Music in the Mind

An Introduction to the Psychology of Music

by

SIU-LAN TAN, ROM HARRÉ AND PETER Q. PFORDRESHER

Chapter Eleven

THE PSYCHOLOGY OF MUSICAL PERFORMANCE: FROM INTENTION TO EXPRESSION AND BACK AGAIN

> *After we know the notes, then we have to read between the notes, something which is not printed, something which is not there. That is when your process starts as an interpreter, when you can show what you understand in the piece, what you feel, what makes a good interpretation…"*
> *(Concert pianist Rudolf Firkusny, quoted in Noyle, 1987, p. 85)*

Arguably the most common way a musical performance is conceptualized goes something as follows: In a large concert hall, the lights grow dim and various quiet conversations and rustling of programs quickly diminishes to near silence. The performer strides across the stage to the sound of spirited but restrained applause. She takes her seat at the piano, lifts up her hands in a dramatic gesture, and then plunges them toward the piano keys to begin an instrumental performance. The music is emotional, yet subtle. The tempo varies slightly to yield just the amount of emotional expressivity that is characteristic of the historical period of the piece. The audience sits silently and attentively, almost motionless until the end of the performance, whereupon they erupt into another round of enthusiastic yet polite applause. At a few pauses prior to the ending a few naïve audience members begin to applaud, but cease quickly in response to the disapproving looks of others.

Both historically and anthropologically the account above is an exception to the rule. Music performance is more commonly a collective activity, shared by all persons who are simultaneously perceivers and producers of music. Most music of the world is also an invitation for participants to dance. One immediately thinks of the use of music in traditional cultures in which music performance serves the purpose of bringing together a village or army. But one need not stop there. Compare the example listed above to a typical rock and roll concert, in which not only standing up and moving to the music, but singing along with the musicians is necessary for the performance to be considered a 'success'. Consider also more intimate uses of music performance. The musicologist

Ellen Dissanayake (2000) has proposed that the most important use of music occurs during interactions between mother and child when mothers sing lullabies (as we discussed in Chapter Eight).

For these reasons, the current chapter attempts to address music performance in all its guises, expert and novice, collective and solo, ad hoc and pre-planned. Nevertheless, because the literature to date has typically followed the model of solo, expert, concert performance you will find that much of the research on performance as a whole assumes this model.

Why performance?

Modern introductory psychology text books typically do not include a chapter concerning action. Although the study of action has a long history in psychology, dating back to the ideo-motor theory of William James (1890), action has always received far less scrutiny than the study of perception. In a recent article, the psychologist David Rosenbaum (2005) suggested that this prioritization of perception stems from the philosophical perspective of empiricism that drives most of psychology. Put simply, given that we presumably gain knowledge from what our senses extract from the world around us, how better to know the workings of the mind than to study how that information enters? However, this simplified view ignores a likely corollary, put forth by Piaget among others (see Chapters Seven and Eight) that we may gain much of our knowledge from acting on the world. Philosophical underpinnings aside, it is demonstrably true (by any casual survey of psychology’s major journals) that research on perception far outweighs research on action.

Music psychology mirrors psychology as a whole, given that the training of most music psychologists has been that of cognitive psychology (true to some extent even for music theorists within music cognition). Thus the flagship journal of music psychology, for instance, is called *Music Perception*, which occasionally features an article or two on performance. Thus, one may ask, of what use is the study of performance? An assumption one may make, proposed in the same article by Rosenbaum, is that performance is nothing more than the execution of mental ideas, and that although the study of this execution may be of interest to kinesologists, the mere coordination of muscles leaves little of interest to the cognitive scientist, less even to those interested in socio-cultural issues.

There are several reasons to place the study of performance in a more central role. First, musical experiences are a result of performance. People occasionally do compose melodies in their minds through musical imagery (composers in particular), but for most of us even a melody that is ‘stuck in one’s head’ originates in a performance. Musical experience thus relies on some kind of episodic experience of a performance. Second, performance, involves more than ‘just’ remembering a melody. It involves a certain kind of memory, one that involves the transformation of perceptual data into motor commands. Moreover, as will be explained below, it appears that the kind of memory recall used in performance is of a special variety, possibly different from the kind of recall that goes on when we engage in mental imagery for music. Third, music performance involves the acquisition of skill. Though the degree to which music performance requires ‘expertise’ may be exaggerated, performing music always involves very fine-grained motor coordination and many persons spend years acquiring the kind of motor precision needed to perform, resulting in re-organization of the brain (see also Chapter Ten).

This chapter describes one iteration within a continuously recurring cycle that must occur as an individual performs. The genesis of a performance lies in the acquisition of musical information, from long term memory, from musical notation, or generated extemporaneously as one composes or improvises. Thus the first section will consider how such musical ‘intentions’ are generated. Then muscles must be controlled and guided in a way that realizes these intentions such that intentions are communicated to the listener in a coherent fashion. The same intention may be expressed through vastly different forms of movement, such as the difference between playing a melody on the keyboard versus singing it. Finally, the performer must monitor what he or she has done, possibly comparing this result with that of surrounding musicians, to determine whether the musical results of actions match the intended outcome.

For background to this chapter, we drew on other recent reviews of music performance, specifically those written by Alf Gabrielsson (1999, 2003) and Caroline

Palmer (1997). These may be consulted for more in-depth accounts of the literature we discuss.

The retrieval of musical ideas

Music performance is a kind of memory task. The actions that give rise to musical events relate to a string of musical ideas that lie within the performer's mind. That being said, the nature of memory demands can vary widely. The prototypical classical soloist described at the beginning of the chapter has arguably the most onerous memory task. Such performers usually memorize long and complex pieces from notation over hours of practice. Likewise, some conductors are known for possessing vast stores of musical memories. The conductor Toscanini was famously reported to inform a bassoonist with a broken key on his instrument that the key would not be used in that evening's concert, drawing on a precise memory not only for the music of that evening but for the relationship between the music and the instrument for a single instrument in the orchestra. However, any musical performance involves some kind of memory.

Different forms and uses of memory

The structure of memory constitutes perhaps the most central topic to cognitive psychology. A tremendous amount of detailed research on the structure and function of memory has been conducted over the years. Unfortunately, the proliferation of theories and approaches has led to a confusing picture of the way memory works. Suffice it to say that memories can be stored over the long term or the short term, can be consciously accessible at the current moment or never consciously accessible.

The classical performer described above likely engages in what would be described as *procedural long-term memory*. Procedural memory involves memories that are inextricably linked to motor actions, and are not consciously accessible. An example related to piano performance is touch typing. If you are able to touch type, try to recall what finger is used to press the 'e' key. It's likely that this task was more difficult for you than typing the word 'the', and more to the point it's likely that you recalled this letter by imaging your fingers striking the keys. Similarly, performers (of the piano or otherwise) do not need to know consciously the identity of each note they perform, and may even disengage conscious thought from the performance, a possibility supported by recent evidence that the frontal lobes effectively ‘shut down’ during performances from memory (Parsons, Sergent, Hodges, & Fox, 2005), as well as evidence from the domain of sport, that skilled performers do better when they do not ruminate consciously on the task at hand (e.g., Beilock, Bertenthal, McCoy, & Carr, 2004). More recent data suggest that when performers hear music that they know how to perform, the brain shows activity in areas related to motor control (Lahav, Saltzman, & Schlaug, 2006). However, other music does not activate these areas, even when the music consists of the same pitches presented in a different order. It should be noted that performers of music not associated with notation (e.g., rock musicians) still draw on a long-term procedural representation; all that differs is the initial form of the information to be memorized. The example of Toscanini is somewhat different in that the memory is *explicit* rather than procedural. Conductors rely on conscious awareness of the different instruments that come together to create music, and in this way may differ from performers.

At this point you may be wondering about two other examples: performers who read from notation, and performers who improvise. Both use memory, albeit of different types than our first examples. Performances from notation rely on what might be called *short term working memory*. Although the musical notes are presented in front of them, musicians must maintain this information in consciousness for some period of time prior to their performance, and may even maintain the information in memory for some time afterwards (as discussed later). More to the point, classic research on the role of eye movements suggests that performers do not look at the notes they are immediately about to play, but instead read music in an anticipatory manner, attending instead to notes in the future (e.g., Sloboda, 1974). Thus, at any point in time, the performer probably produces notes that have persisted in memory for at least a few milliseconds and that are not providing immediate sensory data.

Improvised performances provide the least obvious connection with memory, in that these performances, at first glance, are completely spontaneous. This problem is compounded by the fact that the cognitive bases of improvisation are not well understood. However, anecdotal observations and analyses of improvised solos do suggest some use of memory-based processes. First, any practicing jazz musician will tend to disabuse the starry-eyed theorist of the notion that jazz solos are entirely spontaneous, by the simple fact that jazz musicians *practice* certain characteristics of solos. Although no two solos are different, solos tend to result from permutations of musical motifs that are well practiced. In this way, jazz solos mirror language. Although most statements are completely creative when taken as a whole, they typically consist of shorter phrases that we speak quite often. A second way in which jazz solos are memorial stems from the fact that jazz solos incorporate notes that abide by a certain rule system, referred to as a *grammar,* that fits within the genre. Jazz musicians thus rely on *schematic* information, which is commonly considered to reside in memory (cf. Johnson-Laird, 2002; Toiviainen, 1995).

We now turn to different ideas about how musicians use memory. Here we discuss research on the use of working memory during performance.

Working memory and planning of serial order

One of the most influential papers concerning the cognitive bases of production was a paper by Karl Lashley entitled “the problem of serial order in behavior” (1951). At the time, Lashley was lashing out against behaviorists (e.g., B.F. Skinner) who had suggested that retrieval during sequence production resulted from a *serial chain* of events in memory. Serial chaining approaches suggest that people associate event #1 (e.g., the note C) with event #2 (the note D) and so on when retrieving events during a performance. Lashley claimed that such theories are untenable. First, people can produce sequences much more quickly than a chaining approach would predict. Pianists can produce sequences of notes in which the time interval between notes is as rapid as 16 notes per second (about 62 ms per note, Palmer, 1997); the timing of neural events, given refractory periods in neural impulses, could not support such rapidity.

Lashley’s second point is more germane to this chapter. He also pointed to a certain kind of error known as the *serial ordering error,* as evidence against serial chaining. A serial ordering error is one in which people misplace the order of a particular event while maintaining the appropriate content. Here is an example from speech, heard once by one of your authors (PQP):

Spoken:
“We are here to celebrate the wedding of Kelley and Sheith…”
Intended:
“We are here to celebrate the wedding of Shelley and Keith…”

In this instance, the priest at a wedding swapped the position of two speech sounds (phonemes), the “Sh” sound from “Shelley”, and the “K” sound from “Keith”. Thus in the spoken example the priest produced all the correct phonemes but misplaced the order of two phonemes.

Why do serial ordering errors contradict serial chaining approaches? The reason is simple: serial ordering errors usually span multiple positions. Take the example above. The sounds that get swapped do not come from adjacent phonemes. Instead the sounds travel a distance of approximately 7 phonemes (?– e – l –y – a – n – d - ?). Instead of serial chaining, the error above (and other errors like it) show how speakers relate sounds to the structure of language. It is significant that the priest swapped two sounds that were associated with the beginnings of proper names. The psycholinguist Stephanie Shattuck-Hufnagel (1979) suggests that when we plan speech sequences we use a *slot-filler* mechanism, whereby the appropriate content (e.g., phonemes) is applied to slots associated with the structure of language.

Serial ordering errors occur in music performance as well. Research to date has focused on piano performance and it appears as though ‘slips of the finger’ do not merely result from hitting a nearby key, suggesting just sloppy playing. As in speech, slips of the finger often originate from note events that are many positions away. Just like errors in speech, musical errors seem to reflect mental associations of content (the notes) with the global structure of music. One difference between speech and music is that musical errors rarely ‘trade’ places as in the example above (referred to as an ‘exchange’ error). Instead a misplaced musical event is typically repeated in its correct location, as in the following example:

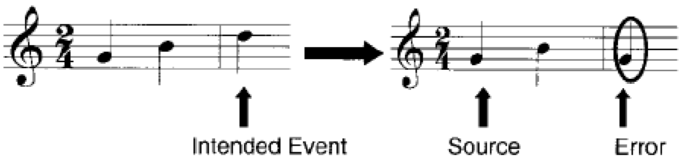

Figure 1 (Source: Palmer & Pfordresher, 2003)

As can be seen, this example shows a situation in which the performer wrongfully repeats the G note from position 1 when they reach position 3. The error event thus travels a distance of 2 positions (from the target to error location). Such an error is also referred to as a *perseveratory* error, as opposed to *anticipatory* errors that involve playing a note in advance of its correct location.

Some of the pioneering research on this topic (Palmer & van de Sande, 1993, 1995) suggested that serial ordering errors in music are constrained by the structure of music. That is, the degree to which notes ‘travel’ to incorrect locations in production depends on how the performer conceptualizes the structure of the music. For instance, Palmer and van de Sande (1993) found that in piano performances for which each hand plays an independent melodic line (i.e., ‘polyphonic’ music), notes from within the right hand’s melody tend not to replace notes from within the left hand’s melody. In later work Palmer and van de Sande (1995) demonstrated that errors only travel within a single musical phrase (analogous to a clause in speech). The implication of both findings is that at any point in time, the performer has on her mind a certain subset of the notes in a melody – for instance, the notes within a phrase – and that notes may be further encoded with respect to a particular melodic line out of which they do not travel. The relationships may be conceptualized as follows:

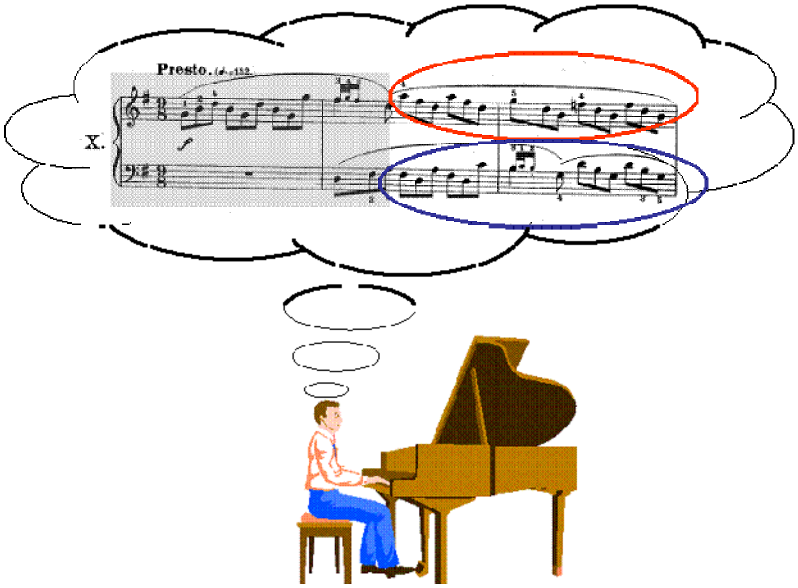

Figure 2

This figure shows a pianist engaged in memory retrieval during the performance of Bach's 10th 2-part invention. The pianist is currently producing the second phrase, hence phrase #1 is shown in grey. Within the 2nd phrase the two parts are conceptualized as separate and are thus shown within different ovals.

A broader implication of these findings is that although a musician must have the entire piece stored in memory, the process of retrieval during performance does not result in the performer retrieving the entire piece. Nor does the performer retrieve the piece one note (or one chord) at a time, as might follow from a serial chaining approach. Instead, memory retrieval is *incremental* in nature. At a given point in time the performer holds some, but not all, notes from a performance in working memory (cf. Baddeley, 1986; Miller, 1956). The notion of incrementality, like analyses of serial ordering errors, originated in speech production research as a way of accounting for the fact that people seem to plan future utterances (e.g., words) as they continue to produce the current utterance (e.g., Kempen & Hoenkamp, 1987).

At this point, you may have noticed that the depiction shown in Figure 2 suggests that the performer may have equal accessibility to all the notes within the second phrase, that these notes come as a block or 'chunk.' Indeed working memory theories have often invoked the idea that items or events may all be treated as a common 'chunk' of information. However, such a proposal seems untenable as an account for how people retrieve events in a sequence. It would predict that performers produce any note from within a phrase with equal probability. Such a situation would lead to very low accuracy, and performers are typically highly accurate (less than 5% of all performed notes are usually errors (e.g., Palmer & Pfordresher, 2003). Furthermore, serial ordering errors often reveal some notes within a phrase to be more likely error sources than others. Thus, an incremental approach has to account for why some notes within a phrase are more accessible than others.

Such an account has been developed in recent years by one of us (PQP) and Caroline Palmer (Palmer & Pfordresher, 2000, 2003; Pfordresher, Palmer, & Jungers, 2007). We have developed a model we refer to as the 'range' model because it accounts for the range of events that are accessible in working memory as one plans and produces a melody. It also accounts for degrees of accessibility across events. Although the model was described mathematically in the original papers, a verbal description will suffice here. First, the model presupposes that one maintains some knowledge of the current position. The model is also limited in that it does not account for the specific notes associated with positions but rather simply accounts for how mentally accessible each position is (in the memory jargon, the model includes 'order' but not 'item' information).

Within these constraints, the range model proposes that accessibility of notes within a phrase is governed by two components. First, notes that are closer to your current position are more accessible than notes that are far away. For instance, if the current note is note #5, then notes at positions 6 and 4 will be more accessible than notes at positions 8 and 2. Second, the model predicts that notes are more accessible if they are associated with a similar 'metrical accent' to other notes. Thus, when one plans to produce a note or chord on a 'weak' beat, notes and chords on other weak beats are more accessible in the brain than are notes and chords on strong beats. The predictions of the model can be illustrated using the following example:

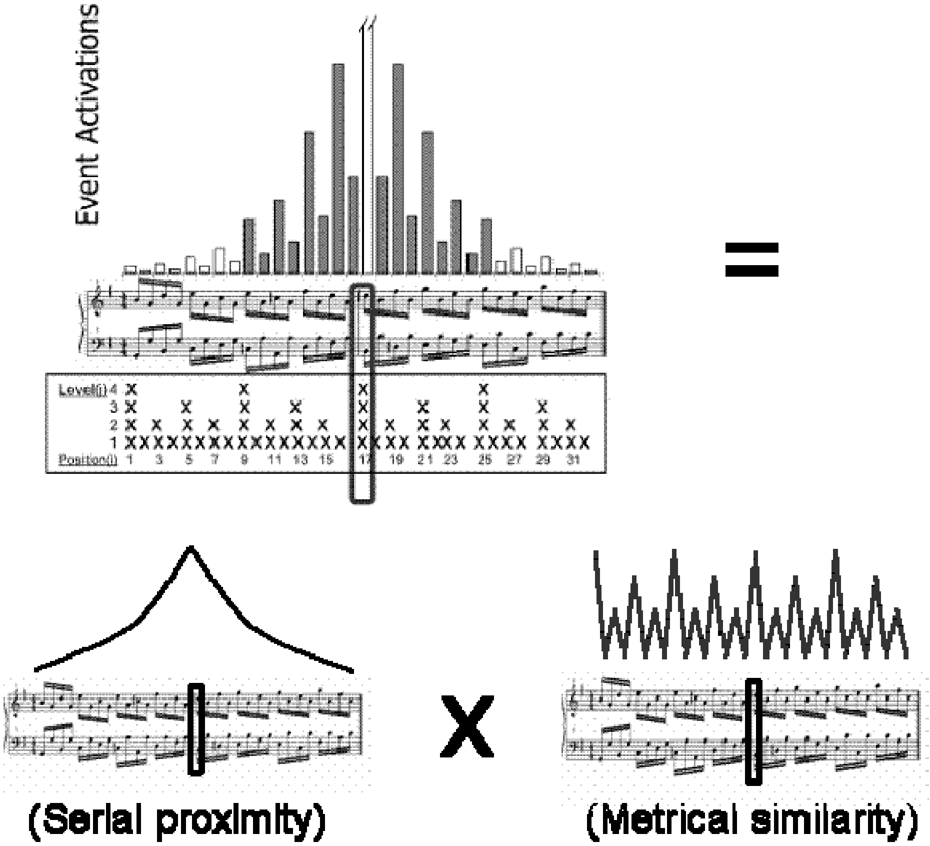

Figure 3 (Adapted from Pfordresher et al., 2007)

The top panel illustrates what the performer may have on her mind during a performance (an elaboration on the callout shown in Figure 2). The grid of x's below the music illustrates the pattern of metrical accents at each position (more x's = more strongly accented). The current position (highlighted by the rectangle superimposed on the notation) is a strong metrical position. The bars in the graph above the notation illustrate how accessible or active each note is – how likely it is that each note will be performed. The current note is most highly active, which relates to the fact that performers usually produce the correct note (the bar extends above the plot). However, other positions are also active by virtue of both their proximity to the current event, and how similar their position is to the current position with respect to metrical accent. What this means for the performer is that if he or she produces an error when performing the D#-note in the right hand (at the highlighted position), that error is more likely to be F# (2 events in the future) or D-natural (2 events in the past) than it is to be either B or A (1 event in the future or past, respectively). These predictions have been tested and verified across many studies of piano performance from memory or from notation, at different tempos (speeds), and for both adult and child pianists (reviewed in Palmer & Pfordresher, 2003).

How do these predicted error patterns reflect the two components mentioned earlier? The two components are shown separately in the lower plots; their product generates model predictions. As can be seen, the way in which ‘event activations’ taper off as events grow increasingly distant from the current position stems from the proximity of notes from the performer’s current serial position. Second, the predicted pattern of errors has a jagged, alternating form that results from the metrical similarity component.

The research on the role of working memory in performance is quite recent, and many questions remain. Two particularly compelling issues relate to how well these findings generalize to performance of instruments other than the piano and to singing, as well as how people plan sequences during improvisation. We now consider what happens after a performer retrieves the notes that make up music and puts those notes into action. How do performers control the actions that are used to bring a performance to life?

Controlling movements during performance

Music performance is most obviously a form of motor control. Memory feats aside, it is the fine-grained control of rapid movements that most commonly draw the respect of people for expert performers. We will consider movements from two perspectives. First, we will consider the basic issue of how it is that performers can control movements so that the produced music matches a communicative intent. This issue will be addressed with respect to the way in which pianists control the timing and velocity of key presses, how performers of stringed instruments control hand position, and how singers control the length and tension of vocal folds to generate pitch. Second, we will consider the way in which movements are modulated to give the performance aesthetically appealing ‘expressive’ nuances. There are of course many other ways in which performers augment the performance for expressive purposes, such as variations in loudness and timbre. We focus on the qualities listed above both because they represent the dominant focus in the literature, and because they provide a set of core principles for performance that likely apply across various expressive devices.

Timing of actions

Minimally, musical events (notes, chords) must satisfy two criteria: they must occur at the correct time and they must be correct in content. The former criterion relates to the time at which actions are initiated and/or a particular endpoint of the action is reached. The latter criterion can be realized in different ways depending on the type of performance. In piano performance, demands on ‘tuning’ are eliminated due to the fixed nature of pitches. In many other instruments, however, fine-grained control is necessary to maintain tuning. Violinists, for instance, must position fingers of the left hand very precisely on the finger board (without the guidance of frets that one sees on the guitar) to maintain tuning. Singers face an even more daunting task, in having to maintain both a certain tension and length of the vocal folds (also called ‘vocal cords’) – a muscle that is invisible unless one uses very expensive equipment (a laryngoscope, also very uncomfortable to use). These are the criteria that separate ‘good’ from ‘bad’ musicians in the judgment of the general population, although it is not entirely clear that people are quite as ‘bad’ as many in western cultures presume, or that most people we hear on the radio are quite as ‘good’ as we think they are. In both cases, as in so much else, literally nobody is prefect.

Maintaining regularity in timing. There are many ways to conceptualize timing of musical events. The most intuitive is duration: the time that elapses between the beginning of an event and its end. In the piano, duration is of little importance because the loudness of each sound decays rapidly after the key is pressed, though on other instruments duration can be of great importance. Nevertheless, duration on its own does little to convey rhythm. A better cue for rhythm is the so-called *inter-onset interval,* which is the time that elapses between successive note beginnings. A major task of the performer is to produce inter-onset timing in a way that successfully maintains the sense of a regular underlying ‘beat’, which gives the listener a regular sense of how rapidly musical events are proceeding (the ‘tempo’). How and how well do performers do this?

Imagine a melody comprising a sequence of notes with equivalent durations, and no expressive nuances to the timing of these notes. Certainly this would constitute a boring piece. Nevertheless an ‘ideal’ performance would entail equivalent, *isochronous* timing. Oddly enough, nobody can do this precisely. Inter-onset timing always varies from event to event. Good performers may minimize this variability, but no human can eliminate it. Even machines typically have some variability (Levitin, personal communication).

Why do these inaccuracies occur? The dominant view at present comes from the work of the psychologists Alan Wing and Alan Kristofferson (1973). Their original research addressed isochronous finger tapping (tapping the same beat repeatedly), but is applicable to piano performance as well. Their model, referred to as the ‘two-level timing model’ suggests that timing inaccuracies are the sum of two components, and both are inaccurate. The first component is an internal time-keeper, or ‘clock’. This clock’s job is to maintain a certain time interval. How it accomplishes this task isn’t clear; some have suggested that the internal clock counts internal, randomly generated pulses (e.g., Kileen & Weiss, 1987), others suggest that a regular internal oscillator is used (e.g., Large & Jones, 1999). Either way the brain somehow keeps track of time. Because nobody is perfect, the internal clock is inaccurate. In addition, the *production* of time intervals requires the use of muscles, and the precision with which we use muscles also leads to inaccuracies.

The nature of these inaccuracies is worth mentioning. The two-level model is referred to as a *stochastic* models, meaning that both components vary randomly in a way that forms a bell-shaped (or *Gaussian*) distribution. Take the internal clock. Assume one attempts to tap a beat in which each tap is 500ms away from the other. If so, the model would genera clock ticks centered around the desired500ms interval, but with some variability. Plotting the distribution of all clock intervals would look like this:

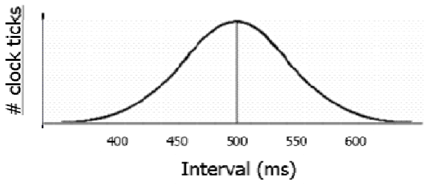

Figure 4

According to this framework, the most frequent inter-tick interval is the correct interval (500ms). However, more often than not, the clock is incorrect, though not by a lot. In the example above, for instance, the majority of ticks fall within +/- 50 ms of the correct interval and only rarely do intervals overshoot or undershoot by 100ms or more. Of course this is just a single example. A more variable clock could result in inaccuracies that extend to a couple hundred milliseconds ore more. A less variable clock could result in inaccuracies ranging only a few milliseconds.

Typically, musical training focuses on reducing variability at both the clock and peripheral levels. Percussionists arguably focus the most on reducing internal clock variability in addition to peripheral variability, whereas other instrumentalists focus more on peripheral variability. One intriguing prediction from this model is that when a musician switches to a new instrument that requires the use of different muscles (e.g., clarinet to drums), internal clock variability should remain consistent but peripheral variability should increase.

This model has been very influential and is disarmingly simple. However, simplicity can come with a cost and there are at least two respects in which this model's limitations may be too great.

One weakness of this model is that it implies that musicians are similarly accurate when producing any time intervals. In contrast to this claim, much evidence suggests that people are most accurate at producing time intervals within the range of 200 and 1,000 ms. This finding comes from the seminal work of the psychologist Paul Fraisse (e.g., 1982), who spent much of his time asking participants to tap any interval that felt comfortable. Although this 'task' may initially sound ludicrous, people are surprisingly consistent from one day to the next if you ask them to tap in this manner. In general, people seem to favor time intervals around 600 ms, although a good deal of variability exists across individuals and older individuals may favor longer (slower) time intervals than do young people (e.g., McAuley, Jones, Holub, Johnston, & Miller, 2006). Despite these individual variations, there does seem to be a consistent tendency for people to have great trouble producing intervals that are very rapid (e.g., grade notes, trills) in a rhythmic way, and the timing of these nuances do not seem to vary with tempo (a hallmark of rhythmic organization) as do the durations of musical notes in general (Desain & Honig, 1994). In addition, time intervals of greater than 1 second are very difficult to estimate and produce accurately without mentally subdividing the longer interval into a series of shorter intervals (like counting “one-one-thousand, two one-thousand…”,). In general, the mental organization of time appears to be something like this:

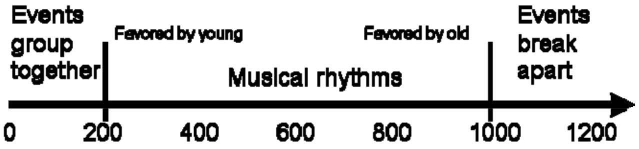

Figure 5

Mixed support has arisen for the view that timing in piano performance is not consistent across tempos. Repp (1994, 1997) found some evidence for consistency across tempos, whereas MacKenzie and Van Eerd (1990) showed that even for the production of simple scales, fine-grained aspects of timing vary across tempi. Also, performers seemed to play slower tempi too fast and play faster tempi too slow, suggesting assimilation to a favored tempo (although their tempi were much faster than 600ms per interval).

A second critique of the two-level model focuses on the independence of the central timekeeper and muscle movements. In the context of music, this claim suggests that the timing of singing is driven by the same ‘clock’ as the timing of piano playing. In fact, recent evidence suggests that different motor tasks may not be linked to the same internal timekeeper (Robertson et al, 1999; Zelaznick, Spencer, & Ivry, 2002). However, this research focused on tapping versus drawing, not on musical tasks. Although in both cases the tasks were performed rhythmically (in time with a metronome), it may be that greater commonality arises when the goal of a task is to produce musical rhythms. More recent evidence suggests that pianists may conceptualize the timing of key presses (notes) in a way that is independent of the biomechanical constraints regulating movements (Loehr & Palmer, in press), which is consistent with the two-level framework.

Lawful variations in timing: rhythms. Although everybody has an intuitive sense of what the term ‘rhythm’ means, it ends up being a difficult concept to define rigorously. We all have the sense that rhythms are ‘regular’ whereas ‘arrhythmias’ are ‘irregular.’ Certainly this definition works for the case of isochrony, discussed in the previous section. However, inter-onset times in music are often at the same time variable but not arrhythmic. In fact, some research defines ‘rhythm’ as a certain kind of variability in timing that still maintains some kind of regularity. How do performers realize this kind of structure?

The seminal work in this area has been done by Dirk Povel and colleagues in the Netherlands and Mari Riess Jones and colleagues in the US. Two characteristics seem to govern the degree to which varying time intervals form a ‘rhythm’. First, adjacent intervals are heard as rhythmic to the extent that their intervals form simple ratios. Here are examples of intervals that do and do not form such ratios.

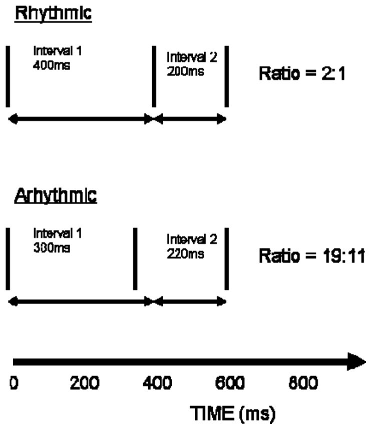

Figure 6

The second characteristic is more complex. As people listen to sequences of time intervals, they begin to extract a sense of regular ‘patterning’ in which subsets of the time intervals form groups. In this way, a listener forms the sense of a regular ‘beat’ even for non-isochronous sequences. The ‘beat’ is typically longer than the shortest interval present but shorter than the longer interval that is present, and there is not surprisingly a tendency for the beat to be around 600ms in length (Parncutt, 1994). Povel & Essens (1985) showed that different orderings of the same sets of durations can lead to simple or complex rhythms, depending on whether the orderings allowed the listener to extract a 'beat.'

With respect to the performer, the ability of performers to reproduce rhythms is determined in part by how complex the rhythm is. A simple rhythm – one comprising proportionally related intervals that form reliable patterns – will be easily learned and reproduced, but not so for a complex rhythm. Take, for instance, the *Dance of the virgins* from Igor Stravinsky's "rite of spring". The music here presents a very clear rhythmic pattern of sorts, but it is hard to extract any sense of regularity from this pattern. As a result, reproducing this rhythm is quite difficult. In fact, Povel (1981) found that when people try to reproduce rhythms consisting of complex ratios (as in Figure 6) their produced rhythms 'drifted' toward the closest simple ratio. Of course, there are also cultural factors at play. The following rhythmic pattern is ubiquitous in western African drumming, and can be easily reproduced there, but is very difficult for most European-descended listeners to learn and reproduce even though all the time intervals are related by a 2:1 ratio with each other.

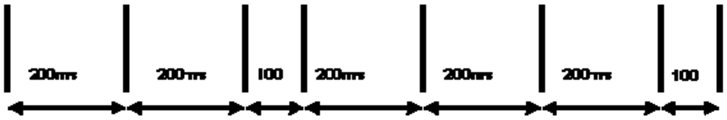

Figure 7

You may have noticed by now that the research summarized above focused mostly on simple tasks like finger tapping rather than on the production of 'real' music. Psychologists use these simplified tasks because they try to reduce behavior to its simplest components. In so doing, psychologists hope to find results that generalize to many complex tasks (e.g., music from both the jazz and classical genres), whereas a task that involved the production of rhythms for a certain standard piece might be limited to that piece. At the same time, limiting one's investigations to such simple tasks brings up the obvious question of whether the results actually do generalize in the intended fashion.

Expressive variations in timing. So-called ‘expressive timing’ probably constitutes the most well-studied aspect of musical performance. Its tradition extends back at least to the detailed analyses of performance carried out at the University of Iowa by Carl Seashore and colleagues (Seashore, 1938/1967). In their lab, performances were analyzed and plotted on ‘performance scores’ that displayed variations in pitch and intensity in real time. Here is an example:

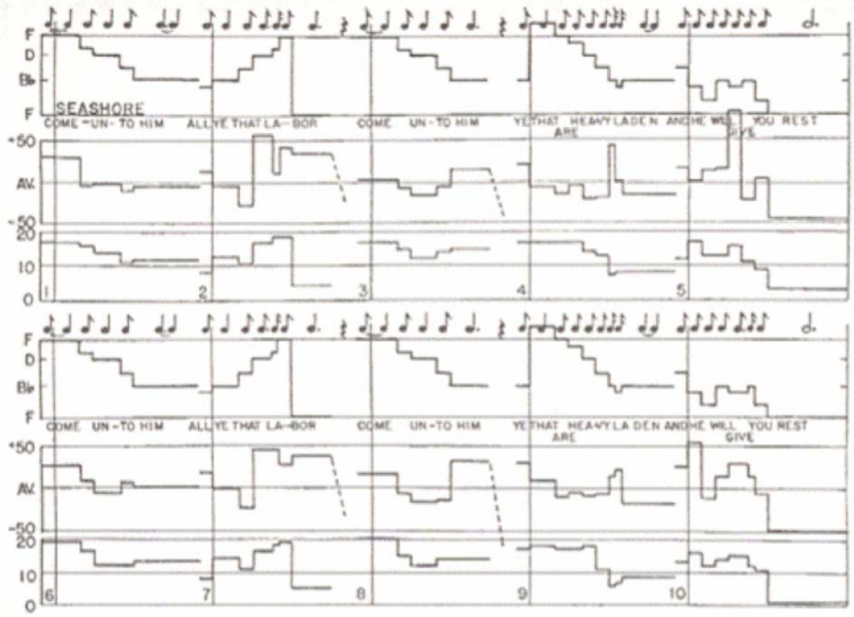

FIGURE 1 Graphical representation of a female singer's performance of the beginning of "Come Unto Him" from *Messiah* by Handel. Top to bottom: pattern of note—length, melody, words, relative duration, tonal power, phrase numbers. (Reproduced from H. G. Seashore, 1937.)

Figure 7

This research revealed that performers systematically deviate from the instructions listed in musical notation, as shown by the middle below each row of musical notes. Here’s how it works. Suppose a performer is playing a melody at a tempo of 100 beats per minute in 4/4 time, which means that each quarter note should be held for 600 ms. A performer may choose to lengthen the note ever so slightly, say to 660 ms, based on her expressive intentions, or may choose to shorten it. It is important to note that these deviations from regularity in timing are not based on the same factors as those discussed earlier with respect to the 2-level model. Expressive deviations are linked to intentions of the performer, whereas the variability accounted for by Wing & Kristofferson (1973) reflect the inherent noise in the system – the kind of thing a performer wants to ‘reduce.’

A logical question regarding expressive timing is how idiosyncratically it is assigned. Seashore and colleagues showed that individuals can be very consistent in their expressive timing. This finding has often been replicated (see reviews by Gabrielsson, 1999; Palmer, 1997). However, how consistent is one performer from the next? Certainly, individuals have varying interpretations, but some common principles hold across performers. Perhaps the most highly generalizable principle is *phrase final lengthening*. Put simply, people tend to slow down toward the end of a musical 'group.' In addition, people tend to speed up at the beginning of a phrase. The idea is similar to language. People tend to pause and to slow down when the reach the end of a clause, such as in:

Mary went to the store (pause), I went to the bank

Similarly, there are phrases in music, as discussed earlier. For instance, the first two lines of the tune *Mary had a little lamb*:

Mary had a little lamb, little lamb, little lamb
Mary had a little lamb whose fleece was white as snow

would constitute two musical phrases, the first a 'question', the second an 'answer.' When performing a these two lines, the performer would likely time each inter-onset interval in a manner that matches the phrases. Here's an example, using a very simple melody:

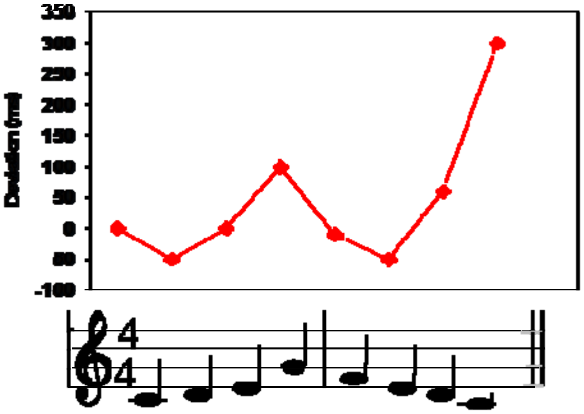

Figure 8

The swooping lines above the notation indicate how much each note deviates from what its length would be if the person were performing exactly as the notation dictates. A ‘mechanical’ performance would result in all performed notes matching a length of 0. Note that this is not the same thing as plotting note durations, which could introduce variations in inter-onset intervals that are determined by the notation. Notice that at the beginning of each phrase (each bar) the performer begins playing metronomically. Then she speeds up to about the middle of the phrase, and slows down to the end of it. Note also that the performer slows down more after the second ‘answer’ phrase than after the first ‘question’ phrase.

Although performers do vary one from the other, there is evidence for an overall tendency to time performances in this way, with more complexities arising for more complicated pieces. Todd (1985) proposed a mathematical model that predicts the degree of slowing a performer exhibits to the amount of ‘closure’ associated with a phrase. Degree of closure in his model was predicted by principles from the extensive model of Lerdahl and Jackendoff (1983), discussed elsewhere in this book. Though the predictions of this model do not perfectly fit every situation (see Repp, 1991 for detailed analyses of a number famous performances), it provides a starting point. Moreover, some evidence suggests that people can imitate a performance with ‘typical’ (i.e., Todd-like) timing than an ‘atypical’ performance (Clarke & Baker-Short, 1987; but see Repp, 2000 for a slightly different interpretation). Thus, performers can more readily imitate performances of others that conform to standard conventions.

Thus, performers use variations in timing to express musical ideas. In so doing the performer gives the listener an idea about her interpretation of the music, and ‘gives life’ to the performance. Some have suggested that the lively nature of expressive performance comes about because the variations in timing are reminiscent of patterns of physical motion. Expression thus helps propel the music forward. Anders Friberg and Johan Sundberg (1999 see also Sundberg & Verillo, 1980) tested a model of expressive timing based on the way in which runners stop themselves after the fall phase of a step. The resulting curves are similar to those predicted by Todd’s (1985) model, though based on slightly different principles. Despite the intuitive appeal of the music-as-motion analogy, this view has detractors. A specific example is a recent paper by Henkjan Honig (2005, 2006) who argued that expressive nuances may be better predicted by a model based on performer's need to follow the rhythmic pattern than on patterns typical of large-scale physical motion (like running and dancing).

Other aspects of timing may be used to communicate the performer's interpretation. Consider, for instance, timing among many notes that are notated as sounding simultaneously. Performers typically don't perform all these notes at once, and these slight deviations may be linked to expressive intentions. Specifically, performers appear to use *voice leading*, which occurs when one note is produced slightly before the others. Voice leading has been studied in solo piano performance (as has expressive timing), in which context one can make the easiest case for a link between voice-leading and intentions. Caroline Palmer (1989, 1996) has reported evidence suggesting that voice leading is used to highlight which of the pitches that are notated as being simultaneous is most important, i.e., which one constitutes the melody. In other words, voice leading helps one part 'stand out' to the listener. In addition, there is evidence from perception that asynchronous events are less likely to sound as though they belong to the same Gestalt group than are synchronous events (Bregman, 1990). Some have suggested that voice leading may be linked more to the velocity with which certain fingers strike the keys (related to the loudness of notes) than to timing per se (Goebl, 2001; Repp, 1996), though this alternative account still links voice leading to expressive intentions.

Additional expressive information may be communicated by performers visually. Jane Davidson (1993) has recorded performer's movements while playing the piano, focusing on movements of the trunk, arms and head. Her data show that performances associated with more movement are perceived as being more expressive than those with less movement (e.g., when a performer's movements are constrained). This perceived expressivity was found even when people saw point-light displays instead of the entire film. In point light displays, the viewer sees little dots of lights positioned at individual joints. Thus a pattern of moving dots is all the viewer has to reconstruct the underlying motion.

To summarize, performers use intentional deviations from the music notation in order to convey their interpretation of the musical structure, and otherwise 'give life' to a performance. The way performers do this may allude to physical motion in general, and/or may help the listener follow along.

Tuning of actions

It is worth noting just how dominated research on music performance is by the piano. Much less research has involved other instruments. This focus has much to do with the ease of measuring timing in piano performances. Current technology allows one to record the exact time at which a piano key contacts the key bed. Calculating timing for other instruments typically requires more complex analyses based on the acoustic waveform. However, examining the piano is somewhat limited with respect to the next factor in motor control during performance, the ‘tuning’ of a performance.

Tuning refers to the precision with which a performer generates musical events that match intended pitch categories. Consider the prototypical ‘bad singer.’ Such individuals have been found to sing with appropriate timing (Hyde & Peretz, 2004), but may sing the wrong notes entirely. This is clearly a case of poor ‘tuning.’ In other respects, tuning may be used expressively as in the case of vibrato or slight deviations from a pitch class. In contrast to research on timing, research on tuning concerns instruments other than the piano, which is tuned prior to performance. The most critical instruments from the standpoint of tuning are unfretted stringed instruments (e.g., the violin) and the human voice.

Mechanisms for tuning. We focus on singing, as a ubiquitous musical behavior and it poses perhaps the most difficult task with respect to tuning. Singing in tune is accomplished by varying the tension and length of the vocal folds (or cords) in one’s larynx. One cannot see these muscles without the aid of a laryngoscope (which involves inserting a tube down one’s throat); furthermore, it is impossible to feel the muscle tension that you generate, rendering both visual and proprioceptive feedback useless. By contrast, violinists can observe the position of their fingers (though experts do not do this), and trumpeters can feel the tension of their lips (embouchure) to control tuning.

A good deal of work has been done on the way in which experts accomplish vocalization (see Sundberg, 1987 for a review). Although studies of experts do not address sources of mistuning, they do offer thoughts on how performers accomplish the feat of tuning. There has recently been a burst of activity on research examining the nature of inaccuracy in singing as a way of understanding tuning by virtue of what happens when the system breaks down. These inaccuracies are commonly referred to as 'tone deafness', but as will be discussed later it is not clear that inaccurate singing actually results from an inability to perceive tones.

Vocal production involves three basic processes: respiration, phonation, and articulation. Respiration refers to the control of breath (inspiration, breathing in, and expiration, breathing out). You may have been told about the virtues of singing or speaking 'from the diaphragm.' This statement refers to the use of lower abdominal muscles to push air up out of the lungs. The voice works best when a strong column of air is sent up through the trachea (the air tube through your throat). Many untrained singers, by contrast, sing by tightening their chest muscles, which results in a more strained sound.

Phonation contributes the most to tuning. The process is quite complex and we will not offer the complete details here (Sundberg, 1987 and Chapter Five of Handel, 1989 are a good sources for these details). Inside your larynx are a set of muscles and bones that govern two flaps of skin known as the 'vocal folds' (or 'vocal cords). Vocal folds can be brought together (adduction) or held apart (abduction). The folds are held apart when we whisper or breathe, and a pitch-like quality results when folds are held together (adduction). During adduction, air from the lungs causes the vocal folds to vibrate back and forth, and the resulting frequencies of vibration generate pitch. Phonation refers to the generation of this pitch quality. The length, tension, and massiveness of vocal folds are all varied when we produce different pitches. In general, high pitches are created when the folds are tenser, less massive, or (counter-intuitively) longer (and vice-versa). Falsetto voice, for instance, is produced when the vocal folds are stretched out and, coincidentally, thinner.

The final process in vocal production, articulation, arguably is crucial for the communication of words, but not quite so much for tuning. This stage involves the manipulation of the tongue, lips, and jaw in order to articulate speech sounds. More generally, these alterations can influence the sound quality (timbre) and can make the sound of one’s voice rich and full (if sound is directed out of the mouth) or nasal (if sound is directed out of the nose.

Individual differences in the tuning of actions. Of course, once we mention singing, individual differences leap to mind as an important issue. We’ve all had the painful experience of hearing a well-intentioned friend sing a ‘melody’ that bears little resemblance to the intended tune, and likewise we can all think of professional singers who perform astounding vocal feats. Indeed, a large percentage of people, at least 17% by a recent estimate (Cuddy, Balkwill, Peretz, & Holden, 2005) consider themselves ‘tone deaf’, and when one of us (PQP) surveyed 1,105 intro to psychology students, 59% claimed to be unable to imitate melodies by singing (Pfordresher & Brown, 2006). Is incompetence in vocal tuning as prevalent as we think it is?

First of all, it is important to distinguish mistuning of the voice during singing from other vocal dysfunctions that may result in less-than-ideal vocal quality. Many people, for instance, have ‘rough’ sounding voices. This quality occurs when the vocal folds are not held close enough, and some of the air stream gets through. A typical cause of this problem is from calluses (‘nodes’) that form on the vocal folds, often after years of abuse (singing too loud). Many rock stars who are apparently able to scream for year with little effect on their voice have learned special screaming techniques (believe it or not) so that they, like a baby, can wail for hours without damage. But the untutored screamer is unfortunately destined for a brief singing career.

The cause of the kind of mistuned singing that we have all experienced is, surprisingly enough, elusive. Some interesting possible answers have emerged only recently, but we are far from fully understanding the phenomenon. An intuitive explanation focuses simply on motor control: bad singers may be unable to control their vocal folds. However, evidence in favor of this view is lacking, and some reflection can determine why. Even bad singers are typically adequate at using pitch to communicate in language, as in differentiating a question from an answer. Moreover, bad singers may be no less precise in controlling pitch spontaneously than are good singers. The differences between people appear in situations where one has to imitate rather than generate the structure of a sequence, particularly musical sequences.

If motor control doesn't explain poor singing, what does? The common usage of the term 'tone deaf', as an indicator of poor singing (Sloboda, Wise, & Peretz, 2005), suggests a perceptual basis. Plainly put, inaccurate singers may also be inaccurate listeners. True tone deafness would suggest that inaccurate singers have a reduced ability to hear pitch differences, resulting in a higher *difference threshold* (the minimal pitch difference one can detect) than found in the normal population. Recent highly influential research by Peretz and colleagues (e.g., Peretz et al., 2002; cf. Allen, 1878) has in fact revealed a population with such a deficit, and this population is also less accurate at singing than the rest of the population (Peretz , Champod, & Hyde, 2003). This deficit in pitch perception has been termed *congenital amusia*, suggesting an inherited disorder of music processing.

Thus, there are likely to be at least some inaccurate singers who suffer from a perceptual deficit. But is this true of most bad singers? Probably not. Other research has identified inaccurate singers who are apparently unimpaired when given perceptual tasks (Bradshaw & McHenry, 2005). If bad singers do not suffer from either a motor or a perceptual deficit, what is the basis of their impairment? Recent research by one of us (PQP) suggests that the cause may be linked the ability to link perception with action (Pfordresher & Brown, 2006). That is, bad singers may distort the structure of music when generating motor plans, while showing no impairment for production tasks that are non-imitative or for perceptual tasks.

Monitoring the outcomes of performance

Finally, we turn to a topic often overlooked in accounts of performance: the role of performer as perceiver. We refer to the perception of the results of one performance as *perceptual feedbac*k. Perceptual feedback can span many modalities, including audition, vision, haptics (touch) and prorprioception (the feeling of one's own movement based on feedback from the muscles and joints). In practice, almost all the research on self-perception in performance has focused on auditory feedback, which we focus on here.

It may be taken as a truism that the performer has to hear what he or she has done in order to continue performing effectively. This makes sense from what has been called an ‘error monitoring’ account (e.g., Levelt, 1989): people use feedback to gauge whether they have done the right thing. But this ‘truism’ may not be true, at least not for performances of well-learned melodies. In fact, when pianists perform with the sound turned off, some subtle differences in a performance result but the changes to performance would not constitute ‘disruption’ per se (Repp, 1999). Singing appears to be more reliant on auditory feedback, in that intonation suffers when auditory feedback is masked (Mürbe, Pabst, Hofmann, & Sundberg, 2003), although changes to intonation caused by masked feedback are not nearly as large as those differences found between accurate and inaccurate singers when both sing with normal auditory feedback. Thus, again the removal of auditory feedback causes at best modest disruption.

Though the presence of auditory feedback may not be crucial, it is crucial that auditory feedback – when present – match the sequence of executed actions in a performance. Alterations of auditory feedback can be, though they are not always, devastating to a performance, such that a skilled pianist sounds like a beginner. The most widely studied alteration to auditory feedback is delayed auditory feedback (DAF). First explored in the domain of speech (by Black, 1951; Lee, 1950), DAF simply involves adding a temporal gap between the time at which an action is executed (e.g., a key press) and the time at which the resulting sound begins. It is like playing in a large echoic room (e.g., a cathedral), only in this room you can only hear the echo, not the initial sound of your instrument (incidentally, pneumatic pipe organs function a bit like this). Anecdotally, performers often mention that the experience of DAF is like certain very difficult performance experiences they’ve had. The influence of DAF on performance is profound. It can slow down timing, cause timing variability to increase, and increase error rates. DAF’s strongest effects, however, may be on timing of actions more so than whether the correct note is played (Pfordresher, 2003). The effect does not appear to be related to attention; attempts to manipulate attention to sound conducted in unpublished research by one of us (PQP) have not influenced DAF’s effect.

Importantly, not all delays are created equal. Certain delays cause no disruption at all. Many people have found that delays around 200ms in length are more disruptive than others (e.g., MacKay, 1987). However, it is unlikely that such an absolute time metric pertains to rhythmic behaviors like music. Instead, it seems as though disruption relates to rhythmic relationships between actions and sound (Finney & Warren, 2002). Recent data suggest that produced timing slows down when auditory feedback sounds as one's finger is moving up from a key (after it is produced) and speeds up when it coincides at a time when the finger moves back down toward the key (Pfordresher & Benitez, in press).

Another characteristic of auditory feedback that has been investigated more recently is pitch. Interestingly, simply altering pitch in a random or random-like way does not disrupt production (Finney, 2001; Pfordresher, 2005). So disruption does not result from hearing an artificially induced 'error.' However, hearing altered pitches that form a sequence like the one you're playing, but that does not line up appropriately with your sequence, can be very disruptive. For instance, hearing a melody with the same contour, but having all the upwards intervals in the melody sound when you are producing downwards pitch changes on the keyboard increases errors significantly, though it does not similarly disrupt accuracy (reviewed in Pfordresher, 2006).

These findings lead to an important implication regarding the way in which performers use perceptual feedback. Put simply, performance and perception draw on the same neural resources. Recent neuroscience research has revealed so-called *mirror neurons* that are responsible for the mental simulation of action (reviewed in Chapter Four). According to this view, auditory feedback that presents a similar, though displaced, sequence to the one a musician tries to perform may be activating mirror neurons at the wrong time, leading to the generation of inappropriately timed action plans.

Many remaining questions exist regarding the use of feedback in performance. Research to date is largely limited to audition, as opposed to other modalities, and to feedback from oneself rather than others. The situation is beginning to change. For instance, recent research has investigated the effect of performing with an ensemble in which information from other ensemble members is delayed (Bartlette, Headlam, Bocko, & Velikic, 2006), and the effect of singing in a chorus when other parts' are mismatched relative to one's own (Fine, Berry, & Rosner, 2006). Both studies document disruptive effects comparable to those found with self-feedback. Finally, we should not forget the audience, who can be a substantial source of feedback. Any performer knows that when the audience is engaged and large, the effect on one's performance can be profoundly good, and that the opposite can occur for small and/or disinterested audiences. There is much room for future research on this important topic.

Coda

Music performance is a complex, fascinating, though under-represented topic. Part of its omission may have to do with preconceptions about action being somehow ‘thoughtless.’ Research on the cognitive bases of music performance contradicts this assumption. The performer must coordinate many cognitive, perceptual, and motor functions including planning, execution, the use of feedback, and on-line adjustments of planning. Furthermore, performers expressively deviate from what might be considered a ‘typical’ performance in order to ‘breathe life’ into music. Performance is thus a compelling area from both a cognitive and an aesthetic perspective.

Despite the intricate research that has already been carried out, much more needs to be done. In particular, current research on music performance suffers from a limited focus on individual, expert performers within the Western music tradition. Future research will be needed to explore music making across broader populations and cultures, as has been done in other disciplines (such as Ethnomusicology). Such efforts have just begun in music cognition. To date, research has made a compelling case that the mind of the musical performer presents a rich domain for exploring the role of cognition in action. Populations of expert musicians provide examples of the extremes to which these intricacies may be taken, but we hasten to emphasize that musical communication is not limited to experts. The minds of people singing chants at British football matches warrant as much intrigue as those of the concert pianist.

References

Allen, G. (1878). Note-deafness. *Mind, 10,* 157-167.

Baddeley, A.D. (1986). *Working memory*. London: Oxford University Press.

Bartlett, C., Headlam, D., Bocko, M., & Velikic, G. (2006). Effect of network latency on interactive musical performance. *Music Perception*, *24*, 49-62.

Beilock, S. L., Bertenthal, B. I., McCoy, A. M., & Carr, T. H. (2004). Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills. *Psychonomic Bulletin & Review, 11*, 373-379.

Black J. W. (1951). The effect of delayed side-tone upon vocal rate and intensity. *Journal of Speech and Hearing Disorders, 16,* 56-60.

Bradshaw, E., & McHenry, M. A. (2005). Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *Journal of Voice, 19*, 431-439.

Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, MA : MIT Press.

Clarke, E. [F.], & Baker-Short, C. (1987). The imitation of perceived rubato: A preliminary study. *Psychology of Music, 15*, 58-75.

Cuddy, L. L., Balkwill, L., Peretz, I., & Holden, R. R. (2005). Musical difficulties are rare: A study of “tone deafness” among university students. *Annals of the New York Academy of Sciences, 1060*, 311-324.

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music*, *21*, 103-113

Desain, P., & Honig, H. (1994). Does expressive timing in music performance scale proportionally with tempo? *Psychological Research, 56*, 285-292.

Dissanayake, E. (2000). Antecedents of the temporary arts in early mother-infant interaction. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 389-410). Cambridge, MA: MIT Press.

Fine, P., Berry, A., & Rosner, B. (2006). The effect of pattern recognition and tonal predictability on sight-singing ability. *Psychology of Music*, *34*, 431-447.

Finney, S. A. (1997). Auditory feedback and musical keyboard performance. *Music Perception, 15*, 153-174.

Finney, S. A., & Warren, W. H. (2002). Delayed auditory feedback and rhythmic tapping: Evidence for a critical interval shift. *Perception & Psychophysics, 64,* 896-908.

Freiberg, A., & Sundberg, J. (1999). Does music performance allude to locomotion? A model of final *ritardandi,* derived from measurements of stopping runners. *Journal of the Acoustical Society of America, 105*, 1469-1484.

Gabrielsson, A. (1999). The performance of music. In D. Deutsch (ed.) *The psychology of music* (pp. 501-602). San Diego, CA : Academic Press.

Gabrielsson, A. (2003). Music performance research at the new millennium. *Psychology of Music, 31*, 221-272.

Goebl, W. (2001). Melody lead in piano performance: expressive device or artifact? *Journal of the Acoustical Society of America, 110,* 563-72.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events.* Cambridge, MA : MIT Press.

Honig, H. (2005). Is there a perception-based alternative to kinematic models of tempor rubato? *Music Perception, 23*, 79-85.

Honig, H. (2006). Computational modeling of music cognition: A case study on model selection. *Music Perception, 23,* 365-376.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15,* 356-360.

James. W. (1890). *Principles of psychology* (volume 2). New York : Holt.

Johnson-Laird, P. N. (2002). How jazz musicians improvise. *Music Perception, 19,* 415-442.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science, 11,* 201-258.

Killeen, P. R., & Weiss, N. A. (1987). Optimal timing and the Weber function. *Psychological Review, 95,* 274-295.

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *The Journal of Neuroscience, 27,* 308-314.

Large, E. W. and Jones, M. R. (1999) The dynamics of attending: How people track time-varying events. *Psychological Review, 106*, 119-159.

Lashley, K. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112-136). New York : Wiley.

Lee, B. S. (1950). Effects of delayed speech feedback. *Journal of the Acoustical Society of America, 22*, 824-826.

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA : MIT Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA : MIT Press.

Loehr, J. D., & Palmer, C. (in press). Cognitive and biomechanical influences in pianists' finger tapping. *Experimental Brain Research.*

MacKay, D. G. (1987). *The organization of perception and action.* New York: Springer-Verlag.

MacKenzie, C. L., & VanEerd, D. L. (1990). Rhythmic precision in the performance of piano scales: Motor psychophysics and motor programming. In M. Jeannerod (Ed.) *Attention and performance XIII* (pp. 375-408). Hillsdale, NJ : Earlbaum.

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., & Miller, N. S. (2006). The time of our lives: Live span development of timing and event tracking. *Journal of Experimental Psychology: General, 3*, 348-367.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Mürbe, D., Pabst, F., Hofmann, G., & Sundberg, J. (2003). Effects of a professional solo singer education on auditory and kinesthetic feedback – A longitudinal study of singers' pitch control. *Journal of Voice, 18,* 236-241.

Noyle, L. J. (1987). *Pianists on playing: Interviews with twelve concert pianists.* London: Scarecrow Press.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human perception and Performance, 15,* 31-346.

Palmer, C. (1996). On the assignment of structure in music performance. *Music Perception, 14*, 21-54.

Palmer, C. (1997). Music Performance. *Annual Review of Psychology, 48,* 115-138.

Palmer, C., & Pfordresher, P.Q. (2000). The faces of meter in performance and perception. In C. Woods, G. Luck, R. Brochard, F. Seddon & J. A. Sloboda (Eds.) *Proceedings of the 6th International Conference on Music Perception and Cognition.* Keele, UK : Keele University Department of Psychology.

Palmer, C., & Pfordresher, P.Q. (2003). Incremental planning in sequence production. *Psychological Review, 110,* 683-712.

Palmer, C., & van de Sande, C. (1993). Units of knowledge in music performance. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19,* 457-470.

Palmer, C., & van de Sande, C. (1995). Range of planning in skilled music performance. Journal of *Experimental Psychology: Human Perception & Performance, 21,* 947-962.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception, 11,* 409-464.

Parsons, L. M., Sergent, J., Hodges, D. A., Fox, P. T. (2005). The brain basis of piano performance. *Neuropsychologia, 43*, 119-215.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: The Montreal battery of evaluation of amusia. *Annals of the New York Academy of Sciences, 999,* 58-75.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron, 33,* 185-191.

Povel, D. J. (1981). Internal representations of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 3-18.

Povel, D. J. & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2,* 411-440.

Pfordresher, P. Q. (2005). Auditory feedback in music performance: The role of melodic structure and musical skill. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1331-1345.

Pfordresher, P. Q. (2006). Coordination of perception and action in music performance. *Advances in Cognitive Psychology, 2*, 183-198. [www.ac-psych.org]

Pfordresher, P. Q. & Benitez, B. (in press). Temporal coordination between actions and sound during sequence production. *Human Movement Science.*

Pfordresher, P. Q., & Brown, S. (2006, April). Is "Tone Deafness" Really Tone Deafness? *Poster presented at the Cognitive Neuroscience Society Meeting*. San Francisco, CA.

Pfordresher, P. Q. Palmer, C., & Jungers, M. (2007). Speed, accuracy, and serial order in sequence production. *Cognitive Science, 31,* 63-98.

Repp, B. H. (1991). Diversity and commonality in music performance: An analysis of timing microstrutcure in Schumann's "Träumerei." *Journal of the Acoustical Society of America, 92,* 2546-2568.

Repp, B. H. (1994). Relational invariance of expressive microstructure across global tempo changes in music performance: An exploratory study. *Psychological Research, 56,* 269-284.

Repp, B. H. (1996). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America, 100*, 3917-3932.

Repp. B. H. (1997). Variability of timing in expressive piano performance increases with interval duration. *Psychonomic Bulletin & Review, 4,* 530-534.

Repp, B. H. (1999). Effects of auditory feedback deprivation on expressive piano performance. *Music Perception, 16*, 409-438.

Repp, B. H. (2000). Pattern typicality and dimensional interactions in pianists' imitation of expressive timing and dynamics. *Music Perception, 18,* 173-212.

Robertson, S. D., Zelaznik, H. N., Lantero, D., Bojczyk, K. G., Spencer, R. M., Doffin, J. G., & Schneidt, T. (1999). Correlations for taming consistency among tapping and drawing tasks: Evidence against a single timing process for motor control. *Journal of Experimental Psychology: Human perception and Performance,25,* 1316-1330.

Rosenbaum, D. A. (2005). The Cinderella of psychology: The neglect of motor control in the science of mental life and behavior. *American Psychologist, 60*, 308-317.

Seashore, C. A. (1938/1967). *The Psychology of Music*. New York : Dover.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in stentence production. In W. E. Cooper & C. Walker (Eds.) *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295-342). Hillsdale, NJ : Earlbaum.

Sloboda, J. [A.] (1974). The eye-hand span – An approach to the study of sight reading. *Psychology of Music, 2*, 4-10.

Sloboda, J. A., Wise, K. J., & Peretz, I. (2005). Quantifying tone deafness in the general population. *Annals of the New York Academy of Sciences, 1060*, 255-261.

Sundberg, J. (1987). *The science of the singing voice.* Dekalb, IL : Northern Illinois University Press.

Sundberg, J., & Verillo, V. (1980). On the anatomy of the ritard: A study of timing in music. *Journal of the Acoustical Society of America, 68*, 772-779.

Todd, N. [P. M.] (1985). A model of expressive timing in music performance. *Music Perception, 3,* 33-58.

Toiviainen, P. (1995). Modeling the target-note technique of bebop-style jazz improvisation: an artificial neural network approach. *Music Perception, 12*, 399-413.

Wing, A. M. & Kristofferson, A. B. (1973). Response delays and the timing of interresponse intervals. *Perception & Psychophysics, 14*, 5-12.

Zelaznik, H. N., Spencer, R. M. C., & Ivry, R. B. (2002). Dissociation of explicit and implicit timing in repetitive tapping and drawing movements. *Journal of Experimental Psychology: Human perception and Performance, 28*, 575-588.