867

Quatrièmes Journées Internationales d'Etude :
Perception et Production du Rythme

Fourth Rhythm Workshop:
Rhythm Perception and Production

Actes du Congrès
Proceedings

8 - 12 Juin / *June* 1992
Bourges, France

EDITEURS: C. AUXIETTE, C. DRAKE, C. GERARD

DETECTABILITY OF RHYTHMIC PERTURBATIONS IN MUSICAL CONTEXTS: BOTTOM-UP VERSUS TOP-DOWN FACTORS

Bruno H. Repp

Haskins Laboratories, New Haven, Connecticut

Abstract

Two experiments examined the hypothesis (Repp, 1992) that the relative detectability of slight rhythmic perturbations in an otherwise mechanically regular music performance reflects listeners' expectations of local performance microstructure: Expected variations are more difficult to detect. An alternative possibility is that variations in detectability are caused by psychoacoustic contextual factors. Experiment 1 assessed the relative detectability of timing perturbations (lengthening of single tones) in two tunes and found high negative correlations between the detection accuracy profiles and the timing profiles of expressive performances of the tunes, which strongly supports the expectancy hypothesis. However, Experiment 2 found a much weaker relation between the detection accuracy profiles for intensity increments and the performance intensity profiles, suggesting that the perception of intensity variations is less dependent on top-down expectancies. As to the role of bottom-up factors, they may have contributed to performance variations in Experiment 1, though their effects coincided with those of top-down expectancies. The bottom-up factors that may have governed performance in Experiment 2 remain to be identified.

Introduction

In music, as in language, there is rich and systematic subcategorical variation that contributes vitally to naturalness, expressiveness, and individuality. These aesthetic and communicative qualities are likely to be an integral part of the mental representations of music and speech. The very fact that certain normative subcategorical properties are *expected* for speech and music alike presumably accounts for listeners' failure to consciously notice their presence. Just as listeners' expectations of phonetic and prosodic microstructure of speech can be probed and demonstrated, it is meaningful to ask what the normative microstructure for a particular piece of music might be and how its mental representation could be probed.

In a recent exploratory study of the detectability of timing perturbations in music performance (Repp, 1992), I presented listeners with musical excerpts rendered with isochronous timing (i.e., equal tone interonset intervals or IOIs) on a digital piano. In each presentation of each excerpt, a single IOI and the tone occupying it was lengthened by a small amount, and the listeners' task was to detect and report the position of that tone. All IOIs in each musical excerpt were probed in this way, and the percentages of correct responses (hits) were plotted as a function of IOI position to yield a *detection accuracy profile* for each excerpt. There was a significant negative correlation between this profile and the temporal microstructure of an expressive performance of the same music: Lengthening was more difficult to detect in positions where musicians were likely to apply it.

These findings led me to hypothesize that listeners have expectations about the temporal microstructure of performed music which interact with their perception of actual timing. When lengthening is expected in a certain position, actual lengthening in that position will sound normal, as it were, and hence will be difficult to detect. When lengthening occurs where it is unexpected, listeners will readily detect it as a timing irregularity. This "top-down" explanation of the findings was contrasted with a potential "bottom-up" explanation, according to which the complex acoustic structure of the music was directly responsible for the variations in detectability of lengthening. The two explanations are not mutually exclusive; that is, effects of expectations based on purely musical factors (such as hierarchical grouping, metric structure, and harmonic progression) may be overlaid on effects due to psychoacoustic factors (such as changes in pitch, intensity, and duration), and they may even be congruent and thus indistinguishable.

The purpose of the present study was to replicate and extend my earlier findings, and to attempt to dissociate some potential bottom-up influences from potential top-down influences. Using simpler materials than the previous study, the present experiments were designed to examine the detectability of both duration and intensity increments, as well as possible perceptual interactions between timing and intensity microstructure.

I composed two monophonic melodies (Tune A and Tune B; Figure 1), in which similar pitch sequences (I and II in Figure 1; the transposition is considered irrelevant here) occur in very different metrical contexts. These pitch sequences were expected to yield similar results if bottom-up factors predominate, but different results if top-down (metric and grouping) factors govern detection performance.

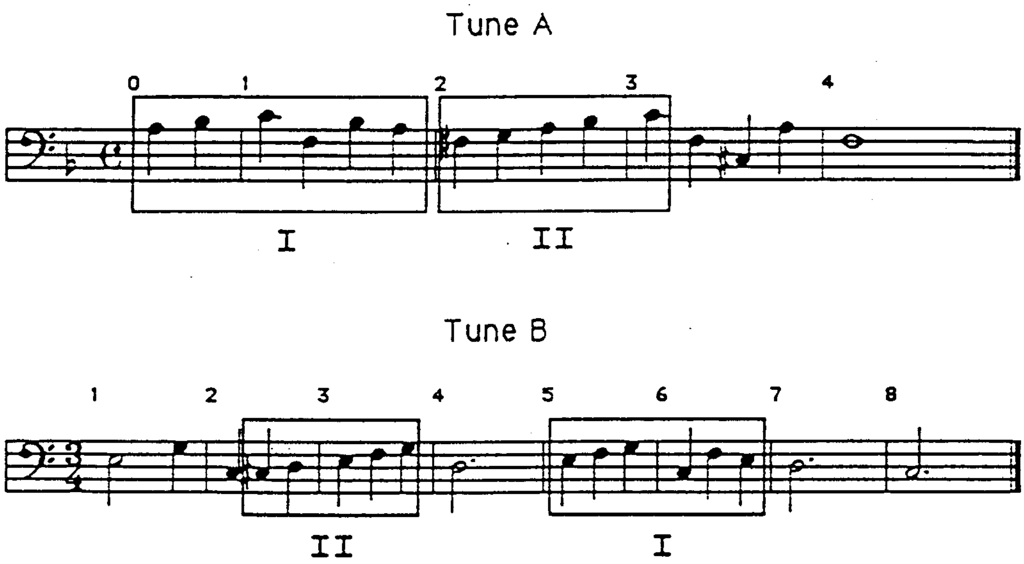

Fig. 1: Musical materials.

To produce a reference performance, I played the two tunes three times in alternation with the right hand, at a moderate tempo and "with expression", on a Roland RD250S digital piano. Figure 2 shows the average IOIs and MIDI velocities (relative tone intensities) for the two tunes. The timing patterns show the expected *ritardandi* at the ends of structural units (half-phrases), as well as *accelerandi* at their onsets. The intensity pattern shows accentuation of metrically strong tones as well as of preceding tones in Tune A, but not consistently in Tune B. If these performances are at all representative, and if listeners have performance expectations that interact with their perception of musical rhythm, then the accuracy profiles in the detection tasks should be inversely related to the performance patterns in Figure 2.

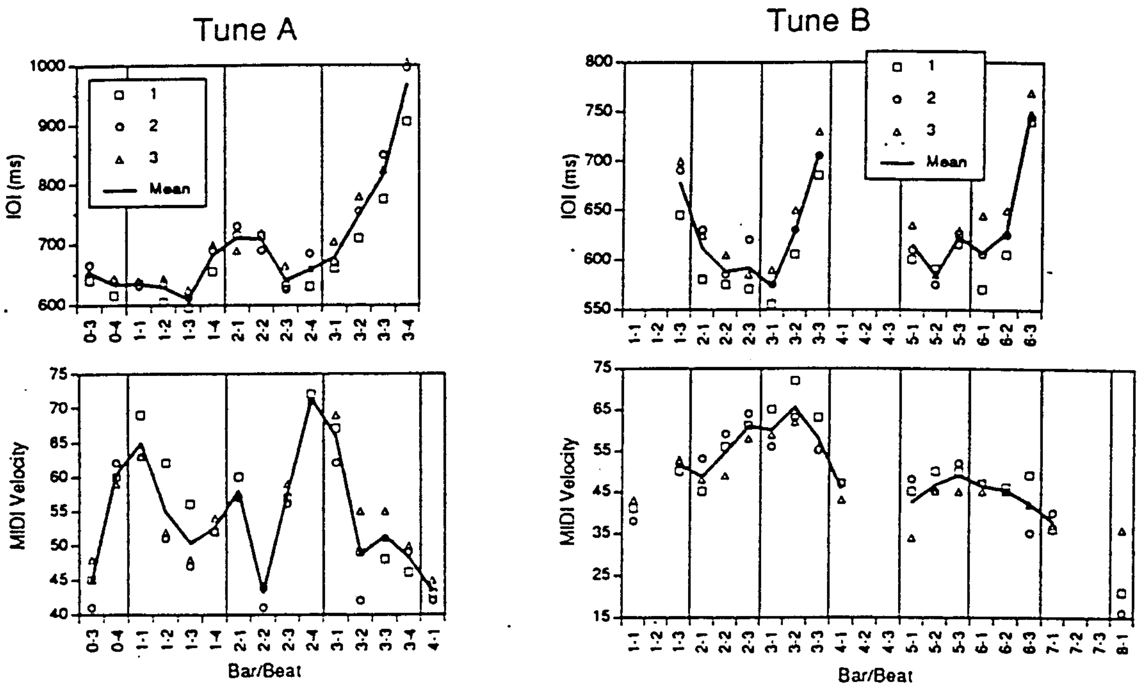

Fig. 2: Performance measurements.

Experiment 1: Detection of Duration Increments

Twelve subjects participated, most of whom had only limited musical training. The materials were produced under MIDI control on the digital piano. "Deadpan" versions of the two tunes, used for familiarization only, had quarter-note IOIs of 600 ms and constant MIDI velocities for all tones. The offset of each tone coincided with the onset of the following tone. In each of the experimental stimuli, a single IOI (and the tone occupying it) was lengthened by a certain amount. This could be any of the quarter-note IOIs, except for the first one in Tune A. The longer IOIs were never lengthened, and the subjects knew that. Five amounts of lengthening were used, ranging from 14% (for familiarization only) to 12%, 10%, 8%, and 6%. On the experimental tape, Tunes A and B always occurred in alternation. There were four blocks of experimental trials, each consisting of 13 pairs of tunes, so that each position in each tune was lengthened once, in random order. The amount of lengthening decreased across successive blocks.

The experiment included two tests of this kind, which were identical except that in one the tunes had no intensity microstructure (A, B), whereas in the other they were assigned the average MIDI velocity profile of the reference performance (AV, BV).

The subjects' task was to detect and report which tone had been lengthened in each playing of each tune. They marked their answers by circling the appropriate quarter-note in the notated music. If they did not hear any lengthened note at all, they were to enter a question mark; in fact, however, they often circled a wrong note, which I considered a false alarm response. Responses were scored as correct if they fell within one position of the lengthened note, as there was a pronounced tendency among subjects to circle the note following the correct position.

Subjects' performance level was well above chance, even in the most difficult condition. As expected, accuracy declined as task difficulty increased. Subjects were slightly less accurate when intensity microstructure was present than when it was absent. Intensity microstructure also selectively increased false alarms responses to tones with relatively high intensity. Figure 3 shows the average error profiles for the two parts of the test. The average correlation between the error profiles and the performance timing profile was -0.87 for Tune A/AV and -0.86 for Tune B/BV.

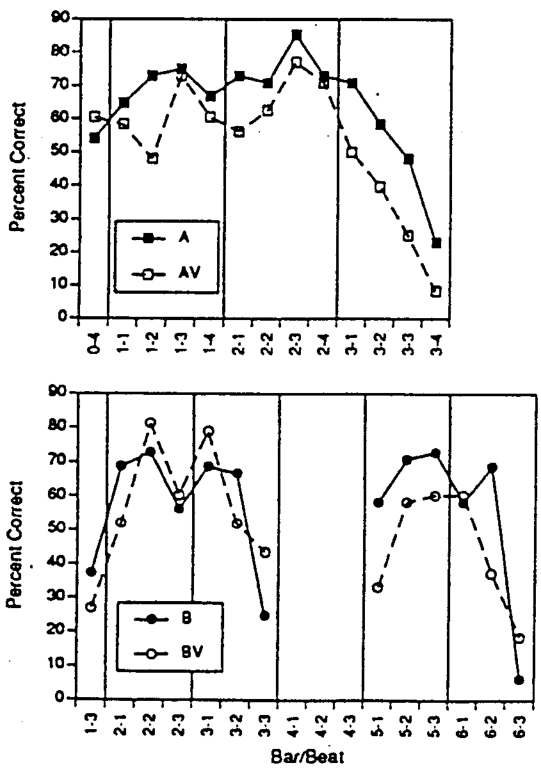

Fig. 3: Results of Experiment 1.

This close correspondence could have been partially due to psychoacoustic factors that affected both perception and (indirectly) performance. For example, lower detection performance may be expected in positions adjacent to a long IOI (Monahan & Hirsh, 1990). This bottom up explanation could account for most of the variation in the accuracy profile for Tune B/BV, and for the poor performance in the final position of Tune A/AV. However, the progressive decline in detection performance over the preceding five positions in Tune A/AV is not easily attributed to bottom-up factors. It is likely to reflect subjects' expectation of an extended final *ritardando*--a top-down factor. Comparisons of the error profiles for the similar pitch sequences (I and II) in the two tunes did not add much to these observations.

In summary, Experiment 1 replicates the earlier finding (Repp, 1992) of a correlation between detection accuracy and performance timing profiles, and strongly supports the hypothesis that lengthening is more difficult to detect where it tends to occur in performance. The results are less strong with regard to separating top-down and bottom-up contributions to this correspondence. However, the correspondence is so close that a purely top-down explanation seems most parsimonious.

Experiment 2: Detection of Intensity Increments

Experiment 2 was exactly analogous to Experiment 1, with intensity variation taking the place of timing variation as the primary variable, and the converse for the variable of secondary interest. Twelve listeners participated. "Deadpan" versions of the two tunes without any timing or intensity microstructure were again used for initial familiarization. In each of the experimental stimuli, the MIDI velocity of a single quarter note tone was increased by a certain amount, corresponding to increases in peak rms sound level of 3 dB (familiarization only), 2.5, 2, 1.5, and 1 dB. Again, there were two tests, one without any timing microstructure in the stimuli (A, B), and the other with the original average timing microstructure being present (AT, BT). The instructions asked subjects to circle the note that was louder than the others.

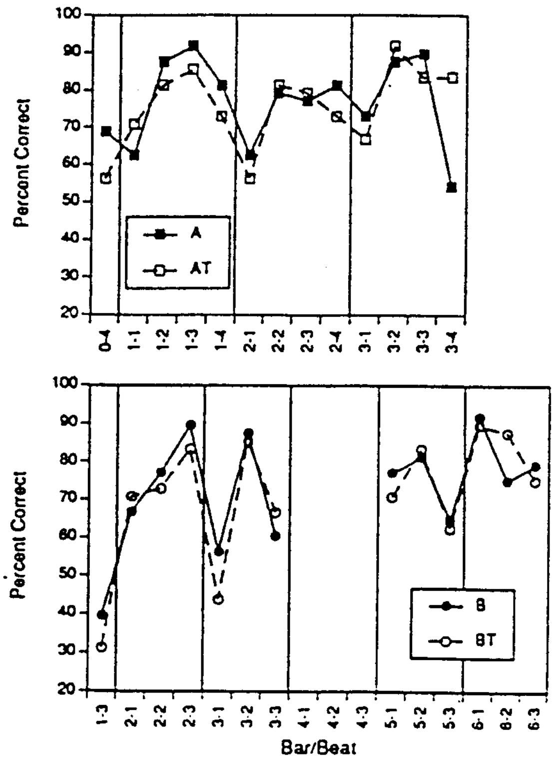

Fig. 4: Results of Experiment 2.

As in Experiment 1, responses were scored as correct if they fell within one position of the correct note, though there was no bias to circle the following note here. As expected, accuracy declined across the four test blocks but was still well above chance in the most difficult condition. There was no effect of timing microstructure; detectability of intensity increments seemed to be independent of timing variations. Similarly, false alarm responses were not systematically affected by timing microstructure.

Figure 4 shows the average error profiles for the two tunes. Detection accuracy varied significantly across positions for both tunes. However, there was no close correspondence with the performance intensity profiles. The average correlation was -0.47 ($p < .10$) for Tune A/AT and -0.10

(n.s.) for Tune B/BT. Comparison of similar pitch sequences in the two tunes revealed a substantial difference in only one position.

These results seem to indicate that detection of intensity increments depends more on bottom-up than on top-down factors. However, the psychoacoustic effects that could account for the observed variation in accuracy across positions are unknown at present, as relevant psychophysical research is virtually nonexistent. The results might also indicate that the performance intensity profiles were atypical. However, the obtained error profiles do not suggest a plausible performance structure either, as they should if they reflected top-down expectations. Thus, the results of this experiment are not well accounted for at present and call for further research.

Conclusions

The present experiments explore a middle ground between traditional psychoacoustics and the perception of real music. They simultaneously deal with situations that are unusually complex from a psychoacoustic perspective and unusually primitive 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.

I suspect that music and its performance are well adapted to what appear to be more primitive determinants of auditory perception. Just as the evolution of speech and of its articulation seems to be grafted upon more basic perceptual and motoric capabilities, so the development of music and its performance, though telescoped into a much shorter historic time span, was probably built upon foundations provided by the human auditory, motoric, and emotional equipment. Thus there may be an inherent congruence or harmony between top-down and bottom-up factors, which makes them difficult to separate. This will have to be considered in future research.

References

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

Repp, B. H. (1992). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition* (forthcoming).

[Research made possible through the generosity of Haskins Laboratories and NIH BRSG Grant RR-05596.]