Expressive Timing in the Mind’s Ear

Bruno H. Repp

Introduction

Musical imagery, defined here as the vivid imagination of musical sounds that are not physically present at the moment, may occur in at least four different real-life situations. First, a composer may imagine novel music without the aid of notation or a musical instrument. Such creative imagery is likely to be fragmentary and exploratory in nature (see Mountain, this volume). Second, a musically literate person may imagine music as he or she is reading an unfamiliar score. This ability varies widely even among professional musicians (see Brodsky, Henik, Rubinstein, & Zorman, 1999), and in most cases the resulting imagery is probably incomplete and at a much slower tempo than the one intended by the composer. Third, a previously heard or imagined piece of music may be recalled from memory, with or without the aid of a score. Although this process also is often fragmentary and discontinuous, it can be carried out at approximately the correct tempo and in a continuous fashion if the memory is strong enough and if attention is devoted to the task. Fourth, a musician may use musical imagery during performance to achieve the desired sound and expression, and to compare the musical image with the immediate feedback received from the instrument. This imagery must of course be continuous and at the correct tempo, though it can be intermittent or absent when performance or listening is on automatic pilot, as it were. An involved listener thoroughly familiar with a piece of music may engage in an analogous process of what Levinson (1997, p. 16) has called ‘vivid anticipation’.
It may be methodologically difficult, however, to distinguish this form of anticipatory imagery from the perception of the simultaneously occurring sounds.

This chapter is concerned mainly with the third of these situations ("playing back" a familiar piece from memory at an appropriate tempo) and to some extent with the fourth (imagery during performance). Moreover, it addresses only one aspect of musical imagery and that is its timing. Specific properties of the imagined musical sounds, such as their vividness, pitch, timbre, or intensity, will not be dealt with here. The question of interest is whether imagined music is (or can be) expressive in the same way that a good performance is expressive, and specifically whether the musical image is (or can be) paced in the same way as an expressive performance.

Expressive timing consists in a continuous modulation of the local tempo of a performance, which is most obvious in compositions from the Romantic period. This modulation has three largely independent aspects: First, there is the mean value around which the modulation occurs, which corresponds roughly to the basic or global tempo. Second, there is the magnitude of the modulation, which may be measured either in absolute terms by a standard deviation or in relative terms by a coefficient of variation (the standard deviation divided by the mean). Third, there is the specific pattern of the modulation, which is the focus of attention in the present research. The expressive timing pattern is constrained but not fully determined by the musical structure, and different patterns are possible for the same music, although a typical timing pattern can usually be identified by analyzing large samples of performances (Repp, 1998a).

To the best of the author's knowledge, there are no previous studies of musical imagery that have examined whether imagined music is expressive, and specifically whether it is expressively timed. In fact, the question may not even have been raised previously. However, several authors have attempted to determine whether music is imagined at the same global tempo as it is performed. Some of these studies compared keyboard performance with and without auditory feedback. Playing on a silent keyboard clearly requires musical imagery, though not necessarily to a greater extent than normal performance does; the imagery may just be more conscious in the absence of heard sound. More than a century ago, Ebhardt (1898) observed that pianists played more slowly when the piano action was separated from the strings, but more recent studies with electronic keyboards by Gates and Bradshaw (1974), Finney (1997), and Repp (1999c) have not found any significant tempo difference. Gabrielson and Lindström (1995) asked musicians to play simple tunes on a keyboard and then to tap rhythmically on a sengraph (a silent pressure-sensitive button) while imagining the tunes. They, too, found only a nonsignificant tendency for the sengraph tapping to be slower than the keyboard performances. Clynnes and Walker (1982) report an experiment in which musicians were asked to repeatedly play or merely imagine ("think") various short pieces. The imagined performances were found to be significantly slower and also more variable in tempo than the actual performances. However, Halpern (1988) observed that the tempi of imagined songs were not significantly slower than the preferred tempi for the same songs in an adjustment task, though she did find a regression to the mean, with slow songs being imagined faster and fast songs being imagined slower than their preferred versions.
In his book, *Sentics: The Touch of the Emotions*, Clynes (1977) reports informal experiments with several famous musicians in which they were asked to tap rhythmically on a sestograph while imagining works by different composers. Analysis of the averaged sestographic pressure curves revealed different shapes for different composers, which lent support to Clynes's theory of a composer's distinctive 'inner pulse'. These findings demonstrate that musical imagery can give rise to characteristic motor kinematics, but they do not imply that the timing of the finger taps imitated the expressive timing patterns of performances; on the contrary, the finger taps were paced by a regular visual signal and thus must have been fairly evenly timed.

One study that came close to addressing the question of expressive timing is the aforementioned one by Gabrielsson and Lindström (1995), in which musicians were asked to play simple tunes on a keyboard and then to tap out the rhythmic patterns of the tunes on a sestograph with the intention of conveying specific emotions. The timing patterns in the two conditions turned out to be similar, which provides evidence that the different emotional performance styles were preserved in the musical imagination, particularly with regard to their timing. However, Gabrielsson and Lindström were mainly concerned with gross differences in tempo and rhythm between contrasting emotional styles, not with the subtle temporal inflections of typical expressive performance.

Of course, one might ask: Why should imagined familiar music not have all the expressive features of a real performance, including its expressive timing? After all, listeners are exposed only to expressive performances, and musicians normally produce only expressive performances, so that one may reasonably assume that this is the format in which music is stored in people's memories. Certainly, musicians must be able to generate expressive intentions from long-term memory in order to play expressively, and experienced listeners presumably derive expectations from a similar memory representation in the process of appreciating the expressive details of a performance. However, psychologists and music theorists commonly assume that what musicians and listeners store in memory is a categorized and schematic musical structure similar to what can be seen in a printed score, and that expressive nuances are generated at the time of performance or listening from implicitly learned rules and conventions that are applied to the stored structure (see, e.g., Clarke, 1985, 1988; Todd, 1985). Such a theoretical division between abstract representation and concrete realization, reminiscent of Chomsky's well-known competence-performance distinction in linguistics, leaves open the theoretical possibility of not applying the expressive rules to the remembered structures and of imagining (or indeed producing) a deadpan performance. Therefore, it may be wrong to assume that imagined music is automatically and necessarily expressive. Rather, the expressiveness of an imagined performance, as of a real performance, may well be under conscious strategic control and may depend on the requirements of a particular task. Thus it makes sense to ask whether an imagined performance even can be as expressively timed as a real performance. The remainder of this chapter summarizes some suggestive findings from the author's research.
Evidence for expressive timing in imagined music

The relevant empirical evidence comes from several recent experiments on timing perception or production that included conditions in which musical imagery was required (Repp, 1998c, 1999b, in press). These experiments all made use of the same musical excerpt, the opening of Chopin's Etude in E major, op. 10, No. 3, which is shown in Figure 1 on the facing page. The final chord, which is not in the original music, was added to give the excerpt maximal closure. The melody, in the highest voice, is divided into a number of segments or rhythmic groups, each ending with a long note, as indicated above the score. An accompaniment in continuous sixteenth-note values is provided by the alto voice. The lower voices, played by the left hand, provide a rhythmic and harmonic underpinning. With the exception of the initial upbeat (an undivided eighth note), there is at least one note onset at every sixteenth-note metrical subdivision, so that the timing of a performance of the excerpt can be described in terms of nominally equal inter-onset intervals (IOIs), as indicated below the score.

A typical performance of the excerpt lasts about 20 seconds and contains large expressive timing modulations. The graph below the musical score represents a typical expressive timing pattern (or timing profile), which was obtained by averaging the timing patterns of performances by 18 advanced student and amateur pianists (Repp, 1999a). The pattern represents the average durations of the IOIs between successive 'primary' tones, defined as the highest-pitched tones in the sixteenth-note positions in the score. (The initial eighth-note upbeat is not included in the graph or in any of the following comparisons and correlations.) This average timing pattern is quite representative of the individual performances in the sample; radical departures from the typical pattern are observed only in performances of some experienced concert pianists (Repp, 1998a). For further discussion of this typical timing pattern, which is also aesthetically pleasing, see Repp (1997, 1998a, 1998b).

The first set of results relating to imagery comes from a study in which 6 skilled pianists were requested to carry out a number of different tasks, each of which was repeated 10 times in immediate succession (Repp, 1999b). Variation across repetitions was small, and therefore the data were averaged over all repetitions, separately for each pianist. The first task was normal expressive performance of the Chopin excerpt on a digital piano (Roland RD-250s). The data were recorded in MIDI format on a Macintosh Quadra 660AV computer using MAX software. Figure 2 on page 190 compares the average expressive timing pattern of each pianist (heavy line) with the typical timing pattern (thin line), previously shown in Figure 1. The correlations between the profiles are shown as well. It is evident that all individual patterns were highly similar to the typical pattern, although there were individual differences in some details and in the magnitude of the timing modulation. One pianist (T.C.), a specialist in 20th century music, unexpectedly produced a rather flat timing profile, and a second pianist (M.S.) also showed somewhat reduced timing modulation, probably due to a rather fast tempo. The other four pianists, however, produced large expressive timing modulations, as expected.

The second task in the experiment again required expressive performance, but with the sound of the digital piano turned off. Although auditory imagery may well be involved in normal expressive performance, it seems more essential in silent perfor-
Figure 1. Top: The opening of Chopin’s Etude in E major, op. 10, No. 3, with melodic-rhythmic groups and nominal inter-onset intervals (IOIs) indicated. Bottom: A typical expressive timing profile. [Reproduced from Repp (1999a: Fig. 1) with permission of The Psychonomic Society.]

mance. Figure 3 on page 191 compares the expressive timing patterns produced with (heavy line) and without (thin line) auditory feedback. They were extremely similar. Only the final IOI was substantially shortened by two pianists. It can also be seen that there were no consistent differences in overall tempo between the two tasks. This result is in agreement with several other studies that have compared keyboard performance with and without auditory feedback (Gates & Bradshaw, 1974; Finney, 1997; for additional detailed analyses of the present data, see Repp, 1999c), and it suggests that auditory imagery can effectively pace expressive performance. However, the pianist’s physical interaction with the keyboard may also contribute to the expressive timing pattern.

The third task of the experiment required the pianists to tap with the index finger on a quiet response key (the ‘enter’ key of the computer keyboard) in synchrony with every sixteenth note of one of their own expressive performances of the Chopin Etude excerpt. However, it is the fourth task which is of particular interest here. Here the
pianists were asked to tap in synchrony with an imagined expressive performance of the excerpt. In this condition, not only was there no sound, but the physical interaction with the piano keyboard was also absent. This, then, was a pure musical imagery condition, and the finger taps reflected the temporal unfolding of the auditory image. Figure 4 on page 192 compares the pianists' tap IOIs (thin line) with the timing profiles of their silent performances in the second task (heavy line). The four pianists who had played with much expressive timing (top and center panels) were able to imagine and tap out their expressive timing patterns quite well, though not with perfect accuracy. By contrast, the two pianists who had played with less expressive timing (bottom panels) tapped almost metronomically. (The overall correlations for these two pianists are misleading because they reflect mainly the lengthened final or initial IOIs; they
are much lower when these IOIs are omitted.) It may be that, even in their expressive performances, these two pianists intended to play metronomically, and that the small timing nuances in their performances were an unintended byproduct of their physical interaction with the piano keyboard (cf. Drake & Palmer, 1993; Penel & Drake, 1998; Repp, 1999b, 1999d). Thus their deadpan musical imagery may not reflect a general inability to imagine music with expression but rather may be a reflection of their expressive intentions in performances of this specific musical excerpt. Of course, there remains also the possibility that these two pianists had somehow misunderstood the instructions. In that case, their results indicate that musical imagery is not automatically expressive. Indeed, the author (B.R.) as a participant felt that the task required considerable attention. It may be easier, after all, to imagine a deadpan performance.
Figure 4. Average timing profiles of six pianists in Tasks 2 (expressive performance without auditory feedback) and 4 (tapping in synchrony with an imagined expressive performance). [Reproduced from Repp (1999b: Fig. 4) with permission of The Helen Dwight Reid Educational Foundation (Heldref Publications).]

In fact, conditions in the second half of the experiment required the participants to do just that (see Repp, 1999b).

Two further tasks from the first half of the same experiment are pertinent to musical imagery. First, as in the third task, the pianists were required to tap their finger in synchrony with an expressive performance of the Chopin excerpt (Task 5). However, here the model performance was computer-generated and had a typical timing pattern (based on performances from nine pianists) very similar to the one shown in Figure 1. Figure 5 on the facing page compares the model timing pattern (heavy line) with the tap timing pattern of each pianist (thin solid line) and also shows the correlations, labeled '5', between these two patterns. After only three practice trials, not included in the figure, all pianists were quite successful in anticipating the timing variations in
the model, even though there was a general tendency to underestimate long IOIs and to overshoot the following IOIs in compensation. What is especially noteworthy is that at least one of the two pianists who did not seem to have intentions or imagery of expressive timing (bottom panels) was as accurate as the others in anticipating what to him must have seemed unusually large timing variations. To the extent that musical imagery was involved in this task, and there seems little doubt that it was, at least as far as timing is concerned, the image presumably was based on a memory of the repeatedly heard model performance, or more likely on a conflation of that memory with the individual pianist’s preferred timing pattern.

The final task of the first half of the experiment (Task 6) was again a pure musical
imagery task. Here the pianists had to synchronize their finger taps with a series of 'clicks' (actually, very high-pitched digital piano tones) while imagining the music in synchrony with the clicks. The clicks followed exactly the same expressive timing pattern as the music in the preceding task. As can be seen in Figure 5 (dotted line, '(6)’ correlations), the pianists were about as accurate in this task as in the preceding one. Thus they were able to anticipate the timing pattern of the clicks to a considerable extent by generating an appropriately timed musical image from memory. It seems unlikely that they would have been able to maintain this level of synchronization accuracy without imagining the music.

A more recent synchronization experiment using the same musical excerpt (Repp, in press) included a baseline condition in which click sequences were presented initially without musical imagery. Four different timing patterns were presented: three very different patterns (T1, T2, T4) derived from an analysis of a large sample of expert performances (Repp, 1998a), and a random pattern (R1), generated by scrambling the IOIs of the T1 pattern. Twelve musically trained undergraduate students, mostly string instrument players, participated. They tried to tap in synchrony with the click sequences instantiating the different timing patterns, without having been told about the music. There were 10 successive trials for each pattern. Subsequently, the participants tapped in synchrony with the same click sequences, which now were superimposed on the identically timed Chopin excerpt. The presence of the music was expected to facilitate synchronization with the musical patterns, but not with the random pattern. Finally, the participants again tapped to the click sequences without hearing the music, but with instructions to imagine the music in synchrony with the clicks.

The results are summarized in Figure 6 on the next page in terms of an index of anticipation accuracy, R0*, which ranges from 0 to 1. The index represents the correlation between the model timing pattern and the tap timing pattern, corrected for a hypothetical minimum correlation that would obtain if the taps tracked rather than anticipated the model pattern (see Michon, 1967). Somewhat surprisingly, the musical T2 pattern was more difficult to synchronize with than the random pattern, R1. However, as predicted, anticipation accuracy improved substantially for all three musical patterns when the music was added to the clicks, whereas the R1 pattern showed only a small improvement, probably due to general task practice. Moreover, the level of performance achieved in the music condition was essentially maintained in the imagery condition, at least for the T1 and T4 patterns. These results indicate that synchronization accuracy is similar when music is present and when it is merely imagined, and they also demonstrate that synchronization with at least some musical timing patterns is facilitated in both conditions, relative to a condition in which music is neither heard nor imagined. (Unfortunately, a subsequent replication of this experiment with a modified design did not yield clear evidence of imagery; see Repp in press.)

The same experiment also included completely isochronous stimulus sequences with IOIs of 500 ms. Even though they served mainly as a practice condition, they yielded another very interesting result pertaining to imagery, shown in Figure 7 on page 196. The constant IOIs of the stimulus sequence are represented by the dotted horizontal line in each panel. The data points with double standard-error bars represent the average IOIs between the participants’ synchronized finger taps. In the
Figure 6. Average indices of anticipation accuracy ($R^2$) for four different timing patterns in three conditions: clicks only, clicks accompanied by music, and clicks accompanied by musical imagery. [Data from Repp (in press).]

top panel are the results for clicks without music. After an initial 'tuning in' to the sequence tempo, the tap IOIs did not deviate significantly from 500 ms, as expected. The center panel shows the results for isochronous clicks accompanied by isochronous music. Here, as observed in several earlier studies in which participants had tapped in synchrony with isochronous music, albeit without superimposed clicks (Repp, 1999a, 1999b, 1999d), there was a systematic pattern of deviations from isochrony in the tap IOIs. That pattern bears some relation to the musical structure and to the typical expressive timing pattern (Fig. 1); it is believed to represent a combination of typical expressive tendencies and automatic error correction processes (see, e.g., Pressing, 1998). The result of principal interest here is shown in the bottom panel of the figure. This is the condition in which the participants tapped in synchrony with an isochronous click sequence while only imagining the music. Here, too, significant deviations from regularity occurred in the tap timing. The deviations were smaller but extremely similar in pattern to those produced when the music was heard ($r = .91$, or .84 if the initial three data points are omitted). These results were replicated in the second experiment of Repp (in press). Thus, musical imagery can not only facilitate synchronization with appropriate expressive timing patterns, but it can also interfere with synchronization with a mechanically regular sequence by introducing involuntary and subconscious timing modulations in the motor activity.

Finally, one result should be mentioned that failed to show an effect of musical imagery. The author has conducted many perceptual studies in which participants
Figure 7. Average tap timing profiles (with double standard errors) for synchronization with an isochronous timing pattern in three conditions: (a) clicks only; (b) clicks accompanied by music; and (c) clicks accompanied by musical imagery. [Data from Repp (in press).]
were required to detect small hesitations (lengthenings of single IOIs) in an otherwise isochronous rendition of a musical excerpt (Repp, 1992, 1998b, 1998c, 1998d, 1999a, 1999d). The detectability of these timing perturbations always varied greatly with position in the music, as shown for the Chopin Etude excerpt by the filled circles and solid line in Figure 8 on the next page. These variations were found to be strongly related to the typical expressive timing pattern for the music (Fig. 1), such that a slight lengthening of an IOI was difficult to detect when that IOI tended to be lengthened in typical expressive performances. In one of these earlier experiments (Repp, 1998c: Exp. 2), participants were given the task of detecting similar deviations from isochrony in a sequence of clicks while imagining the music in synchrony with the clicks. Each click sequence was immediately preceded by an expressively timed rendition of the Chopin excerpt (with a typical timing pattern similar to that shown in Fig. 1), so that the participants heard the music many times in the course of the experiment. They also read along in the score as they imagined the music and marked their detection responses in the score. The results are shown as the open circles and dotted line in Figure 8. Even though there was unexpectedly large variation in the detectability of timing perturbations across positions in the click sequence, the pattern of this variation did not at all resemble the pattern obtained with music, apart from the poor detection scores in the initial and final positions which may be attributed to psychophysical causes. Thus, musical imagery did not seem to affect timing perception, at least not in a structurally specific way. By contrast, results very similar to those for the music alone were obtained when the isochronous click sequence was superimposed on the Chopin excerpt, even though participants had been instructed to focus on the clicks and ignore the music (Repp, 1998c: Exp. 3).

Conclusions

The results summarized here show that music can be imagined as expressively timed. However, they do not prove that music is always or necessarily imagined in that way. The author’s introspections and the finding that two out of six pianists tapped metronomically when synchronizing with imagined music (Fig. 4) suggest that expressive timing is optional and possibly a luxury in the mind’s ear. To achieve expressive timing, it may be necessary to imagine not just the sound of the music but the bodily activity of performing it as well. This may require extra effort. However, it may also be the case that the ongoing physical activity (the finger tapping), which is naturally inclined towards isochrony, inhibits the imagination of musically relevant performance movements. Perhaps the imagination of expressive timing is easier without finger tapping, but then a different measure of its occurrence would have to be found.

Clearly, the imagination of expressive timing is under conscious control, at least to the same extent as expressive timing in performance. However, musical imagery apparently also has consequences for timing that are below the level of awareness. As we have seen, imagining music without expressive timing while tapping in synchrony with a metronome leads to small but systematic variations in tap timing, similar to the variation found when isochronous music is actually heard in synchrony with the metronome. One might then also predict that tapping directly in synchrony with an
Figure 8. Percent correct hits as a function of sequence position in a task requiring detection of small local increments in IOI duration in an isochronous sequence, in two conditions: music, and clicks with imagined music. [Reproduced with slight modifications from Repp (1998c: Fig. 7) with permission of Springer-Verlag.]

imagined metronomic music performance, without an auditory metronome, would reveal a similar pattern of timing variations in the finger taps. Repp’s (1999b) study actually included this task as well, but the results were not so clear. This calls for further research.

So far, all the positive evidence for effects of musical imagery on timing comes from the timing of concomitant motor activity. There are no results so far suggesting that musical imagery interacts with the perception of the timing of simultaneous auditory input. Although the imagination of musical sounds is a form of internal perception, the pacing of that imagination may be viewed as a form of internal action—a performance on the mind’s synthesizer, as it were. This is a potentially rich area for further exploration by cognitive psychologists.

Acknowledgements

The author’s research and preparation of this chapter were supported by NIH grant MH-51230. Thanks to Amandine Penel for helpful comments on the manuscript.
References


Repp, B. H. (in press). The embodiment of musical structure: Effects of musical context on sensimotor synchronization with complex timing patterns. In W. Prinz & B. Hommel (Eds.), *Attention and...*