a broader level, a study of pianists found that working memory span predicted sight-reading (playing a piece for the first time) abilities above and beyond years of experience and hours spent practicing (Meinz & Hambrick, 2010). Again, this suggests that implementing perceptual information (in this case the notes on the musical score) into an action plan depends on working memory.

Examining the Relationship Between Cognition, and Perception and Action. Taken together, these studies indicate that a large working memory span should be highly related to auditory imagery self-report and performance. Both imagery and working memory should enable accurate sensorimotor synchronization, particularly in expressively timed music because of its constantly changing temporal information. Furthermore, working memory might be related to strategic synchronization by way of anticipatory timing. That is, if one is able to assimilate action plans and feedback into the motor plan (the theoretical effector of action) fast and accurately enough, he or she could learn to precede beats with taps, especially as the beat sequence becomes more familiar after repeated exposure.

To test the relationship between cognition and SMS, Experiment 1a tested non-musicians and employed a battery of tasks used to assess working memory span, auditory imagery abilities, and synchronization to expressively timed music. Two working memory tests were used (digits backwards and a verbal operation span task), and results were combined into a composite working memory score. To test auditory imagery, two separate tasks were used to measure pitch imagery and tempo imagery, with the hypothesis that tempo imagery should be a better predictor of synchronization than pitch imagery. This is because synchronization (in this task, at least) does not require producing a pitch. Therefore, the ability to modulate temporal information will supersede the ability to modulate pitch information in generating an inverse model prediction. In addition to these empirical tests of imagery, the BAIS was included because of its ability to capture individual differences in imagery along two dimensions: imagery vividness (BAIS-V) and imagery control (BAIS-C). Although the subscales are highly correlated, they often show dissociations in predicting outcomes (Halpern, in press). Lastly, a synchronization-tapping task was used to in order to examine SMS. To test SMS in a musically realistic situation, the task used expressively timed piano music.

The use of non-musicians as participants fills a gap in the synchronization literature, as it will elucidate how relationships among working memory, imagery, and SMS exist without potential effects of extensive musical training. Trained musicians might use other strategies that do not involve imagery or working memory. For example, they might have superior selective and sustained attention to the beat (Keller & Burnham, 2005), eliminating distracting auditory information and lessening the need for working memory. Alternatively, they may have had experiences with similarly fluctuating beat sequences, resulting in practice effects. To test for musician/non-musician differences in cognition and SMS, Experiment 1b replicated the tasks of 1a in a small sample of musicians. If musicians are indeed employing strategies that lessen the role of working memory and imagery, then their SMS performance should be just as good or better than in non-musicians, but should not correlate as strongly with performance on the cognitive tests.

Lastly, to test for the possibility that differences in SMS might be due to noise in the motor system and/or peripheral differences, Experiment 2 added a test of synchronization perception in a sample of non-musicians. Some participants in Experiment 1a may have had very accurate perception of the tempo irregularities and may have learned the timing profiles very well. However, they may have had difficulty transferring this information to the motor effector. Thus, the relationships established in Experiment 1 will have overlooked how simply perceiving dynamic time patterns with interacting with them is related to working memory and imagery. Experiment 2, therefore, adapted the Beat Alignment Test (BAT; Iverson & Patel, 2008) to see how well a second sample of non-musicians could judge the placement of a click track relative to the beats in expressively timed piano excerpts.

Experiment 1a

Methods

Participants. Subjects (*N* = 45, 27 female) were recruited from Bucknell University's research subject pool and given course credit for participating. Age ranged from 18-22 (*M* = 19.4). Years of musical experience ranged from 0-15 (*M* = 3.29).

Materials and Stimuli. A musical background questionnaire was used to assess years and types of musical experiences. The self-report index of auditory imagery used was the Bucknell Auditory Imagery Scale, which consists of two 14-item subscales, one for vividness and one for control.

Stimuli for the pitch imagery test were synthesized in Finale (MakeMusic inc.), using the grand piano midi sample. Pitches ranged from G3 to E5 and were all 400ms in duration. The metronome beats for the tempo imagery test came from sampled bell sounds on a Roland SPD-S MIDI percussion pad (originally used by Pecenka & Keller, 2009). Lastly, the excerpts used in the SMS tapping task came from MIDI recordings of well-rehearsed pianists, and were edited slightly to remove incorrect and add missing notes (Repp, 1998; Repp 1999; Repp, Keller, & Knoblich, 2007).

Procedure

Musical Background and BAIS. Participants first filled out the musical background questionnaire. They were then directed to the lab computer to complete the BAIS, which had been computerized by the experiment to facilitate the presentation of the questionnaire, and analysis of the responses. The first half of the BAIS comprised the vividness (BAIS-V) subscale. Each item had two parts, "a" and "b." Part "a" instructed participants to consider a particular piece of music or situation (e.g. “consider the start of the tune ‘Happy Birthday’”). Part “b” then stated a particular acoustic quality of the piece or situation (e.g. “the sound of a trumpet playing the beginning”). Participants then had to rate how vividly they could imagine that sound in its context by clicking on a number from 1 (“no image generated”) to 7 (“as vivid as actual sound”). Parts “a” and “b” were presented concurrently as text blocks, with “a” positioned above “b.” The second half the BAIS comprised the control (BAIS-C) subscale. Each item contained the same “a” and “b” pairs as BAIS-V, but in a new order. Three seconds after “a” and “b” were presented, part “c” appeared, describing a transformation of the sound in part “b” (e.g. “the trumpet stops and a violin continues the piece.”). Participants then rated how easily they could make that change in their head by clicking on a number from 1 (“no image generated”) to 7 (“Extremely easy to change the image”).

Working Memory Tests. After the BAIS, participants completed the two working memory tests. For the digits backwards test, the experimenter read lists of numbers. After reading the list, the experimenter said, “recall,” and the participant recited the list backwards. If the participant correctly recited the list, they were read another list of the same length. If he or she recited the second list correctly, the next list was one digit longer. If he or she incorrectly recited, the list would be once more. Two incorrect attempts ended the test, and the length of the last correct recitation constituted the participant’s score. The number of list items started at 2 and reached a maximum of 8.

The operation span test was presented using Microsoft PowerPoint. Participants were shown short sentences that were either semantically correct (e.g. “Paris is in

France") or semantically incorrect/nonsensical (e.g. "the lamp washed itself"). Each slide contained one sentence. A "++" slide was shown in between each sentence slide. When the "++" slides were presented, participants were asked to circle "true" if the preceding sentence was correct, or "false" if incorrect on a response sheet. They were also instructed to try and remember without writing down the last word of each sentence while making their true/false judgments. At the end of a trial, a slide asked them to recall the last words of the sentences from that trial on their response sheet. Trial 1 contained a sequence of 2 sentences. Each trial increased the sequence length by 1, up to Trial 6 (7 sentences). The sentence and "++" slides progressed automatically every 2 seconds. The slides instructing participants to recall the words progressed to the next trial after 4 seconds x Trial # (e.g. recall time allotted for Trial 3 was 12 seconds) to ensure adequate time to write down the words. Participants were given one point for each word recalled in the correct serial position

Auditory Imagery Tasks. The order of the two auditory imagery tasks was counterbalanced. The Pitch Imagery Arrow Task (PIAT) was adapted (it was made slightly shorter) from a previous study that tested imagery performance in relation to the BAIS (Gelding et al, 2014). For a visual depiction of the task, see Appendix A. Participants were seated at a lab computer. Each trial was presented randomly in one of 5 major keys (C, C#, D, E*b*, E). The trial started with a scale in the current key to establish a tonal context. Scale tones were presented steadily with a 500ms inter-onset interval (IOI) and participants were alerted to the start with a single flash of the "!" character. After the scale, the "!" character flashed twice to indicate the start of the test phase. A sequence of pitches then played with a steady IOI of 1,000ms. The first pitch was always the tonic, and then moved in increments of 1 scale step, randomly up or down. The range of possible pitches was three below tonic to four above tonic (i.e. if the pitch reached the dominant scale tone below tonic it would move up on the next step, and if it reached the dominant above tonic it would move down on the next step).

At each pitch onset, a black arrow appeared for 800ms to indicate the direction (up or down) the current pitch had moved from the previous pitch. After a certain number of pitch/arrow pairs (described below), the pitch stopped playing, but the arrows continued, now colored grey. The participants' task was to imagine the continuation of the pitch sequence according to the grey arrows. After a designated number of grey arrows (initially three), there was a 1 second pause, and then a probe tone played without an arrow. Participants then had to indicate if the probe was correct (meaning it matched where the pitch would be if it had continued to play according to the grey arrows) or incorrect (it did not follow the grey arrows). An incorrect probe tone had four possible pitches: one or two steps below or above the correct pitch.

All participants started at level 1, which contained three pitch/arrow pairs, and one silent, grey arrow. They moved up a level if they responded correctly six consecutive times, or if their proportion correct at the current level exceeded .60. They moved down a level if they answered incorrectly three consecutive times. This differed from the original PIAT, which contained sub-stages at each level, making it more difficult to advance. This change was made to ensure participants had time to complete all tasks, given the limited amount of time in an experimental session. The maximum level was 5, and each level number indicated how many silent grey arrows were included in each trial at that level. The number of pitch/arrow pairs increased by one at levels 2, 3, and 5.

The Tempo Imagery (TI) test was adapted from a previous study of imagery and sensorimotor synchronization (Pecenka & Keller, 2009). Again, participants were seated at a lab computer. In this test, participants heard five beats with IOIs either increasing from 400ms to 500ms (slowing down) or decreasing from 500ms to 400ms (speeding up). Participants then imagined this pattern continuing for two beats during which time there was no actual sound, and then judged the placement of a final beat (too early or too late). The last beat was never in time with the preceding pattern. The error on the first trial was always +/- 25% of the correct beat placement. The percent error decreased as the test progressed, making it harder to distinguish early from late. The test ended once participants reached their discrimination threshold. Responses were made on the left (early) and right (late) arrow keys of the computer keyboard. The beats were presented as 100ms long piano notes. The beat sequences were pitched at E4, and the probe beat was pitched at F4.

Sensorimotor Synchronization. The stimuli for the synchronization task consisted of a Mozart excerpt (Repp, Keller, & Knoblich, 2007), and a series of excerpts of a Chopin etude that came from a previous study by Repp (see Repp, 1998, 1999). The Mozart was selected as a baseline measure of participants' abilities to tap along to actual music. It is acoustically similar to the Chopin (polyphonic piano music), but has a fairly regular beat sequence. The Chopin was selected because typical performances of Chopin's music are very expressive, containing irregular beat sequences. Thus, using the two pieces also offers a comparison between how participants tap along to two different types of beat sequences. Only one performance of the Mozart piece was used. Because the main interest in the SMS task was tapping to expressively timed music, three different Chopin performances were used, each with a unique timing profile.

For the duration of the test, participants were seated at a MIDI keyboard next to the lab computer. This way, they could read the instructions while positioned for the task. Before tapping, participants watched the experimenter demonstrate the task with the Mozart excerpt. They were told not to tap along while the experimenter demonstrated. The purpose of the demonstration was to ensure all participants tapped at the same rate, and to make the target beats as clear as possible. After the demonstration, participants tapped along to four takes of the Mozart excerpt.

Following the four Mozart takes, the experimenter demonstrated how to tap along to the Chopin excerpt. Participants then tapped along to four takes of three different performances of the excerpt. The order the performances was randomized, but blocked such that a given performance occurred four times in a row, rather than being interspersed with other performances. The purpose of this design was to give participants a chance to learn the timing profile of each piece. Complete randomization of the 12 trials would have confounded analysis of the progress participants made in synchronizing to each piece over multiple takes.

Participants' first taps started the music to ensure that their first tap always matched the first beat in the music. Participants were told to maintain contact with any white piano key using the index finger of their dominant hand and to not stop tapping until the music stopped. After a three second pause, a "!" character appeared on the computer screen to alert participants that they could then start the next trial.

Dependent Measures and Analyses. For a reference list of dependent variables, see Appendix B. Digits Backward was scored as a span (the longest sequence a participant could remember without error). The operation span task was scored as the number of correctly remembered words (maximum = 27). These were summed into a single, aggregate working memory (AWM) score to simplify analysis. Each sub-scale of the BAIS was scored as the mean of all 14 items. The PIAT was scored as the proportion correct from all trials (minus practice trials). The TI was scored according to participants' thresholds. The threshold score represents an estimate of the lowest percentage of error in a stimulus that a participant could still discern as early or late. Thus, lower scores indicate better performance.

The primary dependent measure in the SMS tapping task was the median of the absolute values of asynchronies. Asynchronies were computed as the difference in milliseconds between a tap and its target beat. Thus, each take contained a series of asynchronies equal in length to the number of beats/taps. Asynchrony series were converted to absolute values because measures of central tendency of signed asynchrony scores will grossly underestimate the amount of error. Median was chosen over the mean to account for uncharacteristically large asynchronies. Thus, each participant produced four asynchrony scores for the Mozart excerpt, and 12 for the Chopin excerpt (one for each take). Lower asynchrony scores indicate better performance (i.e. greater synchrony).

Asynchrony, however, is only a coarse measure of how accurately participants can synchronize. It does not necessarily indicate the strategy they use to synchronize. Thus, a secondary dependent measure assessed the degree to which individuals anticipate upcoming tempo changes (prediction), or respond to past tempo chances (tracking). The predicting vs. tracking tendencies can be determined by computing a cross-correlation (CC) at different lags between ITIs and inter-onset intervals IOIs of the beats. A high lag-0 CC indicates that the tap sequence is highly related to the beat sequence, showing a predicting tendency. A high lag-1 CC indicates that the tap sequence is highly related to the shifted beat sequence (e.g. the time between taps two and three is similar to the time between beats one and two), showing a tracking tendency. With these correlations, a predicting/tracking index (P/T) was established by subtracting the lag-1 CC (tracking measure) from the lag-0 CC (prediction measure). Thus, a P/T index is highly positive to the extent that a participant is predicting, and highly negative to the extent that he or she is tracking. As with asynchrony scores, a P/T index was produced for all takes on both excerpts.

Lastly, as a measure of how participants changed their asynchronies and P/T over the course of four takes on each excerpt, the slope of each of these scores was calculated for each excerpt. Thus, each excerpt (one Mozart and three Chopin) had one slope score, for a total of four asynchrony slopes, and four P/T slopes per person.

Results

Working Memory. The mean span on the digits backwards task was 5.20 (*SD* = .94), with a minimum score of 3 and a maximum score of 8. The operation span task had a mean of 20.40 (*SD* = .4.20), a minimum of 7 and a maximum of 27 (the maximum possible score was also 27). These two working memory scores—digits backwards span and operation span—were summed to create an aggregated working memory score (henceforth AWM). This was done to facilitate analyses of working memory in relation to other variables. Both working memory tests correlated with the derived AWM measure (digits backwards: $r_{(40)} = .56, p <. 001$; operation span: $r_{(40)} = .96, p < .001$).The mean of the AWM was 25.40 (*SD* = 4.68), the minimum score was 11 and the maximum 33.

Auditory Imagery. In all analyses, the BAIS was analyzed as two separate subscales (vividness and control), never as a single scale. The means, standard deviations, and ranges for the two subscales were similar (BAIS-V: *M* = 4.57, *SD* = .76, range = 3.14-6.5.0; BAIS-C: *M* = 4.94, *SD* = .80, range = 3.57-6.71). The PIAT is normally scored according to the highest level reached. However, modifications to the PIAT—which, due to time constraints, made it easier to advance through levels and removed the use of "stages" within each level—resulted in nearly every participant reaching the maximum level. Thus, the proportion correct was used as the measure of pitch imagery (*M* = .71, *SD* = .14, range = .42-1.00). Rhythm imagery was scored according to the threshold of temporal discrimination, calculated as the lowest percentage of error at which participants could still correctly detect the direction of error (early vs. late) (*M* = 30.18, *SD* = 3.64, range = 25.30-46.59), meaning a lower score indicated better performance.

Relationships Among Musical Background, Working Memory, and Imagery. See **Table 1** for a summary of the correlations among years of musical experience, AWM, and the measures of imagery. Most participants had fewer than 4 years of musical training and had not engaged in music performance or practice in over 5 years. There was a small but significant correlation between years of music and AWM, $r_{(40)} = .37$. Years of musical training also correlated moderately with proportion correct on PIAT.

The imagery measures showed several relationships with each other. As expected, BAIS-V and BAIS-C were moderately related, $r_{(43)} = .67$. Both BAIS subscales correlated with proportion correct on the PIAT, though BAIS-C showed a slightly stronger relationship. Only BAIS-C correlated with rhythm imagery, $r_{(40)} = -.31$. The negative correlation reflects the fact that lower scores on the rhythm imagery task reflect better performance. Thus, high BAIS-C was related to acute rhythm imagery. Interestingly, there was no correlation between the PIAT and rhythm imagery suggesting that the two tasks successfully captured different forms of imagery.

AWM showed significant correlations with several forms of imagery. The strongest imagery correlate of AWM was rhythm imagery, $r_{(39)} = -.41$. AWM also correlated with the PIAT ($r_{(37)} = .40$), and with BAIS-C ($r_{(40)} = .33$). BAIS-V did not correlate with working memory.

Table 1

Correlation Matrix of Auditory Imagery and Working Memory Scores

Measure	1	2	3	4	5	6
1. Years of Music	1.00	.37*	.09	.26	.43**	.03
2. Working Memory (AWM)	--	1.00	.26	.33*	.40*	-.41**
3. BAIS-V	--	--	1.00	.63**	.41**	-.09
4. BAIS-C	--	--	--	1.00	.48**	-.31*
5. Pitch Imagery (PIAT)	--	--	--	--	1.00	.07
6. Tempo Imagery (TI)	--	--	--	--	--	1.00

*significant at the .05 level (2-tailed)
**significant at the .01 level (2-tailed)

SMS: Asynchrony and P/T Index. Analysis of asynchrony and P/T requires that both time series (the beat sequence and tap sequence) have the same number of events (e.g. an excerpt containing 35 beats must be paired with 35 taps). However, not all participants produced tap sequences equal in length to the corresponding beat sequence. In order for a file to be included in the analysis, the following criteria were used. Data were considered unusable if there were five fewer taps relative to beats in one or more of a participant's takes. If a participant was only missing 4 or fewer taps, an interpolation procedure was used to fill in the missing taps. The procedure worked as follows: the data were processed in a Matlab script that identified ITIs that were greater than twice, but less than three times the size (in milliseconds) of the corresponding IOI. The script divided the original ITI by 2, placed the halved values sequentially in the tap sequence, and then shifted the remaining taps to fill in the missing space. See Appendix D for a depiction of the process. Nine participants produced data that were unusable for the SMS analysis, reducing the sample size on SMS measures to 36.

To get a general sense of how well participants could synchronize with the Mozart excerpt (which had a mostly steady beat sequence) and the Chopin excerpts (which had irregular, very expressive beat sequences), the median asynchrony scores were averaged across all takes for the Mozart, and for the Chopin, and compared with a paired-sample *t*-test. As expected, asynchrony was significantly lower for the Mozart excerpt (M = 189.92, SD = 81.16) compared to the averaged Chopin excerpts (M = 365.41, SD = 112.92), $t_{(70)} = -7.57$, $p < .001$ (see **Figure 1**). Because there were three versions of the Chopin excerpt, a one-way ANOVA was conducted to see if there were differences in asynchrony among the versions. The ANOVA revealed no significant differences, suggesting that Chopin excerpts were all equally difficult for participants. For this reason, the mean score across all takes of all three versions was used in future correlation and regression analyses.

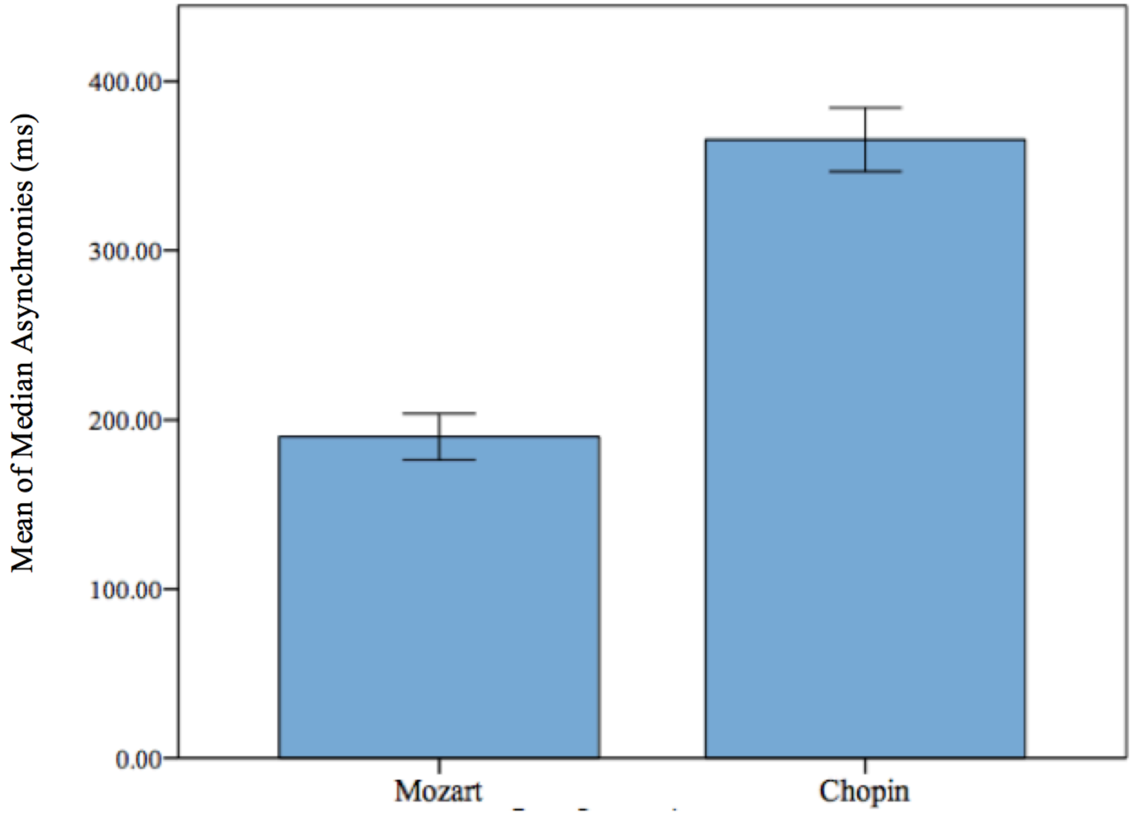


Figure 1. Comparison of the median asynchrony scores between the two pieces (Mozart and Chopin). Error bars represent SEM.

To see if participants differed in anticipatory timing on the two pieces, another paired-sample *t*-test was again used, this time comparing the means of the P/T indices from all takes of the Mozart excerpt, and all takes of the Chopin excerpts (see **Figure 2**). Scores greater than 0 indicate prediction, whereas negative scores indicate tracking. Generally, participants predicted more on the Mozart excerpt ($M = .08$, $SD = .08$) than on the Chopin excerpts ($M = -.01$, $SD = .19$), $t_{(70)} = 2.59$, $p < .05$. According to a one-sample *t*-test, the average P/T index for the Mozart takes was significantly higher than a test value of 0 (zero representing no tendency towards prediction or tracking), $t_{(35)} = 6.09$, $p <$

.001. The mean P/T index from all Chopin takes, however, was not different than 0, $t_{(35)}$ = -.22, p = .83. This suggests that, as expected, participants were better able to predict the regular beats of the Mozart excerpt, whereas prediction was not as common, or more difficult during the Chopin excerpts. In other words, on the Mozart trials the participants were capable of beat prediction, and the Chopin trials show that there was no trend in either direction when the beat was irregular. Again, to check for differences among the three Chopin excerpts a one-way ANOVA was used to compare the mean P/T indices of each version. There were no significant differences.

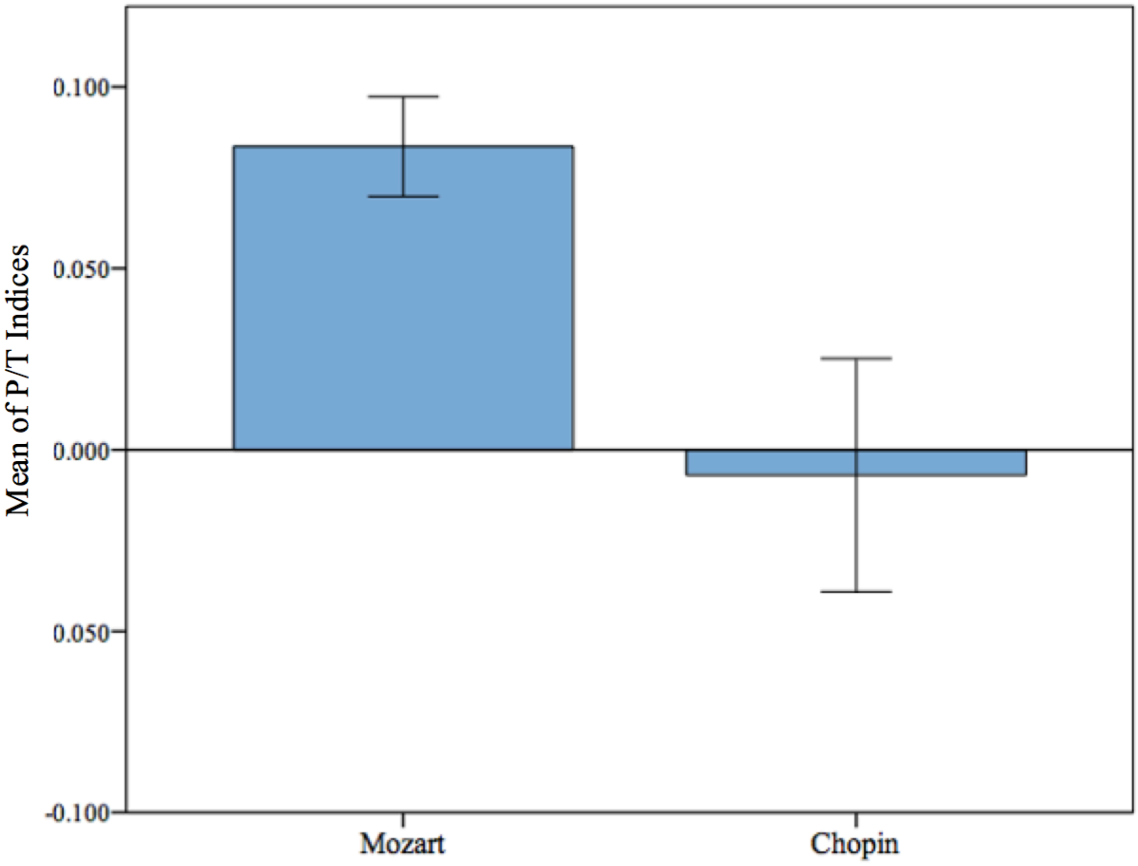


Figure 2. Comparison of the mean P/T indices between the two pieces (Mozart and Chopin). Error bars represent SEM

Correlations Among SMS, Working Memory, and Auditory Imagery. To see how AWM and the imagery variables related to asynchrony and P/T, bivariate correlation were used to test only participants with usable SMS tapping data ($N = 36$). Importantly, even with the smaller sample size, the correlations between working memory and imagery described above were maintained. Asynchrony (see **Table 2** for a summary of asynchrony correlates) on the Mozart takes correlated only with BAIS-V, $r_{(34)} = -.46$. Because lower asynchrony scores indicate better performance, the negative correlation means that BAIS-V predicted better synchronization. Unlike the Mozart takes, asynchrony on the Chopin takes did not correlate with BAIS-V, but did with BAIS-C, $r_{(34)} = -.40$. Asynchrony on the Chopin takes also correlated with both imagery tasks.

Table 2

Correlations Between Asynchrony and Cognitive Variables

	AWM	BAIS-V	BAIS-C	PIAT	TI	YM
Mozart	-.04	-**.46****	-.19	-.28	.05	-.11
Chopin	-.12	-.21	-**.40***	-**.56****	**.35***	-.01

*significant at the .05 level (2-tailed)
**significant at the .01 level (2-tailed)

Working memory did not relate to asynchrony in either excerpt. It did, however, correlate with the P/T index (see **Table 3** for a summary of P/T correlates) of the Chopin takes, $r_{(34)} = .41$. Another interesting correlation related to P/T on the Chopin takes involved TI. Although rhythm imagery did not relate to the mean P/T index, it did correlate with the slope of the P/T index over the course of the four takes from each version, $r_{(34)}$ = -.41. This means that better rhythm imagery was associated with a greater degree of improvement on the P/T index. In other words, people high in rhythm imagery were more likely to become high predictors over the course of four takes on a given excerpt. This was the only significant correlation among the measures of slope. There were no correlates of P/T for the Mozart takes.

Table 3

Correlations Between P/T and Cognitive Variables

	AWM	BAIS-V	BAIS-C	PIAT	TI	YM
Mozart	.02	.16	.25	.12	-.01	-.12
Chopin	**.41***	-.02	.14	.02	-.19	-.16

*significant at the .05 level (2-tailed)

To better understand the relationship between SMS, and working memory and imagery, hierarchical regression models were used. Auditory imagery and AWM were the predictor variables, and asynchrony (Table 4) and P/T (Table 5) were the predicted variables. Only the SMS measures from the Chopin takes were tested in these models, as the SMS measures from the Mozart takes showed only one significant bivariate correlation (BAIS-V). With asynchrony as the dependent variable, and BAIS-C in step 1 of the model, R^2 was .20, $p < .01$. The addition of the two imagery tasks (PIAT and TI) at step 2 explained an additional 26% of the variance in asynchrony for a total R^2 of .46, $p < .01$. Adding AWM in step 3 did not produce a significant change in R^2.

Table 4

Summary of Hierarchical Regression Analysis for Variables Predicting Asynchrony

Measure	β	*t*	sr^2	R	R^2	ΔR^2
Step 1				.45	.20	.20
BAIS-C	-.46	-2.81**	.21			
Step 2				.68	.46	.26
PIAT	-.57	-3.45**	.22			
TI	.38	2.48*	.11			
Step 3				.69	.47	.01
AWM	.14	.90	.01			

*significant at the .05 level (2-tailed)
**significant at the .01 level (2-tailed)

A second model was used to predict P/T indices from the Chopin excerpts. Step 1 contained AWM, the only significant bivariate predictor of P/T, and produced an R^2 of .17, $p < .05$. Step 2 added BAIS-C, which explained an additional 10% of the variance in P/T, $p < .05$. Lastly, a third step included rhythm imagery, which added an additional 16% of explained variance for a final R^2 of .42, $p < .01$.

Table 5

Summary of Hierarchical Regression Analysis for Variables Predicting P/T

Measure	β	*t*	sr^2	*R*	R^2	ΔR^2
Step 1				.41	.17	.17
AWM	.41	2.58*	.17			
Step 2				.51	.26	.09
BAIS-C	.33	1.98*	.09			
Step 3				.64	.41	.16
TI	-.44	-2.87**	.16			

*significant at the .05 level (2-tailed)
**significant at the .01 level (2-tailed)

Discussion

Experiment 1a had participants complete several tests of auditory imagery and working memory, and then tap along to expressively timed music that was characterized by a dynamic, irregular beat sequence. Synchronization (as measured by the amount of asynchrony in milliseconds) on the tapping task was predicted by self-reports of imagery control, an objective test of pitch imagery, and an objective test of TI. Strategic synchronization by means of anticipation of beat onsets also implicated imagery control and TI, but the strongest predictor of P/T index was working memory.

These correlations could be explained in terms of the two internal models of motor control: the forward model, and the inverse model. First, the role of imagery control in both synchronization and prediction could reflect the need to constantly update the image contained in the inverse model. Because the necessary motor commands are not identical when synchronizing to a dynamic sequence, it follows that being able to easily change the image informing the model—as measured by BAIS-C—would be related to better synchronization, and also anticipation of beat onsets. If—as theorized in studies of motor control (Wolport et al, 1995)—the inverse model contains instructions regarding the timing and force of an action, then there must also be a mechanism through which those instructions change to adapt to new action demands. Controlling images that are related to forthcoming actions could be a part of such a system. Difficulty modifying an image—reflected by a low BAIS-C score—would therefore complicate synchronization and prediction when the necessary motor commands are changing. Interestingly, high BAIS-V but not BAIS-C predicted good synchronization in the Mozart excerpt but not the Chopin. Because the Mozart piece contains fewer beat irregularities than the Chopin, it could be the case that once one has established a vivid image of a necessary motor command, there is no need to change it, as it will continue to function appropriately.

TI predicted both asynchrony and the P/T index, such that a low threshold in the TI task predicted low asynchrony, and high P/T (high prediction). This could reflect the need to perceive the temporal irregularities underlying expressively timed music in order to interact with it. In other words, timing the innervation of motor effectors (in this case, flexion and extension muscles in the index finger) could be more accurate among those with a superior ability to internally discern temporal properties. Thus, if one can imagine both when an action should occur and the commands needed to realize that timing, the actual output should be more accurate (as indicated by low asynchrony), and possibly slightly in advance of the external stimulus (i.e. beat, as indicated by a high P/T score). TI did not predict synchronization in the Mozart excerpt, suggesting that entrainment to regular beat sequences requires minimal internal time processing. Instead, regular time sequences might rely on feedback (via the forward model) from a comparison between the action and an efferent copy. TI would not be relevant here, as the timing of innervations is probably coming from internal clocks governed by neural oscillators (Ranking, Large, & Fink, 2009) that automatize motor commands. In other words, the commands are generated automatically once the musical beats and neural oscillations entrain, leaving the forward model to correct motor noise by providing feedback.

Pitch imagery predicted asynchrony, but not the P/T index. More specifically, people with accurate pitch imagery tended to synchronize better, regardless of whether they were predicting or tracking beat onsets. It makes sense that internal pitch processing would not relate to anticipatory timing, as pitch imagery—as measured by the PIAT—is not based on dynamic time sequences. Instead, good pitch imagery might facilitate learning and remembering the melody of the excerpts, which would give participants a general idea of when their taps should occur. It would not, however, facilitate predicting beat onsets, which is a more temporally based skill. Through this process, superior pitch imagery would generate a better outcome prediction, which would be factored into the forward model for corrective feedback.

The results also suggest a role for working memory in SMS, specifically in anticipatory timing. The AWM score was not a significant predictor of asynchrony, but was the only significant bivariate correlate of P/T. This is partially in line with the hypothesis that working memory would relate to both levels of SMS. Given the effortful nature of synchronization to expressively timed music seen in previous studies (and seen anecdotally, as several participants reported the need to focus intently on the music), one would expect working memory to be implicated in measures of asynchrony. However, there could be other forms of conscious processing needed for synchronization, such as selective attention. Synchronizing to a dynamic time series might not require maintenance and manipulation, but instead depend on rigorous monitoring of the stimulus.

Anticipatory timing, on the other hand, could very reasonably require working memory. In order to predict beat onsets, as indicated by a highly positive P/T index, one must maintain the inverse model dictating present motor output, while simultaneously updating the model or forming a new one in preparation for forthcoming motor output. In other words, working memory could allow one to stay ahead of the beat by processing what he or she knows (or thinks) is coming next, while executing an action related to the current stimulus.

If images are in fact forming the contents of the inverse model (Pfordresher & Halpern, 2013), then working memory—specifically the central executive system—could be the controller that is driving the transition from one image to the next. This is supported by the significant bivariate correlation between BAIS-C and AWM, and by the fact that BAIS-C explained additional variance beyond AWM in the regression model predicting P/T. Thus, the two functions are closely related, and auditory imagery probably plays a role in anticipatory timing, but working memory as the more generalizable ability better predicted ones P/T index.

Experiment 1b

In order to better understand SMS in relation to imagery and working memory, Experiment 1a was replicated in a recruited sample of musicians. This was done to see if there were group differences in imagery, working memory, and SMS between musicians and non-musicians, and also to see if there were different relationships among those variables in individuals with extensive musical training. Because musicians have experience with musical synchronization and have likely developed several strategies to facilitate interactions with dynamic time series, they should show superior performance on the tapping task, but fewer relationships between the cognitive variables tested here and SMS.

Methods

Participants and Procedure. Thirteen experienced musicians were recruited using flyers, classifieds, and Facebook. The qualifications for participation were: 1) must have 10 or more years of musical experience either through ensemble performance or private lessons; 2) must be currently practicing/performing. All subjects were paid $10 for their participation. The procedure was identical to Experiment 1a.

Results

The same bivariate correlations examined in Experiment 1a were again examined in Experiment 1b. The only significant correlation in the sample of musicians was between asynchrony on the Chopin trials and BAIS-V, $r_{(11)} = -.65$, $p < .05$. The null results in the correlational analysis could be due partially to the low sample size. Even with the small sample the musicians showed several group differences when compared to the sample of nonmusicians (See **Figure 3**). Musicians performed significantly better on the PIAT (*M* = .86, *SD* = .10;) than non-musicians (*M* = .69, *SD* = .12). Musicians also had significantly higher AWM scores, (*M* = 28.87, *SD* = 2.49), than non-musicians (*M* = 25.59, *SD* = 1.82). On measures of SMS, musicians produced significantly lower asynchrony scores (*M* = 237.79, *SD* = 56.50) than non-musicians (*M* = 365.41, *SD* = 112.92), $t_{(45)} = -3.59$, $p < .01$. There was no group difference in P/T.

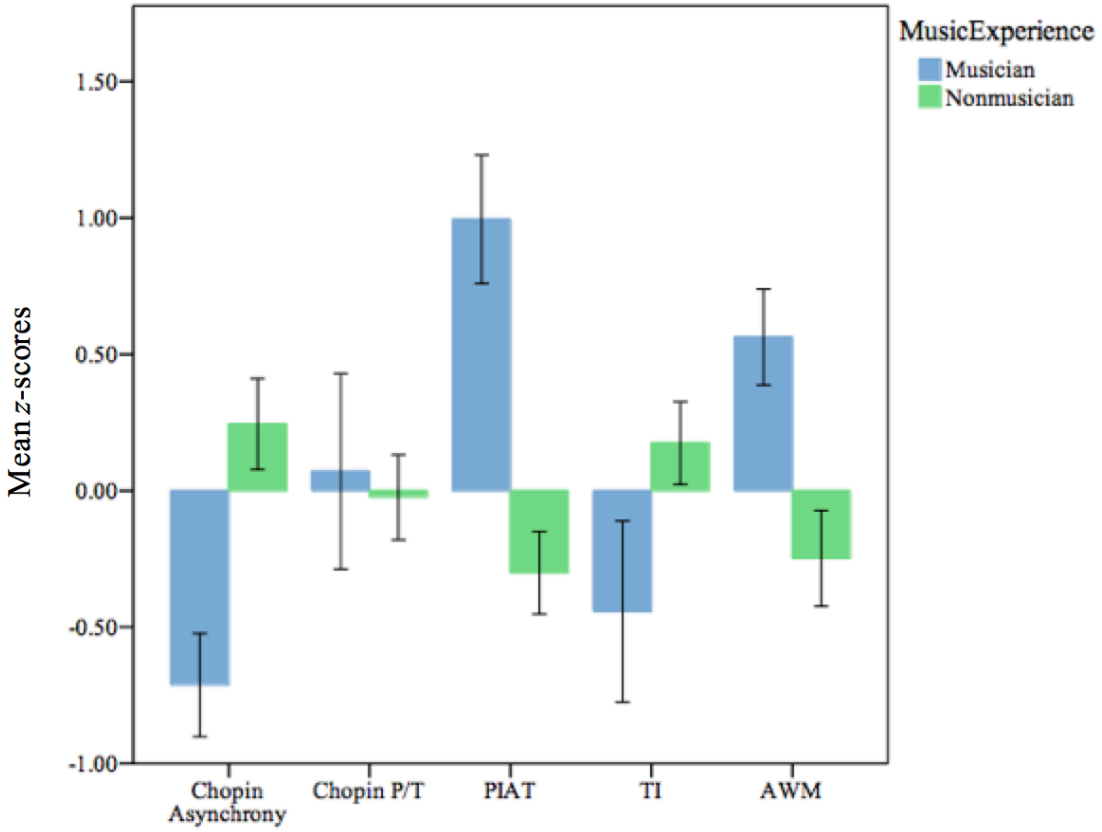


Figure 3. Group differences between nonmusicians (Experiment 1a) and musicians (Experiment 1b) on the SMS and imagery tasks. Scores were standardized to *z*-scores. Error bars represent SEM.

Discussion

The most interesting finding from the sample of musicians is that BAIS-V predicted asynchrony in the Chopin trials. This relationship was not found in Experiment 1a, suggesting that musicians may benefit from vivid auditory imagery when synchronizing to dynamic timing profiles, whereas non-musicians benefit from other types of imagery, namely control as measured by the BAIS-C self report. Vivid auditory images could facilitate internalizing novel sequences, which would in turn improve synchronization. Musicians have plenty of experience matching actions to sounds in a variety of temporal patterns. Thus, the motor demands may not be an issue. The biggest challenge for musicians might instead be expecting what will come next in the sequence. Thus, if they can clearly imagine expected sounds in the context of typical expressive music, they could synchronize better, even to novel sequences.

One confound that could be limiting the analysis of correlations is that musicians performed significantly better on the PIAT, had significantly higher AWM scores, and were generally better at synchronization. Although there were no ceiling effects, it might be beneficial to develop a more challenging pitch imagery task, and more variable timing profiles in the excerpts in order to challenge the musicians more. In order to address familiarity with typical expressive techniques, a useful manipulation might be randomizing (with a set of boundaries) the IOI sequences in the excerpts, requiring them to learn essentially a new style of timing.

Experiment 2

In order to see how the cognitive variables that were tested in Experiments 1a and 1b might relate to perception without action, Experiment 2 replaced the tapping the task with a test of expressive timing perception. This was done to better understand how dynamic time series are processed, as the act of synchronization must start with perceiving the auditory sequence and understanding its musical properties. Experiment 2, therefore, was intended to test for cognitive variables that would predict expressive timing *perception*, as it is possible that the relationships found in Experiment 1 were realized at the perceptual level, and not necessarily related to action. Given the need to process the musical properties of these types of time series, it was hypothesized that the PIAT and TI test would relate to expressive timing perception without action, as they capture the most musical imagery abilities.

Methods

An adaptation of the Beat Alignment Test (BAT) developed by Iverson and Patel (2008) was used instead of the SMS tapping task in Experiment 2. Other than that, all materials and procedures were the same as in Experiment 1. The BAT includes a perceptual test in which participants make judgments of how well clicks imposed over music match the actual beat of the music. The clicks can be early or late (phase error), or correct, relative to the beat of the music. Participants answer “yes” or “no” as to whether the clicks are on the beat or not.

The present study used expressively timed music instead of the rock, jazz, and show tune excerpts of the original BAT. Thus, the task was renamed the Expressive Beat

Alignment Test (EBAT). Click track were imposed over the Chopin and Mozart excerpts, as well as four additional expressively timed music excerpts. The clicks were played in one of three trial types: 1) before the beat, for an early phase error condition; 2) after the beat, for a late phase condition; 3) on the beat, for an on-time condition. Participants were asked to judge if the clicks matched the beat and also if the clicks fell before or after the beat. Based on pilot testing, early and late click tracks were set to play 25% of the median IOI in advance of or later than the first beat in order to have varying levels of performance without ceiling or floor effects.

The click tracks started on the fifth IOI (i.e. with/just before/just after the sixth beat) to ensure that participants had time to identify the musical beat. There were six different pieces (see Appendix C for a list) shortened to ~10-second excerpts. If participants responded before the end of the excerpt, the music was stopped. Each excerpt was played in the three conditions (early clicks, late clicks, on time clicks), and there were two repetitions of all pieces in all conditions for a total of 36 trials.

The order was randomized for all participants. The randomized trials were preceded by practice trials. After responding during practice trials, the correct answer was given to the participant. Feedback was not given in the actual test. The excerpts for the practice trials were two variations of *Twinkle Twinkle Little Star*. One variation was isochronous, and the other was expressively timed. Participants were scored on their proportion correct.

Results

Overall, performance on the EBAT (see **Figure 4**) was slightly worse than the results reported by Iverson and Patel (2008) on the original BAT. This is to be expected, as the expressive version is probably more difficult for most people. The mean proportion correct across all conditions (on-time, early, late) was .47 (*SD* = .10). This is significantly above chance ($t_{(21)} = 6.71$, $p < .001$; chance = .33 proportion correct). A one-way ANOVA of proportion correct broken down by condition showed that people were significantly better at the on-time trials (*M* = .59, *SD* = .12) than both types of phase error trials (early/late), and better at the early click trials (*M* = .48, *SD* = .15) than the late click trials (*M* = .36, *SD* = .16), $F_{(2, 63)} = 15.41$, $p < .001$. Performance on the late click trials was at chance ($t_{(21)} = 1.11$, $p = .28$, chance value = .33), suggesting that in that condition participants mostly guessed. Performance on the early click trials, unlike the late trials, was above chance, $t_{(21)} = 4.75$, $p < .001$.

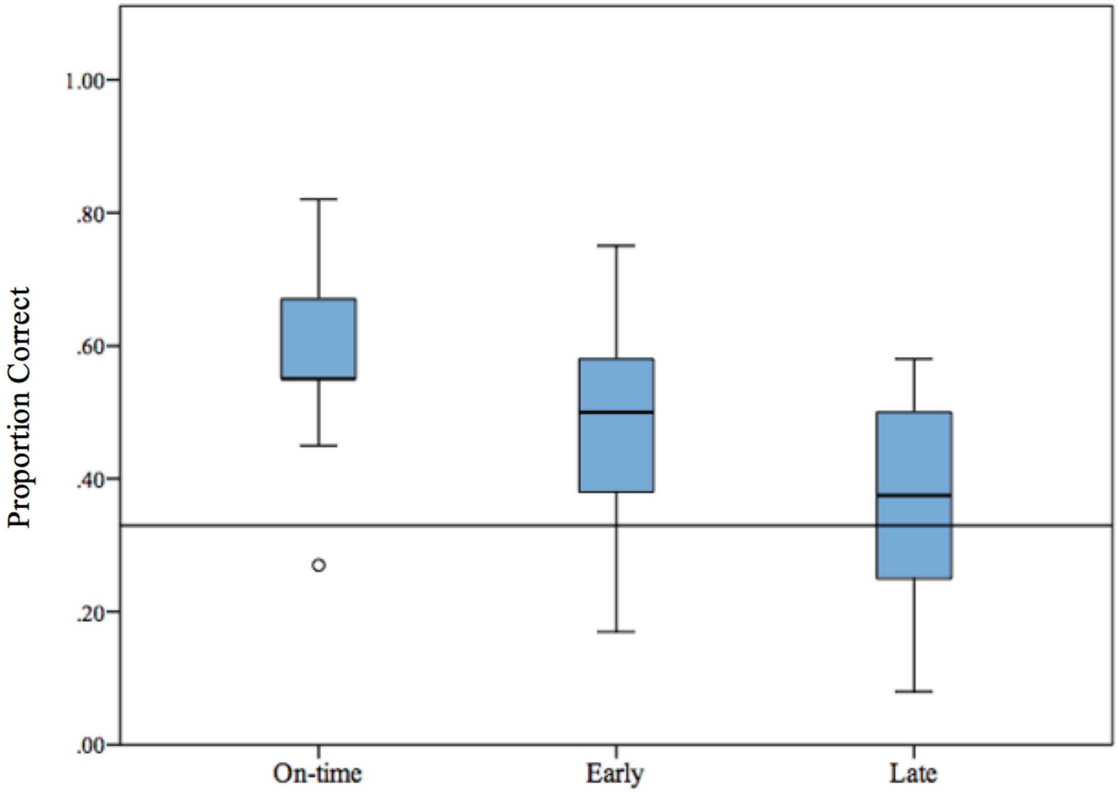


Figure 4. Proportion correct for each trial type of the EBAT. The horizontal black line represents chance performance at .33 on the y-axis.

In order to look at how overall performance was related to specific trial types, several bivariate correlations were used to compare aggregate proportion correct, to proportion correct on each of the three trial types. Overall EBAT scores averaged across all three trial types were positively correlated with performance on the early ($r_{(20)} = .71$, $p < .001$) and late ($r_{(20)} = .61$, $p < .01$) trials, but not the on-time trials, showing that people varied most on their ability to distinguish early from late, but showed good performance with little variability for the on-time trials. In other words, overall performance was explained by phase error trials and not by the on-time trials.

Next, performance on the EBAT was correlated with the same measures of working memory and imagery used in the SMS tapping task. Performance on the on-time trials was positively correlated with BAIS-V ($r_{(20)} = .50$, $p < .05$). The early trials correlated negatively with rhythm imagery ($r_{(20)} = -.48$, $p < .05$), meaning a lower threshold of rhythm imagery (i.e. good rhythm imagery) predicted success on the early click EBAT trials. Interestingly, the late trials correlated *negatively* with working memory, $r_{(20)} = -.69$, $p < .01$. However, in general, people were performing at chance levels in the late click trials, making this result difficult to interpret. Nevertheless, given that working memory was associated with prediction in the SMS tapping task, it is possible that individuals high in working memory are predisposed to perceive off-beat clicks as early, as part of a generalized mechanism underlying anticipatory timing. To test this possibility, the proportion of "early" responses on the late click trials was calculated. Indeed, this variable correlated directly with AWM, $r_{(20)} = .48$, $p < .05$. This means that people higher in working memory were more likely to perceive a late click as early, whereas people with lower AWM scores were equally likely to perceive a late click as early or on-time.

Discussion

The purpose of Experiment 2 was to see how the same cognitive variables examined in Experiment 1 related to perceptual discernment of expressive time patterns without action In Experiment 2, a new set of participants completed the same tests of imagery and working memory, but instead of the SMS tapping task they took the EBAT, a test of expressive timing perception. This was motivated by the fact that the tapping task required peripheral engagement, which could introduce motor noise into the measures of synchronization. Thus, some participants may have been able to efficiently process dynamic time patterns, but had difficulty executing the sequences as taps. The EBAT showed results fairly consistent with the tapping task, such that there were correlations among imagery types, and working memory was a significant correlate of one task condition.

Interestingly, unlike the tapping task where BAIS-C was a good predictor of performance, BAIS-V but *not* BAIS-C predicted performance on the on-time trials in the EBAT. On the one hand, the excerpts used in the EBAT all contained dynamic time sequences as in the Chopin piece from Experiment 1. However, a fundamental difference between the tapping task and the EBAT is that the former required consistently changing motor output, whereas the latter required making a comparison between the onsets of two auditory events. From this perspective, a role for imagery vividness over control is not so surprising, as vividly imagining one sequence while attending to the other could help identify correct alignment. This would fit into the forward model described above, such that imagery vividness contributes to a feedback loop.

Regarding the phase error trials (early/late clicks), TI predicted performance on the early click trials and working memory was implicated in the late click trials. Furthermore, early click trials were performed above chance, and late clicks were at chance. It is somewhat surprising that the two types of phase errors produced markedly different results. One would expect both types of phase errors to be equally difficult, or related to the same cognitive mechanisms. As explained previously, it is possible that some participants are more sensitive than others to early phase errors. Evidence for this possibility was found in the positive correlation between working memory and proportion of "early" responses on the late click trials. This could influence perception of the late-clicks: given that participants were quite good at the on-time trials, it follows that they could easily identify phase error trials, but then tend to categorize most errors as early because of the apparent (and currently unexplained) preference for, or sensitivity to events preceding the beat. Future studies might examine the relationship between anticipation and perception of the order of auditory events.

The chance performance on the late clicks trials could be a matter of overactive stream segregation. It is probable that most participants perceive the clicks and the piano music as two separate streams, due to the disparate timbres of the two sounds (Iverson, 1995). Thus, at any given time participants' selective attention can focus only on one stream, rather than on both as a single unit. This could be problematic because making a judgment about the direction of error would require rapidly shifting attention from one stream to the other. In doing so, participants' attention may be captured more easily by the click, which has a much shorter attack and decay time than the piano. This could lead to the click being perceived as the initial event more often, leading one to more commonly respond "early." Future studies might consider varying the acoustic properties of the click to make it more or less similar to the piano sound. It is possible that performance on all trial types would be improved by a like-timbre "click" (where the click is actually a piano note) because it would not divide the listener's attention as much as the current task design.

General Discussion

Experiment 1a, conducted in non-musicians, showed that imagery predicts motor behavior during synchronization with dynamically timed, expressive music. Furthermore, working memory was related to several measures of imagery, but also explained most of the variance in the use of anticipation during synchronization, suggesting that it has both direct and indirect roles in SMS. Experiment 1b tested the same relationships in a small sample of musicians, showing that imagery vividness might be more beneficial to musicians than to non-musicians, and that musicians are generally better at synchronization. Additionally, Experiment 2 showed that self-reported imagery vividness and a test of TI predict one's ability to perceive expressive timing patterns.

The differences between musicians and non-musicians are indicative of potentially qualitatively different approaches to synchronization between the two groups. The non-musicians exhibited relationships between SMS and tempo imagery, pitch imagery, imagery control, and working memory. This could be because without extensive musical training, one must use several methods of imagery to plan and update actions. Musical training, however, might result in a generalized ability of inverse models to adapt quickly to dynamic time series. Thus, fewer factors related to good synchronization in musicians because there is not as much need to manage the inverse model using dynamic imagery.

This raises the question of what is driving the superior synchronization in musicians. If their inverse models are indeed more adaptable and faster to assimilate temporal information, what cognitive abilities underlie that difference? One explanation could attribute their superior performance to the higher AWM scores observed in Experiment 1b. However, if that were the case, then one would expect to have seen a correlation between AWM and asynchrony in the non-musicians (which was not found, even though non-musicians showed a large range of AWM scores). Perhaps the combination of robust working memory and years of training on a working memory-intensive task such as music contributes to superior sensorimotor synchronization. Or, there could be other forms of executive function involved that were not tested here.

These potential roles for auditory imagery and working memory in SMS might challenge aspects of current theories of motor timing. For example, the concept of central pattern generators (CPGs; Hooper, 2001) maintains that behaviors that are rhythmic and periodic are instantiated by automatic oscillators. Because oscillatory activity is automated once initiated, there should be no relation to effortful cognitive processes, such as imagery and working memory. However, the time series generated by the beats in the Chopin excerpts were not periodic, as they were not regularly spaced. Thus, oscillators alone probably could not account for the necessary control of motor timing, relying instead on effortful processing. This would also explain why performance on the Mozart excerpt was generally better than performance on the Chopin, and showed fewer associations with imagery: beat regularity, as seen in the Mozart, facilitated the initiation of CPGs.

Experiment 2 tested the perception of expressive music without related action. Imagery vividness was related to good judgment of phase matching between two auditory sequences. When compared to the results of Experiment 1b, this reinforces the idea that synchronization may be more of a perceptual challenge than a motor challenge for musicians. This is because BAIS-V, which predicted good phase matching judgment in non-musicians, also predicted low asynchrony in musicians. In other words, musicians may have approached the SMS task as a matter of phase judgment. They developed an image of expected auditory events, which in turn contributed to a feedback loop while tapping.

Regarding the role of imagery between Experiments 1 and 2, results showed that TI was the only consistent predictor of performance on the SMS task and EBAT. The TI task assessed how well participants could judge the placement in time of beats in a short, irregular sequence by imagining the continuation of the sequence. It is understandable that the ability to internalize changing time patterns could relate to both perceiving and acting upon expressively timed music: without accurate internal timing, generating actions on a schedule would be complicated (as in SMS), as would comparing external events on two separate schedules (as in the EBAT).

This brings up a limitation in the two experiments: two different tests (the SMS tapping task and the EBAT) were administered to two separate samples of participants. The results were still informative, implicating separate roles for the four types of imagery tested in a perception-action task, and a perception task. However, there was no direct comparison of the EBAT to the tapping task, meaning there is no evidence that the perceptual abilities to detect temporal errors relate to successful motor output.

Another potential limitation was the modified version of the PIAT that was used to test pitch imagery. Due to time constraints within testing sessions, the PIAT had to be shortened. Although it still yielded significant results, it may not have been as sensitive of a measure as it was in its original format. Perhaps with all levels and stages included in the task design, there would have been stronger correlations between the PIAT and related measures of imagery and SMS.

As described previously, Experiment 1b may have been limited by its small sample size. This is being addressed, as the experiment is ongoing. However, low variability and non-normal distributions of the variables tested in the musicians could, at this time, limit the efficacy of the parametric comparison that were made.

It is important to keep in mind that the relationships described here are only correlational, and their causal contributions to the internal models of motor control are largely speculative. However, this paves the way for future studies. One approach is to use dual-task paradigms that require participants to use either working memory or imagery in a context irrelevant to SMS while trying to synchronize in the tapping task. Presumably, dual-tasks would worsen synchronization to expressively timed music. However, if such an experiment is extended participants might eventually automate their execution of a given timing profile, and no longer be impaired by a concurrent task. This could elucidate how dynamic time sequences are learned.

Other future studies should move towards refining the EBAT. As described above, a version of the EBAT that varies the timbre of the click could be used to examine the role of stream segregation in perceiving synchrony (or asynchrony). Also, the current version of the EBAT uses only a 25% shift in phase alignment for the error trials. This shift resulted in a range of performances without ceiling or floor effects, but it still leaves the possibility that participants developed strategies for responding that were separate from their perceptual abilities. For example, one participant reported only paying attention to the first and last clicks, which she claimed were the most informative. Thus, she was not judging the offset of one sequence relative to another, but instead ordering two auditory events (click and beat) during two very small windows (first click and last click). A threshold design in which the phase shift gradually decreases in size might be a preferable alternative.

Neural underpinnings of the perception-action links involved in expressively timed synchronization could be established using transcranial magnetic stimulation (TMS). Particular interest should be given to the supplementary motor area (SMA). In the current experiment, SMS was used as the predicted variable, suggesting that it is the outcome of cognitive and perceptual processes. However, in theory, a perception-action link in SMS should be bidirectional. Thus, impairing motor planning by targeting the SMA with TMS might not only increase asynchrony, but also decrease performance on a perception task such as the EBAT. Before that, however, the EBAT and expressive timing tapping task should be tested in a single sample to see if performance on the two tasks will correlate.

Overall, the experiments presented here revealed a potential role for auditory imagery and working memory in sensorimotor synchronization, specifically when synchronizing with real, expressively timed music. There is also evidence that different types of imagery are implicated at different stages of perception and action, such that clearly forming an image might facilitate judgments of incoming sound, whereas changing an established image (imagery control) and being able to imagine temporal relationships (TI) could help update action plans. Furthermore, Experiment 1b offered burgeoning evidence that musicians relative to non-musicians might employ qualitatively difference strategies when synchronizing to expressively timed music. These findings add to an extensive body of literature on SMS that use similar tapping tasks by showing that auditory imagery is a significant correlate of synchronization, and that working memory may be necessary for anticipatory timing in dynamic musical sequences.

References

Conway, A.R.A., Cowan, N., & Bunting, M.F. (2001). The cocktail party phenomenon revisited: The importance of WM capacity. *Psychological Test and Assessment Modeling, 8*, 331.

Croom, A.M. (2015). Music practice and participation for psychological well-being: A review of how music influences positive emotion, engagement, relationships, meaning, and accomplishment. *Musicae Scientiae, 19*, 44-64.

Flanagan J.R., Vetter, P., Johansson, R. S., and Wolpert, D. M. (2003). 'Prediction precedes control in motor learning'. *Current Biology*, *13*. 146-150.

Gelding RW, Thompson WF, Johnson BW, (In Press), The Pitch Imagery Arrow Task: Effects of musical training, vividness, and mental control, *PLOS ONE*, Accepted 23rd February, 2015.

Grahn, J.A. (2012). See what I hear? Beat perception in auditory and visual rhythms. *Experimental Brain Research, 220*, 51-61.

Halpern, A.R. (1988). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 434-443.

Halpern, A.R. (1988). Perceived and imagined tempos of familiar songs. *Music Perception, 6*, 193, 202.

Halpern, A.R., & Bartlett, J.C. (2002). Aging and memory for music: A review. *Psychomusicology: A Journal of Research in Music Cognition, 18*, 10-27.

Halpern, A.R., Zatorre, R.J., Bouffard, M., Johnson, J.A. (2004). Behavioral and neuralcorrelates of perceived and imagined musical timbre. *Neuropsychologia, 42*, 1281-1292.

Hambrick, D.Z., Tucker-Drob, E.M., & Elliot, M. (2015). The genetics of music accomplishment: Evidence for gene-environment correlation and interaction. *Psychonomic Bulletin & Review, 22*, 112-120.

Haueisen, J., & Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, *13* , 786-792.

Hodges, D.A. (1996). Human musicality. In D.A. Hodges (Ed.), Handbook of music psychology (2nd ed.) (29-68). San Antonio, TX: IMR Press.

Hooper, S.L (2001). “Central Pattern Generators”. *Encyclopedia of Life Sciences*. John Wiley & Sons. doi: 10.1038/npg.els.0000032. ISBN 978-0-470-01590-2.

Hubbard, T.L. (2013). Auditory aspects of auditory imagery, in *Multisensory imagery*, by Lacey, Simon (Ed); Lawson, Rebecca (Ed), 51-76. New York, NY, US:Springer Science & Business Media.

Ilari, B.S. (2004). Music cognition in infancy: Infants’ preferences and long-term memory for complex music. *Information & Learning, 6*, 7-20.

Ito, M. (2008). Control of mental activities by internal models in the cerebellum. *Nature Reviews Neuroscience*, *9*, 304-313.

Iverson, J.R. & Patel, A.D. (2008). The Beat Alignment Test (BAT): Surveying beat processing abilities in the general population. *Proceedings of the 10th International Conference on Music Perception and Cognition (ICMPC10). Sapporo, Japan,* 465-468.

J.A. Grahn, & J.B. Rowe. (2012). Finding and feeling the musical beat: Striatal dissociations between detection and prediction of regularity. *Cerebral Cortex*, doi:10.1093/cercor/bhs083

James, W. (1890). Principles of psychology. New York: Holt.

Janata, P., Tillman, B., & Bharucha, J.J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective & Behavioral Neuroscience, 2*, 121-140.

Keller P.E., & Koch I. (2006) Exogenous and endogenous response priming with auditory stimuli. *Advances in Cognitive Psychology, 2*, 269-276.

Keller P.E., & Koch I. (2008). Action planning in sequential skills: relations to music performance. *Quarterly Journal of Experimental Psychology*, *61,* 275–287.

Keller, P. E. (2012). Mental imagery in music performance: underlying mechanisms and potential benefits. *Annals of the New York Academy of Sciences, 1252* , 206–213.

Keller, P.E. & Burnham, D. (2005). Musical meter in attention to multipart rhythm. *Music Perception, 22*, 629-661.

Keller, P.E., Pecenka, N., Fairhurst, M.T., & Repp B.H. (2011). Relations between basic temporal error correction processes and the quality of interpersonal coordination. Paper presented at the 4th Joint Action Meeting. Austrian Academy of Sciences, Vienna, Austria. 7-9 July, 2011.

Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. D. (2010). Follow you, follow me: Continuous mutual prediction and adaptation in joint tapping. *Quarterly Journal of Experimental Psychology, 63*, 2220-2230.

Lamont, A. (1998). Music, education, and the development of pitch perception: The role of context, age, and musical experience. *Psychology of Music, 26*, 7-25.

Large, E. W. (2008). Resonating to musical rhythm: Theory and experiment. In Simon Grondin (Ed.), *The psychology of time* (pp. 189-231). West Yorkshire: Emerald.

Lin, Y. (2014). Understanding the benefits of an Asian music therapy student peer group.*Australian Journal of Music Therapy, 25*, 28-44.

Maes, P., Leman, M., & Palmer, C., & Wanderley, M.M. (2014). Action-based effects on music perception. *Front. Psychol. 4*:1008. doi: 10.3389/fpsyg.2013.01008.

Maes, P., Wanderley, M.M, & Palmer, C. (2015). The role of working memory in the temporal control of discrete and continuous movements. *Experimental Brain Research, 233*, 263-273.

McDermott, J. & Hauser, M.D. (2006). Nonhuman primates prefer slow tempos but dislikemusic overall. *Cognition, 104*, 654-668.

Meinz, E.J., & Hambrick, D.Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory. *Psychological Science, 21*, 914-919.

Merker, B. (2014). Groove or swing as distributed rhythmic consonance: Introducing the groove matrix. *Frontiers in Human Neuroscience, 8*.

Naoumenko, S.I. (1982). Individual differences in musical ability. *Voprosy Psychologii, 5*, 85-93.

Nombela, C., Hughes, L.E., Owen, A.M., & Grahn, J.A. (2013). Intro the groove: Can rhythm influence Parkinson's Disease? *Neuroscience & Biobehavioral Reviews, 37*, 2564-2570.

Nowicki, L., Prinz, W., Grosjean, M., Repp, B. H., & Keller, P. E. (2013). Mutual adaptive timing in interpersonal action coordination. *Psychomusicology: Music, Mind, and Brain, 23,* 6-20.

Patel, A. (2008). *Music, language, and the brain*. Oxford: Oxford University Press.

Pecenka, N., & Keller, P. E. (2009b). The relationship between auditory imagery and musical synchronization abilities in musicians. In Jukka Louhivuori, Tuomas Eerola, Suvi Saarikallio, Tommi Himberg, & Päivi-Sisko Eerola (Eds.), *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009)* (pp. 409-414). Jyväskylä: University of Jyväskylä, Finland.

Pecenka, N., & Keller, P.E. (2009a). Auditory pitch imagery and its relationship to musical synchronization. *Annals of the New York Academy of Sciences, 1169*, 282-286.

Pfordresher, P. Q. & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. Psychonomic Bulletin and Review. On-line, 15 February, 2013.

Phillips-Silver, J., Toiviainen, P., Gosselin, N., Piché, O., Nozaradan, S., Palmer, C., & Peretz, I (2011). Born to dance but beat deaf: A new form of congenital amusia. *Neuropsychologia*, *49*, 961-969.

Prioir, M., & Troup, G.A. (1988). Processing of timbre and rhythm in musicians and non musicians. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior, 24*, 451-456.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372-422.

Rentforw, P.J. (2012). The role of music in everyday life: Current directions in the social psychology of music. *Social and Personality Psychology Compass, 6*, 402-416.

Repp, B. H., Keller, P. E., & Jacoby, N. (2012). Quantifying phase correction in sensorimotor synchronization: Empirical comparison of three paradigms. *Acta Psychologica, 139*, 281-290.

Repp, B.H., & Su, Y. (2013). Sensorimotor synchronization: A review of recent research (2006-2012). *Psychonomic Bulletin & Review, 20*, 403-452.

Schmidt-Kassow, M., Heinemann, L.V., Abel, C., Kaiser, J. (2013). Auditory-motor synchronization facilitates attention allocation. *Neuroimage, 82*, 101-106.

Sloboda, J.A. (2000). Individual differences in music performance. *Trends in Cognitive Sciences 4*, 397-403.

Tarr, B., Launay, J., Dunbar, R. (2014). Music and social bonding: "Self-other" merging and neurohormonal mechanisms. *Frontiers in Psychology, 5*, 1096-2005

Van der Steen, M. C. & Keller, P. E. (2013). The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Frontiers in Human Neuroscience, 7*.

Vorberg, D., & Wing, A. M. (1996). Modeling variability and dependence in timing. In H. Heuer & S. W. Keele (Eds.), *Handbook of perception and action: Motor skills*. San Diego, CA: Academic Press.

White, K., Ashton R., & Brown R. (1977). The measurement of imagery vividness: normative data and their relationship to sex, age, and modality differences. *British Journal of Psychology, 68*, 203–211.

Wing, A. M. (1997). 'The role of internal models in motor planning and control: evidence from grip force adjustments during movements of hand-held loads'. *Journal of Neuroscience*, 17.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society B, 358*, 593–602 .

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, *2*, 338–347.

Zatorre, R.J., & Halpern, A.R (2005). Mental concerts: Musical imagery and auditory cortex. *Neuron 47*, 9-12.

APPENDIX A: Pitch Imagery Arrow Task Example (Level 1)

Auditory Sequence*	**C**	**D**	**C**	**-**	**F (probe)**
Visual Sequence	**-**	↑	↓	↑	**?**
Correct Imagery Sequence	**-**	**-**	**-**	**C**	**Incorrect probe**

1. A sequence of three notes paired with arrows matching direction of pitch motion

2. Arrows without sound play, during which time imagery is used

3. A probe tone plays, user compares it to last imagined pitch. Responds: correct/incorrect

*It is not expected that participants will have any knowledge of the pitch names (C,D,E, etc…) and for this task they don't need to. The note names are used here simply to orthographically represent pitch changes

APPENDIX B: List of Dependent Variables

Variable	**Abbreviation**	**Description**	**Interpretation of measure**
Years of Music	YM	Self-reported years of music training	High value indicates more training
Working Memory Sum	AWM	Sum of participants' scores on digits backwards and operation span tests of working memory	High values indicate greater working memory span
Temporal Imagery	TI	Threshold estimate on the temporal imagery test	*Lower* scores indicate finer temporal discernment, i.e. better rhythm imagery
Pitch Imagery	PIAT	Proportion of correct trials on the Pitch Imagery Arrow Task	Higher scores indicate better pitch imagery, i.e. more correct trials at higher difficulty levels
Asynchrony	---	The median asynchrony score from the expressive timing synchronization task, converted into an absolute value	Lower values indicate better synchronization, or, less asynchrony.
Prediction/Tracking index	P/T	A measure of anticipatory timing derived from tapping patterns (lag-0 – lag-1)	A positive value indicates prediction, negative tracking
BAIS Vividness	BAIS-V	Self reported vividness of auditory images	Higher values indicate more vivid images
BAIS Control	BAIS-C	Self reported control of auditory imagery	Higher values indicate ease of changing images

APPENDIX C: List of Pieces Used as Stimuli in the SMS and EBAT Tasks

Etudes, No. 3 Etude in E major, Op. 10 by Frédéric Chopin (SMS and EBAT)

Nocturnes, No. 1 Nocturne in B major, Op. 32 by Frédéric Chopin (EBAT)

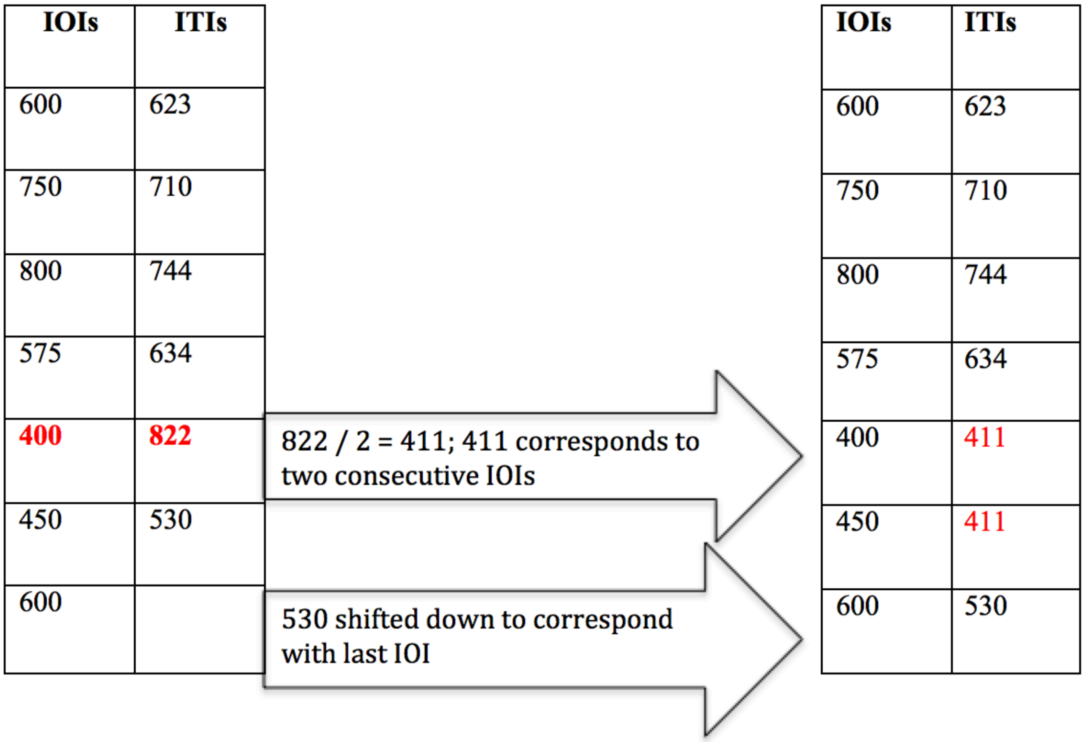
Nocturnes, No. 1 Larghetto in B*b* minor, Op. 9 by Frédéric Chopin (EBAT)

Sonata in F major, K. 533, III. Allegretto by W.A. Mozart (SMS and EBAT)

Sonata in G major, K. 283, I. Allegro by W.A. Mozart (EBAT)

Sonata in e minor, Op. 90, I. Con vivacità by L. Beethoven (EBAT)

APPENDIX D: Depiction of Interpolation Procedure Used in SMS Analysis



IOIs	ITIs
600	623
750	710
800	744
575	634
400	**822**
450	530
600	

822 / 2 = 411; 411 corresponds to two consecutive IOIs

530 shifted down to correspond with last IOI

IOIs	ITIs
600	623
750	710
800	744
575	634
400	411
450	411
600	530