Detectability of duration and intensity increments in melody tones: A partial connection between music perception and performance

BRUNO H. REPP
Haskins Laboratories, New Haven, Connecticut

Two experiments demonstrate positional variation in the relative detectability of, respectively, local temporal and dynamic perturbations in an isochronous and isodynamic sequence of melody tones, played on a computer-controlled piano. This variation may reflect listeners' expectations of expressive performance microstructure (the *top-down hypothesis*), or it may be due to psychoacoustic (pitch-related) stimulus factors (the *bottom-up hypothesis*). Percent correct scores for increments in tone duration correlated significantly with the average timing profile of pianists' expressive performances of the music, as predicted specifically by the top-down hypothesis. For intensity increments, the analogous perception–performance correlation was weak and the bottom-up factors of relative pitch height and/or direction of pitch change accounted for some of the perceptual variation. Subjects' musical training increased overall detection accuracy but did not affect the positional variation in accuracy scores in either experiment. These results are consistent with the top-down hypothesis for timing, but they favor the bottom-up hypothesis for dynamics. The perception–performance correlation for timing may also be viewed as being due to complex stimulus properties such as tonal motion and tension/relaxation that influence performers and listeners in similar ways.

Music played by human performers, Western tonal art music in particular, exhibits rich and finely differentiated variation that cannot be captured by conventional notation. This variation contributes vitally to the naturalness, expressiveness, and individuality of a performance. Collectively, it is known as *expressive microstructure* (Clynes, 1983). Its two most important dimensions are *agogics* and *dynamics*. The agogic (or timing) microstructure represents continuous modulations in local tempo or tone interonset intervals, whereas the dynamic (or intensity) microstructure represents the pattern of relative tone intensities (see Todd, 1992, 1995). The variation is not random but, to a large extent, rule governed, despite much individual variability (see, e.g., Gabrielsson, 1987).

The primary purpose of performers' expressive devices is to elucidate the musical structure (Clarke, 1985; Palmer, 1989) and to create an allusion to physical or biological motion within this structural organization (Todd, 1992, 1995). Musically experienced listeners have corresponding expectations about how an expressive performance of a particular composition should be shaped. Experienced musicians' tacit knowledge of the rules governing expressive microstructure enables them to play expressively even when sightreading a new piece; experienced listeners' analogous knowledge enables them to appreciate and evaluate a performance, even of music not heard previously (as long as it is in a familiar style). The mental sound image of music imagined, remembered, or read from a score is almost certainly expressive, not mechanically rigid. Musical listeners' specific expectations about the expressive microstructure of a specific piece of music may account for the fact that expressive variation in well-performed music is usually not noticed as such; attention is drawn to the agogics or dynamics only when the variation is excessive or goes in unexpected directions. Researchers working on computer synthesis of expressive performance have also observed this informally.

On the basis of these observations and considerations, Repp (1992b) devised an experimental method to assess listeners' specific microstructural expectations. He presented listeners with multivoiced excerpts from the piano literature, which were played with isochronous timing (i.e., with mechanically regular tone interonset intervals [IOIs]) and legato (i.e., without any silent intervals between tones) on a computer-controlled digital piano. In each of the repeated presentations of an excerpt, one or two nonadjacent IOIs (as well as the tones filling them, to maintain legato articulation) were lengthened by a small amount, and the musically trained listeners' task was to detect and report the position of the lengthened tone(s). All IOIs in each musical excerpt were probed in this way, and the percentages of correct responses were plotted as a function of position to yield a detection ac-

curacy profile (DAP) for each excerpt. A false-alarm profile (FAP) based on incorrect responses was also derived. False-alarm rates were thought to be a more direct (though less reliable) measure of subjects' expectations: An (unchanged) IOI expected to be relatively short should sound relatively long, and hence attract false alarms. Finally, a representative performance-timing profile was obtained from measurements of the expressive timing patterns of expert performances of the music. Repp's hypothesis was that the relative difficulty of detecting lengthening and the relative frequency of false alarms for each tone would both be inversely related to its relative degree of lengthening in a typical expressive performance.[1]

Correct response and false-alarm rates indeed varied dramatically across positions and were positively correlated, indicating variable expectations or perceptual biases. Moreover, the predicted negative correlation between the DAP (and the FAP) and the performance-timing profile was obtained: Lengthening was more difficult to detect in those positions in which musicians were likely to slow down. Since performance timing was related to the musical structure, so was perception: Lengthening was observed in performance *and* was more difficult to detect (in an isochronous context) in metrically accented positions, close to the end of the excerpt, and at the ends of structural units (phrases and subphrases); moreover, both the extent of observed lengthening and the difficulty of detection increased with the depth of the nearest boundary in the hierarchical grouping structure (Lerdahl & Jackendoff, 1983).

These findings seemed to provide impressive evidence for the existence of microstructural expectations, at least with regard to timing (and lengthening in particular). However, the mechanisms by which these expectations reveal themselves in the laboratory remain a matter of speculation. It must be assumed that an isochronous musical excerpt, despite its deadpan quality and repeated presentation, automatically and instantly accesses a mental representation of which the expected microstructure is an integral part. Moreover, the expectations thus generated must interact immediately with veridical perception of timing, either directly by distorting the perceived durations of the IOIs or indirectly by affecting response decisions at some early, unconscious stage. These assumptions, although they are in the spirit of popular interactive processing models, are not without problems. For example, it is not clear why listeners do not establish a deadpan mental representation of the music after hearing it many times in the course of the experiment. In Repp's (1992b) study, there was no indication that the positional effects decreased over time. Also, according to his hypothesis—the *top-down hypothesis*— musically inexperienced subjects should not have well-defined microstructural expectations; yet, his experiments did not reveal a clear effect of musical experience.

A possible alternative account of his findings must therefore be considered—namely, a *bottom-up hypothesis*, according to which the variation in the DAP and FAP arises from psychoacoustic stimulus factors, without any reference to higher level knowledge about musical structure and microstructure (see Drake, 1993; Monahan & Hirsh, 1990). Repp (1992b) made an attempt to assess the role of simple stimulus factors (pitch height and distance, absolute and relative intensity, tone density) in his materials via multiple regression analysis, but without any clear result. Yet such bottom-up variables deserve further attention in view of recent findings that even young infants are sensitive to major phrase boundary cues in music (Jusczyk & Krumhansl, 1993; Krumhansl & Jusczyk, 1990).

The bottom-up and top-down hypotheses are not mutually exclusive, and they are difficult to separate conceptually and methodologically when the music is complex, because there are many bottom-up cues to the higher level structural representations that constrain observed microstructural variations in performance as well as listeners' expectations about these variations. In fact, it is arguable to what extent musical structure is in the sound pattern and to what extent it is a cognitive construct of performers and listeners (see General Discussion).

The purpose of the present study was to reexamine the two hypotheses using simpler musical materials, in which potential bottom-up accounts for variations in detection performance were more limited and could be defined more clearly. Instead of original piano compositions played with expressive dynamics (Repp, 1992b), the present experiments used simple monophonic tunes composed of piano tones of equal duration and intensity (except for tones that were detection targets), at the risk of attenuating the microstructural variations and expectations elicited by the materials and thus undermining the top-down hypothesis. In Experiment 1, the task was again the detection of duration increments, whereas in Experiment 2, the investigation was extended to the detection of intensity increments.

Because the only variable stimulus property (apart from the change to be detected) was pitch, specific bottom-up hypotheses were restricted to effects that pitch may have on perception of relative IOI duration or on the relative loudness of piano tones. In principle, such effects can take two forms: The pitch variation can result in variations in *sensitivity* across positions in the tune, such that changes in duration or intensity are more difficult to detect in some positions than in others; or it can introduce position-specific *perceptual bias*, such that some IOIs (tones) are perceived as a priori longer or louder than others. Both effects will affect the DAP, but only bias will affect the FAP as well. Thus the bottom-up hypothesis can account for a positive correlation between the DAP and the FAP (indicating variation in bias), but it is also compatible with the absence of such a correlation (indicating variation in sensitivity only).[2] The top-down hypothesis, on the other hand, necessarily implies a directional bias and hence is only compatible with a positive DAP–FAP correlation.

The major prediction of the top-down hypothesis is the negative correlation between the DAP and the per-

formance profile. The bottom-up hypothesis does not predict such a correlation and has difficulty accounting for it. The correlation would have to be either coincidental or due to performers' attempts to compensate for perceptual biases introduced by bottom-up factors (Drake, 1993). At first blush, this seems implausible: Agogic and dynamic variation in expressive performance is generally much larger than seems necessary from this viewpoint, and performers do not generally have the goal of making their performance seem mechanically precise, as a compensation account would imply. However, it could be that bottom-up perceptual effects provide the seeds from which expressive strategies sprout as a form of deliberate exaggeration or overcompensation. While this suggestion is quite speculative, the bottom-up hypothesis deserves attention precisely because it could offer *explanations* for some expressive conventions. The top-down hypothesis, within the present experimental context, takes these conventions as given and inherent in cognitive structural representations of the music.

Specific bottom-up hypotheses for the present tasks can be derived from the existing psychoacoustic literature, despite considerable differences in stimuli and methodology. Psychoacoustic research characteristically uses extremely simple stimuli and highly practiced listeners. For example, studies of duration discrimination typically present silent intervals delimited by the onsets of very brief sounds, so that IOI and silent-gap duration covary. In the present materials, however, long (600-msec), gradually decaying tones of varying pitch followed each other without intervening silence, and tone duration covaried with IOI. Moreover, the present listeners received no special training and were faced with high uncertainty about the location of the change to be detected.

An increase in the difficulty of duration discrimination with the pitch distance between two marker tones has been obtained in many studies using very short silent intervals (e.g., Collyer, 1974; Divenyi & Danner, 1977; Fitzgibbons, Pollatsek, & Thomas, 1974; Formby & Forrest, 1991; Neff, Jesteadt, & Brown, 1982; Perrott & Williams, 1971; Williams & Perrott, 1972). These intervals were an order of magnitude shorter than the filled IOIs in the present musical paradigm; also, short gaps are generally perceived as interruptions (i.e., as offset–onset intervals) rather than as onset–onset intervals. Moreover, Divenyi and Sachs (1978) found that the effect of pitch distance on the discrimination of silent intervals decreased with interval duration and was essentially absent at durations beyond 50 msec. Therefore, the relevance of these results to the present study is questionable.[3] However, there are some indications that pitch distance effects also occur at longer IOIs.

Hirsh, Monahan, Grant, and Singh (1990, Experiment 2) presented their subjects with sequences of six 20-msec tones at IOIs of 200 msec and determined the just-detectable delay in the onset of a single tone, which sometimes also differed in pitch from the other tones. The pitch difference tended to raise the discrimination threshold, but not consistently so; there were complex interactions with position in the sequence, and with the direction and magnitude of the pitch change. In a much earlier study with three-tone sequences, Divenyi (1971) already observed that the detectability of timing perturbations was not a simple function of the frequency separation between tones. In particular, he found that lengthening of the silence between tones was more difficult to detect when the frequencies formed a simple ratio (i.e., a common musical interval). These effects, however, again tended to wash out at slower rates of presentation.

In a recent experiment similar in motivation to the present study, Drake (1993) presented listeners with simple melodic sequences composed of six 50-msec pure tones at IOIs of 300 msec. The sequences contained either two pitch jumps (C-C-G-G-C-C) or a pitch turn (D-E-F-E-D-C). The (untrained) subjects' task was to detect and locate changes in duration (both increments and decrements) in any of the five IOIs. Changes at pitch-jump locations were more difficult to locate (but not more difficult to detect) than changes in other positions; there was no effect in the pitch-turn sequence. Thus there is no very clear evidence so far that pitch distance has an effect on *sensitivity* to temporal change at relatively long IOIs.

More convincing evidence that pitch distance can create a *perceptual bias* comes from studies of the auditory kappa effect (Crowder & Neath, 1995; Shigeno, 1986, 1993). In this paradigm, listeners are asked to compare the durations of two time intervals delimited by three tones of different frequency. The consistent finding is that when the frequency of the second tone is closer to that of the first tone than to that of the third tone and the two time intervals are equal, subjects perceive the first time interval to be shorter than the second. The interval durations in these tasks were comparable to those employed in the present study, but the tones were separated by silence rather than being contiguous. It is not clear whether the kappa effect applies to the IOIs or to the silences between tones.

Despite these difficulties of generalization, there seems to be only one reasonable bottom-up hypothesis for the present duration-increment detection task: An increment in IOI duration may be more difficult to detect if the tones delimiting it are widely separated in pitch than if they are close in pitch. This could be due either to reduced sensitivity or to a bias (namely, the kappa effect), with different consequences for the DAP–FAP correlation. There should be a negative correlation between the DAP and the absolute pitch distances (i.e., regardless of direction) between successive tones in the tune.

As to the possible effect of pitch distance on intensity discrimination, there is surprisingly little relevant psychoacoustic literature. Nearly all intensity-discrimination tasks have used carriers with identical spectral characteristics. Dai and Green (1992) have demonstrated that intensity differences between two successive pure tones are more difficult to detect when the tones fall into different critical bands. However, it is not known whether this finding would generalize to complex tones differing

in fundamental frequency, whose spectra overlap extensively.[4] Drake (1993) included an intensity-discrimination task in her study (referred to above) and found poorer detection of intensity increments, but better detection of intensity decrements, on the high notes in her pitch-jump sequence. This suggests a perceptual bias to perceive higher tones as less loud than lower tones, but again, the generalizability of these results to complex musical sounds is not guaranteed. It is noteworthy, however, that the generative rules for music performance developed by Sundberg and his collaborators include a 3-dB/octave increase in sound level with pitch (see Friberg, 1991; Sundberg, 1988). This rule could represent either a deliberate effort to compensate for a reduced perceived loudness of higher complex tones (a bottom-up effect) or an attempt to satisfy listeners' expectations about typical performance dynamics (a top-down effect). Such expectations could derive from a correlation of pitch and dynamics in music performance.

It may be hypothesized, then, that intensity increments will be more difficult to detect in a high tone than in a low tone, and perhaps also that detection scores will be in inverse proportion to the pitch distance from the preceding tone. These bottom-up hypotheses predict negative correlations between the DAP and the absolute pitches of, or the absolute pitch distances between, successive tones in the tune. While pitch distance may affect sensitivity, absolute pitch may cause a bias to hear lower tones as louder. However, if there is also a tendency to play higher tones louder in performance, the resulting negative DAP–performance correlation would be compatible with either a bottom-up or a top-down account.

Additional, local bottom-up effects predicted for each task are that detection of a change in the first and last IOI or tone of a melodic sequence should be impaired, due to the absence of one adjacent IOI or tone for comparison. In duration-increment detection, the final IOI should suffer especially (Hirsh et al., 1990; Monahan & Hirsh, 1990). In intensity discrimination, the initial tone may be more affected. (Moreover, the final long tone was not probed, as explained below.) A gradual increase in detection scores over the first four or five positions may be predicted on the basis of increasing perceptual definition of the standard IOI duration (Drake & Botte, 1993; Ivry & Hazeltine, 1995).

The melodies used included two features that were intended to provide additional fuel for the bottom-up and top-down accounts, respectively. As will be seen shortly, each melody was composed of three similar parts, each reaching an apex at a successively higher pitch and with a larger upward jump to that pitch. This systematic variation of pitch height represented a bottom-up factor that might affect duration and/or intensity increment detection. The two melodies also had almost identical pitches, but they had different metrical properties, induced by differences in the notation and in the exact sequence of pitches. Metrical structure (by which is meant here the placement of the theoretical downbeats in the pitch sequence) was a pure top-down variable, since no tempo-

ral or dynamic accents were present in the stimuli; therefore, any effect of metrical structure on the DAP and FAP was going to be an additional indicator of top-down expectations, provided that metrical structure also affected performance. If performance microstructure happened to be unaffected by metrical structure, then, according to the top-down hypothesis, listeners should be insensitive to it also. Effects of metrical structure on performance but not on perception, or vice versa, would be inconsistent with the top-down hypothesis. The bottom-up hypothesis, of course, does not predict any effects of meter.

Finally, an important factor relevant to both hypotheses was reexamined in the present study—namely, the role of listeners' musical experience. According to the top-down hypothesis, musically experienced listeners should have more clearly defined expectations about performance microstructure and thus should show a more finely differentiated DAP that is more highly correlated with the relevant performance profile than the DAP of musically untrained subjects. The hypothesis makes no predictions about absolute accuracy, as strong expectations may actually hinder detection. The bottom-up hypothesis, on the other hand, predicts only higher accuracy for musically experienced subjects, because of their training or superior auditory abilities, but no difference in the DAP. As mentioned earlier, Repp (1992b) did not find any very clear effects of musical training on either overall accuracy or on the DAP, despite a wide range of accuracy scores. In precursors to the present experiments, Repp (1992a) found a correlation between musical experience and the overall intensity increment detection score, but no effect on the DAP in either duration or intensity increment detection, which was problematic for the top-down hypothesis. These previous studies used heterogeneous groups of subjects and examined correlations with questionnaire measures of musical experience. Here, a more focused approach was taken by sorting subjects into groups according to musical training and comparing them by means of analysis of variance (ANOVA).

PERFORMANCE MEASUREMENTS

Musical Materials

The two experimental tunes (which may also be considered as two versions of a single tune), labeled Tune A and Tune B, are shown in Figure 1 in musical notation. They were composed by the author with the intention of providing melodies that invited expressive performance. Both began with a staccato note which served to mark the first downbeat in the perceptual experiment. This initial note was followed by a quarter-note rest and a double upbeat in Tune A, but by two quarter-note rests and a single upbeat in Tune B. By defining the duration of the rest, the initial tone thus served as a prime for the metrical structure of the tunes. All subsequent notes were quarter notes, except for the final long note.

The pitches in the two tunes were almost identical; they described three cycles of an up–down pitch motion,
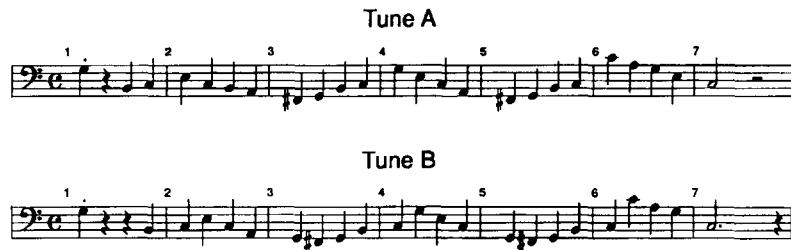
Tune A

Tune B

Figure 1. Musical materials for both experiments. The small digits are bar numbers.

corresponding to three subphrases. As can be seen, however, the turning points in the pitch contour were shifted by one beat in Tune B relative to Tune A: In Tune A, they always occurred on a downbeat; in Tune B, they always occurred on the metrically weak second beat. Similarly, the subphrase boundaries occurred at different metrical points in the two tunes: following Positions 3-2 and 5-2 in Tune A, but following Positions 3-3 and 5-3 in Tune B.[5] However, the subphrase boundaries occurred at the same points within the pitch structure of each tune. In other words, the grouping structure was aligned with the pitch structure, but the metrical structure was shifted with respect to these two.

The three subphrases differed in the height of the upper turning point, which represents an upward excursion of a third in the first subphrase, of a fifth in the second subphrase, and of an octave in the third subphrase. The following note, however, was always a (major or minor) third lower, so the downstep from the pitch peak was essentially held constant. The focus thus was on upward pitch jumps.

The small differences in pitch structure between Tunes A and B, together with the notation and the initial priming tone, served to force listeners into a particular metrical framework. If the pitches had been identical in the two tunes, their metrical structure would have been ambiguous. In the present materials, although a listener could start out hearing one tune with the metrical structure of the other, this interpretation would lead to what are arguably less well-formed melodies, and at the end there would either be an extra note or a missing note. It was expected, therefore, that such an awkward metrical interpretation would be abandoned after a few hearings.

Method

Five pianists performed the experimental tunes. Four of them were graduate students of piano performance at the Yale School of Music, and the fifth was the author, a serious amateur. After a short practice period, each pianist played the two tunes from the notation (Figure 1) three times in alternation, with the right hand, at a moderate tempo, legato, and "with expression." The expressive shaping was done intuitively, without conscious deliberation of the musical structure or microstructure. The different metrical structure of the tunes was obvious from the notation and was not pointed out specifically.

The graduate students played on a Yamaha MX100A Disclavier, which is a real (i.e., mechanical-acoustic) upright piano with added electronic components that enable computer recording and play-

back of performances. The author played on a Roland RD-250s digital piano with Piano 1 sound, monitored over earphones. The onset times and velocities of all keystrokes (as well as their offset times, which are irrelevant here) were registered by a microcomputer in MIDI format. IOIs were calculated from the onset times. Velocities were represented by numbers between 0 and 127 which, in the relevant midrange, correspond to steps of about 0.25 dB in peak rms sound level (Repp, 1993b).

Results and Discussion

The IOIs and velocities were averaged, first over the three repetitions and then across the five pianists' performances of each tune.[6] Figure 2 shows these average timing and intensity profiles. In view of the extensive pitch commonality of the two tunes, their profiles have been superimposed in each panel of the figure. The subphrase boundaries are indicated by vertical dotted lines. The abscissa shows the sequence of musical pitches. Pitches that occur in only one tune are represented by gaps in the other tune's profile. Of course, these pitches had to be omitted from statistical analyses comparing the two profiles.

The average timing profiles (Figure 2a) were fairly flat, except for a pronounced ritardando at the end. Presumably, this was one price to be paid for using such simple materials. The most striking difference between the timing profiles for the two tunes occurred on the last shared pitch (g), which occupied the penultimate IOI in Tune A but the final IOI in Tune B. This difference thus occurred because the final ritardando was differentially aligned with the common pitches, due to the extra note (E) in Tune A. The time course of the ritardando was in fact quite similar in the two tunes. Therefore, the last shared note (g) was omitted from the ANOVA.

A two-way repeated measures ANOVA on the IOIs (with fixed factors of tune, position, and random factor of pianists) revealed significant effects of position $[F(17,68) = 3.30, p < .0003]$ and of tune $[F(1,4) = 33.46, p < .005]$, as well as a position × tune interaction $[F(17,68) = 3.43, p < .0003]$. The position main effect indicates that there was reliable timing variation, apart from the final ritardando. The tune main effect indicates that Tune B tended to be played faster than Tune A, which is of little interest. The interaction, which is the effect of prime interest, seems to be mainly due to a shortening of the IOIs immediately following the subphrase boundaries in Tune A, but not in Tune B. In Tune A, the
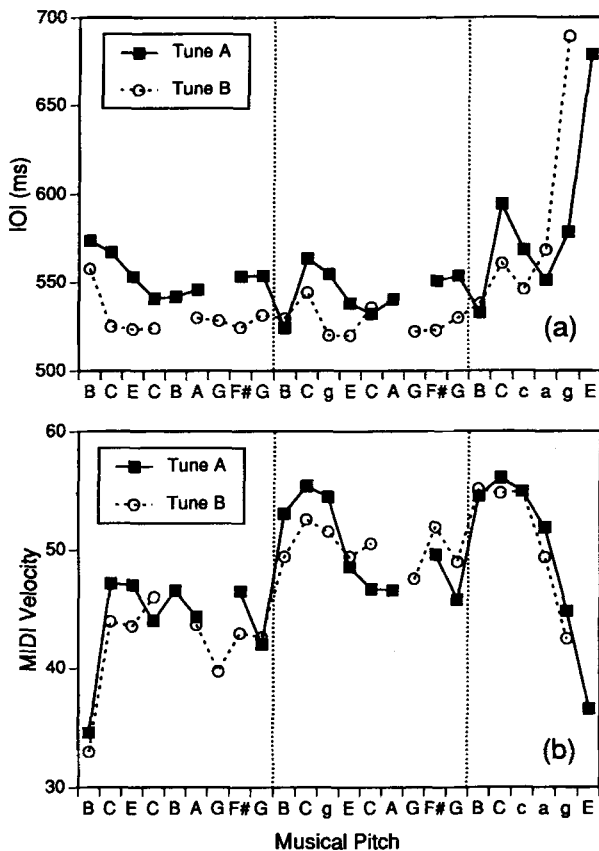
Figure 2. Average interonset intervals (IOIs) and MIDI velocities of five pianists' performances of the two tunes used in Experiments 1 and 2, aligned according to musical pitch (assuming the key of C major). Vertical dotted lines indicate subphrase boundaries. Lowercase letters indicate "an octave higher," not a particular octave.

It appears that the pianists made greater use of dynamics than of agogics in performing the experimental melodies. The intensity profiles (Figure 2b) show striking variation across positions [$F(18,72) = 5.40$, $p < .0001$], but there was no significant tune × position interaction [$F(18,72) = 1.00$], and hence no effect of metrical structure.[7] The profiles show a pronounced peak on the tones involved in the upward pitch step at the beginning of each subphrase, and this peak increases in height from the first to the second subphrase, but little thereafter. Correlations between absolute pitch height and dynamics were positive but small. Somewhat larger, but still nonsignificant, positive correlations were obtained with the absolute pitch distance from the preceding tone. The measure that correlated most strongly with dynamics was the *directional* pitch distance from the preceding tone (Tune A: .57, $p < .01$; Tune B: .43, $p < .05$). Thus there was a tendency to play louder when the pitch went up than when it went down.

To the extent that these performance profiles are representative, they provide an estimate of the expectations that the top-down hypothesis attributes to musically experienced listeners. The virtual absence of effects of metrical structure on performance was surprising, given that the pianists (including the author!) were well aware of the difference and played the tunes in alternation, which should have encouraged contrasting interpretations. Sloboda (1983, 1985) found that pianists could convey metrical structure through performance parameters, but more so through dynamics than through timing. However, his bouncy melodies were of a very different character than the present expressive tunes, which moved much more slowly, at a beat rate close to the optimal pulse (Fraisse, 1982; Parncutt, 1994), with only one tone per beat. This slow event rate and the resulting absence of a hierarchical rhythmic structure may have been responsible for the near absence of metrical effects in the present case. It was predicted, therefore, that meter would have little effect in perception also. Even without metrical effects, there was enough variation in the performance profiles, especially in the intensity profile, to permit a fair assessment of the top-down hypothesis.

EXPERIMENT 1

Method

Subjects. Twenty-four paid volunteers participated in the study. They were divided into three groups of 8: musicians (M), amateur musicians (A), and nonmusicians (N). The subjects in Group M had had at least eight years of formal training on an instrument and still played that instrument.[8] They included 4 graduate students at the Yale School of Music and four Yale undergraduates. They represented various instruments and ranged in age from 18 to 25. The subjects in Group A had had some formal musical training (2–10 years) and were able to read music, but most of them did not play an instrument anymore. They were mostly Yale undergraduates and ranged in age from 18 to 28 years, except for one subject who was 43. The subjects in Group N had had little or no musical training, and only one of them could read music. They ranged in age from 19 to 33 years and were Yale undergraduates or employees.[9]

notes corresponding to these IOIs (Pitch B) initiated double upbeats on the third beat, whereas in Tune B, they formed single upbeats on the fourth beat (see Figure 1). Their shortening in Tune A suggests that they were treated as upbeats to upbeats, as it were. A tendency present in both tunes is the progressive lengthening of the tone preceding the upward pitch jump in each subphrase (Pitch C), as if it took longer to reach a higher pitch. There was also a slight acceleration at the beginning of each tune, which seemed to last longer in Tune A than in Tune B. Somewhat unexpectedly, there was no noticeable ritardando preceding subphrase boundaries. Thus the effects of metrical structure on performance timing were quite limited, being restricted to the onsets of subphrases.

The relationship between IOI duration and the absolute pitch distance between the tones defining the IOI was also examined. These correlations were positive but fell short of significance (.37 and .40 for Tunes A and B respectively). Omission of the extended final IOI made the correlation significant in Tune A (.60, $p < .01$) but not in Tune B (.34). Thus there was a weak tendency to lengthen the IOI between tones far apart in pitch.

Materials. The tunes (Figure 1) were realized on a computer-controlled Roland RD-250s digital piano with Piano 1 sound, legato articulation, a standard IOI duration of 600 msec, and a constant MIDI velocity. The initial staccato tone only helped to establish the meter; the relevant IOIs were those between the subsequent tones. Completely isochronous versions were used for initial familiarization only. Experimental stimuli contained one or two IOIs that were lengthened by delaying the nominal offset of the tone occupying it and the onsets of all following tones in the MIDI instructions. The IOI following a target IOI thus remained unchanged. The lengthening of the tone filling the target IOI was necessary to maintain legato articulation; otherwise, there would have been a salient alternative cue in the detection task.There were 22 target IOIs in Tune A and 21 in Tune B; these were probed in 15 and 14 trials, respectively. Thus about half the trials had two target IOIs that always occurred in different subphrases and were never very close to each other; also, they always occurred in different metrical positions. The assignment of target IOIs to trials was random. The duration increment to be detected was 8.3% (50 msec), 6.7% (40 msec), 5% (30 msec), and 3.3% (20 msec) during four consecutive test blocks, each containing 15 or 14 trials. The four test blocks of progressive difficulty were preceded by three completely isochronous examples of the tune and three demonstration trials with 10% (60-msec) lengthening. There were separate tests for Tune A and Tune B; these tests were identical except for the difference in number of trials. The trials were separated by a silent interval of about 4 sec. To create variety, the tune was randomly transposed from one trial to the next within an octave range centered on the pitches shown in Figure 1. The test sequences were recorded directly from the digital piano onto digital tape.

Procedure. The subjects were tested individually in a quiet room. They listened over Sennheiser HD 540 II earphones and entered their responses on answer sheets that showed the tune in musical notation for each trial. The notated tune was always in C major, but the random transposition of the test stimuli was pointed out to the subjects. Those subjects who could not read music were told that note height represented pitch height, and were asked to follow the score with their pencil as they listened. For the three demonstration trials, the correct responses had already been filled in. If subjects had difficulty hearing the lengthened tones, they were allowed to listen to these trials again. For the subsequent test blocks, the subjects were informed that there could be either one or two lengthened tones on each trial (never the first or last tone), and were asked to circle the note(s) corresponding to the lengthened tone(s). They were specifically asked not to circle the following note (a very common occurrence in the earlier studies of Repp, 1992a, 1992b) and not to guess randomly, but to place a question mark at the end of the line if they could not hear any lengthened tone. Half the subjects in each group listened to the Tune A test before the Tune B test; the others listened in the reverse order. Before the second test, the different metrical structure of the new tune was explained carefully. Each test took about 22 min, and there was a break in between.

Results and Discussion

Overall accuracy. Despite the explicit instructions, the subjects again had a strong tendency to circle the note following the correct one (*late responses*), though individual differences were very large in that respect, and no subject gave late responses exclusively. Therefore, responses were accepted as correct if they were adjacent to the correct position. Overall, among the 49.6% responses scored as correct, there were 32.8% direct hits, 14.5% late responses, and only 2.3% early responses.

There was a difference between the musicians and the other subjects in the incidence of late responses: Late re-

sponses were less than half as frequent in Group M than in Groups A and N, and since musicians gave more correct responses, this difference was even larger in terms of the average percentage of correct responses that were late: 14.8% (range: 1%–48%) in Group M versus 36.9% (range: 8%–89%) in Group A and 36.8% (range: 1%–79%) in Group N. Because of the enormous individual variability, the group difference did not reach significance in an ANOVA. However, 5 of the 8 subjects with rates below 10% were musicians, whereas none of the 6 subjects with rates above 50% was a musician. Two factors may underlie the late-response tendency: (1) Listeners obviously hear the following tone when they realize that a lengthening has occurred, and they may circle the tone they hear instead of backtracking on the answer sheet; (2) they may attribute the perceived hesitation to the following tone because of its delayed onset, or possibly because the delayed onset makes it seem slightly accented (cf. Clarke, 1985).[10] Recent experiments (Repp, 1995c) suggest that both factors play a role.

The overall percentages of correct responses, averaged across positions and tunes, are shown in Figure 3 as a function of duration increment (test block), separately for the three groups of subjects. A repeated measures ANOVA was carried out with the fixed factors of group, block, tune, and order; subjects nested within groups constituted the random factor. Not surprisingly, performance declined significantly across test blocks [$F(3,18) = 168.49, p < .0001$]. There was also a significant main effect of subject group [$F(2,18) = 4.16, p < .04$]: Musicians performed better than amateurs and nonmusicians, with little difference between the latter two groups.[11] Since the guessing probability in this task is very small, it is clear that subjects can detect 3.3% increments with better-than-chance accuracy.[12]

Detection-accuracy profiles (DAPs). Figure 4 shows the DAPs for Tunes A and B, separately for the three groups of subjects. Separate repeated measures ANOVAs were conducted on each tune, with the fixed factors of
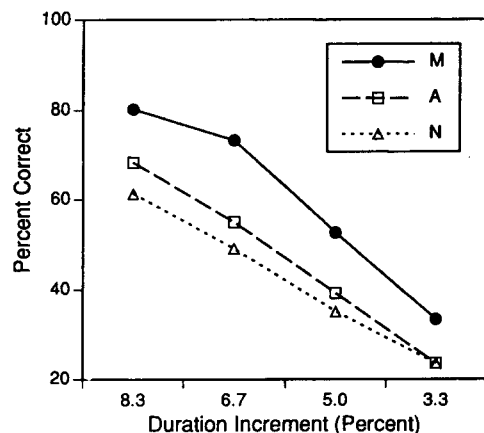

Figure 3. Percent correct scores for the three subject groups in Experiment 1, averaged across positions, as a function of duration increment (test block). M, musicians; A, amateurs; N, nonmusicians.
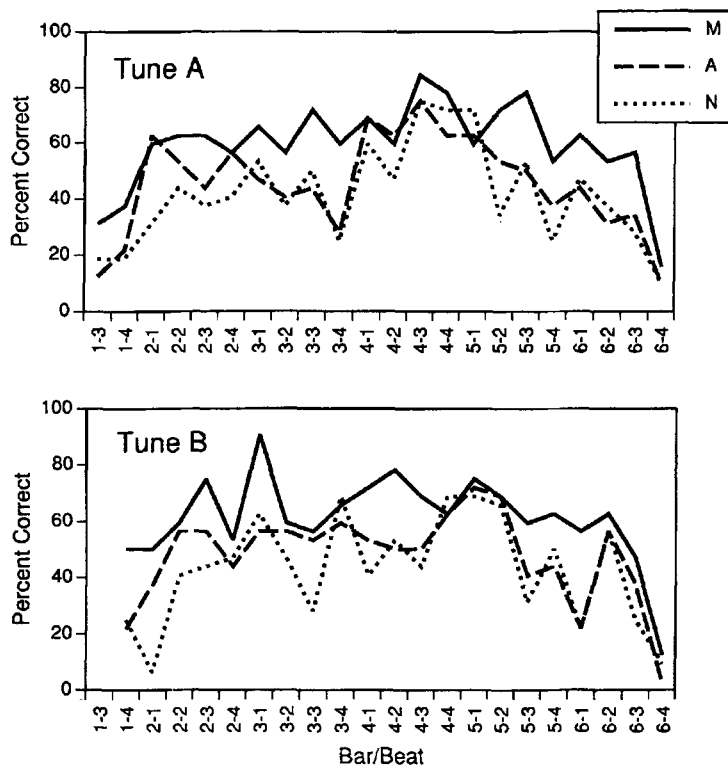
Figure 4. Detection-accuracy profiles for Tunes A and B in Experiment 1, separately for each of the three subject groups. M, musicians; A, amateurs; N, nonmusicians.

group and position and the random factor of subjects nested within groups. Detection scores varied significantly as a function of position in both Tune A [$F(21,441)$ = 11.71, $p < .0001$] and Tune B [$F(20,420)$ = 11.32, $p < .0001$]. There was a significant main effect of subject group for each tune, but the group × position interactions were far from significant. Figure 4 shows that, contrary to the prediction of the top-down hypothesis, the average DAP of the musicians was not more differentiated than that of the nonmusicians—if anything, it was a little flatter.

Figure 5 compares directly the average DAPs for Tunes A and B. The data are collapsed here over all subjects, as there were no significant interactions involving subject groups. The two tunes are aligned here by pitch normalized to the key of C, as in Figure 2. Any pitches that occur in one tune but not in the other are visible as gaps in the graph; they were omitted from the ANOVA reported below. The common subphrase boundaries are indicated by vertical dotted lines.

The similarities between the two profiles are more striking than the differences: Accuracy was low at the beginning, was highest during the second subphrase, and declined during the final subphrase. The final decline replicates earlier findings (Repp, 1992a, 1992b) and may be attributed to listeners' expectation of a final ritardando, in agreement with the top-down hypothesis. The bottom-up hypothesis predicts only the poor performance

in the final position, and not the preceding decline. The poor performance in the initial two positions is consistent with a bottom-up explanation, though a more gradual increase in accuracy was expected. There are small dips in the profiles at analogous points in the second and third subphrase, immediately preceding the upward pitch jumps (Pitch C); the low performance in the second position may also reflect this bottom-up factor. The mag-
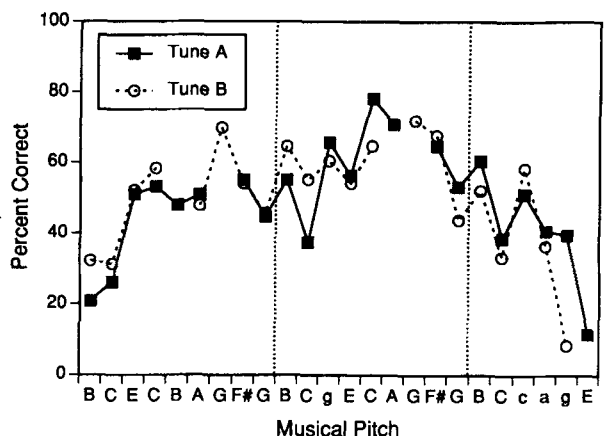

Figure 5. Average detection-accuracy profiles for Tunes A and B in Experiment 1, aligned according to pitch. Vertical dotted lines indicate subphrase boundaries.

nitude of the jump, however, did not seem to matter. Note also that the direction of the effect is not consistent with the kappa effect, according to which IOIs at pitch-jump locations should sound longer to begin with, thus making duration increments easier to detect.

The most obvious difference between the two DAPs is at the last common pitch, g, where detection scores were 31% lower in Tune B than in Tune A. This tone fills the final IOI in Tune B, but it is followed by another IOI in Tune A. For that IOI, performance was as low as it was for the final IOI in Tune B. Note the corresponding (inverse) difference in the average performance-timing profiles (Figure 2a), and its discussion, above. As in the performance analysis, the last common pitch was omitted from the ANOVA.

A repeated measures ANOVA (with fixed factors of group, tune, position, and random factor of subjects nested within groups) yielded, in addition to the expected main effects of position and subject group, a marginally significant tune × position interaction [$F(17,357) = 1.72$, $p < .04$]. The largest remaining difference between the profiles occurred in the second IOI of the second subphrase (Pitch C), in which performance was 18% lower in Tune A than in Tune B. This looks like an effect of metrical structure (the tone falls on a downbeat in Tune B but on the weak fourth beat in Tune A), and it also mirrors an effect noted in the performance-timing profile (Figure 2a); however, the analogous positions in the first and third subphrases show no such difference. Thus, there are no consistent effects of metrical structure. Since metrical effects were also rather weak in performance, this finding does not upset the top-down hypothesis.[13]

As predicted by the top-down hypothesis, the average DAPs (Figure 5) were significantly correlated with the average performance-timing profiles (Figure 2a). The correlation was $-.74$ ($p < .001$) for each tune. It seems that the last data point must have contributed substantially to these correlations. With that data point omitted, however, the correlations were still significant: $-.66$ ($p < .001$) for Tune A and $-.57$ ($p < .01$) for Tune B. A local inverse correspondence may also be seen in the performance and accuracy profiles for Tune A at the beginnings of the second and third subphrases. However, while the performance data suggested a possible effect of metrical structure at this point because Tune B showed less of a timing perturbation than Tune A, the DAPs show similar results for the two tunes. Thus the perceptual effect is probably not metrical in nature; it may reflect expectation of lengthening preceding an upward pitch jump. The height of the jump seemed to matter little, however.

The top-down hypothesis also predicted that musicians would show stronger perception–performance correlations than nonmusicians. A tendency in that direction was in fact obtained, but only when the final data point was included. In that case, the correlations for the three subject groups were $-.79$, $-.66$, and $-.64$ for Tune A, and $-.80$, $-.76$, and $-.52$ for Tune B. With the final

data point omitted, however, this tendency disappeared. Thus it cannot be given much weight.

The main prediction of the bottom-up hypothesis was that the relative detectability of lengthening would be inversely related to the pitch distance between the tones delimiting an IOI. The correlations of the DAPs with this variable (expressed in semitones) were indeed negative, but were significant only for one tune (Tune A: $-.27$, n.s.; Tune B: $-.62, p < .01$). This gives only weak support to the hypothesis. Moreover, since a weak positive correlation between absolute pitch distance and IOI duration was observed in performance, the present negative correlation is compatible with a top-down as well as a bottom-up account.

False-alarm profiles (FAPs). Because of the possibility of two target IOIs on a trial, the subjects were free to give two responses on each trial, if they wanted. However, false alarms (circling of any note not adjacent to a correct position) were relatively infrequent: Relative to the total number of target positions, only 8% false-alarm responses were given; relative to the number of missed targets (50.4%), the false-alarm rate was 16%.[14] Individual differences were considerable, ranging from 0.6% to one exceptional case of 36.9% false alarms relative to the number of target positions. There were no pronounced differences among the three subject groups, but for some reason Tune A elicited consistently more false alarms (8.9%) than Tune B [7.2%; $F(1,21) = 5.83, p < .03$].

The FAPs are shown in Figure 6. Note that this figure plots numbers of responses, not percentages. The profiles for the two aligned tunes were quite similar, except for the very high false-alarm rate on the first g in the second subphrase in Tune A. It is possible that the coincidence of that tone with a downbeat enhanced the false-alarm tendency, but there are no such meter-related differences in other corresponding locations. It appears that the high-pitched notes following pitch skips attracted false alarms, but the octave jump in the third subphrase caused fewer false alarms than the jump of a fifth in the second subphrase. Apart from these locations, false alarms were
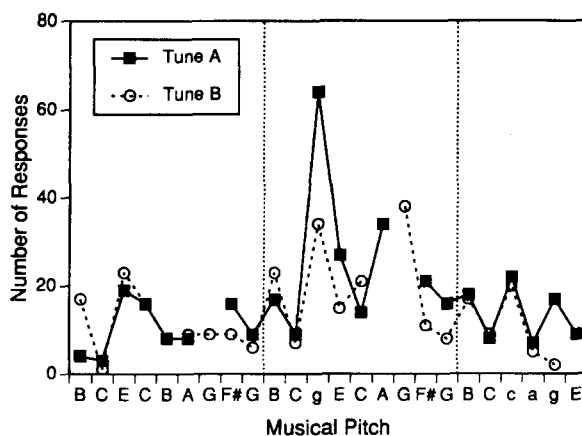

Figure 6. Average false-alarm profiles for Tunes A and B in Experiment 1, aligned according to pitch.

also frequent on Pitches A and G in the second subphrase, where G fell on a downbeat in Tune B, but A fell on the weak fourth beat in Tune A. Thus, again, there is no clear evidence for any effect of metrical structure. Because of the large individual differences in false-alarm rates, no ANOVA was conducted on the FAPs.

The peaks and valleys in the FAPs match those in the DAPs (Figure 5) rather well, though their relative heights differ. The correlations were significant: .57 for Tune A and .60 for Tune B, both $ps < .01$. These correlations are consistent with the top-down hypothesis: False alarms tend to occur on those IOIs that, because they are not expected to be lengthened, sound relatively long to begin with. The correlations would also be consistent with the bottom-up hypothesis if false-alarm rates showed a positive correlation with the absolute pitch distance between tones delimiting the IOIs (i.e., the kappa effect). However, these correlations were *negative* and nonsignificant (Tune A: $-.11$; Tune B: $-.31$). Thus, if anything, the subjects perceived IOIs as being longer when they were between tones close in pitch. This finding suggests that the kappa effect did not operate in the present tunes, and hence it provides no support for the bottom-up hypothesis.

Summary. The results of Experiment 1, obtained with much simpler materials than those used by Repp (1992b), provide additional support for the main prediction of his top-down hypothesis: Detectability of lengthening was negatively correlated with lengthening in performance, and false-alarm frequencies (a more direct but less precise index of listeners' expectations) correlated positively with detection accuracy. Those subjects with extensive musical training performed better overall but did not have more differentiated DAPs, contrary to a secondary prediction of the hypothesis. While disappointing, the finding that effects of metrical structure were generally absent in both perception and performance is not inconsistent with the top-down hypothesis. The bottom-up hypothesis is consistent with these secondary results, but does not provide a convincing account of the main findings. In particular, it provides no explanation of the gradual decline in detection performance toward the end of a tune, and it provides no rationale for the false-alarm distribution because the kappa effect seems to be absent.

EXPERIMENT 2

This experiment was quite analogous to Experiment 1, except that intensity increments rather than duration increments were to be detected. The very distinctive performance-intensity profile (Figure 2b) provided a good basis for re-examining the top-down hypothesis for intensity-increment detection, which had not received much support in a pilot study (Repp, 1992a).

Method

Subjects. Twenty new subjects were recruited from the Yale community and divided into three groups according to the same criteria as in Experiment 1. There were 8 musicians (20–33 years old), 6 amateurs (21–38 years old), and 6 nonmusicians (29–45

years old). The differences in mean age among the groups [$F(2,17) = 4.52$, $p < .03$] were inadvertent (volunteers were taken as they came), but were not considered a serious problem.[15] The subjects were paid for their services.

Materials. The tunes and the test arrangement were the same as in Experiment 1; even the same random sequences were used. The only difference was that intensity increments occurred instead of duration increments.[16] The initial staccato tone and the final long tone were not possible targets. The increment was 11 MIDI velocity units (about 2.75 dB) in the three demonstration trials, and 9 (2.25 dB), 7 (1.75 dB), 5 (1.25 dB), and 3 (0.75 dB) velocity units in the four test blocks. Tunes were again randomly transposed from trial to trial, which was especially important in this experiment, as there was some random variability in peak sound level among digital piano tones of different pitches (Repp, 1993b).

Procedure. The procedure was the same as in Experiment 1, except that the subjects were asked to circle the notes corresponding to tones that seemed louder than the others. The possibility of two targets occurring on one trial was pointed out. Instructions not to circle the following note were omitted, as no such tendency was expected in this task. The order of the two tunes was nearly counterbalanced: In the nonmusician group, 4 subjects happened to listen in one order and 2 in the other.

Results and Discussion

Overall accuracy. As in the previous experiments in this series, a liberal scoring criterion was adopted, with responses to adjacent positions being accepted as correct, even though they were quite rare: Of the 45.7% correct responses, 2.6% were early and 1.9% were late. The decision to count these responses as correct seems justified, for although they probably included some random guesses (as did direct hits, of course), individual and group differences in their frequency suggested that they were at least partially made up of misplaced correct responses. Individual percentages of such responses ranged from 0 to 10.5. They were twice as frequent in Groups A and N as they were in Group M, though that group difference was not significant. The difference in frequency of early and late responses was not significant either. Surprisingly, however, there was a main effect of tune and a tune $\times$ group interaction [$F(2,17) = 7.27, p < .006$]. These effects were due to the fact that nonmusicians were much more likely to misplace their responses in Tune B than in Tune A, for unknown reasons.

The overall percentages of correct responses are shown in Figure 7. Not surprisingly, detection performance declined as the increment got smaller [$F(3,42) = 92.82, p < .0001$]. However, even in the most difficult condition, scores were still above chance, certainly for the musicians.[17] As in Experiment 1, there was a significant difference between subject groups, with musicians performing best and nonmusicians worst [$F(2,14) = 7.58, p < .006$]. Because of the unintended confounding of musical experience with age, age was entered as a covariate in a follow-up analysis. The group difference remained significant [$F(2,14) = 6.56, p < .009$], and the correlation between age and accuracy scores was nonsignificant. (Among the nonmusicians, the oldest subject had the highest score.)

Detection-accuracy profiles (DAPs). The DAPs of the three subject groups are shown separately for each tune
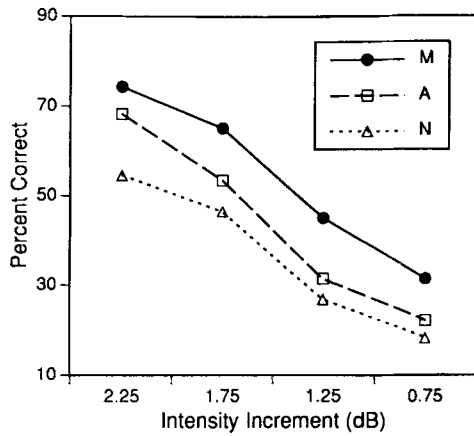
Figure 7. Percent correct scores for the three subject groups in Experiment 2, averaged across positions, as a function of intensity increment (test block). M, musicians; A, amateurs; N, nonmusicians.

in Figure 8. It can be seen that there was dramatic variation in accuracy across positions; the position main effect was highly significant for both Tune A [$F(21,357) = 14.12, p < .0001$] and Tune B [$F(20,340) = 12.34, p < .0001$]. In addition, each tune showed a group main effect [Tune A: $F(2,17) = 4.39, p < .03$; Tune B: $F(2,17) = 8.92, p < .003$]. However, even though Figure 8 seems to suggest that group differences were larger during the

second half of each tune than during the first half, the position × group interaction was far from significant in each case. Again, there is no evidence that musicians showed a more differentiated DAP than amateurs or nonmusicians.

The average DAPs of the two tunes are shown aligned according to pitch in Figure 9. As in Experiment 1, the two profiles were very similar. The subphrase boundaries are indicated by vertical dotted lines. In each tune, intensity increments were difficult to detect at the beginning of a subphrase and much easier during its second half (which was missing in the third, abbreviated subphrase).

The ANOVA on the aligned profiles naturally showed a highly significant effect of position [$F(18,306) = 30.03, p < .0001$], as well as a main effect of subject group [$F(2,17) = 10.37, p < .002$], with no main effect of tune. However, the tune × position interaction was significant [$F(18,306) = 3.67, p < .0001$], indicating that the two functions, while similar, were not identical. Could these differences have been due to effects of metrical structure? Even though such effects were not evident in performance, a possible prediction was that intensity increments should be more difficult to detect on downbeats or in strong metrical positions in general (i.e., on the first and third beats in each bar). Tune A shows a zigzag pattern during the last subphrase that is consistent with this prediction, as the valleys in the DAP coincide
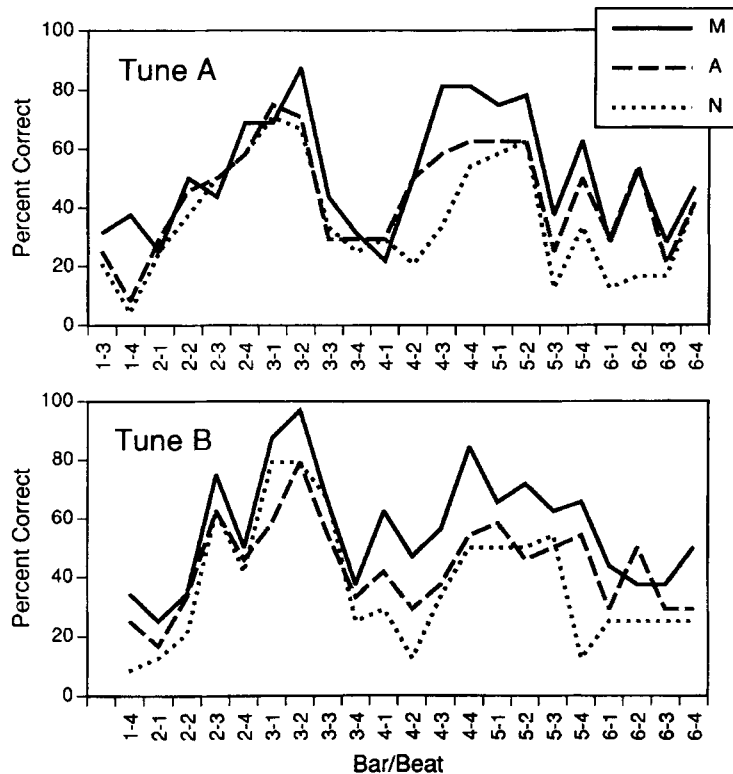

Figure 8. Detection accuracy profiles for Tunes A and B in Experiment 2, separately for each of the three subject groups. M, musicians; A, amateurs; N, nonmusicians.
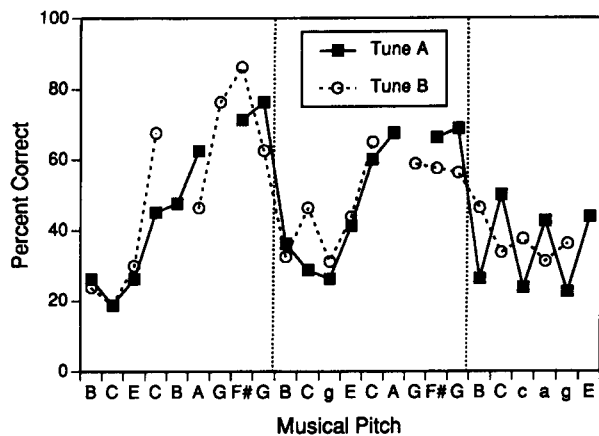
Figure 9. Average detection-accuracy profiles for Tunes A and B in Experiment 2, aligned according to pitch.

with strong metrical positions. However, Tune B does not show the complementary pattern. Moreover, there is no evidence of metrical effects elsewhere in the tunes. Consider, for example, the adjacent pitches C and E in the first subphrase (second and third data points): E fell on the downbeat in Tune A, C on the downbeat in Tune B, yet there is no interaction. The analogous pair of pitches in the second subphrase, C and g, shows a difference on C in favor of Tune B; that difference is in the wrong direction, however, as C was on the downbeat in Tune B and should have shown lower performance than in Tune A. Only the third analogous pair in the last subphrase, C and c, shows the predicted interaction. Overall, however, there is no convincing evidence for effects of metrical structure, which—in view of the absence of metrical effects in performance—is consistent with both hypotheses under consideration. The origin of the significant differences between the accuracy profiles for Tunes A and B is not clear at present.

There appears to be an inverse correspondence between the DAPs and the performance-intensity profiles (Figure 2b) in qualitative terms. However, the perception–performance correlations were small and nonsignificant: −.23 for Tune A and −.05 for Tune B. This was due to the absence of increased detection accuracy at the beginnings and ends of the tunes, where intensity was markedly reduced in performance. Also, the enormous increase in detection accuracy during the first subphrase has no correspondence in performance. Nevertheless, the two broad peaks in the accuracy profiles correspond to valleys in the performance-intensity profile. It is difficult to conclude, therefore, that perception and performance are unrelated, but the parallelism certainly is less striking than in the case of timing.

The partial correspondence between the DAP and the performance-intensity profile could have been mediated by bottom-up factors: either absolute pitch height or the pitch difference between adjacent tones. Accuracy scores indeed correlated negatively with absolute pitch (Tune A: $r = -.68, p < .001$; Tune B: $r = -.49, p < .05$) and with

directional pitch change (Tune A: $r = -.52, p < .05$; Tune B: $r = -.44, p < .05$). Intensity increments thus were more difficult to detect on higher pitched tones and/ or when the pitch went up rather than down. Moreover, it can be seen in Figure 9 that accuracy scores *kept increasing* as the pitch went down (middle portions of first and second subphrases), whereas they were low and relatively stable when the pitch went up. In a stepwise regression analysis, however, pitch change made no significant contribution beyond the effect of absolute pitch height, which showed the stronger correlations with detection scores. The performance intensity profiles, however, had shown stronger correlations with pitch change than with pitch height. Thus pitch height seemed to be more of a factor in perception than in performance, whereas pitch change affected both. Effects of pitch height may be due to lower piano tones being perceived as louder than higher tones of equal intensity, due to their larger spectral bandwidth and slower decay times. This was confirmed in a recent study in which listeners were required to adjust the relative loudness of piano tones differing in pitch (Repp, 1995b), and it probably represents a genuine psychoacoustic effect that is not mediated by musical experience. Effects of pitch change, on the other hand, may reflect a more complex relational stimulus property of *pitch motion*, which may or may not be purely bottom-up in nature (see General Discussion, below).

False-alarm profiles (FAPs). False-alarm responses were much more frequent in this experiment than in Experiment 1.[18] The average percentage, relative to the number of target positions, was 24.8; relative to the number of misses (54.3%), it was 45.7%. That is, almost every other time that a target was missed, a false-alarm response was given. Individual differences were considerable: Individual false-alarm rates ranged from 5.2% to 47.7% (relative to the number of target positions). A few subjects actually had more false alarms than hits. (Note that this does not imply chance performance in the present paradigm.) Musicians gave somewhat fewer false alarms than the other subjects, but the difference was not significant. Surprisingly, however, there was a main effect of tune: False alarms were more frequent in Tune A than in Tune B [$F(1,17) = 4.71, p < .05$], just as in Experiment 1.

The aligned FAPs for the two tunes are shown in Figure 10. It can be seen that the profiles were quite similar, though response rates in some positions were elevated in Tune A relative to Tune B. The two sharp peaks on pitches F♯ (second subphrase) and c (third subphrase) in Tune A correspond to downbeats. This cannot be evidence for a role of metrical structure, however, because the effect goes in the wrong direction: Downbeats should be expected to be louder, if anything, and hence should not attract false-alarm responses. The most striking fact about the FAPs, of course, is their similarity to the DAPs shown in Figure 9. The correlations between these profiles were .69 for Tune A and .72 for Tune B, both $ps < .001$. False alarms were frequent precisely in those positions in which detection of intensity increments was
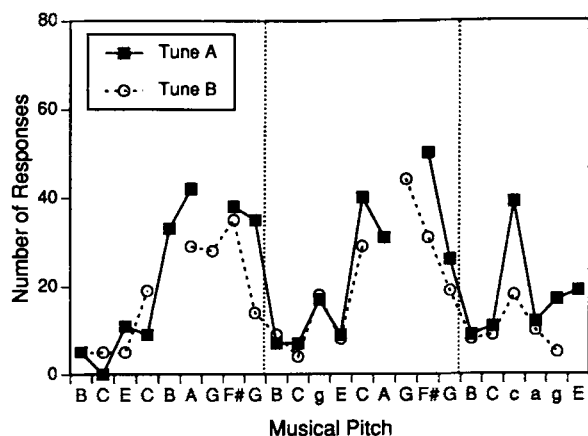
Figure 10. Average FAPs for Tunes A and B in Experiment 2, aligned according to pitch.

easy. This is in accord with the top-down hypothesis, but it could also represent a psychoacoustically based bias. Since pitch height and pitch change have already been identified as relevant stimulus factors, it seems plausible that the false-alarm distribution reflects the operation of these factors as well. Their correlations with the FAPs were not impressive, however, due in part to the lower reliability of the FAPs, which reflected mainly the results of a few subjects with very high false-alarm rates. Still, there were tendencies to give false alarms to low tones (Tune A: $r = -.30$, n.s.; Tune B: $r = -.49$, $p < .05$) and to tones lower than the preceding tone (Tune A: $r = -.13$, n.s.; Tune B: $r = -.37$, n.s.). The strong DAP–FAP correlation thus cannot be taken as unambiguous support for the top-down hypothesis; it is consistent with the bottom-up hypothesis as well.

Summary. As already suggested by Repp (1992a), it appears that detection of intensity increments is not governed by top-down expectations of performance microstructure to the same extent as may be the detection of duration increments. The perceptual results seem to reflect primarily a perceptual bias due to an interaction between pitch and perceived loudness, which may be specific to piano tones. The relation between perception and performance remains unclear in the case of dynamics.

GENERAL DISCUSSION

The present experiments explore a middle ground between traditional psychoacoustics and music perception. They deal with situations that are unusually complex and informal from a psychoacoustic perspective, but unusually primitive and constrained from a musical perspective. Thus they are open to criticism from both sides. Yet it is necessary to begin to fill the yawning gap between these different research traditions in order to better understand the relevance of psychoacoustic findings to the perception of realistic music. The present research used melodies that, even though they were quite simple, lent themselves to expressive and aesthetically pleasing per-

formance and could be listened to in a correspondingly musical mode. Thus they were not merely auditory sequences but also could be expected to call on listeners' aesthetic sensibilities and musical experience. The experimental task, although it was demanding and repetitive, also deliberately deviated from typical low-uncertainty psychoacoustic paradigms in order to get somewhat closer to realistic music perception.

To some extent, the very simplicity of the musical materials counteracted these efforts. However, the purpose of Experiment 1 was to attempt to replicate Repp's (1992b) findings on duration-increment detection with materials that gave only limited room for purely stimulus-based (bottom-up) effects. The experiment was moderately successful in achieving this goal. Although neither the performance-timing profile nor the DAP was as intricately structured as in the earlier study, due to the reduced complexity of the music, they were negatively correlated, and both were positively correlated with the FAP, suggesting an underlying bias. The auditory kappa effect could have provided an explanation for this bias, but there was no convincing evidence that this effect operated in the present melodies. In the absence of such a bottom-up explanation for the perceptual results and of a bottom-up rationale for a connection between perception and performance, the data lend support to the top-down hypothesis that listeners expected a certain timing microstructure, and that these expectations interacted with perception of duration increments. The absence of any effect of musical experience on the shape of the DAP or FAP is troublesome from that perspective, for it suggests that musically untrained people, too, have expectations about expressive timing, even though they have had little exposure to expressively played music. It is possible, however, that film music and the softer styles of popular music provide fairly universal exposure to the basic phenomena of expressive performance. The presence of corresponding effects of metrical structure in both perception and performance would have given a further boost to the top-down hypothesis, but the general absence of clear metrical effects, attributed to the slow tempo and primitive rhythmical structure of the materials, is at least not inconsistent with the hypothesis.

The conclusions from Experiment 2 are different. It appears that perception of intensity increments is governed more by the bottom-up factors of pitch height and pitch distance than by any top-down expectations about dynamic performance characteristics. The absence of a strong perception–performance correlation, the obtained significant correlations of the perceptual results with the relevant stimulus factors, and the absence of effects of metrical structure and musical experience are all consistent with a bottom-up explanation. This does not mean that top-down expectations about dynamics never play a role, only that they could not be convincingly demonstrated in this experiment, despite a highly varied performance-intensity profile that provided a good basis for such expectations. In view of the strong DAP–FAP correlation, the observed bottom-up effects (greater detectability of

intensity increments in low tones, and in tones following higher tones) probably represent variations in perceptual bias (directional) rather than variations in sensitivity (nondirectional). Subsequent studies bear more directly on this distinction (Repp, 1995c).

One reason why perception of timing may be governed more by top-down processes than perception of intensity is that agogics provide a more important correlate of musical structure than do dynamics. Performers use a local slowing of tempo to mark structural boundaries, often in proportion to the depth of the boundary in the grouping hierarchy (Repp, 1992b; Todd, 1985). Dynamics, although they often are correlated with tempo modulations (Todd, 1992), may be only a secondary and less reliable indicator of structure. Since listeners' microstructural expectations presumably are driven by a structural representation of the music heard, they would then naturally carry much stronger biases with regard to timing than to dynamics. Performers, too, presumably have more freedom in varying the dynamic shape of a performance, without obscuring the structure in the process. Constraints on dynamics may arise primarily from the linear pitch sequence as such, rather than from a higher level hierarchical representation.

For heuristic reasons, the attempt has been made in this research to draw a clear distinction between bottom-up (stimulus) and top-down (cognitive) factors. However, this distinction can easily become blurred. For example, as already mentioned, whenever there is a consistent stimulus correlate of some perceptual effect (such as the apparent effect of pitch height on perceived loudness), the correlation can become part of listeners' expectations. Any bottom-up effect thus can simultaneously be a top-down effect, and other arguments (such as parsimony or even more elementary stimulus factors, such as the distribution of energy across critical bands and its effect on the auditory computation of loudness level) need to be invoked to establish the heuristic priority of the bottom-up account. Conversely, what seems like a top-down effect may actually reside in the stimulus structure, because this structure may be richer than has been granted. Ultimately, the division between bottom-up and top-down, like the analogous one between perception and cognition, is a *variable* determined by how much structure a theorist is willing to impute to the stimulus. The more structure is said to be in the world, the less needs to be attributed to the listener's cognitive processes and past experience; the structure is simply picked up, rather than constructed or inferred (Gibson, 1966).

Jones (1976, 1987, 1990; Jones & Boltz, 1989) has been a prominent advocate of such a stimulus-structure-oriented approach to music. According to her, the pitch, loudness, and time dimensions of auditory patterns (including music) are "inextricably bound together and cannot be evaluated separately" (Jones, 1976, p. 329). More recently, she has proposed a two-component model of music performance which combines a rigid "vertical" timing component with a flexible "horizontal" compo-

nent that generates expressive timing variation (Jones, 1987, 1990). The horizontal component "captures the shape and pacing of a melodic line, the velocity profiles of melodic phrases" and is controlled by a "motion generator" (Jones, 1990, p. 227). Jones does not discuss *how* velocity profiles are aligned with the pitch structure of the music, but she leaves little doubt that the alignment is not arbitrary and that it is to a large extent governed by the pitch structure of the melody within structural units ("melodic phrases"). Indeed, the notion that sequences of musical tones have inherent dynamic properties such as tension/relaxation or "tonal motion" has long been a staple of musicologists and philosophers of music (see, e.g., Repp, 1993a; Shove & Repp, 1995; Truslit, 1938; Zuckerkandl, 1956).

Such a view of music—as being invested with holistic and relational properties that go beyond the mere sequence of pitches and that appeal directly to the kinematics of the human body—provides a new perspective on the relationship between perception and performance that is at the heart of the present research. The traditional top-down and bottom-up hypotheses contrasted here basically assume that a causal relationship underlies any observed perception–performance correlation: According to the top-down hypothesis, expectations about performance govern perception; according to the bottom-up hypothesis, perceptual distortions presumably underlie expressive performance strategies. Without the assumption of such causal connections, parallels between perception and performance cannot be explained by these hypotheses. However, an enriched view of the music itself makes it possible to see perception and performance as two parallel kinds of reactions (the listener's and the musician's) to the same complex information; hence the causality goes from the music to both, not from one to the other. Thus, for example, the gradual decline in the DAP toward the end of the musical excerpt as well as the performer's ritardando may both be reflections of the buildup of tension in the music as it approaches its end, which is signaled both temporally and melodically (by the falling pitch and the harmonic cadence implied by the monophonic melody). In other words, the performer slows down because the music *asks for it*, and the listener expects the ritardando for the same reason. Of course, the musical structure does not completely determine perception and performance, but it exerts significant constraints on them, perhaps more so on perception than on performance, and—as the present research suggests—more so on timing than on dynamics.

REFERENCES

BREGMAN, A. S. (1990). *Auditory scene analysis.* Cambridge, MA: MIT Press.

CLARKE, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 209-236). London: Academic Press.

CLYNES, M. (1983). Expressive microstructure in music, linked to living qualities. In J. Sundberg (Ed.), *Studies of music performance*

(Publication No. 39, pp. 76-181). Stockholm: Royal Swedish Academy of Music.

COLLYER, C. E. (1974). The detection of a temporal gap between two disparate stimuli. *Perception & Psychophysics*, **16**, 96-100.

CROWDER, R. G., & NEATH, I. (1995). The influence of pitch on time perception in short melodies. *Music Perception*, **13**, 379-386.

DAI, H., & GREEN, D. M. (1992). Auditory intensity perception: Successive versus simultaneous, across-channel discriminations. *Journal of the Acoustical Society of America*, **91**, 2845-2854.

DIVENYI, P. L. (1971). The rhythmic perception of micro-melodies: Detectability by human observers of a time increment between sinusoidal pulses of two different, successive frequencies. In E. Gordon (Ed.), *Experimental research in the psychology of music* (Vol. 7, pp. 41-130). Iowa City: University of Iowa Press.

DIVENYI, P. L., & DANNER, W. F. (1977). Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. *Perception & Psychophysics*, **21**, 125-142.

DIVENYI, P. L., & SACHS, R. M. (1978). Discrimination of time intervals bounded by tone bursts. *Perception & Psychophysics*, **24**, 429-436.

DRAKE, C. (1993). Perceptual and performed accents in musical sequences. *Bulletin of the Psychonomic Society*, **31**, 107-110.

DRAKE, C., & BOTTE, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics*, **54**, 277-286.

FITZGIBBONS, P. J., POLLATSEK, A., & THOMAS, I. B. (1974). Detection of temporal gaps within and between perceptual tonal groups. *Perception & Psychophysics*, **16**, 522-528.

FORMBY, C., & FORREST, T. G. (1991). Detection of silent temporal gaps in sinusoidal markers. *Journal of the Acoustical Society of America*, **89**, 830-837.

FRAISSE, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149-180). Orlando, FL: Academic Press.

FRIBERG, A. (1991). Generative rules for music performance: A formal description of a rule system. *Computer Music Journal*, **15**, 56-71.

GABRIELSSON, A. (1987). Once again: The theme from Mozart's Piano Sonata in A major (K.331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-103). Stockholm: Royal Swedish Academy of Music.

GIBSON, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

HIRSH, I. J., MONAHAN, C. B., GRANT, K. W., & SINGH, P. G. (1990). Studies in auditory timing: 1. Simple patterns. *Perception & Psychophysics*, **47**, 215-226.

IVRY, R. B., & HAZELTINE, R. E. (1995). Perception and production of temporal intervals across a range of durations: Evidence for a common timing mechanism. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 3-18.

JONES, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, **83**, 323-355.

JONES, M. R. (1987). Perspectives on musical time. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 153-175). Stockholm: Royal Swedish Academy of Music.

JONES, M. R. (1990). Musical events and models of musical time. In R. Block (Ed.), *Models of cognitive time* (pp. 207-240). Hillsdale, NJ: Erlbaum.

JONES, M. R., & BOLTZ, M. (1989). Dynamic attending and responses to time. *Psychological Review*, **96**, 459-491.

JUSCZYK, P. W., & KRUMHANSL, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 627-640.

KRUMHANSL, C. L., & JUSCZYK, P. E. (1990). Infants' perception of phrase structure in music. *Psychological Science*, **1**, 70-73.

LERDAHL, F., & JACKENDOFF, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

MONAHAN, C. B., & HIRSH, I. J. (1990). Studies in auditory timing: 2. Rhythm patterns. *Perception & Psychophysics*, **47**, 227-242.

NEFF, D. L., JESTEADT, W., & BROWN, E. L. (1982). The relation between gap discrimination and auditory stream segregation. *Perception & Psychophysics*, **31**, 493-501.

PALMER, C. (1989). Mapping musical thought to musical performance.
Journal of Experimental Psychology: Human Perception & Performance, **15**, 301-315.

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

PERROTT, D. R., & WILLIAMS, K. N. (1971). Auditory temporal resolution: Gap detection as a function of interpulse frequency disparity. *Psychonomic Science*, **25**, 73-74.

REPP, B. H. (1992a). Detectability of rhythmic perturbations in musical contexts: Bottom-up versus top-down factors. In C. Auxiette, C. Drake, & C. Gérard (Eds.), *Fourth rhythm workshop: Rhythm perception and production* (pp. 111-116). Bourges, France: Imprimérie Municipale.

REPP, B. H. (1992b). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition*, **44**, 241-281.

REPP, B. H. (1993a). Music as motion: A synopsis of Alexander Truslit's *Gestaltung und Bewegung in der Musik* (1938). *Psychology of Music*, **21**, 48-72.

REPP, B. H. (1993b). Some empirical observations on sound level properties of recorded piano tones. *Journal of the Acoustical Society of America*, **93**, 1136-1144.

REPP, B. H. (1995a). Acoustics, perception, and production of *legato* articulation on the piano. *Journal of the Acoustical Society of America*, **97**, 3862-3874.

REPP, B. H. (1995b). *Relative loudness of piano tones differing in pitch.* Manuscript in preparation.

REPP, B. H. (1995c). *Variations on a theme by Chopin: Relations between perception and production of deviations from isochrony in music.* Manuscript submitted for publication.

SHIGENO, S. (1986). The auditory tau and kappa effects for speech and nonspeech stimuli. *Perception & Psychophysics*, **40**, 9-19.

SHIGENO, S. (1993). The interdependence of pitch and temporal judgments by absolute pitch possessors. *Perception & Psychophysics*, **54**, 682-692.

SHOVE, P., & REPP, B. H. (1995). Musical motion and performance: Theoretical and empirical perspectives. In J. Rink (Ed.), *The practice of performance* (pp. 55-83). Cambridge: Cambridge University Press.

SLOBODA, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology*, **35A**, 377-396.

SLOBODA, J. A. (1985). Expressive skill in two pianists: Metrical communication in real and simulated performances. *Canadian Journal of Psychology*, **39**, 273-293.

SUNDBERG, J. (1988). Computer synthesis of music performance. In J. A. Sloboda (Ed.), *Generative processes in music* (pp. 52-69). Oxford: Oxford University Press, Clarendon Press.

TODD, N. (1985). A model of expressive timing in tonal music. *Music Perception*, **3**, 33-58.

TODD, N. P. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America*, **91**, 3540-3550.

TODD, N. P. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America*, **97**, 1940-1949.

TRUSLIT, A. (1938). *Gestaltung und Bewegung in der Musik.* Berlin-Lichterfelde: Chr. Friedrich Vieweg.

VAN NOORDEN, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands.

WILLIAMS, K. N., & PERROTT, D. R. (1972). Temporal resolution of tonal pulses. *Journal of the Acoustical Society of America*, **51**, 644-647.

ZUCKERKANDL, V. (1956). *Sound and symbol: Music and the external world.* Princeton, NJ: Princeton University Press.

NOTES

1. The sign of the correlation is due to the preferred representations of the data: percent correct rather than percent errors for the DAP, and IOI duration rather than local tempo for the performance profile.

2. While the magnitude of the DAP-FAP correlation provides an indication of whether presumable bottom-up effects are due mainly to

variations in bias or to variations in sensitivity, a better way of distinguishing these two possibilities is to compare increment and decrement detection. This was done in subsequent experiments (Repp, 1995c). The present paradigm does not lend itself to a signal-detection-theory analysis because the false-alarm rates are too unreliable and difficult to convert into the proportions needed for calculation of d' and β.

3. Similarly, the well-known demonstrations by van Noorden (1975) and others of stream segregation as a function of pitch distance (see Bregman, 1990) typically involved faster rates of presentation than those used here, simple rather than complex tones, and silent gaps between tones. Stream segregation is not likely to occur in the present paradigm.

4. Results of a recent study do suggest that two piano tones are easier to compare with respect to loudness when they have the same pitch than when they differ in pitch (Repp, 1995b).

5. Positions are denoted by a bar number followed by a beat number. It is assumed here that the subphrases have the same upbeat structure and the same length.

6. Naturally, there were individual differences among the pianists in their timing and velocity profiles. These differences, which were more quantitative than qualitative, will not be discussed here in detail. The average profile is an estimate of what all performances have in common, and hence an estimate of what the average musical listener may expect to hear.

7. Some individual pianists showed differences between the two profiles, but they were different in nature for each pianist and thus averaged out.

8. The second criterion was relaxed for one subject who had had 12 years of piano and 9 years of violin instruction but did not currently play either instrument.

9. Eight additional subjects were excluded because of difficulties in performing the task. Three were excused from the experiment because they could not hear any lengthening in the initial examples; of the five who completed the test, three responded randomly, one was close to chance and often gave more than two responses per trial, and one performed near chance during the first but not the second half of the test (which was highly atypical). Of these five subjects, three would have been classified as amateurs and two as nonmusicians. It is noteworthy that they ranged in age from 29 to 57; thus age may be a handicap in this task. Repp (1992a) similarly observed that musically untrained listeners often have great difficulty hearing timing differences.

10. When an IOI is lengthened, the tone occupying it has extra time to decay before the next tone comes on, and the final decay following its nominal offset (the simulated key release) is correspondingly shortened (see Repp, 1995a). Thus there is less energy of the preceding tone at the onset and during the initial portion of the following tone, which may well enhance its perceived loudness.

11. The superiority of the musicians would have been even clearer if one subject in that group had not performed rather poorly.

12. In a forced-choice task, chance level would be about 1/7, considering that responses in three adjacent positions were accepted as correct. However, the subjects responded only 16% of the time when they could not detect a target (see false alarms, below), so the actual chance level was about $(1/7)^2$, or 2% correct.

13. Starting with the last two tones in the first subphrase, every pair of successive IOIs in Tune A shows higher performance in the first (metrically strong) position than in the second (metrically weak) position. (This is seen more clearly in Figure 4.) However, since Tune B generally shows a similar (pitch-aligned) pattern (Figure 5), it is difficult to attribute it to metrical alternation, which was just the opposite in Tune B from that in Tune A.

14. Incorrect responses could not simply be divided into false alarms and misses, because of occurrences such as a single-target trial with two false-alarm responses or with one correct response and one false-alarm response.

15. No subject had to be excluded because of inability to hear the intensity increments or because of random performance on the test, even though the task was quite difficult. (The same was true in Repp, 1992a, Experiment 2.) It appears that intensity-increment detection is a more straightforward task for untrained subjects than is duration-increment detection.

16. There was also a change in spectrum along with the intensity change, as the instrument used modeled the natural covariation between dynamics and spectral structure in piano tones. However, the resulting change in timbre (increase in brightness) was almost certainly too small to be detected as such.

17. Given a basic guessing probability of about 1/7 and an average false-alarm rate of close to 50% (see below), the chance level in this experiment was roughly 7%.

18. It seems unlikely that this higher incidence of false alarms was due to *random* variations in the perceived loudnesses of the different piano tones as a function of pitch (even though some random variability in peak intensity was found by Repp, 1993b). In that case, the FAP distributions (Figure 10) should have been flatter, and absolute detection accuracy should have been considerably worse than it was. The cause of the frequent false alarms probably was a *systematic* effect of pitch variation on perceived loudness.