Tapping to Ragtime: Cues to Pulse Finding

Joel Snyder & Carol L. Krumhansl
Cornell University

Two experiments investigated cues to pulse finding using a relatively unconstrained, naturalistic paradigm. Participants tapped what they felt was a comfortable pulse on a keyboard playing a percussive sound. The stimulus materials were based on ragtime excerpts, played metronomically (i.e., without expressive timing or tempo variation). The first experiment, with 8 musically experienced and 8 musically inexperienced subjects, played each excerpt in two versions: a pitch-varied version (the original excerpt) and a monotonic version (with all tones changed to middle C) that was designed to remove all melodic and harmonic cues to pulse. Neither the absence of pitch information nor musical experience significantly affected performance. The second experiment tested 12 musically experienced subjects on shorter excerpts from the same ragtime pieces. Full (right-hand and left-hand parts together) and right-hand-only versions of the excerpts were each played in pitch-varied and monotonic versions. Removing the left-hand part significantly affected tapping performance on a number of measures, causing a lower percentage of tapping on the downbeat, more off-beat taps, more aperiodic taps, more switches between tapping modes, a higher variability of the intertap interval, and larger deviations from the beat. As a whole, these indicate a negative effect of removing the left-hand part. Again, differences between pitch-varied and monotonic versions were generally small. Analysis of the music revealed the following cues to pulse finding: a predictable alternating bass pattern in the left-hand part and a majority of notes on metrically strong positions in both the right-hand and left-hand parts. These results suggest that, for piano ragtime music, temporal cues are prominently available for finding and following the pulse and that pulse finding is largely independent of pitch information. Implications of the experimental measures and music-analytic techniques for models of pulse perception are considered.

Received March 30, 2000; accepted December 21, 2000.

Pulse finding is a basic ability used during musical listening, performance, and dancing. Pulse, which can be established after 2–4 cycles (Desain & Honing, 1999; Fraisse, 1982), serves as a basic rhythmic unit allowing
listeners, performers, and dancers to synchronize with music and each other. Evidence that such periodic behavior is ubiquitous comes from studies of tapping to simple acoustic patterns (Essens & Povel, 1985; Helmuth & Ivry, 1996; Ivry & Hazeltine, 1995; Jones & Pfordresher, 1997; Mates, Radil, & Pöppel, 1992; Parncutt, 1994; Povel, 1981, 1984, 1985; Povel & Collard, 1982; Povel & Essens, 1985; Robertson et al., 1999; Vos, Mates, & van Kruysbergen, 1995), tapping to music (Drake, Penel, & Bigand, 2000; Scheirer, 1998; van Noorden & Moelants, 1999; Vos, van Dijk, & Schomaker, 1994), tapping to polyrhythms (Deutsch, 1983; Klapp, Hill, Tyler, Martin, Jagacinski, & Jones, 1985; Peper, Beek, & van Wieringen, 1995), speech rhythms (Cummins & Port, 1998), limb coordination (Diedrich & Warren, 1995; Schmidt, Beek, Treffner, & Turvey, 1991; Schmidt, Bienvenu, Fitzpatrick, & Amazeen, 1998), time perception (Large & Jones, 1999; McAuley, 1995; McAuley & Kidd, 1998), and infants’ postural adjustments (Bertenthal, Rose, & Bai, 1997; Schmuckler, 1997). In addition, theorists have hypothesized about the possible connections among pulse perception, time perception, motor behavior, and locomotion (Large & Jones, 1999; Parncutt, 1987; Todd, 1994, 1999; Todd, O’Boyle, & Lee, 1999). Computational approaches to pulse include applying a set of rules for updating the current pulse hypothesis (Desain & Honing, 1999; Longuet-Higgins & Lee, 1982; Steedman, 1977), using equations that calculate the pulse strength on the basis of musical features (Parncutt, 1994), implementing mechanisms that oscillate in synchrony with the pulse (Eck, Gasser, & Port, 2000; Gasser, Eck, & Port, 1999; Large, 1994; Large & Kolen, 1994; Large & Jones, 1999; McAuley, 1995; Toiviainen, 1998), and simulating filtering mechanisms that extract periodicities in the music (Scheirer, 1998; Todd, 1994, 1999; Todd et al., 1999).

Past experimental and computational studies of pulse finding in music generally assume that only temporal information is important for pulse perception (but see Steedman, 1977). It is clear from behavioral and computational studies that temporal information is often sufficient for pulse finding, but any additional contribution of pitch information is largely unknown. Few behavioral experiments use musical stimuli to address this question, and even fewer make experimental manipulations to determine what cues listeners use. Another limitation of some past studies is that their experimental measures do not assess multiple aspects of performance. In other words, few studies combine measures of tapping period, tapping phase, and timing variability, thus giving a limited description of pulse-tracking performance. The following review focuses on investigations of meter perception and synchronization with auditory patterns, particularly focusing on the possible role of pitch in perceiving temporal structure, the performance measures used, and benchmarks suggested by the studies to date. Studies using musical and nonmusical patterns will be considered, in addition to studies using perceptual and perceptual-motor tasks.
The research of Dawe, Platt, and Racine (1993, 1994, 1995) has pointed to the importance of harmonic changes in rhythm perception. The first two studies pitted harmonic, temporal (Dawe et al., 1993, 1994), and melodic (Dawe et al., 1994) cues against one another by varying the position of a chord change in a repeating sequence consisting of four short durations and a long duration. Subjects identified the perceived rhythmic pattern (Dawe et al., 1993) and identified the meter (Dawe et al., 1994). Harmonic factors primarily determined the judgments in each study, and musicians weighted harmonic information more heavily than did nonmusicians. The final study in this series (Dawe et al., 1995) quantitatively modeled the contributions to meter perception of temporal accents, the onset of the long-duration note, the termination of the long-duration note, and the position of the harmonic change. The latter factor showed a consistent difference between nonmusicians and musicians, with the harmonic change being the primary influence on musicians’ judgments of rhythm. These results suggest that removing pitch information would diminish the ability to find and follow the pulse in music, especially for musicians. Such a finding would be a major challenge to models of pulse finding that rely solely on temporal information.

To address the problem of synchronizing with performed music, Drake et al. (2000) asked musicians and nonmusicians to tap to six classical piano works, each in three different versions. The versions were mechanical (with no timing or amplitude variations), accented (with 10-dB-amplitude accents for notes on the first beat of the measure), and expressive (performed by a professional pianist in a “natural, musical fashion”). Musicians showed a higher percentage of successful synchronization across pieces (i.e., at least nine consecutive intertap intervals [ITIs] within 10% of the beat duration), a higher percentage of ITIs within 10% of the beat duration in measures 2–11, a higher percentage of ITIs within 10% of the beat duration in the 10 most successful measures for each piece, a tendency to tap at a slower rate, more metrically consistent tapping, and greater flexibility in synchronization rate. In measures 2–11 and in the most successful 10 measures, subjects performed better for the mechanical and accented versions than for the expressive version. These results suggest that expressive music can cause difficulty in some aspects of pulse finding, that musicians have a synchronization advantage over nonmusicians, and that musicians focus on higher metrical levels than do nonmusicians.

In another study of musical synchronization, Vos et al. (1994) had subjects tap at the beginning of the perceived measure for excerpts from solo instrumental pieces by J. S. Bach. Subjects differed in judged measure length (although subjects usually tapped at the first metrical position) and showed increasing judged measure length as a function of tempo. Parncutt (1994) similarly found the phase and period of tapping to differ across subjects for rhythmic monotonic stimuli, suggesting that ambiguity is a general phe-
nomenon of pulse and meter. Vos et al. (1994) also tested whether subjects’ judgments of meter could be explained by autocorrelating the melodic interval series (see also Brown, 1993). The metrical cues for period length found with autocorrelation predicted subjects’ responses poorly, casting doubt on whether this particular melodic cue is used by subjects.

An important early study by Essens and Povel (1985) examined the reproduction of monotonic patterns consisting of short and long interonset intervals (see also Povel, 1981). Subjects accurately reproduced patterns in which the long and short temporal intervals had a 2:1 ratio, but their reproductions of patterns with other ratios gravitated toward interval ratios of 2:1. This was even the case for stimuli with interval ratios of 3:1, which is a relatively simple ratio. This study suggests that subjects can better cognitively represent patterns with 2:1 interval ratios and that patterns may be misrepresented when they contain other ratios. These results could be interpreted as an inability to establish a pulse when time-interval ratios are larger or smaller than 2:1.

A recent study by van Noorden and Moelants (1999) asked subjects to tap to a wide range of music (e.g., 16th century Flemish polyphony, classic jazz, and contemporary popular music from the radio). Analyses of the tapping tempo showed partial dependence on style and a bias for tapping at rates around 500 ms ITI. This result is consonant with Fraisse’s (1982) review of findings showing that preferred tempo, spontaneous tempo, and the indifference interval (i.e., the point of most veridical time perception) each resides in the 500–700 ms range.

In another study of musical synchronization (Scheirer, 1998), subjects tapped to seven performed musical excerpts, each from a different style. The author evaluated two measures, root-mean-square (RMS) deviation from the beat (as judged by the author) and the variance of ITI. These measures respectively index absolute deviation from the beat and consistency of the tapping period. Although there were large differences between subjects and between stimuli on RMS deviation from the beat, a representative value was around 60 ms. Subjects differed from each other on variance of ITI but were fairly consistent across the different stimuli. The range of values for this measure was around 10–100 ms.

Vos et al. (1995) examined the phenomenon known as negative asynchrony, the tendency to tap slightly ahead of a series of isochronous auditory stimuli. The negative asynchrony was reduced when the stimuli were longer in duration and when their rise times were longer. The authors concluded that subjects synchronize to a perceptual center rather than to the physical onset of a tone. Recently, Wohlschlager and Koch (2000) reported that filling the interval between the isochronous pacing tones with additional tones or with bodily movements greatly reduced the negative asynchrony. Thus, one might not expect to observe negative asynchrony while tapping to music containing notes in between perceived pulses.
To summarize, the above studies represent a considerable body of research on synchronization to auditory stimuli. Some studies suggest a possible role of pitch information in temporal perception. Performance measures developed by studies of sensory-motor synchronization include indices of the phase and period of tapping, variability of the tapping period, the magnitude of anticipation, and the magnitude of deviation from the target. They suggest the following generalizations concerning tapping: a preference toward simple ratios, a preferred tapping rate in the range of 500–700 ms, a tapping period that may not correspond to the notated musical meter, a tendency to tap slightly ahead of pacing tones defining isochronous intervals, and a pulse-finding advantage with greater musical experience. With these results in mind, we now turn to the present experiments.

The two experiments reported next asked listeners to tap at a comfortable rate to relatively complex extended musical excerpts based on piano ragtime pieces. Although during its most active period (ca. 1890–1930) ragtime was dominated by vocal music, the piano ragtime of Scott Joplin and his contemporaries is the only form that is widely known today (Berlin, 1980, 1985; Jasen & Tichenor, 1978). Piano ragtime music has a number of features that were desirable for our purposes (see top of Figure 1). It is written in duple meter and is played at a moderate tempo suitable for dancing (originally cakewalks, marches, and two-steps). Its form is relatively obvious, consisting of pairs of 16-measure sections that subdivide into 4-measure phrases. The final of these four phrases typically contains a clear full cadence. The harmony is conventional, based largely on tonic, dominant, and subdominant chords in major. The left-hand part, often consisting of alternating bass notes and chords, defines simple, strong harmonic progressions and a clear duple meter. However, the right-hand melody is often syncopated against the left-hand accompaniment, providing melodic and rhythmic interest.

![Experiment 1 Excerpt Versions](image)

**Fig. 1.** The introduction from *Glad Cat*, showing pitch-varied and monotonic versions. For the monotonic version, all notes from the pitch-varied version are simply changed to middle C (duplicated notes are not shown).
One of our central objectives was to study the contributions of pitch and temporal cues to pulse finding, taking advantage of the characteristics of ragtime just described. The first experiment presented the excerpts in the original form and in a monotonic version (with all notes of the original changed to the same note, middle C). Removing the pitch variation in this way might have a large effect on tapping performance given the very clear harmonic structure of ragtime music. This experiment also compared two groups of listeners, musicians and nonmusicians, to assess possible effects of musical training on performance. The results might differ between these groups because musicians presumably have greater knowledge of harmony, meter, and musical form, as well as possibly more precise motor control than nonmusicians. The second experiment, using musicians only, again presented the music with and without pitch variation. In this experiment, the excerpts either contained both the regular left-hand part and the more syncopated right-hand part, or only the right-hand part.

A second objective of the study was to develop a battery of performance measures. These were all based on the tap time recorded by the computer. The first performance measure was tapping mode, illustrated in Figure 2, for the duple meter of the materials used in this study. There are four possibilities: tapping on the downbeat, tapping on the upbeat, tapping off-beat (periodically but on neither the downbeat nor the upbeat), and tap-

![Diagram of tapping modes](image)

Fig. 2. Subjects could tap in four different modes. Three of these were periodic: tapping on the downbeat (indicated by D), the upbeat (indicated by U), or off-beat. The other mode was aperiodic. Asterisks indicate tap positions.
Tapping aperiodically. To determine if the tapping was to the downbeat or the upbeat, we matched each tap where possible to the closest downbeat position (the first and third eighth-note metrical positions) or upbeat position (the second and fourth eighth-note metrical positions) that was within 100 ms of the tap. Tapping off-beat meant tapping with a steady ITI but not near the downbeat or the upbeat (i.e., not within 100 ms of either). Aperiodic taps included single deviant taps and prolonged tapping with an inconsistent or drifting period. We analyzed each participant’s tapping performance for tapping mode throughout the excerpt, making it possible to determine the percentage of time spent in each tapping mode and the number of switches between tapping modes. We defined switches as changes to a new tapping mode for at least two consecutive taps, and did not include changes to or from aperiodic tapping.

Figure 3 illustrates the other calculated performance measures. For these, we only used periodic taps. The first of these was based on the ITIs, shown in the top of Figure 3. This performance measure, called the coefficient of variability (CV), was the standard deviation of the ITIs, divided by the mean ITI. We excluded ITIs that were 100 ms greater or lesser than the

![Diagram](image-url)

Fig. 3. Intertap interval (ITI) is the average time in milliseconds between taps (top). Coefficient of variability (CV) is the SD of ITI, divided by mean ITI. Beats to start tapping (BST) is the time in milliseconds from the first note of a trial to the first periodic tap, as a proportion of the quarter-note interbeat interval (middle). In the example shown, BST is equal to 2, since downbeats correspond to the quarter-note level. Coefficient of delay (CDe) is the average time deviation in milliseconds of a tap from the nearest matched beat, as a proportion of mean ITI. Coefficient of deviation (CDev) is the average absolute time deviation in milliseconds of a tap from the nearest matched beat, as a proportion of mean ITI (bottom).
tapped interbeat interval (IBI) to reduce the effect of outliers. CV was a measure of the consistency of tapping independent of the relationship between the taps and the events in the music. Shown in the middle of Figure 3 is beats to start tapping (BST), defined as the time in milliseconds from the first note of the stimulus to when subjects began tapping periodically, divided by the quarter-note IBI. This assessed how much musical information the subject needed to find the pulse, controlling for musical tempo. The final measures were based on the deviations of the tap from the corresponding events in the music, shown at the bottom of Figure 3. The first of these, called the coefficient of delay (CDel), was the average time between a tap and the matched beat (downbeat, upbeat, or off-beat), divided by the mean ITI. A positive value meant that the tap occurred after the corresponding beat, whereas a negative value indicated that the tap occurred before the beat. The second, called the coefficient of deviation (CDev), was the average absolute time between the tap and the closest beat divided by the mean ITI, a measure of deviation from the beat independent of sign (i.e., delay or anticipation). For CDel and CDev, we only used taps within 100 ms of a downbeat, an upbeat, or some other metrical position (i.e., off-beat). We normalized the three coefficients (CV, CDel, CDev) using the subjects’ mean ITI to control for tapping rate.

**Experiment 1**

This experiment used the eight ragtime excerpts listed in Appendix A. They were played at slightly different tempos, but each had a quarter-note IBI in a comfortable tapping range. Appendix A also shows the form of the selections, which always ended at a section boundary so that the excerpts were of approximately equal duration and avoided awkward endings. The music was played in a metronomic fashion, with no expressive timing or tempo variations. These are of considerable interest in that they reflect structural characteristics of the music (see, for example, the review by Palmer, 1997). However, timing variability was not included in the current study to avoid complicating our analyses of tapping consistency with respect to previous taps and to events in the music. Two versions of each excerpt were presented, one with pitch variations and the other monotonic. Musically experienced and inexperienced subjects participated in order to assess the effects of musical training.

**METHOD**

**Subjects**

Eight musically experienced (male = 2, female = 6; mean years playing music = 13) and eight musically inexperienced (male = 4, female = 4; mean years playing music = 3) Cornell
University students participated for extra credit in psychology courses. No subject reported hearing problems, and two reported having absolute pitch. On a 7-point scale, the average familiarity with ragtime music was 2.6 for musically experienced subjects and 2.2 for inexperienced subjects. One other subject participated but was unable to perform the task, and two other subjects’ data were incomplete because of computer malfunction.

Materials and Stimuli

We selected eight Musical Instrument Digital Interface (MIDI) piano ragtime pieces in 2/4 meter (see Appendix A), which are available from the corresponding author upon request. The tempo range was 587–697 ms quarter-note IBLs, or 86–102 beats per minute. For each MIDI file, there was a constant time between the offset and onset of adjacent notes (range across excerpts = 3.5–8.5% of the quarter-note IBL). All stimuli were played to subjects using an object-oriented MAX interface on a Macintosh Quadra 680AV computer controlled by the experimenter, a Yamaha TX816 MIDI rack, a Yamaha 1204 mixing console, and AKG K-141 headphones.

We prepared each excerpt using Vision, a MIDI editing program, to have the following characteristics: (1) metronomic timing, (2) equalized MIDI velocity (related to acoustic amplitude and spectrum), (3) MIDI voice set to a synthesized piano timbre, (4) beginning at the start of the original piece, (5) lasting approximately 1.5 min in duration, and (6) terminating at a section end. We created two versions of each excerpt, as illustrated in Figure 1 for the beginning of one of the excerpts: (1) the original version with the previously mentioned modifications, and (2) one identical to the original but with all notes changed to middle C. For this experiment, we term the first of the versions as pitch varied, and the second as monotonic. All melodic and harmonic cues to pulse in the lower and the upper voices were absent in the monotonic versions. Although not shown in Figure 1, we duplicated Cs when multiple tones appeared in the original, so the two versions had the same number of tones. Duplication of simultaneous identical tones results in increased loudness compared with single tones in MAX.

Procedure

The experimenter asked each subject to “tap the beat of the music with the index finger of your dominant hand on the keyboard the way you normally tap your foot while listening to music” and to “find the most comfortable beat but do not begin tapping until you have found the beat mentally.” Subjects tapped on a Yamaha KX88 keyboard, with tap times collected by the MAX interface. A percussive hi-hat sound resulted from subjects’ tapping on the keyboard. Before the experiment began, subjects practiced tapping to a pitch-varied version and a monotonic version of an excerpt from a ragtime piece not used in the experiment. All subjects tapped to both versions of each of the eight ragtime excerpts. Prior to each experimental session, the experimenter chose a random order for the eight excerpts. The MAX interface ran through this order twice, interleaving the two versions of the eight excerpts. For example, if the first two numbers selected were 5 and 3, the first stimuli presented would be the pitch-varied version of excerpt 5 and the monotonic version of excerpt 3. After eight trials, the order would repeat with excerpt 5 presented in its monotonic version, followed by excerpt 3 in its pitch-varied version. The experimenter counterbalanced whether a pitch-varied version or a monotonic version came first, across subjects.

To test the timing accuracy of MAX, we constructed, played, and recorded the output for an isochronous series of tones with 600-ms interonset intervals. Testing showed that the interonset intervals had an average absolute deviation of ~1 ms from the specified value and a maximum deviation of 5 ms. Thus, most of the effect sizes and means we report exceed the temporal resolution of MAX. In addition, MAX played the 600 ms interonset intervals ~1 ms too fast.
RESULTS

A mixed-design analysis of variance (ANOVA) for each of the dependent measures determined the effects of pitch information, musical experience, and excerpt. For effects including excerpt, we adjusted the degrees of freedom using the Greenhouse-Geisser epsilon to correct for violations of sphericity. The results of the statistical tests appear in Table 1.

The top of Figure 4 shows the percentage of taps that were on the downbeat, the upbeat, the percentage of off-beat taps, and the percentage of aperiodic taps. About 80% of the time, subjects tapped on the downbeat, and most of the remaining time they tapped on the upbeat. For percentage of taps on the downbeat, there were no effects of pitch information, \( F(1, 14) = 1.34, p > .05 \), or musical experience, \( F(1, 14) = .13, p > .05 \), but there was a significant effect of excerpt, \( F(7, 98) = 4.95, \varepsilon = .579, p < .005 \). Additionally, pitch information and excerpt significantly interacted, \( F(7, 98) = 4.87, \varepsilon = .658, p < .005 \). Significant main effects and interactions with excerpt are described in more detail later.

About 2% of the taps were aperiodic. Most of these came in small groups; others were single deviant taps. Neither pitch information, \( F(1, 14) = .13, p > .05 \), nor musical experience, \( F(1, 14) = 3.18, p > .05 \), had a significant effect on this variable. Excerpts, however, differed from each other, \( F(7, 98) = 3.55, \varepsilon = .420, p < .025 \). In this experiment, no subject tapped off-beat.

The bottom of Figure 4 displays the average number of switches in tapping mode, which occurred very infrequently. Neither the presence of pitch information, \( F(1, 14) = .53, p > .05 \), nor musical experience, \( F(1, 14) = .23, p > .05 \), affected the number of switches. There was also no significant difference among excerpts, \( F(7, 98) = 2.65, \varepsilon = .345, p > .05 \). To examine possible effects of transitions from one section of the music to the next, we compared the number of switches and aperiodic taps near and far from section transitions. In the music used, 23.3% of the time was near a section transition, defined as within 4 s of a transition from one section of the music to another. Accordingly, 23.3% of the switches and aperiodic taps should occur near the transitions if they are randomly distributed. Significantly more switches, \( \chi^2(1, N = 102) = 12.70, p < .001 \), and aperiodic taps, \( \chi^2(1, N = 753) = 65.15, p < .001 \), were within 4 s of the section transitions than would be expected by chance.

The top of Figure 5 shows the proportion of musically experienced and inexperienced subjects tapping at different numbers of beats after the start of the music for the pitch-varied and monotonic versions. The bottom of Figure 5 shows this information in terms of mean BST. It appears as though musically experienced subjects began tapping earlier, but this effect was not statistically reliable, \( F(1, 14) = 4.45, p = .053 \). Neither was pitch infor-
information significant, $F(1, 14) = 3.04, p > .05$, although there was a significant effect of excerpt, $F(7, 98) = 3.46, \epsilon = .360, p < .05$.

Subjects tapped with an ITI of ~600 ms (corresponding to the quarter-note level) on 87% of the trials, ~300 ms on 9% of the trials, and ~1200 ms on 3% of the trials. The CV, shown at the top of Figure 6, was about 4% of the tapping period or about 24 ms. There was no effect of pitch
information, $F(1, 14) = .98, p > .05$, or musical training, $F(1, 14) = 1.76, p > .05$. Individual excerpts differed from each other on CV, $F(7, 98) = 2.79, \varepsilon = .523, p < .05$. However, no interactions appeared with any combination of pitch information, musical experience, and excerpt.
The middle of Figure 6 shows CDel to be near 1% of the tapping period, corresponding to a tap delay of approximately 6 ms. There were no effects of pitch information, $F(1, 14) = .00, p > .05$, musical experience, $F(1, 14) = .13, p > .05$, or excerpt, $F(7, 98) = 3.12, e = .274, p > .05$. As seen in the bottom of Figure 6, CDev was generally around 4% of the tapping period.
or 24 ms. No significant effects of pitch information, $F(1, 14) = 1.83, p > .05$, or musical experience, $F(1, 14) = 1.02, p > .05$, surfaced on this measure, although excerpt was significant, $F(7, 98) = 3.47, \varepsilon = .395, p < .05$.

**DISCUSSION**

Subjects in Experiment 1 demonstrated the following general pulse-finding behavior. Approximately 80% of all taps were on the downbeat, most of the remaining taps were on the upbeat, and only about 2% were aperiodic taps. Switches between modes of tapping were rare (much less than once per trial on average), and when they occurred they were often near transitions from one section of the music to the next. Subjects tapped about 90% of the time with an ITI of approximately 600 ms, sometimes at twice that rate, and very rarely at half that rate. The BST was about 5 on average for musically experienced subjects and about 9 for musically inexperienced subjects. The variability of the tapping period, CV, was near 4% of the tapping period (~24 ms). On average, tapping occurred just slightly behind the beat, by about 1% of the tapping period (~6 ms), as measured by CDel. Finally, subjects’ average tap deviation from the beat in absolute values, CDev averaged around 4% of the tapping period (~24 ms).

As shown in Table 1, we found no differences between pitch-varied and monotonic versions on any of the performance measures. The lack of differences between pitch-varied and monotonic versions was unanticipated given previous experimental results (Dawe et al., 1993, 1994, 1995) and

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Summary of Experimental Results Showing the Statistical Significance and Direction of Main Effects and the Significant Interactions in Experiment 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance Measure</strong></td>
<td><strong>Pitch Information</strong></td>
</tr>
<tr>
<td>Downbeat</td>
<td>n.s.</td>
</tr>
<tr>
<td>Off-beat</td>
<td>—</td>
</tr>
<tr>
<td>Aperiodic</td>
<td>n.s.</td>
</tr>
<tr>
<td>Switches</td>
<td>n.s.</td>
</tr>
<tr>
<td>BST</td>
<td>n.s.</td>
</tr>
<tr>
<td>CV</td>
<td>n.s.</td>
</tr>
<tr>
<td>CDel</td>
<td>n.s.</td>
</tr>
<tr>
<td>CDev</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

**Note**—n.s. = not significant; BST = beats to start tapping; CV = coefficient of variability; CDel = coefficient of delay; CDev = coefficient of deviation. 

*p < .05.
musical intuitions suggesting that pitch information, especially harmony, is important for the perception of musical rhythm and meter. We will return to this issue when we consider differences between the methods of these experiments and when we consider rhythmic characteristics of the excerpts.

In this experiment, we also found no effect of musical experience on any of the performance measures. Musically experienced subjects did not differ statistically from musically inexperienced subjects on any measures, although some trends favored musically experienced subjects. Whether these results will generalize to other styles of music is questionable given recently reported effects for musical experience using classical piano works (Drake et al., 2000). This divergence of findings suggests that pulse-finding differences as a function of musical experience are not due to general motor control differences. Instead, style-dependent pulse-finding differences as a function of musical experience may indicate the presence of more perceptual difficulties (e.g., temporal periodicity extraction) with less musical experience.

Experiment 2

Experiment 2 further investigated what information subjects use to find and follow the pulse in piano ragtime music. In this music, the left hand usually plays a pattern of alternating low and high notes, often with chords as the second part of the alternation. This steady left-hand part seems a likely cue for finding the pulse. In this experiment, we tested the importance of the left-hand part for pulse finding by measuring the effects of eliminating it (i.e., with stimuli consisting of only the right-hand notes). We also attempted a replication of the findings from Experiment 1 regarding the performance benchmarks and the lack of importance of pitch variations. In addition, these stimuli allowed a test of the hypothesis that pitch variations may be more important in music with relatively weak temporal cues to pulse by looking for effects of pitch within the right-hand versions. As shown in Appendix B, we used shorter excerpts to lessen the performance disruptions due to section transitions observed in the first experiment.

METHOD

Subjects

Twelve musically experienced Cornell University students (male = 5, female = 7; mean years playing music = 12) participated for extra credit in undergraduate psychology courses or received $5. One of these subjects reported having congenital high-frequency hearing loss, and one reported having absolute pitch. Two additional subjects’ data were not in-
cluded because of computer malfunction and another because the subject failed to follow instructions. On a 7-point scale, the average familiarity with ragtime music was 3.9. Although this average familiarity is higher than for the musically experienced subjects in Experiment 1, the difference was not statistically reliable, $t(18) = 1.74, p = .10$.

Materials and Stimuli

The same eight modified pieces from Experiment 1 were used. The length of each excerpt was ~40 s in this experiment, again terminating at the end of a section to avoid awkward endings. Four versions of the ragtime excerpts were used, as illustrated in Figure 7 for the beginning of one of the excerpts: (1) full pitch-varied, (2) full monotonic, (3) right-hand pitch-varied (a pitch-varied version of only the right-hand notes), and (4) right-hand monotonic (a monotonic version of only the right-hand notes). The stimuli were presented using the same equipment as in Experiment 1.

Procedure

The procedure was also similar to Experiment 1. The full and right-hand stimuli were presented in separate blocks, with pitch-varied and monotonic versions interleaved within blocks, as before. We counterbalanced whether subjects heard the full or the right-hand block first.

RESULTS

Repeated-measure analysis of variance (ANOVA) tested three within-subjects factors: pitch information, full/right-hand, and excerpt. We used

---

**Experiment 2 Excerpt Versions**

- **Full Pitch-Varied**
- **Full Monotonic**
- **RH Pitch-Varied**
- **RH Monotonic**

*Fig. 7. The introduction from *Glad Cat*, showing full pitch-varied, full monotonic, right-hand (RH) pitch-varied, and right-hand monotonic versions. The full pitch-varied and full monotonic versions are identical to the pitch-varied and monotonic versions in Experiment 1, respectively. All of the notes normally played by the left hand were absent in the right-hand versions.*
the same degrees of freedom adjustments for sphericity violations as in Experiment 1. The right-hand-only versions of one excerpt, Pine Apple, differed erroneously from the full versions in their construction. Experiment 2 ANOVAs were performed with and without data from Pine Apple. Main effects of pitch information and full/right-hand did not differ according to whether the data from Pine Apple were included. The main effects of excerpt for aperiodic and CDel were significant with all excerpts included but not when Pine Apple was excluded. In addition, the three-way interaction for off-beat tapping was not significant with all excerpts but became significant when we removed Pine Apple. Data from Pine Apple were not used in the results we present, which are summarized in Table 2.

The top of Figure 8 shows the percentage of taps on the downbeat, upbeat, the percentage of off-beat taps, and the percentage of aperiodic taps. Subjects tapped significantly more on the downbeat for the pitch-varied than monotonic versions, $F(1, 11) = 25.90, p < .001$, and more for the full than the right-hand versions, $F(1, 11) = 50.67, p < .001$. We also observed a significant difference between excerpts, $F(6, 66) = 27.08, \epsilon = .478, p < .001$. For percentage of taps on the downbeat, we detected an interaction between pitch information and full/right-hand, $F(1, 11) = 8.82, p < .025$. We also found interactions between pitch information and excerpt, $F(6, 66) = 3.37, \epsilon = .547, p < .05$, and between pitch information, full/right-hand, and excerpt, $F(6, 66) = 4.37, \epsilon = .532, p < .01$.

### Table 2

Summary of Experimental Results Showing the Statistical Significance and Direction of Main Effects, and the Significant Interactions in Experiment 2

<table>
<thead>
<tr>
<th>Performance Measure</th>
<th>Pitch Information</th>
<th>Full/Right-Hand</th>
<th>Excerpt Number</th>
<th>Significant Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downbeat</td>
<td>Pitch&gt;Mono*</td>
<td>Full&gt;right-hand*</td>
<td>*</td>
<td>Pitch x full/right-hand</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Full/right-hand</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pitch x excerpt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Full/right-hand x excerpt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pitch x full/right-hand</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Full/right-hand x excerpt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Off-beat</td>
</tr>
<tr>
<td></td>
<td>n.s.</td>
<td>Right-hand&gt;full*</td>
<td>*</td>
<td>Full/right-hand x excerpt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pitch x full/right-hand</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Full/right-hand x excerpt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Aperiodic</td>
</tr>
<tr>
<td></td>
<td>n.s.</td>
<td>Right-hand&gt;full*</td>
<td>n.s.</td>
<td>Full/right-hand x excerpt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Switches</td>
</tr>
<tr>
<td></td>
<td>n.s.</td>
<td>Right-hand&gt;full*</td>
<td>n.s.</td>
<td>Full/right-hand x excerpt</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BST</td>
</tr>
<tr>
<td></td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CV</td>
</tr>
<tr>
<td></td>
<td>n.s.</td>
<td>Right-hand&gt;full*</td>
<td>n.s.</td>
<td>Full/right-hand x excerpt</td>
</tr>
<tr>
<td></td>
<td>Mono&gt;Pitch*</td>
<td>n.s.</td>
<td>n.s.</td>
<td>Full/right-hand x excerpt</td>
</tr>
<tr>
<td></td>
<td>n.s.</td>
<td>Right-hand&gt;full*</td>
<td>*</td>
<td>None</td>
</tr>
</tbody>
</table>

*Note—n.s. = not significant; BST = beats to start tapping; CV = coefficient of variability; CDel = coefficient of delay; CDev = coefficient of deviation.

*p < .05.
Results for tapping off-beat were similar (but opposite in direction) to those for tapping on the downbeat. Somewhat more off-beat tapping occurred for monotonic than for pitch-varied stimuli, although this effect was not significant, $F(1, 11) = 3.84, p > .05$. Significantly more off-beat tapping occurred for right-hand than full versions, $F(1, 11) = 45.30, p < .001$. Excerpts also differed from each other on this measure, $F(6, 66) = 16.11, \epsilon = .481, p < .001$. Last, we detected interactions between full/right-
Tapping to Ragtime

hand and excerpt, \( F(6, 66) = 18.33, \, \varepsilon = .559, \, p < .001 \), and between pitch information, full/right-hand, and excerpt, \( F(6, 66) = 2.81, \, \varepsilon = .544, \, p < .05 \).

We found slightly more aperiodic tapping on monotonic than pitch-varied versions, but the difference was not significant, \( F(1, 11) = 4.02, \, p > .05 \). Aperiodic tapping also occurred less on the full than on the right-hand versions, \( F(1, 11) = 48.69, \, p < .001 \), but excerpts did not differ from each other, \( F(6, 66) = 2.97, \, \varepsilon = .462, \, p > .05 \). A significant interaction appeared between full/right-hand and excerpt, \( F(6, 66) = 3.66, \, \varepsilon = .497, \, p < .025 \).

The bottom of Figure 8 shows switches in tapping mode. There was no statistical difference between pitch-varied and monotonic versions, \( F(1, 11) = 2.56, \, p > .05 \). Subjects made fewer switches in the full versions than in the right-hand versions, \( F(1, 11) = 12.06, \, p < .01 \), and there were significant differences between excerpts, \( F(6, 66) = 9.59, \, \varepsilon = .402, \, p < .001 \). We also observed a significant interaction between full/right-hand and excerpt, \( F(6, 66) = 3.91, \, \varepsilon = .367, \, p < .05 \).

The top of Figure 9 shows the proportions of subjects tapping at different numbers of beats after the start of the music, and the bottom of Figure 9 shows this as mean BST. We observed no effects on BST for pitch information, \( F(1, 11) = 4.29, \, p > .05 \), for full/right-hand, \( F(1, 11) = .50, \, p > .05 \), or for excerpt, \( F(6, 66) = 2.57, \, \varepsilon = .384, \, p > .05 \). No significant interactions occurred among pitch information, full/right-hand, and excerpt for BST.

On 96% of the trials, subjects tapped with an ITI of ~600 ms, corresponding to the quarter-note level. The rest of the trials consisted of subjects tapping twice as fast, with an ITI of ~300 ms. No subjects tapped at ~1200 ms ITI. As displayed in the top of Figure 10, pitch information did not affect the CV, \( F(1, 11) = .66, \, p > .05 \), though there was a strong effect for higher values on right-hand than full versions, \( F(1, 11) = 52.63, \, p < .001 \). No significant effect of excerpt surfaced for CV, \( F(6, 66) = 1.93, \, \varepsilon = .645, \, p > .05 \). Full/right-hand and excerpt interacted significantly, \( F(6, 66) = 2.92, \, \varepsilon = .568, \, p < .05 \).

CDel, as displayed in the middle of Figure 10, showed a different pattern than in Experiment 1. Subjects tapped slightly more ahead of the beat with pitch-varied versions than with monotonic versions, \( F(1, 11) = 8.28, \, p < .025 \). The anticipation was approximately 4 ms on average. No difference occurred in CDel between full and right-hand versions, \( F(1, 11) = .63, \, p > .05 \), or between excerpts, \( F(6, 66) = 3.09, \, \varepsilon = .391, \, p > .05 \), but we found a significant interaction between full/right-hand and excerpt, \( F(6, 66) = 3.01, \, \varepsilon = .464, \, p < .05 \). As displayed at the bottom of Figure 10, there was no effect of pitch information for CDev, \( F(1, 11) = .01, \, p > .05 \), whereas higher values appeared for right-hand than full versions, \( F(1, 11) = 48.01, \, p < .001 \). Excerpts differed from each other on CDev, \( F(6, 66) = 3.17, \, \varepsilon = .669, \, p < .025 \), but pitch information, full/right-hand, and excerpt did not interact in any combination for this measure.
The general pulse-finding abilities found in this experiment for the full versions were similar to those of Experiment 1. The main additional finding of this experiment was that a number of measures showed poorer per-
formance for the right-hand-only versions. Without the left-hand accom-
paniment, the percentage of tapping on the downbeat significantly decreased, and the percentages of off-beat and aperiodic taps increased. Also, more switches between tapping modes occurred, indicating greater instability of synchronization for right-hand-only versions. Finally, the right-hand-only
versions produced more variable tapping and taps farther from the beat in absolute value. Thus, removing the left-hand notes generally produced decrements in tapping performance as assessed on a number of measures. With CDel, subjects tapped very slightly ahead of the beat in the pitch-varied versions (unlike in the earlier experiment) and very slightly behind the beat in the monotonic versions (as in the earlier experiment).

As in Experiment 1, pitch information generally had little effect. Two measures showed pitch effects in Experiment 2, however. More down-beat taps occurred for the pitch-varied than the monotonic versions (especially for the full versions), and taps occurred somewhat earlier relative to the beat for the pitch-varied versions than for the monotonic versions. These discrepancies are difficult to explain but could potentially arise from exposing subjects to two pitch-varied versions of each excerpt in Experiment 2, but just one pitch-varied version of each excerpt in Experiment 1. With the exceptions of percentage of tapping on the downbeat and tap delay, this experiment replicated the previous finding that removing pitch information had little effect on tapping performance. Moreover, the lack of interactions between full/right-hand and pitch information, except on percentage of down-beat taps, suggested that pitch cues were not used more when the left hand was absent. This one interaction demonstrated that while removing pitch information resulted in less tapping on the downbeat, the effect was larger for the full versions. This was against the prediction that removing the left-hand part would result in greater use of pitch information.

Excerpt Main Effects and Interactions

Since most performance measures in the two experiments showed significant excerpt effects, we determined whether the same excerpts exhibited poor performance across measures. For Experiment 1, American Beauty had the least down-beat tapping, the most aperiodic tapping, the highest BST, the highest CDev, the second most switches, and the second highest CV. Sensation had the most switches, the highest CV, and the second most aperiodic tapping. For Experiment 2, American Beauty showed the least down-beat tapping, the most aperiodic tapping, the most switches, the highest BST, the highest CV, the second most off-beat tapping, and the second highest CDev. Sensation showed the most off-beat tapping, the second least down-beat tapping, the second most aperiodic tapping, the most switches, and the second highest CV. These were the only two excerpts that showed consistently poor performance. American Beauty may show poor performance because the right hand contained frequent long notes and changing rhythmic patterns, and because it was the only excerpt to begin with a
pickup in both left- and right-hand parts. The reason for the poor performance on *Sensation* is less clear, although it is possible that the relatively large number of different sections in this excerpt, or the fact that the right hand began on the second sixteenth-note metrical position, caused difficulty.

In both experiments, we detected interactions between pitch information and excerpt for down-beat tapping. For Experiment 1, this was mostly due to less tapping on the downbeat for *American Beauty* when pitch information was absent. In Experiment 2, *American Beauty* and *Blue Goose* showed less down-beat tapping for the monotonic than the pitch-varied versions. *American Beauty* was the only excerpt beginning with a left-hand pickup on the next to final sixteenth-note position (upbeat), and with a right-hand pickup on the final sixteenth-note position of the measure, thus eliminating the opening note as a cue to the downbeat in all stimulus versions. Accounting for the three-way interaction of pitch information, full/right-hand, and excerpt in Experiment 2, *American Beauty* had a relatively high percentage of down-beat tapping for the full pitch-varied version only. This shows that subjects only overcame the pickups in *American Beauty* when pitch information was present in the left hand. In contrast, *Blue Goose* showed a decrease in down-beat tapping when pitch information was absent regardless of whether the left hand was present. This suggests that right-hand pitch cues carry information about the downbeat in this excerpt. The most obvious right-hand pitch cue in *Blue Goose* is the presence of a melodic pattern that opened many of the measures with three consecutive intervals of \(-1, +1, \text{ and } -5\) semitones (i.e., descending minor second, ascending minor second, and descending perfect fourth).

In Experiment 2, the interaction of full/right-hand and excerpt for down-beat tapping resulted from *American Beauty* and *Sensation* showing larger deleterious effects of removing the left hand than the other excerpts. For *American Beauty*, this is attributable to the right-hand characteristics described earlier. For *Sensation*, this is most likely because the right hand began with a note on the second sixteenth-note, a very weak metrical position. However, the presence of the left hand in *Sensation*, which began on the downbeat, seems to have enabled subjects to find the downbeat in the full versions. The interaction between full/right-hand and excerpt with aperiodic tapping in Experiment 2 occurred largely because each excerpt showed a different increase in aperiodic tapping when the left hand was absent. The largest increase was in *American Beauty*, consistent with the right-hand characteristics described above. With switches, we observed a significant interaction between full/right-hand and excerpt in Experiment 2, mostly due to *American Beauty* exhibiting a more deleterious effect of removing the left hand than the other excerpts. For CV, the interaction between full/right-hand and excerpt in Experiment 2 was mostly due to different sized
increases in CV when the left hand was absent, depending on the excerpt. Finally, the significant interaction between full/right-hand and excerpt for CDel in Experiment 2 resulted from a slight increase when the left hand was absent for *Lily Queen*, while the other excerpts showed a slight decrease. This is difficult to explain because *Lily Queen* does not stand out from the other excerpts, except in its relatively fast tempo.

**Music Analysis**

In order to isolate musical dimensions that may be cues to pulse in the ragtime pieces, we analyzed the full pitch-varied versions from Experiment 2 on a number of features. The first two features were general aspects of rhythmic structure relevant to pulse finding. These were right-hand syncopation (defined as a half-measure pattern in which emphasized notes occur on weak sixteen-note metrical positions and one or more strong sixteen-note metrical positions contain rests or tied notes), and presence of the standard ragtime left-hand pattern (defined as a pattern of alternating low and high eighth notes with no faster notes in between). Across the excerpts, 40% of half-measures in the right hand contained syncopation, and 67% of the left-hand half-measures contained the standard ragtime left-hand pattern. This coarse analysis fits well with the general notion that ragtime usually has a regular left-hand part and frequently contains syncopation in the right hand.

The next analysis considered the distributions of musical events at each sixteen-note metrical position. These may provide cues to the phase of the metrical structure, that is, the location of relatively strong beats within the metrical period. We calculated average event distributions at the sixteenth-note level for the following musical dimensions: number of right-hand note onsets, right-hand chords (0 = absent, 1 = present), melody note duration in sixteenth notes, melody note pitch value (60 = middle C, 1 scale unit = 1 semitone), melodic interval size in semitones, melodic contour change (0 = absent, 1 = present), number of left-hand note onsets, and left-hand chords (0 = absent, 1 = present). For melodic interval size, the interval is attributed to the metrical position of the second tone constituting the interval (e.g., with notes C₅ D₅, the interval occurs at D₅). A melodic contour change is attributed to the metrical position where the middle of three tones defining the change occurs (e.g., with notes C₅ D₅ C₅, the contour change occurs at D₅). The first three rows of graphs in Figure 11 show the average event distributions across excerpts for the right-hand musical dimensions. Since subjects tapped primarily on the first and fifth sixteenth-note metrical positions (i.e., the downbeat), we looked for peaks at these positions. In the
Fig. 11. Mean distribution of musical events by sixteenth-note metrical position across the Experiment 2 full pitch-varied stimuli. Metrical positions 1 and 5 correspond to the downbeat, and positions 3 and 7 correspond to the upbeat. From left to right, starting with the top row: number of right-hand (RH) note onsets, right-hand chords, melody note duration in sixteenth-notes, melody note pitch value, melodic interval size in semitones, melodic contour change, number of left-hand (LH) note onsets, and left-hand chords.

right hand, modest peaks at the first metrical position appeared for right-hand note onsets, right-hand chords, and melody note duration. In addition, right-hand note onsets, right-hand chords, and melodic interval size showed peaks at the fifth metrical position. The large intervals on the fifth metrical position could form part of a larger melodic pattern that has a consistent phase structure in ragtime. Among the right-hand musical dimensions, only the note onsets and chords contained peaks at both the first and fifth metrical positions, and no pitch features showed a similarly clear pattern. For the left hand, the bottom row of Figure 11 displays results of the event distributions for note onsets and chords. In contrast to the right-hand musical analyses, the left-hand note onsets and chords showed a very clear pattern of peaks. Strong peaks appeared at the odd metrical positions
for both musical dimensions. There were very few left-hand onsets and no chords on even metrical positions.

Next we employed auto-correlation analysis as a way of discovering repeating patterns that may serve as cues to the periodic metrical structure (Brown, 1993; Vos et al., 1994) for each of the excerpts (for a theoretical account of autocorrelations, see Cryer, 1986). Autocorrelations detect periodicities in a series of numbers by correlating the series with itself using relative lags (or shifts) of the time series. The results are displayed in autocorrelograms (here averaged across excerpts), which plot correlation coefficients as a function of lag. A peak in an autocorrelogram at a lag of 4 indicates a periodicity in the music at that duration (i.e., a period of 4 sixteenth notes). In the present analysis, correlations at lags of 2, 4, and 8 were of interest because these corresponded to observed metrical distances between taps. We analyzed the same musical dimensions as in the metrical position analysis for each excerpt.

The first three rows of graphs in Figure 12 show the average right-hand autocorrelation coefficients across excerpts for lags of 1–8 sixteenth notes. After Bonferroni corrections for multiple t tests, no peaks significantly different from 0 appeared at two, four, or eight sixteenth notes for note onsets, chords, note duration, pitch value, or contour changes. The only significant peak in the right-hand dimensions was at four sixteenth-notes for melodic interval size, \( t(7) = 6.12, p < .001 \). In contrast, the bottom row of graphs in Figure 12 shows peaks significantly different from 0 at periods two, four, and eight for left-hand note onsets and chords (\( ps < .05 \)) except in the case of left-hand chords at two sixteenth notes, \( t(7) = 3.48, p > .05 \). It is noteworthy that four sixteenth notes is the lag with the highest correlations in the two left-hand autocorrelograms, corresponding to the predominant tapping period of a half-measure.

To summarize, the musical analyses across excerpts demonstrated a large amount of syncopation in the right hand and a high prevalence of the left-hand pattern of eighth notes. This is consistent with conventional descriptions of ragtime music (Berlin, 1980, 1985; Jasen & Tichenor, 1978). It is also consistent with the behavioral finding that removing the left-hand notes caused significant disruptions in tapping performance. Event distributions and autocorrelational results for the right-hand note onsets, right-hand chords, melody note duration, melodic interval size, and melodic contour change indicated that these are sometimes modest cues for where to tap in the measure (phase) and how often to tap (period). In contrast, the event distributions and autocorrelations for left-hand note onsets and chords indicated these to be very strong cues to phase and period.
General Discussion

We took two approaches to understanding the musical cues used for pulse finding. The first approach was to experimentally manipulate the musical materials. Both experiments contained pitch-varied and monotonic versions of the same excerpts. The results showed remarkably little difference in performance between the two versions. The second experiment presented versions with both left-hand and right-hand parts of the music or only the right-hand part. Removing the left-hand part produced noticeable decrements in performance. The second approach analyzed the music for the presence of cues to pulse in the right-hand and left-hand parts, especially those that corresponded to the predominant period and phase of
subjects’ tapping. Together, these approaches uncovered some possible features important to pulse finding, especially in the left-hand part.

The relatively small effect of removing pitch variations for most excerpts directed the focus to temporal cues that would be available in both pitch-varied and monotonic versions. Some candidate cues in the right-hand part were the relatively frequent note and chord onsets on the downbeat (compared with the upbeat), the relatively long duration of these notes, and the relative size of melodic intervals at these positions. The left-hand part was characterized by a succession of events at the eighth-note level, producing periodicities corresponding to the eighth-note, quarter-note, and half-note levels. In combination, these cues appeared sufficient for rapidly finding and maintaining the pulse. However, when the left-hand part was absent, performance deteriorated. This is understandable because the right-hand part was frequently syncopated, and because the right-hand cues found in the musical analyses were extremely weak or nonexistent, depending on the specific right-hand musical dimension, compared with the left-hand cues.

The finding that musically inexperienced subjects performed similarly to musically experienced subjects contrasts with results of one recent study that used classical piano music (Drake et al., 2000). Whether this discrepancy is due to differences in stimulus materials or the performance measures used is currently unclear and therefore deserves systematic investigation. Such an investigation could serve to establish what aspects (e.g., metrically consistent tapping) of pulse finding vary with musical experience and how they might depend on musical style.

The two experiments showed that subjects performed well in absolute terms when tapping to music. Based on the performance of musically experienced subjects on the full pitch-varied versions, we propose the following benchmarks for pulse finding in this style of music: (1) 90% of taps on the downbeat, (2) no off-beat tapping, (3) less than 1% aperiodic taps, (4) an average of 0.1 switch between tapping modes per excerpt, (5) 90% tapping at ~600 ms ITI, (6) 4 beats to start tapping (BST, about 2400 ms), (7) 4% CV (about 24 ms), (8) ±1% CDel (about 6 ms behind or ahead of the beat), and (9) 4% CDev (about 24 ms).

These benchmark values may differ with other styles of music, but the findings are generally in line with previous results. That subjects prefer to tap with a 600 ms ITI is consistent with studies of spontaneous tapping and other rhythmic behaviors (Clarke, 1999; Fraisse, 1982; van Noorden & Moelants, 1999). Subjects in the current study showed BSTs similar to values reported by Fraisse (1982) for human synchronization with isochronous patterns and similar to some rule-based pulse-finding models (Desain & Honing, 1999). The range of typical values across these reports is 2–4 beats. Once the tapping began, it was remarkably stable. Switches in tap-
ping mode occurred infrequently and often appeared at transitions between sections of the music. Other behavioral data on switches in tapping mode are not available for comparison, to our knowledge. However, the relative instability of some pulse-finding models suggests there may be a role for top-down processes in maintaining a tapping mode once it is established (Scheirer, 1998).

The typical variability of ITI observed in this study (CV = 4%, or about 24 ms) is similar in magnitude to other studies of synchronization (Scheirer, 1998). Subjects’ tapping to isochronous auditory patterns typically anticipates the beat by around 40 ms (Mates et al., 1992; Vos et al., 1995), although when the pacing intervals are filled with other tones, anticipation is greatly diminished (Wohlschlager & Koch, 2000). This latter fact is consistent with the current study, which found no clear anticipation in music with multiple notes per tapped pulse. The average tap delay was small, about 6 ms (behind the beat) in Experiment 1 and about −4 ms (ahead of the beat) in Experiment 2.

Although the precise relationship between the perception of musical pulse and other rhythmic behaviors is unclear, pulse perception may share similar or identical mechanisms with limb coordination and time perception (Ivry & Hazeltine, 1995; Large & Jones, 1999; McAuley, 1995; McAuley & Kidd, 1998). It should be noted that others have argued, based on empirical studies of different rhythmic behaviors, that a common timing mechanism does not underlie motor control across tasks (e.g., Robertson et al., 1999). Furthermore, researchers disagree on the precise mechanism supporting temporal cognition in the 300–1500 ms range, especially whether the phase is represented in addition to the period of time intervals (Gibbon, Malapani, Dale, & Gallistel, 1997; Ivry, 1996; Large & Jones, 1999; McAuley, 1995).

We now provide a brief description of pulse-finding models. It is important to note that the current study does not provide evidence for one type of pulse-finding mechanism over another. However, data such as ours might be used to test the various types of models. One early class of models for meter and pulse applied rule-based algorithms that use information such as melodic patterning and the location of long notes to create abstract hypotheses about the period and phase of the pulse (e.g., Desain & Hoving, 1999; Longuet-Higgins & Lee, 1982; Steedman, 1977). These models can be extremely useful for testing the contributions of different types of information to pulse and meter. The main limitation with rule-based models is that they do not address how we accommodate for the variability found in human musical performance (Palmer, 1997). This limitation also applies to the approach of Parnuccett (1994). His psychoacoustic model quantifies the relationship between stimulus characteristics and perceived accent and pulse salience, relationships that can be used by other types of pulse-finding models (e.g., Toivainen, 1998).
An influential theory of rhythmic cognition is that of Jones and her colleagues, known as Dynamic Attending Theory (Jones, 1976; Jones & Boltz, 1989; Large & Jones, 1999). This theory postulates that people synchronize with an external pattern by generating internal rhythmic expectancies for when events will occur. The period and phase of these expectancies, or attentional rhythms, can adapt to correct for temporal variability in the attended pattern. Influenced by this notion, a number of computational models use adaptive oscillators as basic computational units (Eck, Gasser, & Port, 2000; Gasser et al., 1999; Large & Jones, 1999; Large & Kolen, 1994; McAuley, 1995; McAuley & Kidd, 1998; Toiviainen, 1998). Note that these adaptive models are different from the theory of pulse sensation proposed by van Noorden and Moelants (1999), who used a damped linear oscillator (e.g., a mass and spring system). For simulating musical pulse, an adaptive oscillator is particularly attractive because it can adjust its period and phase in response to temporal deviations from isochrony. This property, in principle, would enable listeners, performers, and dancers to synchronize with performed music in real time. However, obtaining direct behavioral evidence for oscillators has been difficult (McAuley, 1995; McAuley & Kidd, 1998; Pashler, 2001; but see Large & Jones, 1999). In addition, some investigators argue that timing may instead be an emergent property of coded trajectories within a distributed neural system or that it depends on spatially coded interval representations (Buonomano, Hickmott, & Merzenich, 1997; Conditt & Mussa-Ivaldi, 1999; He, Hashikawa, Ojima, & Kinouchi, 1997; Ivry & Hazeltine, 1995; Keating & Thach, 1997; Penhune, Zatorre, & Evans, 1998; Scheirer, 1998; Todd, 1994; Todd et al., 1999; Wright, Buonomano, Mahncke, & Merzenich, 1997).

Another approach to pulse finding relies on linear filtering of the acoustic signal in order to extract periodicities (Scheirer, 1998; Todd, 1994; Todd et al., 1999). Such an approach has the advantage of being motivated by physiology, in particular the filtering characteristics of sensory systems. Applications of this approach have shown how the complex temporal structure of music (i.e., grouping and meter) can be recovered by passive filtering mechanisms (Todd, 1994; Todd et al., 1999), how human musical performances can be tracked without assuming nonlinear oscillators (Scheirer, 1998; Todd et al., 1999), and how uninterrupted summation of tone processing can explain durational accent (Todd, 1994; Todd et al., 1999). Since filtering models and oscillator models can both track temporal variability in music and represent pulse and meter, determining which approach best describes human pulse-finding mechanisms will likely require detailed computational analysis and experimentation.

Most models of pulse assume that pitch information is not vital, an approach largely consistent with the present behavioral results. Several efforts (Desain & Honing, 1999; Eck et al., 2000; Gasser et al., 1999; Large & Jones, 1999; Large & Kolen, 1994; McAuley, 1995; McAuley & Kidd,
have shown that computational models that rely on temporal information can simulate human time perception, determine the pulse and meter of music, and follow temporally modulating patterns. Despite the successes of many of these time-based models and the current study’s findings, it remains a possibility that pitch information is used when it is a reliable cue to pulse, as suggested by the behavioral results found for some excerpts in this study.

Indeed, the studies of Dawe et al. (1993, 1994, 1995) showed the importance of harmonic rhythm (i.e., the temporal position of a chord change) in relatively simple stimuli. Their studies showed that this cue predominantly determined the rhythmic organization when it was compared with temporal duration of tones and melodic grouping. In their studies, however, the various cues were put into conflict. With music, in contrast, we would expect that the cues would generally covary, although further music analysis directed toward this point is needed. Another important issue that music analysis could address is whether harmonic accents are generally as regular as temporal accents. If not, we might expect them to receive relatively little attention in pulse finding. On the other hand, if harmonic accents are as useful as temporal cues, one or both of these cues may be used.

In summary, this study suggested some general benchmarks for pulse finding in humans with piano ragtime music. These abilities are impressive and do not show large improvements with musical training. We hope that data such as these play an increased role by informing computational modeling of sensory-motor synchronization (e.g., the current data have been modeled by Large, 2000; Toiviainen & Snyder, 2000) and encouraging music analysis that more fully characterizes the musical structures on which this ability is based. In addition, studies of various musical styles, more specific and theoretically motivated manipulations of the music, and the use of performed musical stimuli are important aims. The latter issue deserves particular attention (as in Drake et al., 2000) because one of the fundamental accomplishments of synchronizing systems is their ability to accommodate temporal noise, tempo change, and expressive timing.2

References

---

2. The preparation of this article occurred while Joel Snyder was a predoctoral trainee on National Institute of Mental Health (NIMH) training grant T32 MH 19389, “Multidisciplinary Training in Developmental Psychology.” We thank Peter Desain, Ed Large, Steve McAdams, and Neil Todd for their valuable comments on previous drafts of this paper. We also gratefully acknowledge the assistance of Angelica Goss-Bley in running subjects and analyzing data, and Anne Briard in analyzing the music.


### Appendix A

#### Experiment 1 Excerpt Information

<table>
<thead>
<tr>
<th>Title</th>
<th>Duration (s)</th>
<th>Tempo (bpm)</th>
<th>Tempo (IBI)</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pine Apple</td>
<td>84</td>
<td>92</td>
<td>651</td>
<td>AABB</td>
</tr>
<tr>
<td>Glad Cat</td>
<td>72</td>
<td>86</td>
<td>697</td>
<td>IntroAAB</td>
</tr>
<tr>
<td>Lily Queen</td>
<td>99</td>
<td>102</td>
<td>587</td>
<td>IntroAABB</td>
</tr>
<tr>
<td>Sensation</td>
<td>100</td>
<td>96</td>
<td>624</td>
<td>ABACD</td>
</tr>
<tr>
<td>Blue Goose</td>
<td>88</td>
<td>94</td>
<td>639</td>
<td>IntroAAB</td>
</tr>
<tr>
<td>American Beauty</td>
<td>100</td>
<td>97</td>
<td>617</td>
<td>PickupAABB</td>
</tr>
<tr>
<td>Nonpareil</td>
<td>84</td>
<td>95</td>
<td>631</td>
<td>IntroAABB</td>
</tr>
<tr>
<td>Chrysanthemum</td>
<td>88</td>
<td>92</td>
<td>651</td>
<td>IntroAABB</td>
</tr>
</tbody>
</table>

**Note:** Tempo is in quarter-note beats per minute (bpm) and interbeat interval (IBI) in milliseconds. Under Form, Intro stands for a four-measure introduction, Pickup stands for an introduction shorter than one measure, and letters stand for unique sixteen-measure sections.

### Appendix B

#### Experiment 2 Excerpt Information

<table>
<thead>
<tr>
<th>Title</th>
<th>Duration (s)</th>
<th>Tempo (bpm)</th>
<th>Tempo (IBI)</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pine Apple$^*$</td>
<td>41</td>
<td>92</td>
<td>651</td>
<td>AA</td>
</tr>
<tr>
<td>Glad Cat</td>
<td>30</td>
<td>86</td>
<td>697</td>
<td>IntroA</td>
</tr>
<tr>
<td>Lily Queen</td>
<td>41</td>
<td>102</td>
<td>587</td>
<td>IntroAA</td>
</tr>
<tr>
<td>Sensation</td>
<td>40</td>
<td>96</td>
<td>624</td>
<td>AB</td>
</tr>
<tr>
<td>Blue Goose</td>
<td>29</td>
<td>94</td>
<td>639</td>
<td>IntroA</td>
</tr>
<tr>
<td>American Beauty</td>
<td>41</td>
<td>97</td>
<td>617</td>
<td>PickupAA</td>
</tr>
<tr>
<td>Nonpareil</td>
<td>45</td>
<td>95</td>
<td>631</td>
<td>IntroAA</td>
</tr>
<tr>
<td>Chrysanthemum</td>
<td>46</td>
<td>92</td>
<td>651</td>
<td>IntroAA</td>
</tr>
</tbody>
</table>

**Note:** Tempo is in quarter-note beats per minute (bpm) and interbeat interval (IBI) in milliseconds. Under Form, Intro stands for a four-measure introduction, Pickup stands for an introduction shorter than one measure, and letters stand for unique sixteen-measure sections.

$^*$Data from this excerpt not included in analyses.