

Categorical perception of short rhythmic events

R Parncutt

Abstract

Literature on the perception and performance of rhythm in music suggests a fundamental perceptual difference between 2:1 and 3:1 time ratios. Alternating IOIs in the intended or notated ratio 2:1 are generally perceived as distinct temporal categories, and produce triple meter or triple subdivision of the beat. If the nominal ratio is 3:1, only the longer duration appears to be perceived as a temporal category; the shorter duration tends to function as a non-metrical anacrusis to the longer duration that follows it. This interpretation is supported by a range of empirical data on rhythmic timing, and may have important consequences for theories of rhythmic structure and models of performance timing.

Introduction

Gabrielsson, Bengtsson, and Gabrielsson (1983) analysed monophonic performances of melodies in ternary metres (3/4, 6/8) and found that nominal durations in the ratio 2:1 were consistently softened, that is, performed in a temporal ratio of less than 2:1 (e.g., 1.75:1). This effect was incorporated into Sundberg's rules of music performance (see e.g. Sundberg, 1988) under the heading *softening of durational contrasts*. According to that rule, a note may be lengthened if it is preceded by another note of twice its duration and followed by a tone of longer duration.

The inverse effect has been observed, although less consistently, for nominal ratios in the ratio 3:1 (dotted rhythms). It appears that such ratios are often performed in a ratio exceeding 3:1. The conventional "double-dotted" of some Baroque dotted rhythms is a familiar example. In Sundberg's rules, this effect may be accounted for under the heading *heightening of durational contrasts*, whereby short note values are slightly shortened.

Recently, Lindström (1992) and Gabrielsson (1993) demonstrated that the timing of short-long duration pairs in performance varies greatly as a function of the emotional content of the music as intended by the performer. For example, durational contrasts may be heightened to give a happy or angry affect, or softened to give a sad or tender effect. Their results suggest that emotional content affects both nominal ratios (that is, intended notation) and expressive deviations from nominal ratios. For example, in a sad or tender performance a performer may interpret nominal 3:1 ratios (dotted) as if they were 1:1 (even subdivision of the beat).

The present paper is concerned only with relationships between notation and timing. The effect of emotional content is neglected. Moreover, I will be concerned only with inter-onset intervals, or IOIs (called D_{ij} values by Gabrielsson et al., 1983). Of course IOI is only one of several performance parameters related to the notation and emotional character of rhythms. Other parameters include articulation (ratio of duration to IOI, or D_{i0}/D_{ij}) and (relative) playing level (dynamic contrast).

Why are some nominal duration contrasts softened, while others are sharpened? A possible explanation is suggested by Clarke's (1987) experiments on the categorical perception of rhythms. In identification and discrimination tasks performed by musically trained listeners, Clarke found a clear category boundary between duration ratios 1:1 and 2:1, but no category boundary between 2:1 and 3:1. He concluded, in general agreement with Fraisse (1956) and Povel (1981), that duration ratios are generally perceived in only two categories - *even* or *uneven* subdivisions of the beat - and that distinctions between different uneven subdivisions (e.g., 2:1, 3:1, 7:1) are of a continuous rather than a categorical nature. The difference between nominal 2:1 and 3:1 ratios may thus be categorically perceptible only if the difference is exaggerated. Perhaps this is why musicians exaggerate the difference between the two ratios in performance. Such an explanation would be consistent with Sundberg's (1988) tentative explanations of the purpose of the rules of performance. However it remains unclear why deviations from 2:1 and 3:1 are maintained in a piece of music long after it has become clear to the listener which of the two possibilities is

intended.

The purpose of the present paper is to propose an alternative explanation for the observed exaggeration of the distinction between nominal 2:1 and 3:1 ratios in music performance. I hypothesise that the shorter of the two events has *metrical accent* in the 2:1 case but not in the 3:1 case, or, more precisely, that the metrical accent of the shorter event is so small in the 3:1 case that it is negligible or imperceptible. I then hypothesise that local tempo slows in the vicinity of metrical accents, accounting for the softening of 2:1 ratios, and that non-metrical events are shortened, accounting for the sharpening of 3:1 ratios.

Definition of metrical accent

The present definition of metrical accent differs somewhat from that generally accepted in music theory, and more recently, in the psychology of music. Music theory texts generally define metrical accent according to the relative strength of the beats of a bar. In 4/4 time, for example, the first beat is the strongest, the middle the second-strongest, and the second and fourth least strong. Palmer and Krumhansl (1990) measured the relative strength of beats and sub-beats in a measure by a probe-tone task.

A definition of metrical accent based on the notated measure appears to enable a good quantitative approximation to perceived metrical accent in the case of unambiguous meters. However metrical accent may also depend on the actual rhythm that creates the meter, that is, on the specific positions and durations of the notes in each measure. I prefer to define meter in a way that follows directly from Yeston's (1976) theory of rhythm. According to Yeston, a meter is composed of *rhythmic strata*. For example, the strata that make up 4/4 meter include the downbeat (first beat) of each bar, with a period of one measure, or one whole-note; the first and third beats, with a period of one half-measure, or one half-note; and the beats themselves, with a period of one quarter-note. A perceived meter may also include hypermetric strata, that is, strata whose periods exceed the notated length of the measure. The perception of rhythmic strata may be emphasised by calling them *pulse percepts* (Parncutt, 1987) or *pulse sensations* (Parncutt, in press) and considering their *perceptual salience*. The relative importance of the four beats of a 4/4 measure then depends on the relative importance of the pulse sensations generated by the rhythm in question. For example, if the pulse sensation of period one measure is much more salient than the pulse sensation of period one half-measure, then the first beat of the measure will be much stronger

than the third. Conversely, if the half-measure pulse is very strong and the measure pulse very weak, or if there is a strong pulse of period one measure coinciding with the third beat of each bar (an example of metrical ambiguity), then the first and third beats may turn out to have roughly equal metrical accent strength. The experimental results of Parncutt (in press), and the model that I developed to account for them, are consistent with these musical intuitions.

A suitable quantitative definition of metrical accent emerging from the above observations is the following: *The metrical accent of an event depends on the perceptual saliences of all pulse sensations including that event.* In a simple linear model, metrical accent may be predicted simply by adding corresponding pulse saliences. Pulse salience depends in turn on the number and phenomenal accent of rhythmic events coinciding with the pulse, and may be modelled by a pattern-matching routine. Note that an event need not be actually present in a rhythm to have metrical accent; a rest at the start of a measure has metrical accent, but no phenomenal accent.

Metrical accent additionally depends on tempo. As the tempo of a piece in 4/4 meter is steadily increased, the downbeats become stronger at the expense of the individual beats, which begin to be perceived instead as subdivisions of beats; and new metrical accents appear, for example at the starts of alternate measures (hypermeter). Conversely, if the tempo is steadily decreased, nominal subdivisions of the beat may begin to be perceived as beats. The present definition of metrical accent takes into account the dependency of metrical accent on tempo by allowing pulse salience to depend directly on tempo, and metrical accent to depend on pulse salience. Fraisse (1982) summarised literature on rhythm and tempo and concluded that musical rhythm is confined to pulse periods in the approximate range 200 to 1800 ms, and that the strongest rhythms are produced by pulses in the approximate range 400 to 900 ms, centred on about 600 ms. I have quantified these findings (Parncutt, in press) by a bell-shaped relationship between pulse salience and pulse period, called the *existence or dominance region of pulse sensation*.

Enhancement of pulse salience

Rosenthal (1992) defined a *parent* rhythmic level (or pulse) to be 2 or 3 times slower than a given reference pulse, and a *child* level to be 2 or 3 times faster. Rosenthal's terminology is convenient for the expression of the following interesting hypothesis: The salience of a pulse

may depend not only on the events of a rhythm that coincide with that pulse (pattern-matching) and on the pulse's tempo, as described above, but also on the presence of child or parent pulse sensations. I call this effect *enhancement of pulse salience*. Evidence for an enhancement effect may be found in historical experiments on *subjective rhythmisation* whereby a simple undifferentiated pulse (isochronous sequence of identical sound events) is spontaneously perceived in groups of two, three, four, or more. The result of interest here is that groups of four occur consistently more often than groups of three. A possible explanation is that 4/4 meter is more common than 3/4 meter, so listeners are more likely to fit a more familiar "4/4 template" to an isochronous sequence. However this does not explain why 4/4 is more common in the first place. According to the hypothesis of pulse salience enhancement, groups of four are perceived more often than groups of three because the salience of the pulse sensation corresponding to every fourth event is enhanced by the simultaneous presence of another strong pulse sensation corresponding to every second event. No parallel situation exists in the case of grouping by three.

The principle of pulse salience enhancement may be applied to durations in the nominal ratio 2:1 and 3:1. A repeating 2:1 rhythm can evoke pulse sensations of period 1 and 3, where every event of the pulse of period 3 corresponds to the onset of the longer of the two events, and both event onsets belong to the pulse of unit period. At an appropriate tempo, the faster of these two pulses is generally salient enough to be perceived, as two out of three events in the pulse correspond to actual rhythmic events, and the two pulses mutually reinforce each other according to the principle of pulse salience enhancement. At first glance one might suppose that a repeating 3:1 (dotted) rhythm might by analogy evoke pulse sensations of duration 1 and 4. However there are two reasons why the pulse sensation of unit duration may be so weak as to be imperceptible. The first is that only two out of four of the events in that pulse correspond to actual rhythmic events. The second is that the two pulse sensations do not, according to the hypothesis of pulse salience enhancement, reinforce each other. Their periods stand in the ratio 4:1, so there is a missing level between the two: using Rosenthal's terminology, one is the grandchild, not the child, of the other. The repeating 3:1 rhythm may therefore evoke only one pulse sensation - the pulse of period 4, corresponding to the onsets of the longer notes.

According to the above definition of metrical accent, an event can have metrical accent only if it

belongs to a pulse sensation. If the previous argumentation is correct, a repeating 2:1 rhythm typically has three metrical accents in each cycle, two of which correspond to the actual notes. But in the 3:1 case, there is only one appreciable metrical accent in each cycle, corresponding to the onset of the longer event. The shorter event has metrical accent in the 2:1 case but not in the 3:1 case.

Metrical accent and timing

How could this state of affairs explain the observed softening of 2:1 relationships and heightening of 3:1 relationships? I propose two general principles that may account for these (and other) timing effects. They are *slowing of local tempo in the vicinity of metrical accents*, and *shortening of events with weak or negligible metrical accent*.

Local tempo is often assumed to vary continuously with time, and fluctuations in local tempo may be assumed to communicate musical structure (Clarke, 1987). It has commonly been observed in the literature on the performance and perception of musical rhythm that important or accented events are either *delayed* or *lengthened*, or both. Combining these two observations, one may suppose that local tempo slows temporarily in the vicinity of accents. Accents may be phenomenal, metrical, or structural; for the present purpose, I am concerned only with metrical accents, and assume that similar principles apply to other varieties of accent. Consider the case of a nominal 2:1 ratio. If tempo slows near the onset of the longer event, the ratio 2/1 will become $(2+d)/(1+d)$, where d is a small time increment (say, less than one). The result is less than 2, consistent with observed softening of 2:1 ratios. In the 3:1 case, one may assume instead that the shorter note is shifted closer to the longer one, to emphasise its weaker status as a non-metrical anacrusis to the note that follows it, or to strengthen the perceived relationship between the two notes. (An interesting analogy in the area of intonation is the sharpening of the leading tone. Perhaps the leading tone is shifted toward the tonic in order to emphasise its tonal weakness by comparison to the strength of the tonic, or to strengthen the perceived relationship by reducing the pitch distance between the two tones.)

The two proposed rules (slowing of local tempo in the vicinity of metrical accents, and shortening of events with weak or negligible metrical accent) may in some cases cancel each other out. A non-metrical upbeat to a particularly strong downbeat may perhaps be *both* shortened

to emphasise its non-metrical nature *and* lengthened to delay and thereby emphasise the downbeat. Which of these two rules will "win out" in a given situation presumably depends primarily on the strength of the metrical accent following the IOI in question. These ideas are consistent with findings of Gabrielsson (1993) that dotted subdivisions of the beat in 3/4 time are sharper in the middle of the measure than at the end of the measure: at the end of the measure, the short note may be lengthened to emphasise the metrical accent (downbeat) that follows it, thus softening the nominal 3:1 ratio. The interaction between the two proposed rules is consistent with the observation that nominal 3:1 ratios are sometimes sharpened, sometimes softened, as found for example by Lindström (1992) - although it is clear that emotional intent had a far greater effect on the timing of performances studied by Lindström than did notation.

In the 2:1 case, the relative robustness of the timing of the shorter event may be accounted for by regarding the shorter event as a *temporal category* (just as chromatic scale-steps may be regarded as pitch categories). The prevailing temporal framework or frame of reference of a rhythm may be thought to consist of temporal categories. Each cycle of a 2:1 rhythm appears to contain three temporal categories, two of which contain real events (phenomenal accents) and one of which does not. In the 3:1 case, the timing of the shorter event is less robust, suggesting that it is not perceived as a temporal category but rather as an anacrusis or neighbour to the longer note that follows it. If this concept of temporal category is correct, then there is a one-to-one correspondence between temporal category and metrical accent. If a note has non-negligible metrical accent, then it will be perceived to fall in a temporal category of its own; otherwise, it will be perceived as no more than a near neighbour to another, stronger event.

These observations suggest a revision to the theory of Lerdahl and Jackendoff (1983), who assumed that every notated event onset corresponds to a beat at the smallest level of metrical structure (Metrical Well-Formedness Rule No. 1). It appears instead that events must have perceptible metrical accent (as defined here) in order to correspond to a beat in the metrical structure. In short, not all notated events are metrical.

It is instructive to apply the two proposed rules of rhythmic timing to the well-known opening theme in 6/8 metre of Mozart's A-major sonata K.331, whose performance was analysed by Gabrielsson (1987). The first note (duration:

dotted 1/8-note) is sometimes lengthened, sometimes shortened by the pianists whose performance Gabrielsson analysed, perhaps because the tempo was not yet established. The second note of the theme (duration: 1/16-note) is generally shortened in performance, as it is a non-metrical accent preceding a weak metrical accent. The third note (duration: 1/8-note) is lengthened due to the slowing of local tempo in the vicinity of the metrical accent on the following (fourth) note. The fourth note (duration: 1/4-note) is shortened due to normalisation of note durations following lengthening of other notes according to the various rules. The fifth note (duration: 1/8-note) is lengthened due to slowing of local tempo near the metrical accent that follows it. Where two 1/16-notes occur at the end of a measure, the second 1/16-note is longer than the first, again due to slowing of local tempo near the metrical accent that follows it. Consistent application of these rules to examples such as this could produce machine performances that are musically more satisfying than has previously been possible.

Summary and conclusion

The ideas presented here may be applied to the automatic generation of musical performances of any notated rhythm by the following tentative procedure. First, estimate the phenomenal, metrical, and structural accent of each event (actual and implied) in the rhythm, as described above. Classify events with negligible metrical accent as "non-metrical." Then apply the principle that local tempo slows in the vicinity of accents, to an extent that depends on the strength of the accent. A tempo curve may be plotted as the reciprocal of local tempo against notated time in beats, such that the area under the curve between any two points corresponds to the physical time interval between the points. In a first approximation, continuous fluctuations in local tempo in the vicinity of metrical events may be modelled by a triangular peak in the tempo curve - the higher the peak, the greater the slowing. The gradient of the two sides of the peak may be different: for example, the slowing before an important event may last for a longer time period than the acceleration after the event. The two gradients of the tempo curve may be regarded as free parameters in a model for expressive timing that depend on musical style. The triangular peak corresponding to a given event may include other events, so that for example a string of less-important events leading to an important event may decelerate in the fashion of a ritardando. A more sophisticated model could incorporate the mathematical model of ritardando

proposed by Kronman and Sundberg (1987), or other curves implying physical motion (reviewed by Repp, 1993). Contributions to the tempo curve from different events may initially be assumed to add linearly. After all contributions are added, the result would be normalised so as to control the overall tempo of the passage. Non-metrical events would then be shifted to be closer to their immediate neighbours, the degree of shifting depending presumably on relative metrical accent and temporal proximity. The other rules in Sundberg's system would then be applied, such as the higher the louder, shortening of start tone in rising intervals, and marking of melodic and harmonic change.

The proposed two rules are intended to replace two of the rules in Sundberg's system, but there is not a one-to-one correspondence between the new and old rules. The first new rule - that of slowing in the vicinity of metrical accents - can have the effect of softening durational contrasts in the case of alternating 2:1 patterns, but in conjunction with the normalisation of durations so as to hold overall tempo constant, the rule may heighten durational contrasts when applied to a string of short events leading to a metrical accent. The second new rule shortens notes with negligible metrical accent, also heightening durational contrasts. As discussed above, both rules may act at the same time, partially or completely cancelling each other out.

The proposed rules have the following advantages. First, they are less *ad hoc* than the rules that they replace, as they are more closely related to fundamental perceptual parameters in rhythmic perception - pulse salience and metrical accent. Secondly, they are intrinsically tempo dependent, as metrical accent depends on pulse salience, and pulse salience is tempo dependent. Third, they are less "local" than the rules that they replace, depending not only on immediately preceding and following notes but on all notes in a time span at least as long as the psychological present (say, 2 to 8 seconds). Pulse salience and metrical accent are determined or affected by time spans at least as long as the psychological present.

A disadvantage of the proposed rules is that the quantitative theory upon which they are based is currently incomplete in several of its details. For example, the quantitative relationship between phenomenal accent and the various parameters that influence phenomenal accent - IOI, duration (articulation), pitch (melodic contour, harmonic change), dynamics, timbre, and interactions between these parameters - is presently only partially understood (Parncutt, in press). The general application of the described model

depends on a reasonably accurate model of such relationships. Until the gaps in the model are filled, it will be difficult to apply the model in a general way.

References

- Fraisse P (1956) *Les structures rythmiques*, Publications Universitaires de Louvain.
- Fraisse P (1982) "Rhythm and tempo," in D Deutsch (ed.), *The psychology of music*, New York: Academic Press, 149-180.
- 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*, Stockholm: Royal Swedish Academy of Music, 81-104.
- Gabrielsson A (1993) "Emotional expression in music performance," paper presented at the Stockholm Music Acoustics Conference (abstract).
- Gabrielsson A, Bengtsson I, Gabrielsson B (1983) "Performance of musical rhythm in 3/4 and 6/8 meter," *Scandinavian Journal of Psychology*, 24, 193-213.
- Kronman U & Sundberg J (1987). "How to terminate a phrase. An analysis-by-synthesis experiment on a perceptual aspect of music performance," in A Gabrielsson (ed.), *Action and perception in rhythm and music*, Stockholm: Royal Swedish Academy of Music, 49-56.
- Lerdahl F & Jackendoff R (1983) *A generative theory of tonal music*, Cambridge, Mass.: MIT Press.
- Lindström E (1992) "5 x 'Oh, My Darling Clementine': The influence of expressive intention on music performance," unpublished undergraduate thesis, Department of Psychology, Uppsala University.
- Palmer C & Krumhansl C L (1990) "Mental representations for music meter," *Journal of Experimental Psychology: Human Perception and Performance*, 16, 728-741.
- Parncutt R (1987) "The perception of pulse in musical rhythm," in A Gabrielsson (ed.), *Action and perception in rhythm and music*, Stockholm: Royal Swedish Academy of Music, 127-138.
- Parncutt R (in press) "A perceptual model of pulse salience and metrical accent in musical rhythms," *Music Perception*.
- Povel D J (1981) "Internal representation of simple temporal patterns," *Journal of Experimental Psychology: Human Perception and Performance*, 7, 3-18.
- Repp B H (1993). "Musical motion: Some historical and contemporary perspectives," paper presented at the Stockholm Music Acoustics Conference (abstract).
- Rosenthal D (1992) "Emulation of human rhythm perception," *Computer Music Journal*, 16 (1), 64-76.

Sundberg J (1988) "Computer synthesis of music performance," in J A Sloboda (ed.), *Generative processes in music*, Oxford: Clarendon, 52-69.

Yeston M (1976) *The stratification of musical rhythm*, New Haven, Conn.: Yale University Press.