Original Article

QJEP

Quarterly Journal of Experimental Psychology
2018, Vol. 71(8) 1781–1796
© Experimental Psychology Society 2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1080/17470218.2017.1366531
qjep.sagepub.com
SAGE

Working memory and auditory imagery predict sensorimotor synchronisation with expressively timed music

Ian D Colley[1,2], Peter E Keller[2] and Andrea R Halpern[1]

Abstract

Sensorimotor synchronisation (SMS) is prevalent and readily studied in musical settings, as most people are able to perceive and synchronise with a beat (e.g., by finger tapping). We took an individual differences approach to understanding SMS to real music characterised by expressive timing (i.e., fluctuating beat regularity). Given the dynamic nature of SMS, we hypothesised that individual differences in working memory and auditory imagery—both fluid cognitive processes—would predict SMS at two levels: (1) mean absolute asynchrony (a measure of synchronisation error) and (2) anticipatory timing (i.e., predicting, rather than reacting to beat intervals). In Experiment 1, participants completed two working memory tasks, four auditory imagery tasks, and an SMS-tapping task. Hierarchical regression models were used to predict SMS performance, with results showing dissociations among imagery types in relation to mean absolute asynchrony, and evidence of a role for working memory in anticipatory timing. In Experiment 2, a new sample of participants completed an expressive timing perception task to examine the role of imagery in perception without action. Results suggest that imagery vividness is important for perceiving and control is important for synchronising with irregular but ecologically valid musical time series. Working memory is implicated in synchronising by anticipating events in the series.

Keywords

Working memory; sensorimotor synchronisation; auditory imagery; expressive music; music cognition

Received: 18 February 2016; revised: 17 May 2017; accepted: 25 June 2017

Introduction

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). At any age or level of musical sophistication, listening to music is often accompanied by some sort of motor output, most evidently in music performance, but also in the simple case of tapping along to a beat. In either situation, internally generated actions must be coordinated with external stimuli such as the sounds of a co-performer, or even a recording when singing along with the radio. This process of coupling sensory inputs with motor outputs is called sensorimotor synchronisation (SMS; Repp & Su, 2013). Importantly, successful SMS depends on accurate timing, such that one's schedule of motor output must be coordinated with the timing of external events. Despite the apparent ease with which even individuals without extensive formal musical training perceive and interact with a musical beat, the act of synchronisation can be complex.

Part of the readiness with which individuals synchronise to musical beats could be explained by the fact that much music has an underlying isochronous beat, or a nearly isochronous beat given human error and variability. Isochrony refers to events, such as beats, that are regularly spaced in time (e.g., one beat every 500 ms). Periodicities associated with isochronous beats evoke neural resonance (Jones & Boltz, 1989; Large, 2008; Nozaradan, 2014), which could minimise the conscious control needed for appropriate motor output (i.e., synchronisation). Thus, isochronous musical synchronisation can be likened to other widespread, periodic motor behaviours such as

[1]Department of Psychology, Bucknell University, Lewisburg, PA, USA
[2]The MARCS Institute, Western Sydney University, Penrith, NSW, Australia

Corresponding author:
Ian D Colley, The MARCS Institute, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia.
Email: i.colley@westernsydney.edu.au

walking or swimming (Hooper, 2000), given the repetitive and automatic execution after initiation of the motor sequence.

However, in many contexts, music is not strictly isochronous. Performers will often employ expressive timing, which is characterised by intentional fluctuations in beat regularity, for artistic interpretation. Examples include slowing down briefly to highlight a particular chord, or speeding up through an especially exciting passage. In either case, most listeners can maintain a sense of the beat and are therefore able to synchronise with approximately ±10% (Repp, 2006) error by tapping along to the beat; this is, however, a larger degree of variability compared to tapping to isochronous music, which would typically yield approximately ±2% error (Repp, 2006). Such cases of systematic expressive timing showcase the complexity of SMS: how can the motor output of a listener who is engaged in the music be timed to match a stimulus when the timing of the stimulus is musically structured, but irregular?

In such cases of expressive or dynamic timing, motor output is likely demanding in terms of planning and control. Theories of motor control often implicate two types of internal models that might explain how perceptual information is integrated and translated into appropriate motor output for SMS, including during expressively timed music (Flanagan & Wing, 1997; Ito, 2008; Keller, Novembre, & Loehr, in press; Pfordresher & Mantell, 2014; Wolpert, Miall, & Kawato, 1998). First, a *forward model* allows comparisons of intended outcomes to actual motor output, enabling efficient error correction. Specifically, motor commands are sent to the periphery to effect action, but also through the forward model to generate an efference copy (a prediction of what will happen). The other model, an *inverse model*, is informed by practice and motor learning, and generates an estimate of the motor commands needed for a particular outcome (as opposed to predicting the outcome itself as in a forward model). A successful inverse model, therefore, will produce accurate motor signals resulting in desired actions. In summary, the forward model is a predictor of action outcomes, and the inverse model is a controller of motor signals. In the context of expressively timed SMS, these two models could help an individual develop expectations about when an action should occur, and accurately execute the action accordingly (van der Steen, Jacoby, Fairhurst, & Keller, in press; Van der Steen & Keller, 2013).

Even with these models in place to regulate SMS, individuals vary in their ability to synchronise both to isochronous and expressively timed music (Repp, 2001, 2002). These individual differences are most apparent in synchronisation-tapping tasks, in which participants tap along to pacing signals such as a metronome, and their tap times are compared to signal times (Repp, 2005; Repp & Su, 2013). The difference between tap times and signal times at each occurrence of a signal provides a quantification of *asynchrony*. This value represents how well a person is able to synchronise (Pecenka & Keller, 2009; Pecenka & Keller, 2011; Repp, 2002). Individual differences in asynchrony raise the question of what could be informing and regulating the internal models that support motor control, as it is unlikely that the internal models alone act as distinct mechanisms that can fully account for variability in performance (Keller, 2014; Keller, Novembre, & Hove, 2014; Repp, 2005). Thus, the present study assessed the extent to which cognitive variables can predict individual differences in SMS. Specifically, we asked what cognitive processes might influence the forward model's prediction, and allow the inverse model to update and adjust motor commands, and how do these processes eventually manifest as individual differences in SMS? Two likely candidates are working memory and auditory imagery, which are processes that help us establish and manipulate internal representations such as time.

Considering the rapidity with which SMS is accomplished—especially in musical contexts—some form of executive function is likely needed to mediate the constant change in action plans (Maes, Leman, Palmer, & Wanderley, 2014). Working memory, as described by Baddeley and Hitch (1974) should facilitate fluid shifts in action planning and execution—specifically by enabling one to update temporal intervals governing musical pulse sequences—and therefore correspond to successful SMS. Indeed, several studies have explored the role of working memory in linking perception to action. For example, a dual-task paradigm showed that a working memory task impaired the regularity of cellists' bow strokes, suggesting that working memory is needed to regulate motor behaviour (Maes, Wanderley, & Palmer, 2015). In a rhythm reproduction task in which participants had to tap non-isochronous rhythms from memory, participants with higher verbal short-term memory maintenance (a measure closely related to working memory) were better at replicating the rhythms (Grahn & Schuit, 2012). This shows that internalising patterns that are not simply a series of isochronous beats seems to necessitate effortful information processing, mediated by working memory. Regarding actual music production, a study of pianists found that working memory span predicted sight-reading abilities (playing a piece for the first time) 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 is related to working memory. Similarly, implementing previously heard temporal patterns and prior knowledge of musical structure into motor plans that anticipate changes in timing regularity could also benefit from robust working memory.

Another cognitive function that might be necessary for accurate SMS is auditory imagery, which is the activated mental representation of sounds in the absence of or in addition to an exogenous auditory stimulus (Halpern, 1988a, 1988b; Halpern, Zatorre, Bouffard, & Johnson, 2004). In terms of motor planning, the importance of imagery is explained in the ideomotor hypothesis, which posits that in order to predict the necessary commands for an action, one must be able to first imagine the action and related outcomes (James, 1890). Indeed, there is evidence that "replaying" internal auditory experiences is part of the process of planning and executing actions. For example, being able to imagine a sound and simultaneously map it to a specific action sequence produces more efficient action execution (Keller & Koch, 2008). Similarly, anticipating an incompatible sound that has been associated with a particular response through learning (e.g., a high pitched sound paired with a downwards movement) results in a slower response time relative to anticipating a compatible sound (e.g., a low pitched sound and a downwards movement; Keller & Koch, 2006). In relation to SMS-tapping tasks, imagery for pitches is inversely correlated with absolute asynchrony scores (where low asynchrony scores reflect accurate synchronisation), suggesting that imagining specific properties of a sound can facilitate motor timing (Pecenka & Keller, 2009). This relationship could be due to the fact that imagining pitch can help one imagine a melody, which contains information about patterns of musical timing. Interestingly, imagery for rhythm is an even better predictor of SMS than imagery for pitch (Pecenka & Keller, 2009), suggesting further that imagery functions can be categorised and dissociated along separable pitch and time dimensions. One goal of the present study is to see whether this holds true when synchronising with expressively timed music that has been generated by a human performer, rather than with a pre-programmed pacing signal.

Another musical behaviour involving coordination of sensorimotor processes is singing, in which one must adjust laryngeal movements to produce correct pitches using auditory feedback. Better pitch matching in a sample of untrained singers was related to higher self-reported auditory imagery, implicating the use of imagery in an inverse model that maps imagery of desired outcomes to laryngeal action (Pfordresher & Halpern, 2013). This relationship has been explicitly modelled in the Multi-Modal Imagery Association Model (Pfordresher, Halpern, & Greenspon, 2015).

Because auditory imagery is often a lucid and distinct mental state, people are generally able to report on the quality of their imagined sounds. Therefore, reliable self-report assessments can effectively capture individual differences in auditory imagery. One such assessment is the Bucknell Auditory Imagery Scale (BAIS; Halpern, 2015), which measures how *vividly* (BAIS-V) an individual can imagine sounds, and also how easily one can *control* (BAIS-C) the image by manipulating its contents. Behavioural tests of auditory imagery have validated the BAIS by showing direct relationships between imagery task performance and self-report (Gelding, Thompson, & Johnson, 2015). Studies using the BAIS and other self-reported imagery questionnaires have also shown that scores vary greatly among individuals, making them suitable instruments to investigate individual differences (Halpern, 2015; Hubbard, 2013; White, Ashton, & Brown, 1977). Furthermore, the BAIS-V and BAIS-C subscales typically correlate moderately ($r = .70$), leaving room for dissociations in predictive power between them (Halpern, 2015). Therefore, the BAIS was used in this study, given its potential to explain how the quality of one's imagery vividness and imagery control might separately predict variability in SMS abilities.

The studies discussed above looked at working memory and auditory imagery as distinct predictors of SMS. However, the two cognitive functions might overlap in their predictive power. Some researchers suggest that working memory might provide a necessary foundation for imagery by enabling the maintenance of sensory information (Baddeley & Andrade, 2000; Baddeley & Logie, 1992). This suggests that imagery and working memory assessments might be capturing a single process. By examining all three discussed variables (working memory, auditory imagery, and SMS), we tested whether imagery and working memory are to some extent separate predictors of variance in SMS. A second experiment examined the same cognitive variables as predictors of the *perception* of expressive timing without the requirement for action. This allowed us to test whether any results observed in the first experiment (a tapping task) were being realised at the perceptual level rather than through the interaction of processes involved in perception and motor execution.

Experiment 1

To test the relationships between working memory, imagery, and SMS, Experiment 1 employed a battery of tasks used to assess working memory span, auditory imagery abilities, and synchronisation to expressively timed music. Two working memory tests were used (a digits backwards task and a verbal operation span task), and results were combined into an aggregated working memory score (WMS). To test auditory imagery, two separate tasks were used to measure pitch imagery and temporal imagery (TI). In addition to these empirical tests of imagery, the BAIS was included because of its ability to capture individual differences in imagery along two general, not explicitly musical dimensions: imagery vividness (BAIS-V) and imagery control (BAIS-C).

Finally, a synchronisation-tapping task was used to examine two SMS variables: mean absolute asynchrony and anticipatory timing. Anticipatory timing is the extent to which a person anticipates the size of temporal intervals between successive events. Someone with good anticipatory timing would tap just prior to the beat onset rather than just after it. It is especially important in synchronisation with irregular time sequences, and is believed to be informed by imagery related to the target sequence (Pecenka & Keller, 2011). To test SMS in a musically realistic situation, the task used piano pieces with low beat variability (a Mozart sonata), and an expressively timed piano piece with high, musically valid beat variability (a Chopin étude), the latter of which was presented in multiple versions by different pianists. Based on pilot testing, the Mozart excerpt was substantially easier to tap along to. Thus, the Mozart was included as practice for the tapping task, and was presented before the Chopin excerpts in order to acquaint participants with the procedure. The main interest and analyses, however, were limited to the Chopin excerpts.

We hypothesised that working memory, as the most foundational process tested here, would predict measures of both mean absolute asynchrony and anticipatory timing. Next, we expected the BAIS—a self-report of general, not explicitly musical imagery abilities—to independently predict SMS over and above working memory. More specifically, we hypothesised that the BAIS-V would predict mean absolute asynchrony, but BAIS-C would predict anticipatory timing because of the need to update and change one's motor plan. Finally, we expected measures of pitch and tempo imagery, which are specifically related to aspects of music and therefore are likely implicated in musical synchronisation, to predict SMS performance over and above working memory and the BAIS. We hypothesised that pitch imagery, which could provide a general sense of the melody, would relate to mean absolute asynchrony, whereas tempo imagery, which measures one's ability to discern temporal relations, would relate to anticipatory timing. These hypotheses were tested using two (absolute asynchrony and anticipatory timing) three-step (working memory, BAIS, pitch and tempo imagery) hierarchical regression models.

These hypotheses were tested in a sample of nonmusicians and individuals with very little music experience. We chose to sample from this population in order to test how cognition relates to motor behaviour without extensive relevant experience. Anecdotally, we see that nonmusicians spontaneously synchronise with music as in tapping along to a performance, and empirical investigations have shown that they can also intentionally synchronise with music. But with minimal prior exposure to expressively timed music and without explicit training in musical synchronisation, will imagery and working memory predict the quality of SMS with a largely novel pattern of timing?

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 to 22 years (M=19.4 years; standard deviation [SD]=1.39 years). Years of musical experience ranged from 0 to 6 years (M=2.46 years; SD=2.13 years), and years since ceasing musical training ranged from 7 to 15 years (M=11.57 years; SD=3.47 years). No participants reported regularly listening to classical music, or being familiar with the chosen musical stimuli.

Materials and stimuli. A musical background questionnaire was used to assess years and type of musical experience. The self-report index of auditory imagery was the BAIS (Halpern, 2015), which consists of two 14-item subscales, one to assess vividness of images and one to assess control of images. The working memory tests were the backward digit span and an operation span task (Borella, Ludwig, Dirk, & de Ribaupierre, 2011).

Stimuli for the pitch imagery test were synthesised in Finale (MakeMusic inc., 2012) and presented as Musical Instrument Digital Interface (MIDI) files using the grand piano sample. Pitches ranged from G3 to E5 and were all 400 ms in duration. The metronome beats for the TI test (Pecenka & Keller, 2009) came from sampled bell sounds on a Roland SPD-S MIDI percussion pad. The excerpts used in the SMS-tapping task came from MIDI recordings of well-rehearsed pianists and were edited slightly to remove incorrect notes and add missing notes. These recordings have been used in a number of studies of expressive timing (Repp, 1998, 1999; Repp & Knoblich, 2004). For a list of tasks and related dependent measures, see Table 1. The BAIS, pitch imagery task, TI task, and SMS-tapping task were all presented in Max/MSP (Cycling '74, version 6.0).

Procedure

Musical background and BAIS. Participants first filled out the musical background questionnaire. They then completed the BAIS, starting with the vividness (BAIS-V) subscale. The questionnaire was presented using Max/MSP. 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 rated 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*"). The second half of 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

Table 1. List of tasks and measures.

Task/metric	Dependent measures	Description	Interpretation of measure
Musical Background Questionnaire	Years of music experience	Self-reported years of music training	Higher value indicates more training
BAIS	Imagery vividness (BAIS-V) Imagery control (BAIS-C)	Average of self-reported scores on each item, separated by subscale	Higher values indicate better self-reported imagery vividness/control
Working memory (digits backwards and operation span)	Aggregate working memory (WMS)	Sum of participants' scores on digits backwards and operation span tests of working memory	High values indicate greater working memory span
Temporal imagery	Temporal error discrimination threshold (TI)	Threshold estimate on the temporal imagery test	Lower scores indicate finer temporal discernment, that is, better temporal imagery
Pitch imagery	Proportion correct (PIAT)	Proportion of correct trials on the Pitch Imagery Arrow Task	Higher scores indicate better pitch imagery, that is, more correct trials at higher difficulty levels
Sensorimotor synchronisation: asynchrony	Asynchrony	The mean of absolute values of asynchrony scores from the expressive timing tapping task	Lower values indicate better synchronisation, or, less asynchrony
Sensorimotor synchronisation: anticipatory timing	Prediction/tracking index (P/T)	A measure of anticipatory timing derived from tapping patterns (lag-0 to lag-1)	A positive value indicates prediction, negative indicates tracking

WMS = working memory score; TI = temporal imagery; PIAT = Pitch Imagery Arrow Task.
Tasks were completed in the order listed from top to bottom, except for the TI and Pitch Imagery tasks, the order of which was randomised for each participant.

from 1 ("*no image generated*") to 7 ("*extremely easy to change the image*").

Working memory span. After the BAIS, participants completed the two working memory tests. For the digits backwards test, the experimenter read lists of digits to the participant at a rate of two digits per second. The participant was then asked to recall the list and recite it backwards. List length started at two digits and reached a maximum of nine digits. However, the task was ended if participants failed twice to recite the list of digits in the reverse order. There were two trials of each list length. The experimenter was trained with a pacing signal to read the lists at a rate of two digits per second. However, no pacing signal was used during testing to minimise distraction. In the operation span test, participants were shown PowerPoint slides, each containing a short sentence that was either semantically correct (e.g., "Paris is in France") or semantically incorrect/nonsensical (e.g., "the lamp washed itself"). After each item, an intervening slide containing a "++" symbol cued participants to circle "true" if the preceding sentence was semantically correct, or "false" if incorrect on a response sheet. They were also instructed to try and remember the last word of each sentence while making their true/false judgements. 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 two sentences. Each trial increased the sequence length (i.e., the number of sentences) by 1, up to Trial 6 (7 sentences). The sentence and "++" slides progressed automatically every 2 s. The slides instructing participants to recall the words progressed to the next trial after 4 s × Trial number (e.g., recall time allotted for Trial 3 was 12 s) 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. To test pitch imagery, we used a modified version of the Pitch Imagery Arrow Task (PIAT; Gelding et al., 2015). Each trial consisted of a set of tones presented randomly in one of five major keys (C, C#, D, E*b*, and E). Trials started with a musical scale in the selected key to establish a tonal context; no imagery was expected of participants in this phase. Scale tones were presented steadily with a 100 ms duration and 500 ms 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 durations of 500 ms and IOIs of 1,000 ms. The first pitch was always the tonic (first note of the scale), and then moved in increments of 1 scale step, randomly up or down. The range of possible pitches was 3 below tonic to 4 above tonic. Thus, if the pitch reached the lowest scale tone in the set range, it would move up on the next step, and if it reached the highest tone in the range, it would move down on the next step.

Table 2. List of compositions used for the SMS-tapping task and the EBAT.

Title	Mean IOI	*SD* of IOIs	Range of IOIs	Notated beat value
Etudes, No. 3 Etude in E major, Op. 10 by Frédéric Chopin, Performance 1 (Experiments 1 and 2)	489.84	147.04	882	16th note
Etudes, No. 3 Etude in E major, Op. 10 by Frédéric Chopin, Performance 2 (Experiment 1)	453.30	105.94	480	16th note
Etudes, No. 3 Etude in E major, Op. 10 by Frédéric Chopin, Performance 3 (Experiment 1)	601.14	166.82	1,127	16th note
Nocturnes, No. 1 Nocturne in B major, Op. 32 by Frédéric Chopin (Experiment 2)	467.30	221.78	1,618	8th note
Nocturnes, No. 1 Larghetto in B*b* minor, Op. 9 by Frédéric Chopin (Experiment 2)	367.53	83.68	430	8th note
Sonata in F major, K. 533, III. Allegretto by W.A. Mozart (Experiments 1 and 2)	386.43	38.81	213	¼ note
Sonata in G major, K. 283, I. Allegro by W.A. Mozart (Experiment 2)	475.61	39.67	162	8th note
Sonata in e minor, Op. 90, I. Con vivacità by L. Beethoven (Experiment 2)	403.15	76.11	495	¼ note

SMS = sensorimotor synchronisation; EBAT = Expressive Beat Alignment Test; IOI = inter-onset interval; *SD* = standard deviation.

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

All participants completed 50 trials and started at Level 1, which contained three pitch/arrow pairs, and one silent, gray 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. The maximum level was 5, and each level number indicated how many silent gray arrows were included in each trial at that level. The number of pitch/arrow pairs before the silent arrows increased by one at Levels 2, 3, and 5.

The TI test was adapted from a previous study of imagery and SMS (Pecenka & Keller, 2009). In this test, participants heard five beats with IOIs either increasing from 400 to 500 ms (slowing down) or decreasing from 500 to 400 ms (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. Responses were made on the left (early) and right (late) arrow keys of the computer keyboard. The test ended once participants reached their discrimination threshold, defined as percent error at which one could no longer distinguish early from late above chance. The beats were presented as piano notes of 100 ms duration. The beat sequences were pitched at E4, and the probe beat was pitched at F4.

SMS. The stimuli for the synchronisation task consisted of the opening bars of a Mozart piano sonata (K. 533, Rondo, Bars 1-12; Repp & Knoblich, 2004), and three different performances of an excerpt of a Chopin piano étude (Op. 10, No. 3, Bars 1-5; for details regarding the recording see Repp, 1998). The Mozart was selected as a training excerpt so participants could practice the procedure before tapping along to the Chopin excerpts. The Mozart excerpt is acoustically similar to the Chopin (polyphonic piano music), but has a mostly regular beat sequence. The Chopin was selected because typical performances of Chopin's music are very expressive, containing irregular beat sequences (see Table 2 for a list of compositions and descriptive statistics of the IOI sequences). Figure 1 shows that the Mozart is notably less variable, reflecting the mostly regular beat. The three Chopin profiles differ from each other, but are all characterised by alternation between increasing and decreasing speed. Only one performance of the Mozart piece was used, as this was sufficient to have participants acclimate to the procedure. Because the main interest in the SMS task was tapping to expressively timed music, three different Chopin performances were used,[1] each with a unique timing profile. The use of multiple

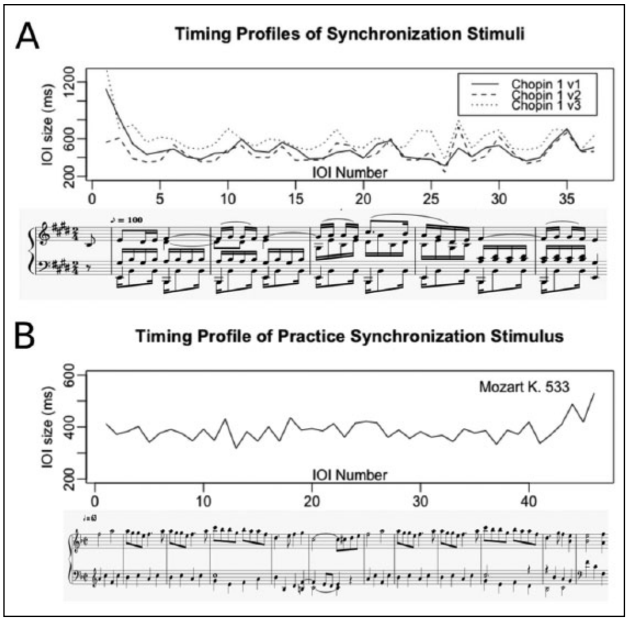

Figure 1. Timing profiles for the excerpts used in Experiment 1 represented as inter-onset intervals (IOIs). The three timing profiles for the Chopin excerpt (Panel A) have the same general shape, but different degrees of fluctuation. The Mozart (Panel B) was a useful excerpt for practicing as it contained some fluctuation, but was not as variable as the Chopin.

performances of the Chopin piece ensures that any observed patterns of synchronisation are not unique to one particular timing profile, but instead generalise to other interpretations of the music. Indeed, the IOI sequences of the three Chopin performances correlate significantly, but not perfectly (Pearson's r ranged from .51 to .81), confirming that the performances were similar in style, but not identical. This is important, as we are interested in expressive timing synchronisation in general, not just for one particular rendition.

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. 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 while avoiding practice effects. After the demonstration, participants tapped along to four trials of the Mozart excerpt.

Following the four Mozart trials, the experimenter demonstrated how to tap along to the Chopin excerpts at the 16th note level (see Figure 1). Participants then tapped along to four trials of three different performances of the excerpt (12 total trials). The order of the performances was randomised, 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 randomisation of the 12 trials would have confounded analysis of the progress participants made in synchronising with each performance over multiple trials, and—based on pilot testing—would have made synchronisation extremely difficult.

Participants' first taps started the excerpts to ensure that the 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 the excerpt and a 3-s pause, a "!" character appeared on the computer screen to alert participants that they could then start the next trial.

Analysis of SMS. The primary dependent measure in the SMS-tapping task was the mean of the absolute values of asynchronies. Asynchronies were computed as the difference in milliseconds between a tap and its target beat. Thus, each trial contained a series of asynchronies equal in length to the number of beats/taps. In order to calculate the average asynchrony for each trial, the asynchrony series (i.e., the difference between each inter-tap interval [ITI] and the corresponding IOI) were converted to absolute values. This is because asynchronies can be positive or negative depending on whether a tap occurs before or after the corresponding beat. Averaging a mix of positive and negative values would produce an average asynchrony that is much lower than the actual magnitude of asynchronies. Thus, each participant produced four absolute mean asynchrony scores for the Mozart excerpt, and 12 for the Chopin excerpts (one for each trial). Lower absolute asynchrony scores indicate better performance (i.e., greater synchrony).

Communications between computer sound cards and Max/MSP are known to produce latencies, such that when a signal to produce a sound is given, the sound will be sent out after a slight delay. While this latency is generally consistent within a system, it can vary considerably between systems, prompting us to test the latency of our lab set-up. To test this, we measured the time between a sent signal (in this case a key tap on the musical keyboard) and the output triggered by receiving the signal (a MIDI note from Max). Thus, t_0 was the key tap, and t_1 was when Max output a sound, both of which were detected by separate microphones, and analysed in Audacity. A test of 1,000 self-paced key taps showed a mean latency of 148 ms, $SD = 10$ ms, meaning on average, the start of the musical stimuli was 148 ms after participants initiated a trial. This latency value was subtracted from the average asynchrony values produced by all participants to account for the delay in the system.

Asynchrony, however, is only a coarse measure of how accurately participants can synchronise. It does not necessarily indicate the strategy they use to synchronise as the tempo fluctuates. Thus, a second dependent measure assessed anticipatory timing, or the degree to which individuals anticipate upcoming tempo changes (*prediction*), or respond to past tempo changes (*tracking*). Such predicting versus tracking tendencies can be modelled by computing a cross-correlation (CC) at different lags between ITIs and IOIs of the beats (Pecenka & Keller, 2011).[2] 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. A prediction/tracking index (P/T) was derived from these correlations 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 trials on both excerpts.

Analysis of asynchrony and P/T required that both time series (the beat sequence and tap sequence) had 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 participant's data to be included in the analysis, the following criteria were used: data from all of a participant's trials were considered unusable if there were at least five fewer taps than beats in one or more of a participant's trials (e.g., a participant with 32 taps in a Chopin excerpt—which had 37 beats—would not be useable). If a participant was only missing four or fewer taps (either consecutive or non-consecutive), an interpolation procedure was used to fill in the missing taps. This was necessary for the P/T analysis, as there needs to be an equal number of ITIs and IOIs to cross-correlate the sequences. 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 gap. Nine participants produced data that were unusable for the SMS analysis, reducing the sample size on SMS measures to 36.

Results

Working memory. The mean span on the digits backwards task was 5.20 ($SD=0.94$, range=3-8). 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). After checking that the two measures were highly correlated, $r(40)=0.78$, $p<0.001$, the WMSs were standardised and summed using z-scores to create an aggregated WMS. This was done to facilitate analyses of working memory in relation to other variables.

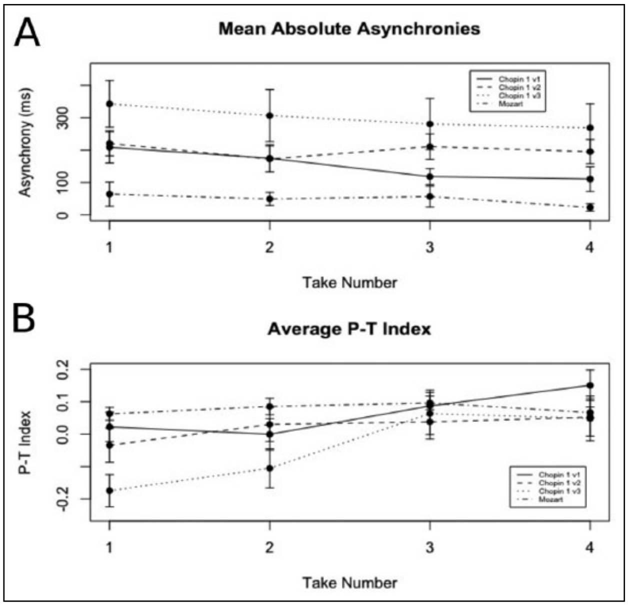

Figure 2. Mean absolute asynchrony and P/T scores averaged across all participants for each trial (1-4) and each performance excerpt (one Mozart and the three Chopin performances). Error bars represent standard error of the mean. Asynchronies have been adjusted for latencies in the processing time of Max/MSP.

Auditory imagery. In all analyses, the BAIS was analysed as two separate subscales (Vividness and Control). The means, SDs, and ranges for the two subscales were similar (BAIS-V: $M=4.57$, $SD=0.76$, range=3.14-6.5; BAIS-C: $M=4.94$, $SD=0.80$, range=3.57-6.71). The PIAT was scored as the proportion correct ($M=0.71$, $SD=0.14$, range=0.42-1.00). Previously, the PIAT has been scored by reporting the maximum level reached by each participant. However, nearly all participants reached the maximum level in this sample, and so proportion correct was a better measure of PIAT performance. TI 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.

SMS: asynchrony and P/T index. To get a general sense of how well participants could synchronise with the Mozart excerpt and the Chopin excerpts, the mean asynchrony scores (see Figure 2) were averaged across all trials of the Mozart ($M=37.92$ ms, $SD=81.16$ ms) and of the Chopin ($M=212.41$ ms, $SD=112.92$ ms). Because there were three versions of the Chopin excerpt, a one-way analysis of variance (ANOVA) was conducted to see whether there were differences in asynchrony among the versions. The ANOVA revealed no significant differences, suggesting that the Chopin excerpts were all equally difficult for participants. For this reason, the mean score across all trials of all three versions was used in subsequent regression analyses.

Table 3. Summary of hierarchical regression analysis for variables predicting asynchrony.

Measure	β	*t*	*sr*^{2}	*R*	*R*^{2}	Δ*R*^{2}
Step 1				.12	.02	.02
WMS	−0.12	−0.70	.01			
Step 2				.46	.21	.20
BAIS-V	0.16	0.70	.01			
BAIS-C	−0.57	**−2.47***	.32			
Step 3				.69	.47	.26
PIAT	−61	**−3.45****	.37			
TI	0.34	**2.15***	.12			

WMS = working memory score; PIAT = Pitch Imagery Arrow Task; TI = temporal imagery.
*p < .05 **p < .01.

To assess the extent to which participants were predicting or tracking beat intervals in the two pieces, the P/T index was averaged across all trials. Positive scores indicate prediction, whereas negative scores indicate tracking. Generally, participants predicted more on the Mozart excerpt (M=0.08, SD=0.08) than on the Chopin excerpts (M=−0.01, SD=0.19), $t(70)=2.59$, $p<0.05$. 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.

One would expect a correlation between asynchrony and P/T such that high predictors have lower asynchronies. Indeed, the correlation trended in this direction, $r(34)=-0.21$, $p=0.15$, but was not significant. This is likely due to the high variability in tapping performance in this sample.

To examine the relationship between SMS, and working memory and imagery, hierarchical regression models were created based on a priori hypotheses regarding the independent predictive power of working memory, the self-reported BAIS, and objectively tested pitch and tempo imagery. Thus, auditory imagery and WMS were the predictor variables, and asynchrony (Table 3) and P/T (Table 4) were the predicted variables. For both SMS measures, we were interested in whether self-reported imagery (BAIS) could add predictive power above and beyond WMS, and whether imagery for musical qualities (pitch and tempo) could add explanatory power beyond the BAIS. Only the SMS measures from the Chopin trials were tested in these models, as the Mozart was intended only as practice, and the primary interest was in predicting performance on the musically expressive time series.

With asynchrony as the dependent variable (Table 3), WMS was included in Step 1 of the model but did not yield significant predictive power, $t(34)=-0.70$, $p>.05$, change in $R^2=.02$. The addition of the BAIS at Step 2 was significant (change in $R^2=.20$). However, a significant contribution came only from BAIS-C, $t(34)=-2.47$, $p<.05$, and not BAIS-V, $t(34)=0.70$, $p>.05$. Finally, pitch and tempo imagery were added in Step 3, which explained an additional 26% of the variance in asynchrony (change in $R^2=.26$; total $R^2=.47$). Both independent variables at Step 3 were significant, but PIAT, $t(34)=-3.45$, $p<.01$, was a stronger predictor than TI, $t(34)=2.15$, $p<.05$.

A second model was used to predict P/T indices from the Chopin excerpts (Table 4). Again, WMS was included in Step 1, and this time was a significant predictor, change in $R^2=.16$; $t(34)=2.44$, $p<.05$. Step 2 added BAIS-V and BAIS-C, which contributed significant predictive power (change in $R^2=.14$). Again, a significant change came only from BAIS-C, $t(34)=-2.31$, $p<.05$, and not BAIS-V, $t(34)=0.88$, $p>.05$. Step 3 included PIAT and TI, which significantly explained further variance in P/T (change in $R^2=.19$; total $R^2=.49$). This time, only TI, $t(34)=-3.24$,

Table 4. Summary of hierarchical regression analysis for variables predicting P/T.

Measure	β	*t*	*sr*^{2}	*R*	*R*^{2}	Δ*R*^{2}
Step 1				.40	.16	.16
WMS	0.40	**2.44***	.16			
Step 2				.55	.30	.14
BAIS-V	0.19	0.88	.04			
BAIS-C	−0.51	**−2.31***	.26			
Step 3				.70	.49	.19
PIAT	0.08	0.46	.01			
TI	−0.51	**−3.24****	.26			

WMS = working memory score; PIAT = Pitch Imagery Arrow Task; TI = temporal imagery.
*p < .05 **p < .01.

$p < .01$, and not the PIAT, $t(34) = 0.46$, $p > .05$, contributed significantly to the model.

Discussion

In Experiment 1, participants completed several tests of auditory imagery and working memory, and then tapped along to expressively timed music that was characterised by a dynamic, irregular beat sequence. Although the observed average asynchronies and SDs were higher than those typically found in a tapping task (Repp & Su, 2013), the participants had very little music training and did not practice tapping with the Chopin excerpt at all. Instead, they watched the experimenter demonstrate the task. Thus, the scores reflect performances on a largely novel task. These asynchrony scores were predicted by self-report of imagery control (BAIS-C), a test of pitch imagery (PIAT), and a test of TI (TI task). Anticipating beat onsets, measured by the prediction/tracking score, also implicated imagery control and TI, but unlike asynchrony, variance in prediction/tracking was also explained by working memory span. It is important to remember that these results were obtained from a sample of people with minimal musical training who had not been practising or performing for at least 7 years.[3] Our results therefore speak to the inherent timing abilities developed by young adulthood and/or the development of those skills in everyday music or motor activities.

BAIS-V (imagery vividness), which was not a significant predictor variable, is a measure of how clearly one can imagine various aspects of sound. This may not be relevant in musical SMS because vividness involves maintaining a static image, whereas synchronisation is a dynamic process. BAIS-C predicted both asynchrony and prediction/tracking. Because BAIS-C is a measure of how easily one can change an established image, this finding suggests that it may be important to quickly adjust a representation of time in order to match a series of changing beat intervals. In other words, predictive auditory images are constantly changing to represent the different IOIs. Thus, easily changing an image would be beneficial.

TI also predicted both asynchrony and the prediction/tracking index, such that a low threshold in the TI task predicted low asynchrony, and a high prediction/tracking score. This could reflect the need to imagine the temporal irregularities underlying expressively timed music in order to form the most appropriate action plan (Pecenka, Engel, & Keller, 2013; Pecenka & Keller, 2009). 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, after some exposure to the music, if one can imagine when an action should occur, the actual output should be accurate, as indicated by low asynchrony. Also, having accurate imagery for tempo could facilitate anticipating the size of the upcoming beat interval—as indicated by a high prediction/tracking score—by providing not just a clear image of onset time, but of the size of the interval.

Pitch imagery predicted asynchrony, but not the prediction/tracking index, meaning people with accurate pitch imagery tended to synchronise better, but not necessarily anticipate upcoming beat intervals. 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 fluctuating time intervals. Instead, good pitch imagery might facilitate encoding and retrieving the melody of the excerpts (Collins, Tillmann, Barrett, Delbé, & Janata, 2014; Lee, Janata, Frost, Martinez, & Granger, 2015), which would give participants a general idea of when their taps should occur (Pecenka & Keller, 2009) and reduce asynchrony. Good pitch imagery would not, however, facilitate anticipation, which is related instead to tempo imagery. Tempo imagery, therefore, is likely the more fundamental skill in timing, but pitch imagery may be beneficial, especially when faced with unfamiliar music.

The results also suggest a role for working memory in SMS, specifically for predicting time intervals when the interval sequence is irregular or “expressively timed.” Given the effortful nature of synchronisation to expressively timed music, we expected working memory to also be related to measures of asynchrony. However, there could be other forms of conscious processing needed for synchronisation, such as selective attention (Keller & Burnham, 2005). Synchronising with a dynamic time series might not require maintenance and manipulation of sensory information as per working memory theory, but instead depend on rigorous monitoring of the stimulus. Anticipatory timing, on the other hand, could very reasonably require working memory (Pecenka et al., 2013), given the need to think ahead to upcoming events while monitoring current actions.

The significant predictor variables found in Experiment 1 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 (measured by BAIS-C) in both synchronisation and prediction could reflect the need to constantly update the image contained in an inverse model. Because the timing of motor commands is not on a fixed interval when synchronising to a sequence of changing interval sizes, it follows that being able to easily change the image informing the model would be related to better synchronisation, and also to anticipation of beat onsets. If the inverse model contains instructions regarding the timing and force of an action—as theorised in studies of motor control (Wolpert, Doya, & Kawato, 2003; Wolpert, Ghahramani, & Jordan, 1995)—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.

Pitch imagery may be involved in maximising synchrony by contributing to feedback from a forward model of the actions of the recorded pianist. Forward models compare an expected outcome to actual outcomes, and so good pitch imagery could help form an accurate expected outcome of the music in real time. Indeed, there is neural evidence of auditory efference copies that are modulated by imagery (Tian & Poeppel, 2010), and evidence for efference copies predicting the intentions of others (Blakemore & Decety, 2001). Such processes could make the difference between expected and actual outcomes more apparent, and therefore easier to incorporate into corrective feedback.

Working memory could operate in tandem with selective attention to inform and update internal models. As participants monitored IOIs of varying sizes using selective attention, they were presumably detecting regularities and patterns in the time series. Individuals with efficient working memory were likely able to use the encoded timing patterns to generate a prediction of the next IOI size in the form of an auditory image, while continuing to monitor the music. An inverse model could use this prediction to update the current motor plan and generate an estimate of necessary motor commands. This estimate is then sent through a forward model that generates its own prediction concerning the outcome of this motor plan. A comparison between the forward model's prediction and the anticipated time interval informing the inverse model allows one to re-update the plan before executing it, if the discrepancy between model predictions is great enough. As this process repeats for the next interval, the forward model contributes feedback from the actual outcome to the next plan. In other words, anticipation, as indicated by a highly positive prediction/tracking index, requires one to maintain attention to present motor output and stimuli while simultaneously updating the contents of internal models in preparation for forthcoming motor output. In other words, working memory could allow one to stay ahead of the beat by processing what should come next, while executing an action related to the current time interval.

Experiment 2

Experiment 2 replaced the tapping task with a test of expressive timing perception in which participants judged whether click tracks were aligned with the beats of musical excerpts. This task, which requires comparing (without tapping along) two external auditory streams, was used to better understand how dynamic time series are processed before any motor activity occurs. The act of synchronisation must start with perceiving the auditory sequence and encoding its temporal properties. Experiment 2, therefore, was intended to test whether working memory and auditory imagery would predict expressive timing *perception*, as it is possible that the relationships found in Experiment 1 were realised at the perceptual level, and not necessarily related to action. Given the need to process the musical aspects of these types of time series, it was hypothesised that the PIAT and TI task tests would predict expressive timing perception without action, as they capture imagery abilities that are related to aspects of music. Regarding the BAIS, we expected vividness but not control to predict accurate judgement of phase alignment. This is because there is no need to update an action plan by changing an established image, but having an especially vivid image of the music might facilitate comparing what proper alignment should sound like, to what is actually being heard. These relationships were expected beyond any variance accounted for by working memory. Therefore, a hierarchical regression model using the same method of entry for the predictor variables as in Experiment 1 (working memory first, then BAIS, then pitch imagery and TI) was used for analysis in Experiment 2.

Methods

Participants. New participants (N=27, 20 female) were recruited from Bucknell University's research subject pool and given course credit for participating. Age ranged from 18 to 21 years (M=19.11 years; SD=1.15 years). Years of musical experience ranged from 0 to 4 years (M=2.21 years; SD=0.79 years), and years since ceasing musical experience ranged from 9 to 12 years (M=10.84 years; SD=0.51 years).

Stimuli and procedure. An adaptation of the Beat Alignment Test (BAT; Iversen & Patel, 2008) was used instead of the SMS-tapping task in Experiment 2. Also, in addition to the two pieces from Experiment 1, four other pieces similar in style were added for the perception test. These are listed in Table 2, and the IOIs and score excerpts are shown in Figure 3. Other than that, all materials and procedures were the same as in Experiment 1. The original BAT includes a perceptual test in which participants make judgements of whether or not 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 named the Expressive Beat Alignment Test (EBAT). Click tracks synthesised from a woodblock sound were imposed over the Chopin and Mozart excerpts used in Experiment 1, as well as four additional expressively timed music excerpts (see Table 2 for a list). 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 error condition; (3) on the beat, for an on-time condition. Participants were asked to judge whether the clicks matched the beat and also whether the clicks fell before or after the beat. Based

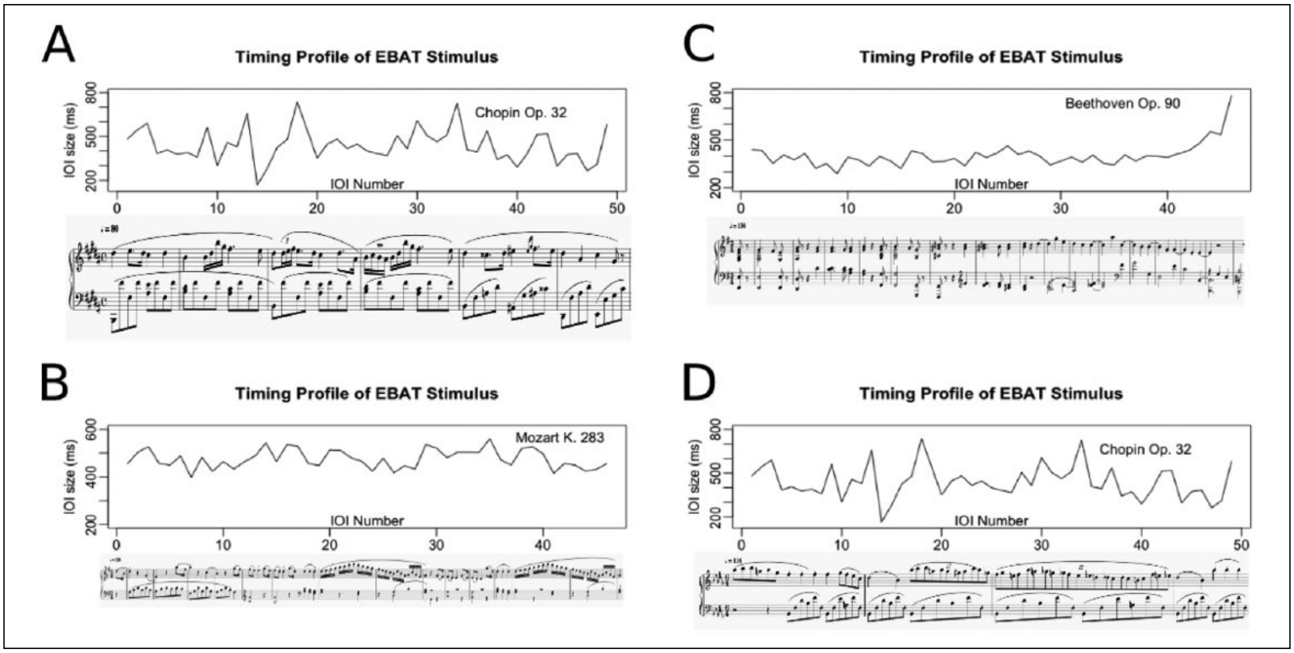

Figure 3. Timing profiles for the additional excerpts used in Experiment 2 represented as inter-onset intervals (IOIs).

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 phase offset was consistent throughout all phase error conditions.

The click tracks started at the fifth IOI (i.e., with/just before/just after the sixth beat depending on trial type) to ensure that participants had time to identify the musical beat. The six pieces used were shortened to ~10 s 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 randomised for all participants. The scored trials were preceded by practice trials that included feedback regarding the correctness of a response. 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, such that each eight-beat passage would gradually slow or hasten. Participants were scored on their proportion correct.

Results

Working memory. The mean span on the digits backwards task was 5.29 ($SD=0.72$, range = 4-7). The operation span task had a mean of 21.05 ($SD=3.14$, range = 15-26). Again, after checking that the two measures were significantly correlated, $r(25)=.46$, $p<.05$, the WMSs were standardised and summed using z-scores to create an aggregated WMS.

Auditory imagery. As in Experiment 1, for all analyses, the BAIS was analysed as two separate subscales (Vividness and Control). The means, SDs, and ranges for the two subscales were similar (BAIS-V: $M=4.48$, $SD=0.72$, range = 3.21-5.90; BAIS-C: $M=4.74$, $SD=0.71$, range = 2.93-5.99). The PIAT was again scored by the proportion correct ($M=.71$, $SD=0.12$, range = 0.50-0.91). TI 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; $M=29.69$, $SD=2.80$, range = 26.05-36.92), meaning a lower score indicated better performance.

The EBAT was scored by proportion correct. There were three conditions—on time ($M=0.59$, $SD=0.12$), early ($M=0.48$, $SD=0.15$), and late ($M=0.36$, $SD=0.14$)—and the mean proportion correct across all conditions was 0.47, $SD=0.10$. This is significantly above chance, $t(26)=6.71$, $p<.001$, chance = 0.33 proportion correct. Also, participants performed significantly better on the on-time trials than on the early, $F_{(21)}=15.41$, $p<.05$, and late, $F_{(21)}=15.41$, $p<.01$, conditions (cf. Jones, Moynihan, MacKenzie, & Puente, 2002; Penel & Jones, 2005).

For the main analysis, we used a hierarchical regression model with the same predictor variables as in Experiment 1, but with proportion correct of EBAT responses as the dependent variable (see Table 5). Step 1 included working memory, which, contrary to our hypothesis, did not predict any variance in the EBAT, $R^2<.01$. Step 2 added the BAIS, which gave the model predictive validity (change in $R^2=.38$). As hypothesised, only BAIS-V, $t(25)=3.66$, $p<.01$, and not BAIS-C, $t(25)=-1.81$, $p>.05$, contributed significantly to the model. Finally, pitch and tempo imagery were added at Step 3, predicting further variance in the

Table 5. Summary of hierarchical regression analysis for variables predicting EBAT.

Measure	β	*t*	*sr*^{2}	*R*	*R*^{2}	Δ*R*^{2}
Step 1				.03	.001	.001
WMS	0.03	0.17	.001			
Step 2				.62	.38	.38
BAIS-V	0.76	**3.66****	.58			
BAIS-C	−0.36	−1.53	.13			
Step 3				.76	.58	.20
PIAT	0.06	0.32	.001			
TI	−0.49	**−3.15****	.24			

EBAT = Expressive Beat Alignment Test; WMS = working memory score; PIAT = Pitch Imagery Arrow Task; TI = temporal imagery.
**p* < .05 ***p* < .01.

EBAT above and beyond BAIS-V (change in $R^2=.20$, $p<.01$; total $R^2=.58$). However, only TI, $t(25)=-3.15$, $p<.01$, and not the PIAT, $t(25)=0.32$, $p>.05$, was a significant predictor variable, and so the hypothesis that both types of musical imagery would contribute was half supported.

Discussion

Overall, compared to the models predicting asynchrony and prediction/tracking scores (a measure of anticipatory timing), the model predicting the EBAT showed several interesting differences. First, working memory did not predict EBAT performance. Working memory did predict anticipatory timing in Experiment 1, but the EBAT does not require anticipation so much as it requires comparison by judging synchrony between clicks and beats. Therefore, attentional systems may be more relevant than working memory to perceiving asynchrony as one does not need to think ahead, but focus on the current beat onset.

Second, unlike the tapping task where BAIS-C (imagery control) was a good predictor of performance, BAIS-V (imagery vividness) but not BAIS-C predicted performance on the EBAT. This was hypothesised, as 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 facilitate the comparison, and help identify correct alignment. This would fit into the forward model described previously, in which imagery vividness contributes to a feedback loop by establishing a robust efferent copy.

Third, TI but not pitch imagery was a significant predictor variable. Although we expected the ability to accurately imagine both pitch and tempo to relate to one's perception of musical alignment, it seems that only one's sensitivity to beat onsets is implicated. On one hand, this seems intuitive, as the task of judging temporal alignment between two auditory streams could require one to imagine what proper timing would sound like. On the other hand, this finding could be a product of the task design. In a more ecologically valid setting in which one is assessing the synchrony of a music ensemble, imagining proper pitches and then comparing those to actual pitch production might be a necessary skill. However, the EBAT requires comparing a percussive sound to actual music, which might require participants to draw exclusively on imagery related to tempo processing.

General discussion

We investigated how working memory and several types of auditory imagery independently related to SMS and perception of expressively timed music. Experiment 1 showed that among people with minimal musical training, imagery predicts motor behaviour during synchronisation with dynamically timed, expressive music. Furthermore, working memory explained variance in the use of anticipation during synchronisation, suggesting that it has a direct role in anticipatory SMS. Experiment 2 showed that self-reported imagery vividness and a test of TI predict one's ability to perceive expressive timing patterns. Together, these studies show that auditory imagery is a mediator of complex timing in motor behaviour, particularly in musical contexts. Thus, our findings add to established theories of the connection between imagery and action, and more broadly to knowledge of the psychology of the arts.

In addition to the lack of extensive or recent musical training in our samples, no participants reported listening to classical music on a regular basis. Most of their musical exposure was contemporary music with an isochronous beat. The novelty of having to synchronise with an irregular but musically structured time series, therefore, lends validity to the predictive power of imagery and working memory found here, as difference in experience was controlled for. Furthermore, all three of the predicted variables across Experiments 1 and 2 (asynchrony, prediction/tracking index, and EBAT performance) are similar to common behaviours: producing and perceiving music. At a higher level, temporal anticipation and EBAT performance could be related to music appreciation considering that one's expectations of a given piece of music and one's ability to perceive synchrony can enhance enjoyment (Krumhansl, 2002; Phillips-Silver et al., 2011). Thus, these results could be indicative of individual differences in cognition that underlie motor and perceptual timing in the general population.

The potential roles for auditory imagery and working memory in SMS found in Experiment 1 are contrary to aspects of current theories of motor timing that posit that SMS is a result of neural oscillators dynamically entraining to periodic external stimuli and is therefore probably independent of cognitive processes (Large, 2000, 2008) such as those measured here. One potential explanation for the role of cognition in musical synchronisation is as follows: central representations of motor plans instantiated by auditory images and working memory may be related to timing in real, expressive music because expression is conscious and goal-directed. Therefore, to imitate it by tapping requires some degree of effortful processing to achieve both corrective and anticipatory timing.

Experiment 2 tested the perception of expressive music without related action. Imagery vividness and imagery for tempo predicted good judgement of phase matching between two auditory sequences. The significance of BAIS-V in Experiment 2 but not Experiment 1 points towards the importance of vivid imagery for perceptual comparisons, whereas control of images may be most relevant in SMS where the objective is motor output. The common predictor between Experiments 1 and 2 was TI, suggesting an internal sense of time passing (Michon, 1967; Stevens, 1886) or a representation of time intervals (Wing, 2002) has roles in both perception and action. Although there is disagreement regarding the nature of perfectly regular motor timing (Wing & Beek, 2002), when a musical time series is characterised by the objective of expressivity, an information processing explanation of timing might be more applicable. In other words, the ability to internalise changing time patterns could reasonably relate to both perceiving and interacting with 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).

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 (Maes et al., 2015; Pecenka et al., 2013) that require participants to use working memory in a context irrelevant to SMS while trying to synchronise in a tapping task. Presumably, dual-tasks would worsen synchronisation 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.

Neural underpinnings of the perception-action links involved in expressively timed synchronisation could be established using transcranial magnetic stimulation (TMS). Particular interest should be given to cortical motor areas (Doumas, Praamstra, & Wing, 2005; Patel & Iversen, 2014). 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 motor cortex with TMS might not only increase asynchrony but also decrease performance on a perception task such as the EBAT.

Overall, the experiments presented here revealed a potential role for auditory imagery and working memory in SMS when synchronising 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 (imagery vividness) might facilitate judgements of incoming sound, whereas a high working memory span and one's ability to change an established image (imagery control) could help plan for and update action plans. The two processes may find common ground in one's ability to imagine temporal relationships (TI), which could be capturing part of a central representation of time. 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 synchronisation, and that working memory may be necessary for anticipatory timing in real, dynamic musical sequences.

Acknowledgements

The authors would like to thank Bucknell University Research Assistants Samantha Santomartino, Melanie Grossman, Adan Rodriguez, and Savannah Straub for their help running the experimental procedure and recruiting participants.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. We elected not to use a mechanical version of the Chopin as training for the task because it would be stylistically inappropriate and detract from the goal of ecological validity.
2. Recent discussions of time series analysis have advocated the use of autoregressive models in place of CCs (Dean & Dunsmuir, 2015). It is worth noting, therefore, that the prediction/tracking (P/T) indices reported here were calculated

using conventional CC analysis, but significantly correlated with the more advanced autoregressive method, $r(34) = .84$, $p < .0001$.

3. Although we were interested in minimal musical training for this experiment, we began recruiting musicians (>10 years of music experience and currently practicing/performing) and accumulated a sample size of 13. The disparate sample sizes meant we could not compare the two groups validly, but we did run bivariate correlations using the musicians' data. The results here showed that only BAIS-V was a significant predictor of asynchrony for musicians. Further data collection is needed, but this finding suggests that musicians might rely on vivid expectations of musical structure followed by corrective feedback, rather than easily updating images.

References

Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, *129*, 126–145.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (*Vol. 8*, pp. 47–89). New York, NY: Academic Press.

Baddeley, A. D., & Logie, R. (1992). Auditory imagery and working memory. In D. Reisberg (Ed.), *Auditory imagery* (Vol. 12, pp. 179–197). Hillsdale, NJ: Lawrence Erlbaum.

Blakemore, S., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*, 561–567. doi:10.1038/35086023

Borella, E., Ludwig, C., Dirk, J., & de Ribaupierre, A. (2011). The influence of time of testing on interference, working memory, processing speed, and vocabulary: Age differences in adulthood. *Experimental Aging Research*, *37*, 76–107. doi:10.1080/0361073X.2011.536744

Collins, T., Tillmann, B., Barrett, F. S., Delbé, C., & Janata, P. (2014). A combined model of sensory and cognitive representations underlying tonal expectations in music: From audio signals to behavior. *Psychological Review*, *121*, 33–65. doi:10.1037/a0034695

Dean, R. T., & Dunsmuir, W. T. (2015). Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. *Behavior Research Methods*, *48*, 783–802.

Doumas, M., Praamstra, P., & Wing, A. M. (2005). Low frequency rTMS effects on sensorimotor synchronization. *Experimental Brain Research*, *167*, 238–245. doi:10.1007/s00221-005-0029-7

Flanagan, J. R., & 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. *The Journal of Neuroscience*, *17*, 1519–1528.

Gelding, R. W., Thompson, W. F., & Johnson, B. W. (2015). The Pitch Imagery Arrow Task: Effects of musical training, vividness, and mental control. *PLoS ONE*, *10*(3), e0121809. doi:10.1371/journal.pone.0121809

Grahn, J. A., & Schuit, D. (2012). Individual differences in rhythmic ability: Behavioral and neuroimaging investigations. *Psychomusicology: Music, Mind, and Brain*, *22*, 105–121. doi:10.1037/a0031188.supp

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

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

Halpern, A. R. (2015). Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain*, *25*, 37–47.

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 neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, *42*, 1281–1292.

Hooper, S. L. (2000). Central pattern generators. *Current Biology*, *10*, 176–177.

Hubbard, T. L. (2013). Auditory aspects of auditory imagery. In S. Lacey & R. Lawson (Eds.), *Multisensory imagery* (pp. 51–76). New York, NY: 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.

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

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

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*, 459–491.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, *13*, 313–319.

Keller, P. E. (2014). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabian, R. Timmers & E. Schubert (Eds.), *Expressiveness in music performance: Empirical approaches across styles and cultures* (pp. 260–282). Oxford, UK: Oxford University Press.

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

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., Novembre, G., & Hove, M. J. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130394. doi:10.1098/rstb.2013.0394

Keller, P. E., Novembre, G., & Loehr, J. (2016). Musical ensemble performance: Representing self, other, and joint action outcomes. In E. S. Cross & S. S. Obhi (Eds.), *Shared representations: Sensorimotor foundations of social life*. Cambridge, UK: Cambridge University Press.

Krumhansl, C. L. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science*, *11*, 45–40.

Large, E. W. (2000). On synchronizing movements to music. *Human Movement Science*, *19*, 527–566.

Large, E. W. (2008). Resonating to musical rhythm: Theory and experiment. In S. Grondin (Ed.), *The psychology of time* (pp. 189–232). Bingley, UK: Emerald Group Publishing.

Lee, Y., Janata, P., Frost, C., Martinez, Z., & Granger, R. (2015). Melody recognition revisited: Influence of melodic gestalt on the encoding of relational pitch information. *Psychonomic Bulletin & Review*, *22*, 163–169. doi:10.3758/s13423-014-0653-y

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

Maes, P. J., 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. doi:10.1007/s00221-014-4108-5

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

Michon, J. A. (1967). *Timing in temporal tracking*. Leiden, The Netherlands: Leiden University.

Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 20130393. doi:0.1098/rstb.2013.0393

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, *8*, 57.

Pecenka, N., Engel, A., & Keller, P. E. (2013). Neural correlates of auditory temporal predictions during sensorimotor synchronization. *Frontiers in Human Neuroscience*, *21*, 380. doi:10.3389/fnhum.2013.00380

Pecenka, N., & Keller, P. E. (2009). Auditory pitch imagery and its relationship to musical synchronization. *Annals of the New York Academy of Sciences*, *1169*, 282–286. doi:10.1111/j.1749-6632.2009.04785.x

Pecenka, N., & Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Experimental Brain Research*, *211*, 505–515.

Penel, A., & Jones, M. R. (2005). Speeded detection of a tone embedded in a quasi-isochronous sequence: Effects of a task-irrelevant temporal irregularity. *Music Perception*, *22*, 371–388.

Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin & Review*, *20*, 747–753. doi:10.3758/s13423-013-0401-8

Pfordresher, P. Q., Halpern, A. R., & Greenspon, E. B. (2015). A mechanism for sensorimotor translation in singing: The Multi-Modal Imagery Association (MMIA) Model. *Music Perception*, *32*, 242–253.

Pfordresher, P. Q., & Mantell, J. T. (2014). Singing with yourself: Evidence for an inverse modeling account of poor-pitch singing. *Cognitive Psychology*, *70*, 31–57. doi:10.1016/j.cogpsych.2013.12.005

Phillips-Silver, J., Toivianinen, 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.

Repp, B. H. (1998). A microcosm of musical expression: I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E Major. *Journal of the Acoustical Society of America*, *104*, 1085–1100.

Repp, B. H. (1999). Relationships between performance timing, perception of timing perturbations, and perceptual-motor synchronization in two Chopin preludes. *Australian Journal of Psychology*, *51*, 188–203.

Repp, B. H. (2001). Phase correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 600–621.

Repp, B. H. (2002). The embodiment of musical structure: Effects of musical context on sensorimotor synchronization with complex timing patterns. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action: Attention and performance XIX* (pp. 245–265). Oxford, UK: Oxford University Press.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, *12*, 969–992. doi:10.3758/Bf03206433

Repp, B. H. (2006). Musical synchronization. In E. Altenmüller, M. Wiesendanger & J. Kesselring (Eds.), *Music, motor control, and the brain* (pp. 55–77). Oxford, UK: Oxford University Press.

Repp, B. H., & Knoblich, G. (2004). Perceiving action identity: How pianists recognize their own performances. *Psychological Science*, *15*, 604–611.

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

Stevens, L. T. (1886). On the time sense. *Mind*, *11*, 393–404.

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, *1*, 166. doi:10.3389/fpsyg.2010.00166

van der Steen, M. C., Jacoby, N., Fairhurst, M. T., & Keller, P. E. (2015). Sensorimotor synchronization with tempo-changing auditory sequences: Modeling temporal adaptation and anticipation. *Brain Research*, *1626*, 66–87. doi:10.1016/j.brainres.2015.01.053

Van der Steen, M. C., & Keller, P. E. (2013). The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Frontiers in Human Neuroscience*, *7*, 253.

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. (2002). Voluntary timing and brain function: An information processing approach. *Brain and Cognition*, *48*, 7–30. doi:10.1006/brcg.2001.1301

Wing, A. M., & Beek, P. J. (2002). Movement timing: A tutorial. *Common Mechanisms in Perception and Action*, *19*, 202–226.

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: Biological Sciences*, *358*, 593–602.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*, 1889–1882.

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